The post-authorisation safety of COVID-19 vaccines:

Real-World Evidence

Xintong Li

Brasenose College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Supervised by:

Professor Daniel Prieto-Alhambra,

Dr Edward Burn,

Dr Antonella Delmestri,

Dr Victoria Strauss

Abstract

Vaccines against COVID-19 were developed, approved, and distributed at an unprecedented speed during the coronavirus disease pandemic. These vaccines have shown high efficacy in preventing severe COVID-19 and acceptable safety profiles in clinical trials. However, potential rare adverse events related to these new vaccines have been reported, and continuous vaccine safety surveillance is needed as mass immunisation against COVID-19 continues. With the ability of capturing information from larger and more diverse populations, routinely collected health data, also known as 'real-world data', can provide valuable real-world insights in post-marketing surveillance, complementing the knowledge gained from clinical trials.

This thesis aims to assess the post-authorisation safety of COVID-19 vaccines by applying epidemiological and statistical methods using real-world data.

I began with a literature review of real-world studies examining the safety of COVID-19 vaccines. I then introduced the methods and different data sources being used in the thesis. Three analytical studies were conducted using multiple electronic health records and claims datasets. In the first study, I characterised the background incidence rates of 15 pre-specified adverse events of special interest associated with COVID-19 vaccines. I observed considerable variability in the rates with respect to age and sex, emphasising the need for standardisation or stratification of the background rates used for vaccine surveillance. This study also found substantial heterogeneity in the estimated rates across databases, suggesting that observed rates among COVID-19 vaccine recipients should be compared with background rates obtained from the same database where possible.

I then assessed the association between COVID-19 vaccines, SARS-CoV-2 infection, and the risk of immune-mediated neurological events. I applied the observed-to-expected analysis and the self-

controlled case series methods using primary care records from the UK and Catalonia, Spain. This study found no increased risk of the included immune-mediated neurological events after COVID-19 vaccination, but increased risk after SARS-CoV-2 infection. These findings reaffirmed the safety of the studied COVID-19 vaccines, underscoring the importance of vaccination.

Finally, I examined the comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with COVID-19 vaccines using datasets from Europe and the US. I compared adenovirus-based COVID-19 vaccines with mRNA-based COVID-19 vaccines, and the secondary analysis compared two brands of mRNA vaccines. This study found an increased risk of thrombocytopenia after the first dose of ChAdOx1 (Oxford-AstraZeneca) compared with BNT162b2 (Pfizer-BioNTech). While the studied events were rare, these observed risks after adenovirus-based vaccines should be considered in planning future immunisation campaigns and vaccine development.

In these analyses, I have demonstrated that real-world data can generate timely and reliable evidence on post-authorisation vaccine safety. These findings have important implications for clinical practice, health policy, and future research. Above all, in light of the well-established benefits of the COVID-19 vaccination, my findings should encourage continued confidence in vaccination.

Acknowledgements

This DPhil would not have been possible without the incredible individuals who have supervised, mentored, inspired, and supported me.

First and foremost, I would like to express my sincere gratitude to my supervisors: Professor Daniel Prieto-Alhambra, Dr Edward Burn, Dr Antonella Delmestri, and Dr Victoria Strauss, for their continuous support and guidance. Their expertise, perspective and scientific rigour have been invaluable to the completion of this work and in shaping me as a researcher.

I have been enormously lucky to work with and learn from the exceptional team at the Pharmaco- and device epidemiology group at the Botnar Institute, both past and present. I would like to thank the many colleagues for the stimulating discussions, encouragements, and the enjoyable moments we have shared over the last three years. I am immensely grateful.

Conducting this research would not have been possible without financial support, and I wish to express my gratitude to the Clarendon Foundation and Brasenose College for their generous funding, which included a scholarship and covered research expenses.

Last but not least I want to express my gratitude to my family for their unwavering support and continuous encouragement, and for always believing in my adventures.

Statement of contributions

My supervisors, Professor Daniel Prieto-Alhambra, Dr Edward Burn, Dr Antonella Delmestri, and Dr Victoria Strauss, provided methodological and scientific guidance throughout all stages of this work. Additional contributors for each chapter are listed below.

The study presented in chapter Five was a network study collaborated with the OHDSI community. I developed the research question and conceived the study together with Professor Daniel Prieto-Alhambra, Patrick B Ryan from Janssen, and Professor George Hripcsak at the university of Columbia. Anna Ostropolets and I developed the phenotypes of the outcomes. Rupa Makadia, Azza Shoaibi, Gowtham Rao, and Anthony G Sena created the analytical package. A call for data partners collaborators was conducted in the OHDSI community. The proposed study was then included in the Vaccine Evidence Workgroup within OHDSI, where collaborators and data partners had weekly meeting to discuss on the study process and interpretation of results. I gathered the results from all data partners, synthesized the results, conducted the meta-analysis, drafted the manuscript including all tables and figures.

For chapter Six, I developed the research questions, conducted the literature searches, designed the study and statistical analysis plan. The analytic codes were developed by me and Edward Burn, Berta Raventós ran the codes in the SIDIAP dataset. Eugenia Martinez-Hernandez provided expertise in reviewing the medical code list for immune mediated neurological events.

The research questions addressed in chapter Seven was part of an EMA funded project, where the EMA had an open call for study proposal on the thrombosis with thrombocytopenia syndrome after COVID-19 vaccines. The study design and data analysis plan were developed by me, Victoria Strauss, and Professor Daniel Prieto-Alhambra. Professor Katia Verhamme from the Erasmus University Medical Center in the Netherland and Catherine Cohet from the EMA commented on the protocol. I built the analytical package,

which was based upon OHDSI packages designed by Profs Marc Suchard and George Hripcsak, Patrick Ryan and Martijn Schuemie on behalf of the OHDSI community. Then the analytical package I developed was run in these datasets with thanks to the following collaborators: Edward Burn for SIDIAP, Mees Mosseveld and Luis H John for ICPI, Can Yin for IQVIA US Hospital CDM, IQVIA Open Claims, France LPD and Germany DA. Assimilation and aggregation of all data, meta-analysis, production of all tables, plots, and the online interactive shiny app were undertaken by me.

I would like to thank all co-authors for their contributions to the manuscript and can confirm that the work produced in this thesis is my own.

Publications and conference proceedings

Publications contributing to DPhil thesis

Li, Xintong, Anna Ostropolets, Rupa Makadia, Azza Shoaibi, Gowtham Rao, Anthony G. Sena, Eugenia Martinez-Hernandez 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*** 373 (2021). doi: 10.1136/bmj.n1435

Li, Xintong, Berta Raventós, Elena Roel, Andrea Pistillo, Eugenia Martinez-Hernandez, Antonella Delmestri, Carlen Reyes et al. "Association between COVID-19 vaccination, SARS-CoV-2 infection, and risk of immune mediated neurological events: population based cohort and self-controlled case series analysis." ***bmj*** 376 (2022). doi: 10.1136/bmj-2021-068373

Li, Xintong, Edward Burn, Talita Duarte-Salles, Can Yin, Christian Reich, Antonella Delmestri, Katia Verhamme et al. "Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different COVID-19 vaccines: international network cohort study from five European countries and the US." ***bmj*** 379 (2022). doi: 10.1136/bmj-2022-071594

Other publications alongside the DPhil

Burn, Edward, Elena Roel, Andrea Pistillo, Sergio Fernández-Bertolín, Maria Aragón, Berta Raventós, Carlen Reyes et al. "Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain." *Nature Communications* 13, no. 1 (2022): 7169. DOI: 10.1038/s41467-022-34668-w

Burn, Edward, Elena Roel, Andrea Pistillo, Sergio Fernández-Bertolín, Maria Aragón, Berta Raventós, Carlen Reyes et al. "Thrombosis and thrombocytopenia after vaccination against and infection with SARS-

CoV-2 in Catalonia, Spain." *Nature Communications* 13, no. 1 (2022): 7169. DOI: 10.1038/s41467-022-34669-9

Burn, E., Li, X., Kostka, K., Stewart, H. M., Reich, C., Seager, S., ... & Prieto-Alhambra, D. (2022). Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. *Pharmacoepidemiology and Drug Safety*, *31*(5), 495-510.DOI: 10.1002/pds.5419

Català, Martí, Xintong Li, Clara Prats, and Daniel Prieto-Alhambra. "The impact of prioritisation and dosing intervals on the effects of COVID-19 vaccination in Europe: an agent-based cohort model." *Scientific reports* 11, no. 1 (2021): 18812. DOI: 10.1038/s41598-021-98216-0

Li, Xintong, Lana YH Lai, Anna Ostropolets, Faaizah Arshad, Eng Hooi Tan, Paula Casajust, Thamir M. Alshammari et al. "Bias, precision and timeliness of historical (background) rate comparison methods for vaccine safety monitoring: an empirical multi-database analysis." *Frontiers in Pharmacology* 12 (2021): 773875. DOI: 10.3389/fphar.2021.773875

Ostropolets, Anna, Xintong Li, Rupa Makadia, Gowtham Rao, Peter R. Rijnbeek, Talita Duarte-Salles, Anthony G. Sena et al. "Factors influencing background incidence rate calculation: systematic empirical evaluation across an international network of observational databases." *Frontiers in pharmacology* 13 (2022). DOI: 0.3389/fphar.2022.814198

Cabezas, Carmen, Ermengol Coma, Nuria Mora-Fernandez, Xintong Li, Montse Martinez-Marcos, Francesc Fina, Mireia Fabregas et al. "Associations of BNT162b2 vaccination with SARS-CoV-2 infection and hospital admission and death with COVID-19 in nursing homes and healthcare workers in Catalonia: prospective cohort study." *bmj* 374 (2021). DOI: doi.org/10.1136/bmj.n1868

Schuemie, Martijn J., Faaizah Arshad, Nicole Pratt, Fredrik Nyberg, Thamir M. Alshammari, George Hripcsak, Patrick Ryan et al. "Vaccine Safety Surveillance Using Routinely Collected Healthcare Data—An Empirical Evaluation of Epidemiological Designs." *Frontiers in Pharmacology* (2022): 2532. DOI: 10.3389/fphar.2022.893484

Lai, Lana YH, Faaizah Arshad, Carlos Areia, Thamir M. Alshammari, Heba Alghoul, Paula Casajust, Xintong Li et al. "Current approaches to vaccine safety using observational data: a rationale for the EUMAEUS (Evaluating Use of Methods for Adverse Events Under Surveillance-for Vaccines) study design." *Frontiers in Pharmacology* (2022): 554. DOI: 10.3389/fphar.2022.837632

Talks

Early-stage researcher talk “Leveraging real-world data for post-authorization safety of COVID-19 vaccines” The UK-RiME Annual Showcase 2023 and Centre for Epidemiology Versus Arthritis 10-year Anniversary Celebration Event June 2023 Manchester, UK

The FDA Center for Biologics Evaluation and Research (CBER) Biologics Effectiveness and Safety (BEST) Initiative Seminar Series “Leveraging real-world data for better health in Europe through collaborations between regulators & academia” May 2023 Online

Plenary Session “Impact of OHDSI’s work in characterising background rates of COVID vaccine adverse events of special interest” The OHDSI 2021 Global Symposium September 2021 Virtual

Conference presentations

Lightning Presentation “The Comparative Risk of Thrombosis with Thrombocytopenia Associated with Different COVID-19 Vaccines: International Network Cohort Study” The 38th International Conference on Pharmacoepidemiology & Therapeutic Risk Management August 2022 Copenhagen, Denmark

Table of contents

List of tables

List of figures

Chapter 1 Introduction

Starting the DPhil program during the COVID-19 pandemic was never going to be easy. Back in early 2020, I received many questions from friends and relatives who were trying to understand what was going on, as they knew that I was trained in epidemiology. While I did try to help interpreting some "basic" epidemiology concepts, the only thing I could do was stay at home and read the numbers and figures from the news. Thanks to the increased availability of data from routine clinical practice, I was able to contribute to the fight against the pandemic.

1.1 Rational and motivation

On 11 March 2020, the World Health Organization (WHO) declared COVID-19, an outbreak caused by the SARS-CoV-2 virus, as a global pandemic. In less than a year after the WHO's declaration, vaccines against COVID-19 were developed using different platforms at unprecedented speed, with phase 3 clinical trials reporting high efficacy in preventing severe COVID-19 (70-95%) [1–4]. Since December 2020 to June 2021, four vaccines have been authorised by regulators, such as the European Medicines Agency (EMA), the Food and Drug Administration (FDA) in the United States, and the UK Medicines and Healthcare Products Regulatory Agency (MHRA). The BNT162b2 mRNA (Pfizer–BioNTech) vaccine was the first to be authorised anywhere in the world and approved by MHRA for use in the UK on 2 December 2020. This was followed by the approval of the ChAdOx1 nCoV-19 (Oxford–AstraZeneca) vaccine on the 30 December 2020, which was an adenovirus-based vaccine. Another mRNA-based vaccine, the mRNA-1273 (Moderna), and another adenovirus-based vaccine Ad26.COV2.S (Janssen/Johnson & Johnson), were approved in the following months during 2021. Large-scale immunisation campaigns have been continuing since then, with additional vaccines approved, and with millions of doses administered world wide.

However, there was still some uncertainty surrounding the safety and effectiveness of these vaccines. Generally, vaccine safety must be continuously monitored, and this was especially true in this case considering that these vaccines were initially approved under Emergency Use Authorization and had only been assessed over a short time period in clinical trials. A list of Adverse Events of Special Interest (AESI) for COVID-19 vaccines had been developed and used in safety monitoring. While spontaneous adverse event reporting systems have served as a foundational component of post-approval pharmacovigilance activities to ensure the safe and appropriate use of medical products, routinely-collected health data such as electronic health records (EHR) and administrative claims had the potential to generate valuable information by providing real-world context about potential adverse events and their rates in populations of interest.

1.2 Aim of the thesis

The overall aim of this thesis was to assess the post-authorisation safety of COVID-19 vaccines by applying epidemiological and statistical methods using national and international real-world data.

1.2.1 Objectives

To this end, my research addressed the following objectives:

1. To estimate the background incidence rates characteristics of Adverse Events of Special Interest for COVID-19 vaccines in databases world-wide.
2. To evaluate the risk of developing adverse events after receiving COVID-19 vaccines using different epidemiological study designs.
3. To compare the risk of developing adverse events between different COVID-19 vaccines.

1.3 Structure of the thesis

First, I performed a literature review to identify studies on vaccine safety using routinely collected health data. I present the key components of this review in Chapter 2. I conducted a study focused on estimating the background incidence rates of adverse events of special interest for COVID-19 vaccines (Chapter 5), which is part of the preparation of COVID-19 vaccine safety surveillance. I then conducted three analytical studies on COVID-19 vaccine safety using data from the UK and other countries. The study presented in chapter 6 investigated the association between COVID-19 vaccines and immune-mediated neurological events, and the study in chapter 7 assessed the comparative risk of thrombosis events between different vaccines. I concluded the thesis with a synthesis of the findings, a summary of the strengths and limitations, and a consideration of implications and the direction for future research.

The remainder of this section will provide a summary of each chapter.

Chapter 2: I introduce the background to this thesis and summarise a literature review on the studies on COVID-19 vaccination safety using real-world data.

Chapter 3: I detail the methods used in the analytic chapters.

Chapter 4: In this chapter I introduce the different data sources being used in the thesis.

Chapter 5: In this chapter, I summarise the results from a study where I used electronic health records and health claims datasets from eight countries and conducted a multinational network cohort study. I estimated the background incidence rates of 15 prespecified adverse events of special interest associated with COVID-19 vaccines.

Chapter 6: In this chapter, I summarise the findings from a population based observed-to-expected analysis and self-controlled case series analysis using primary care records from the UK and Catalonia. The

aim of this study was to study the association between COVID-19 vaccines, SARS-CoV-2 infection, and the risk of immune mediated neurological events.

Chapter 7: In this chapter, I reported the results from an international network cohort study from five European countries and the US. The study aimed to quantify the comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with use of adenovirus based COVID-19 vaccines versus mRNA based COVID-19 vaccines.

Chapter 8: In this chapter, I summarise the findings from the results chapters, discuss the strengths and limitations of my research, and elaborate on the clinical and policy implications of my findings. Additionally, I outline the future work necessary to further enhance understanding around the safety of COVID-19 vaccines while controlling for confounding and bias.

Chapter 2 Background

This chapter provides an overview of rational and core concepts to this thesis. Firstly, what are real-world data and opportunities and challenges in using them in observational studies to generate evidence on real-word safety and effectiveness for medical products. Secondly, the chapter summarises how the safety of vaccines can be assessed. Thirdly, I present a review of studies which assessed the risk of adverse events after COVID-19 vaccines using real-world data.

2.1 Real-world data and common data model

2.1.1 Real-world evidence and routinely collected health data

Real-world data (RWD), defined as data relating to patient health status recorded during routine clinical practice, are collected from a variety of sources outside of traditional phase I–III clinical trials. Routinely collected health data are key sources of RWD, which include electronic health records from primary care or hospital settings, health administrative claims, disease registries, and epidemiologic surveillance systems, and other data collected for purposes other than research or without specific a prior research questions developed before collection.[5] With the increasing accessibility of electronic databases, the recognised limitations of traditional clinical trials such as rising costs and limited generalisability, RWD is increasingly recognised as having the potential to bridge the gap between clinical research and practice. Real-world evidence (RWE) can then be defined as the clinical evidence generated using RWD. [6–11] This thesis primarily focuses on the use of RWD to study the usage and potential benefits and risks of COVID-19 vaccines.

RWE is playing an important role in regulatory decision-making, health technology assessment, and public health in general. [9,12,13] In the United States, the FDA has recognised the value of RWE in drug development and has established a framework for its regulatory use. The FDA's framework aims to

promote the use of high-quality RWE to inform regulatory decision-making, while ensuring that data meet the necessary standards for reliability, validity, and relevance. The FDA also provides guidance on the use of RWE in various contexts, such as post-market surveillance and label expansions.[9,12]

Health Canada has expressed its intent to enhance the utilisation of Real-World Data/Evidence (RWD/RWE) in the regulatory decision-making process in Canada. In collaboration with the Canadian Agency for Drugs and Technologies in Health (CADTH), Health Canada launched an initiative in 2018 to integrate RWE throughout the life cycle of drugs. This was followed by the publication in 2019 of the document "Optimising the Use of Real World Evidence to Inform Regulatory Decision-Making," which acknowledges the growing global trend of using RWE in the assessment of drug safety, efficacy, and effectiveness for regulatory decision-making. [14]

In the UK, the MHRA has also issued a series of guidelines to aid the design of studies aiming to provide evidence suitable for supporting regulatory decisions.[15] Specifically for Europe, where rich healthcare data are available, there are operational, technical, and methodological challenges in the regulatory use of RWD due to the heterogeneous nature of the region, including various healthcare systems (at least one per country) and multiple languages, as described by Cave, et al.[16] The authors also proposed potential solutions to enhance the quality of evidence generated from the perspective of both designing and running a study. These included the use of common data elements, data formats and terminologies, data mapping to a common data model (CDM), implementing quality assurance and control procedures, and complying with the best methodological standards.

Additional and specific concerns were raised about the COVID-19 pandemic and the use of RWD. While the pandemic is thought to have increased awareness of RWD and RWE, it also demonstrated RWD's challenges in identifying the underlying source population, ascertainment of new conditions or treatments, as well as the impact of data quality. This highlighted the significance of rigorous

methodological design in reducing bias, producing robust evidence, and increased reproducibility as well.[17]

2.1.2 Multiple databases study and common data model

The availability of RWD is increasing globally, leading to a rise in studies utilising multiple databases. These multi-database studies leverage diverse sources such as primary care and hospital records, administrative claims, and registries to provide a comprehensive representation of real-world patient populations and healthcare practices. They offer insights into results generalisability and may improve statistical power.

There are several established research networks that utilise routinely collected health data to generate evidence across multiple databases. For example, the Vaccine Safety Datalink (VSD) is a collaborative project involving the US Centers for Disease Control and Prevention, integrated healthcare organisations, and networks across the US. Since its inception in 1990, the VSD has monitored vaccine safety and studied rare and serious adverse events following immunisation using electronic health data from 11 sites. Similarly, the Canadian Network for Observational Drug Effect Studies (CNODES) has obtained access to administrative healthcare data on millions of medication users in Canada and internationally. In Europe, there are European Union-funded networks such as ADVANCE/VAC4EU, EU-ADR, European Health Data Space (EHDS), and the DARWIN EU initiative (launched in 2022). These networks aim to support scientific evaluations and regulatory decision-making.

However, integrating, standardising, and analysing data from multiple sources present challenges. One solution is the adoption of a Common Data Model (CDM) to standardise the structure and relationships among data elements. A CDM, widely used in software engineering and now in health data, facilitates the exchange, pooling, sharing, and storing of data from multiple sources.

There are three types of CDMs being largely used to analyse healthcare databases, including organizing data models, mapping data models, and adaptive rules systems.[18]

The goal of an organizing CDM is to reflect as closely as possible the data found in the source system. Examples include the Sentinel Common Data Model[19] , which is led by the U.S. FDA, standardised US private-payer claims from multi databases; the National Patient-Centered Clinical Research Network (PCORnet) Common Data Model[20] , which collected EHR data from collaborated health centres across the network in the US; and the Informatics for Integrating Biology and the Bedside (i2b2), a framework and web-based tool for de-identified clinical data queries[21].

A mapping CDM applies a set of rules to a given data set in order to define a set with standardised medical constructs. Implementing a CDM involves mapping the source database schema to the standardised CDM schema, transforming the data accordingly, and mapping source terminologies to standard terminologies. This process includes converting data to a common format and applying standardised data transformation rules. CDMs enable researchers to harmonise data and conduct analyses consistently across different sources. Developing common definitions and terminology during the data preparation process further enhances data harmonisation and analytical consistency with CDMs. Examples of mapping CDM is the Observational Outcomes Medical Partnership (OMOP) common data model (OMOP CDM), developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI)[22–24]; and the Study Data Tabulation Model (SDTM) developed by the Clinical Data Interchange Standards Consortium (CDISC), who developed standards in collaboration with experts in pharmaceutical organisations and aims to improve standards for pharmaceutical organisations internationally[25].

An adaptive rules system enables project-specific adaptations of codes and revisions based on the raw data source, an organizing data model, or a mapping data model. The Canadian Network for Observational Drug Effect Studies (CNODES) network is an example[26] .

A study evaluated the usage of common data model among the Sentinel, PCORnet, OMOP CDM, and CDISC SDTM models for a large longitudinal electronic health record data using 11 criteria adapted from previous research, which covered completeness, integrity, flexibility, ease of querying, standards compatibility, and ease and extent of implementation.[27] They found that among the included models, the OMOP CDM best met the evaluation criteria, with superior performance in completeness and integration, for both standards compatibility and terminology coverage.

2.2 Vaccine safety and COVID-19 vaccines

2.2.1 COVID-19 vaccines and concerns on their safety

Vaccine development is a complex and time-consuming process. Traditional vaccines development usually take about 15 years from pre-clinical vitro and in vivo stage, to phase I to III trials, and to approval by regulators.[28,29] In the past, the mumps vaccine developed in the 1960s was the fastest vaccine from development to deployment, which took about four years. Given this historical context, developing a vaccine against COVID-19 within the condensed timeframe of 12–24 months was clearly a big challenge.

The total time spent on clinical trials were reduced by overlapping clinical trial phases. The initial phase I/II trials were followed by rapid subsequent phase III trials once the interim analyses of the phase I/II data were completed. Then, countries like the UK and the USA accelerated the review process of vaccines through Emergency Use Authorization (EUA), and then the vaccines were approved by many more countries and nations.

However, the rapid process of research and development, coupled with limited follow-up time post-vaccination aroused great public concern about the safety profile of COVID-19 vaccines. These concerns were amplified by the fact that no vaccine against coronaviruses had ever been licensed for use in humans

before, and the use of new platforms such as RNA vaccines. A UK-based survey showed that the most common reason given for not intending to receive COVID-19 vaccines was “possibility of the COVID-19 vaccine having side effects”.[30]

On the other hand, vaccines showed high efficacy and effectiveness in preventing COVID-19 infection, which therefore reduces hospitalisation and mortality.[31,32] It is important to demonstrate and summarise the safety information of COVID-19 vaccines for public confidence, and for enabling timely, evidence-based policy decisions for population-level use.

2.2.2 How do we assess the safety of vaccines?

Once a vaccine has been authorised for emergency use or approved, monitoring systems are put in place to track adverse events and side effects in real-world settings. These systems help identify any unexpected or rare reactions that might not have been observed during clinical trials.

2.2.2.1 Passive Surveillance

Passive surveillance using spontaneous report forms is the method most commonly employed in both national and international pharmacovigilance systems. The spontaneous report system relies on healthcare professional or community members reporting an adverse event following immunisation (AEFI). Passive surveillance can detect safety signals and generate hypotheses for further evaluation.

Globally, national regulatory agencies and the World Health Organization (WHO) contribute to the aggregated database Vigibase[33]. This database enables both a central team and individual nations to access and search global data, as well as data specific to their own regions. In Europe, EudraVigilance operated by the European Medicines Agency is used as the system for managing and analysing suspected adverse reactions to medicines that are authorised or being studied in clinical trials.[34] In the US, the Vaccine Adverse Event Reporting System (VAERS) is a national vaccine safety surveillance program co-

monitored by the FDA and the Centers for Disease Control and Prevention (CDC). Its purpose is to detect possible signals of adverse events occurring after the administration of vaccines licensed or authorized in the U.S. This system allows healthcare providers and the public to report any adverse events following vaccination.[35]

However, passive surveillance has notable limitations. One primary challenge is underreporting, as many vaccine-related adverse events go unreported due to factors such as lack of awareness, misattribution of symptoms, or healthcare providers' reluctance to report. Additionally, the data collected through passive surveillance are often incomplete and lack of detailed information necessary to establish causality, making it challenging to differentiate between true vaccine-related adverse events and coincidental health issues. Furthermore, passive surveillance does not provide a denominator of the total number of vaccine doses administered, making it difficult to calculate precise incidence rates, therefore reducing the ability to detect rare adverse events.

2.2.2.2 Adverse events of special interest

Adverse Events of Special Interest (AESI) are pre-specified medically-significant events that have the potential to be causally associated with a vaccine product that needs to be carefully monitored and confirmed by further special studies.[36] Prior to the introduction of COVID-19 vaccines, a list of AESIs for the COVID-19 vaccines was created by the Brighton Collaboration's Task Force for Global Health via the Coalition for Epidemic Preparedness Innovations (CEPI)-funded Safety Platform for Emergency vACcines (SPEAC) project.[37,38] These AESI were selected based on their proven association with vaccination in general, experience with similar vaccine platforms, and theoretical concerns based on immunopathogenesis of COVID-19 diseases, viral replication during COVID-19 infection, and animal models with one or more candidate vaccine platforms.

It's important to note that most AESI can also occur in the absence of vaccination [39]. Therefore, post-authorisation surveillance to distinguish between a true vaccine safety signal and an event that happens coincidentally shortly after vaccination is necessary.

The incidence rate reflects the number of new cases naturally occurring in the population, often referred to as background incidence rates, or background rates. While it is not possible to estimate incidence rates from passive surveillance data due to the lack of denominator, one can use the background incidence rates to estimate the `expected` number of cases in a non-vaccinated population. Numbers of adverse events recorded in spontaneous reporting system (observed) can be then compared to the expected number. This is called the observed-to-expected analysis, which is generally used when a safety concern has been raised. [40] Ideally, background incidence rates should be estimated from populations that have not been exposed to the vaccine of interest but that have similar demographic characteristics to the vaccinated population.

2.2.2.3 Active vaccine safety Surveillance (AVSS)

Active surveillance involves the proactive acquisition and rapid analysis of information from millions of individuals recorded in large healthcare data systems. AVSS aims to verify safety signals identified through passive surveillance or to detect additional safety signals that may not have been reported as adverse events through passive surveillance systems. The “CIOMS Guide to active vaccine safety surveillance: Report of CIOMS working group on vaccine safety” describes the process for selecting the best approach to active surveillance considering key implementation issues, including resource-limited situation such as being confronted with a newly licensed vaccine.[41]

Methods that can be used for monitoring AESIs using the AVSS system include, but are not limited to, cohort event monitoring, sentinel surveillance, and data linkage.[36] For example, in the US, the Sentinel BEST (Biologics Effectiveness and Safety) System is used to conduct active surveillance. The BEST system

comprises of claims data and EHR data across the country with very short data lag, and enables rapid queries to detect or evaluate adverse events as well as studies to answer specific safety questions for vaccines.[42] While cohort event monitoring involves prospective data collection from vaccinated subjects, sentinel surveillance and data linkage can be conducted using routinely collected data.

2.2.3 Methods for vaccine safety assessment using routinely collected data

The increasing availability of real-world data allows safety studies to be conducted for COVID-19 vaccines. It is important to use well-designed epidemiological studies to test and measure the safety concerns on COVID-19 vaccines. Methodology considerations related to vaccine safety have also been discussed in several guidelines, publications, and documents based on experience learnt from other vaccination campaigns. For example, the "Report on appraisal of vaccine safety methods" developed by the Accelerated Development of VAccine beNefit-risk Collaboration in Europe (ADVANCE) reviewed and summarised the methods in vaccine risk assessment, with a focus on methods for signal strengthening and confirmation, rather than signal detection[43]. Baker et al. presented a study design selection framework with a structured decision table on methods for vaccine safety monitoring, from the Post-licensure Rapid Immunisation Safety Monitoring (PRISM) Program, the vaccination safety monitoring component of the US Food and Drug Administration's Mini-Sentinel project standpoint. [44] Commonly used study designs for assessing safety risk of vaccines including cohort studies, case referent studies, and self-controlled methods, which are described below and in the next chapter as well.

2.3 Scoping review: COVID-19 safety studies using RWE

To better understand the methodological approaches being used to study real-world COVID-19 vaccine safety, I conducted a literature review. This review aimed to describe the methods and statistical approaches of COVID-19 safety studies using routinely collected health data, including the types of data

sources. As such, this review primarily focuses on providing a qualitative summary of the included studies to offer insights into future vaccine safety research. No meta-analysis of results was conducted.

2.3.1 Methods

2.3.1.1 Search strategy and article selection process

I performed a scoping review of literature to identify peer-reviewed research studies on the safety of COVID-19 vaccines using routinely collected health data. This review focused on studies of adverse events following COVID-19 vaccination published from 1 January 2020 to 31 May 2022. I executed a search strategy (Appendix) in MEDLINE (via Ovid), EMBASE (via Ovid), and Scopus databases. The search terms were constructed according to the intervention of interest (COVID-19 vaccine), outcomes of interest (adverse events), and data sources (routinely collected health data). Each component included variations of the keywords and were adjusted according to different requirements of these databases. The search strategy was developed in consultation with epidemiologists as well as a librarian at the University of Oxford.

2.3.1.2 Eligibility criteria and selection process

I included manuscripts written in English, published in academic journals or made available on a pre-print server. For publication type, I excluded editorials, opinion articles, reviews (with or without meta-analysis), but included research letters, correspondence, or brief reports if they provided sufficient information on the study question. I also excluded case reports, case series, randomised controlled trials, and modelling studies.

Since my main focus was on methods used in active safety surveillance, I excluded passive surveillance studies such as those using data from spontaneous report systems (e.g. VEARS in the US, VegiBase by WHO, and Euraligance from the Europe). Studies conducted within single centres/tertiary care institutions were excluded as well.

To evaluate the eligibility criteria of studies, I used the PICOS (Population, Intervention/Exposure, Comparator, Outcome, and Study) framework, which provided a structured approach for identifying relevant data from each included paper.[45]

In the context of the review, the PICOS components were defined as follows:

a. People/population: individuals with vaccines against variants of SARS-CoV-2 from routinely-collected health databases. The population of interest included the general population, specific age groups, or individuals with specific health status (e.g. pregnancy, selected comorbidities)
b. Intervention/exposure: COVID-19 vaccination, with no restrictions of type/brand of vaccines.
c. Comparator: no restrictions were applied to the comparator. Comparisons can be made between COVID-19 vaccines and unvaccinated people/period, different type/brand of vaccines, , or SARS-CoV-2 infected patients.
d. Outcome: studies must include at least one safety outcome, which include adverse events or adverse reactions to the vaccines. Studies where outcomes were reported by patients, e.g. using questionnaire or mobile app, were excluded.
e. Study: I included only observational studies, so clinical trials, systematic reviews, cross-sectional, case reports, and case series studies were excluded. Prospective studies where participants were identified based on vaccination status and followed up for outcome information afterward were also excluded. I required absolute comparison (e.g. difference in incidence rates or cumulative incidence) or relative comparison (e.g. rate ratios, odds ratios, hazard ratios) to be estimated. Studies reporting only descriptive incidence rates without comparison were excluded.

After the initial search in each database, all records were imported into ZOTORO (https://www.zotero.org/) software to remove duplicates. I then used Rayyan® to screen all potential papers by title and abstract. Then, I retrieved the full text of remained records and reviewed them.

2.3.1.3 Data extraction and analysis

I used and adapted the structured form from STrengthening the Reporting of OBservational studies in Epidemiology (STROBE), the reporting standard for observational studies[46]. The following items were extracted from each article: the first author, publish date, data sources, study design, methods for addressing for confounding and errors, sample size, study period, characteristics of study population including age and sex, information on vaccine (name, dose, vaccine type, vaccine developer, and number of scheduled doses), comparator, outcome of interests, definition of outcomes, statistical analysis methods, effect estimates, and study limitations.

The extracted data were synthesised and summarised in frequency tables and figures. As this review served as a descriptive synthesis to inform further work, bias or quality of each study was not explicitly assessed.

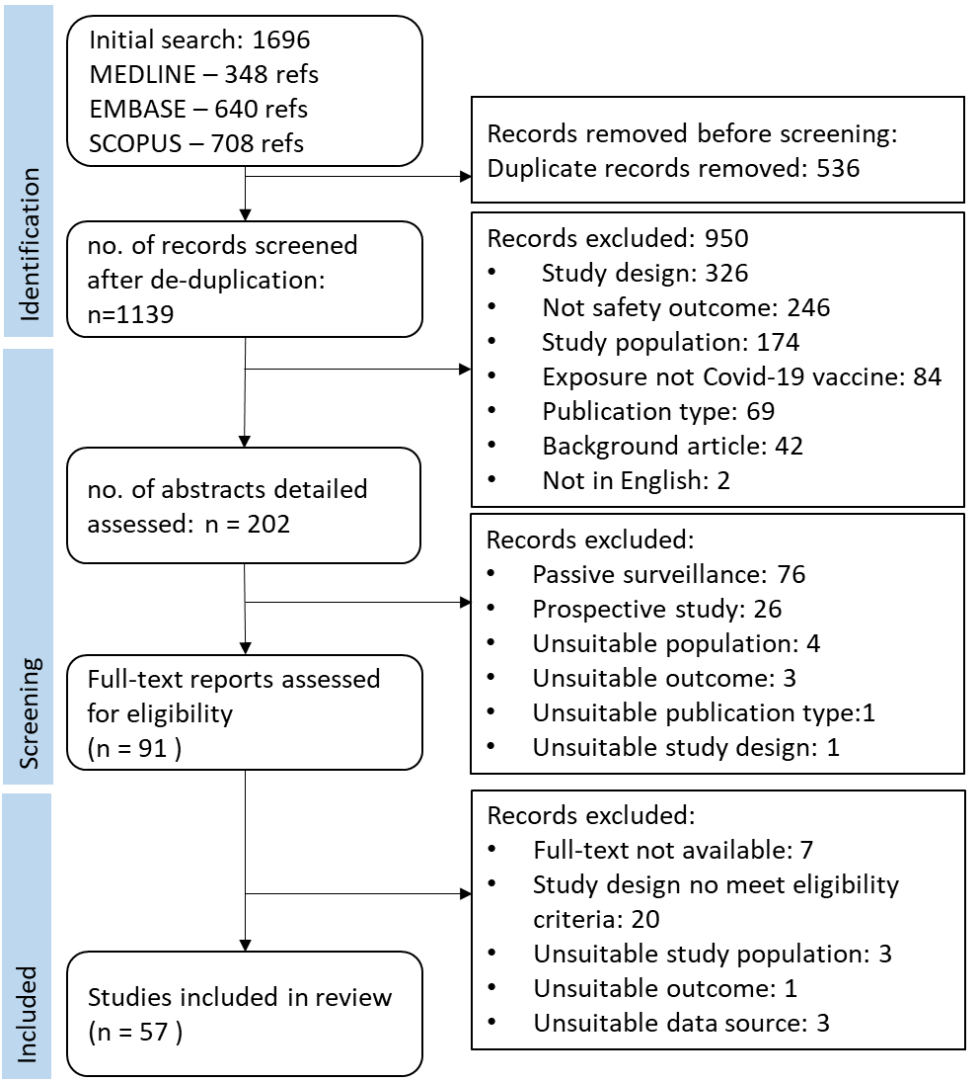
2.3.2 Results

2.3.2.1 Study selection

I identified 1,696 articles from the initial search from the three databases. After removing duplicates (n=536), 1,139 records were screened by title and abstract. I screened the studies based on the criteria listed in the previous section. Firstly, I did a brief screening majorly based on the title, 950 records were excluded in this stage. I then assessed the remaining abstracts in detail, and a total of 91 publications were reviewed in full text. After assessing the eligibility of the full-text manuscripts, 57 were included in the final review(details of included studies were presented in Appendix to Chapter Two). Figure 2-1 shows

details of the selection process based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Figure 2-1 PRISMA flow diagram for study selection process.



2.3.2.2 Characteristics of included studies

Data sources and study populations

The characteristics of the included studies are summarised in Regarding the study population, more than half of the studies (35, 61.4%) targeted the general population. Ten studies were conducted among

individuals with specific medical conditions, while seven studies focused on pregnant women or women who recently delivered.

Table 2-1. From the included publications which assessed the safety of COVID-19 vaccines using routinely collected health data, 11 (19.3%) were published in 2021. Most studies (50, 87.7%) used electronic health records as data sources. Administrative claims and registry data were used in three and four studies, respectively. Most of the studies were conducted using a single data source, while there were six multi databases studies: one used claims data from Italy [47], two used claims data from the USA [48,49], one used multiple registries from Denmark and Norway [50], one used EHRs from the UK [51], and one used EHRs from England and Spain [52].

Geographically, most of the 57 studies were based in Hong Kong (18, 31.6%), followed by the USA (17, 29.8%). Eight studies were from Israel, while seven were from the UK. The remaining studies were conducted in Norway, France, Singapore, Italy, Denmark, Canada, Spain, and Norway. Notably, only three studies used databases from multiple countries [50,52,53]. All the 18 studies from Hong Kong used the Hospital Authority (HA) comprehensive electronic health records system, which is a nation-wide, public health care database with linkage to population-based vaccination records in Hong Kong [54–72].

Five studies included in this review used the Vaccine Safety Datalink (VSD), a collaborative initiative between the CDC and eight integrated healthcare organizations. The VSD collected data from comprehensive electronic medical records of approximately 12 million insured individuals in the United States, with the primary goal of monitoring vaccine safety [73].

Regarding the study population, more than half of the studies (35, 61.4%) targeted the general population. Ten studies were conducted among individuals with specific medical conditions, while seven studies focused on pregnant women or women who recently delivered.

Table 2-1 General characteristics of included studies.

Study characteristics		**n (%)**
Total included		57
Publication year	2021	11 (19.3%)
	2022	46 (80.7%)
Country/ Region (study might include >1 country/ region)	Hong Kong	18 (31.6%)
	USA	17 (29.8%)
	Israel	8 (14%)
	UK	7 (12.3%)
	Norway	2 (3.5%)
	France	2 (3.5%)
	Singapore	1 (1.8%)
	Italy	1 (1.8%)
	Denmark	1 (1.8%)
	Canada	1 (1.8%)
	Norway	1 (1.8%)
	Spain	1 (1.8%)
Data source (study might include >1 type of data source)	Claims	3 (5.3%)
	EHRs	50 (87.7%)
	Registry	4 (7%)
Study population	Adolescents	1 (1.8%)
	General population	35 (61.4%)
	Elderly people	2 (3.5%)
	Pregnancy	7 (12.3%)
	With specific disease condition	10 (17.5%)
	Veterans	2 (3.5%)

2.3.2.3 Study design, exposure, outcome, and risk window

Cohort study was the most frequent study design, used in 31 studies (54.4%). Self-controlled case series (SCCS) design was the second most common, used in 14 studies (24.6%). Observed to expected analysis was used in nine studies, while only one study used a case-control design, and another one used a design similar to real-time surveillance.(Table 2-2)

Among all included studies, BNT162b2 was the most studied vaccine, investigated in 54 (94.7%) of the studies. The mRNA-1273 was the second most common exposure of interest, and was included in 24 (42.1%) studies. The viral-vector based vaccines (ChAdOx1, Ad26.COV2.S), and inactivated virus vaccine

(CoronaVac), were studied as well. Three studies included the booster dose vaccine [74–76]. Most studies included more than one vaccine as exposure. All the 17 studies researched on CoronaVac were from Hong Kong.

Table 2-2 study design, exposure, comparator, and outcome by the selected studies.

Study characteristics		**n (%)**
Study design	Cohort study	31 (54.4%)
	Case-control	1 (1.8%)
	Self-controlled case series	14 (24.6%)
	Nested case-control	4 (7%)
	Observed to expected analysis	9 (15.8%)
	Near real-time surveillance	1 (1.8%)
Vaccine/s studied	BNT162b2	54 (94.7%)
	ChAdOx1	11 (19.3%)
	mRNA-1273	24 (42.1%)
	Ad26.COV2.S	10 (17.5%)
	CoronaVac	17 (29.8%)
Comparator	Self-controlled	14 (24.6%)
	Concurrent unvaccinated or historical cohort	42 (73.7%)
	Active comparator	1 (1.8%)
Outcome of interest	Multiple AESIs, multi area	14 (24.6%)
	Thromboembolic event	12 (21.1%)
	Myocarditis and/or pericarditis	7 (12.3%)
	Pregnancy and birth	7 (12.3%)
	Autoimmune	8 (14%)
	Adverse event/ reaction	2 (3.5%)
	Acute liver injury	1 (1.8%)
	Hearing disorder	1 (1.8%)
	Hematological abnormalities	1 (1.8%)
	Herpes zoster	1 (1.8%)
	Oral Lichenoid Lesions and Oral Lichen Planus	1 (1.8%)
	Thyroid dysfunction	1 (1.8%)
	Mortality	1 (1.8%)

About one in four studies (15 in 57, 26.3%) selected the outcomes of interest based on COVID-19 vaccines AESIs, and included multiple adverse events that covered multiple disease areas. A total of 12 studies

specifically focused on thromboembolic events such as thrombocytopenia, venous thromboembolism, and arterial thromboembolism. Post-vaccination cerebral venous sinus thrombosis (CVST) was investigated in five studies. Autoimmune conditions including Bell's palsy and Guillain-Barré syndrome (GBS) were examined in eight studies, and myocarditis and/or pericarditis were the outcome of interest for seven studies. Maternal, neonatal, and early infant outcomes were included in the studies focused on pregnant women or women who recently delivered. Other outcomes of interest included acute liver injury, hearing disorder, haematological abnormalities, herpes zoster, oral lichenoid lesions and oral lichen planus, and thyroid dysfunction.

I found variations in the risk window for outcomes among the included study. For example, among 12 studies focused on thromboembolic events, 6 studies examined the risk during the 28 days after vaccination, while the remaining studies used 14-day, 21-day, 30-day, and 42-day risk window. For studies which included multiple AESIs as outcomes, risk window definitions also showed variability, ranging from 21 days post-vaccination to 56 days.

Among the studies which were not using self-controlled methods, most compared the vaccine of interest to either concurrent unvaccinated people, or a historical unvaccinated cohort. Only one comparative study was found, where incidence rate ratios of AESIs were estimated comparing CoronaVac to BNT162b2 [58].

Matching was the most widely used method to create comparable vaccinated and unvaccinated groups in cohort studies. In addition to age and sex, studies included other variables such as comorbidity and medication use history. Propensity score methods were used in 6 studies: one used propensity score matching [77], and the other five studies used inverse probability weighting in their analyses.

2.3.2.4 Study limitations

The most discussed limitation of included studies was potential residual or unmeasured confounding. Misclassification of exposures, outcomes, and covariates was mentioned in most studies as well. Other limitations included the imprecision or underpowered analysis due to very low event rates, generalisability of the study population, and database-specific limitations such as the lack of inpatient data. These limitations highlight the challenges and potential sources of bias in the safety studies of COVID-19 vaccines.

2.3.3 Discussion

This review provides a summary of the study designs and methods used to assess the risk of adverse events following COVID-19 vaccinations using real-world data. The majority of the included studies were conducted in developed countries or regions with well-established infrastructure for accessing routinely collected health data and conducting research. These areas also had earlier access to and distribution of COVID-19 vaccinations compared to low- and middle-income countries.

The outcomes of interest for the included study were mainly from the pre-specified AESI list, and those safety signals identified at the early stage of the mass immunisation programs. For instance, thromboembolic events following viral vector-based vaccines and myocarditis after mRNA vaccines were both included as AESIs for COVID-19 vaccines and were identified as safety signals during vaccination campaigns.

One noteworthy observation is that even for the same study outcome, different risk windows were used. The choice of risk window for a specific outcome can be impacted by clinical experience, the nature of disease/outcome, and data sources (for example, data collection process, and the healthcare system in

each country/region). While the Brighton collaboration published peer-reviewed case definitions for a selected set of AESI, there was no “standard” risk window applicable to all events. Moreover, the risk windows used in clinical practice may not directly translate to routinely collected data. Consequently, the differences in outcome definitions and variations in risk windows pose challenges for comparing study results and synthesizing evidence across different studies.

While exact variable matching has been used to identify an appropriate comparison group, other approaches such as the propensity score method were less commonly used. Given the prioritisation strategy at the early stage of vaccine distribution, it is important to apply appropriate methods to address measured confounders. Moreover, while most included studies acknowledged the potential impact of unmeasured confounders as a limitation, none of them explored the extent to which the study might be influenced by unmeasured confounders, such as using negative controls.[78–80]

In conclusion, this review showed the potential to rapidly generate valuable evidence on vaccine safety using real-world data. It also underscored the importance of employing standardized definitions for study variables to facilitate the generation of comparable information across studies. Furthermore, it emphasises the necessity of utilising appropriate study designs and methods to minimize confounding.

Chapter 3 Methods

This chapter considers the design options for the studies that will address the objectives of the thesis.

3.1 Study design

While the COVID-19 vaccines were newly developed, quickly tested and soon after distributed around the world, observational studies on vaccine safety have been conducted over other vaccines for decades, with well-established methods. During my DPhil course, I was involved in a method evaluation project, EUMAEUS (Evaluating Use of Methods for Adverse Events Under Surveillance-for Vaccines). The EUMAEUS project aimed to study the performance of different study designs in vaccine safety research using three claims and one EHR database with negative control outcomes and synthetic positive control outcomes. The project included exposures of vaccines against influenza A virus subtype H1N1, seasonal flu, Herpes Zoster, and Human papillomavirus, and applied 4 study designs: historical rate comparison, cohort study, self-controlled case series (SCCS), and case-control analyses.[81–83] By comparing metrics on type 1 and 2 errors, unmeasured confounders, and timeliness, the study found that while most methods suffered from high type 1 error, case-control and observed-to-expected designs also showed high systematic error. The SCCS method can detect small true effect size associations most rapidly, while historical comparators also perform well for stronger effects.

Based on the literature review, the empirical evaluation, the availability of data, and discussion with my supervisors, I applied the observed-to-expected analysis, cohort study, and the SCCS method in my thesis. The following section provides a brief description of each method used in the DPhil thesis. I summarized the key features, and the relative advantages and disadvantages of each design in Table 3-1.

Table 3-1 Study designs used in the thesis

	Observed-to-expected	**Cohort design**	**Self-controlled case series**
Description	Observed incidence of adverse events vs. expected incidence based on unvaccinated population.	Incidence ratio of adverse events between vaccinated vs. unvaccinated/ persons vaccinated with a control vaccine.	Incidence rates in exposed time periods vs. self-matched unexposed time periods
Level of evidence	Signal strengthening/ Hypothesis generation	Hypothesis testing	Hypothesis testing
Data sources	Spontaneous reports, observational databases*	Prospective cohorts, observational databases	Observational databases
Population	Vaccinated & Unvaccinated persons	Vaccinated & Unvaccinated persons/persons with a control vaccine	Cases
Require completeness in exposures (COVID-19 vaccine)	Y	Y (when comparators are unvaccinated)	N
Requires the Outcome to be recurrent and Independent or non-recurrent and Rare	N	N	Y
Requires Population/demographic information for exposure and control	Y	Y	N
Strength	Greater statistical power; Improved timelines: less affected by data lags.	Use of matching/ stratification to control for confounders.	Automatically control time-invariant confounders.
Weakness	Highly dependent on accurate estimation of background incidence rates; Subject to difference in confounders between current and historical vaccinees	Confounding by indication; unmeasured confounders; Misclassification especially unvaccinated group	Time-varying confounding; Reverse causality bias; Less statistical power; Less suitable for chronic AESIs

*Observational databases include population-based health databases such as administrative billing and electronic health record databases.

3.1.1 Observed-to-expected analysis

Observed-to-expected (O/E) analysis is a widely used method at the signal strengthening and hypothesis generation stage of vaccine safety surveillance.[40] This approach involves comparing the "observed" number of adverse events occurring in vaccinated individuals, as recorded in a data collection system, including both spontaneous reporting system and electronic health care record databases, with the "expected" number of cases that would naturally occur in the same population without vaccination. The expected number is estimated based on the incidence rates observed in a non-vaccinated population. In some cases, this method is referred to as the "historical comparison" method when a historical population serves as the benchmark for comparison.

For O/E analysis, the underlying assumption is that the observed and expected numbers correspond to the same populations, therefore all confounding factors have been controlled for. However, in practice, this is unlikely to be the case. For example, various temporal confounders such as seasonality, changing trends in the detection of adverse events, and variation in diagnostic or coding criteria over time would remain uncontrolled under this analysis. Another assumption is that the background incidence rate in the vaccinated population is the same as the background incidence rate in the population used to calculate the expected. It also requires that the population being used to estimate the background incidence has not been exposed to the vaccine of interest. Therefore, the design is highly dependent on an accurate estimation of background incidence rates of the adverse events for comparison.

Despite certain limitations, the primary purpose of this method is to provide rapid contextualisation of a safety signal, rather than establishing causal relationships. It is not to achieve detailed control but rather to mitigate gross bias by, for example, comparing by age groups. For continuous monitoring, adjustments may be necessary to account for the inflation of type 1 error rates resulting from multiple tests. Since the

observed rates can be calculated with a relatively short follow-up period of data after vaccination, this analysis enables timely detection of potential safety signals and is less affected by data lags.

In the study presented in chapter 5, the observed and expected rates were both stratified by age-sex groups, and incidence rate ratios were estimated within each group.

3.1.2 Cohort design

Cohort studies are a type of longitudinal study where participants are followed over a period of time after cohort entry. Traditionally, participants are recruited under certain protocols, with information on exposures, risk factors, and outcomes collected over time. However, studies with a prospective design are usually time-consuming and require a lot of resources during the follow-up. Cohort studies can be conducted retrospectively using observational databases as well, such as administrative billing and electronic health record databases. In post-marketing evidence generation for vaccine safety, cohort design is used for the hypothesis testing stage.

A cohort study can be conducted between the vaccinated and unvaccinated people, or between people with different types of vaccines, where the latter is also called comparative cohort study. For example, an unvaccinated comparator cohort would generate estimates of the relative risk of AESI occurrence in vaccinated versus unvaccinated populations, whereas comparator groups receiving an alternative COVID-19 vaccine brand would generate estimates of relative risk between different COVID-19 vaccine brands.

To estimate the relative risk between vaccinated and unvaccinated populations, an unvaccinated cohort needs to be defined. However, it is particularly challenging to properly identify a concurrent unvaccinated control cohort for the COVID-19 vaccine. Firstly, in countries without a universal health care system such as the US, vaccinations may be administered outside of typical reimbursement channels during the mass vaccination campaign, and therefore may not be comprehensively captured in administrative claims data.

Besides, due to the vaccine strategy, such as prioritisation of the vulnerable group, the identified unvaccinated control group may be of very different clinical characteristics than the vaccinated one.

Given these limitations and challenges, in chapter 6, I used the active comparator design by comparing users of one COVID-19 vaccine to another vaccine. I used propensity score matching to control for measured confounding and ensure the comparability between the target and comparator group. I then used negative control outcomes, which are further explained in section 3.3, to address the unmeasured confounding.

3.1.3 The self-controlled case series (SCCS) method

The SCCS method is a case-centered method, where only cases are sampled, and risk estimation is conducted by comparing the exposed and unexposed time for the same individual, rather than between individuals.[84] A risk window after exposure needs to be clearly defined. Since individuals serve as their own controls, time-fixed confounding, including both measured and unmeasured, are implicitly controlled.

The SCCS method requires each outcome to be either recurrent and independent, or non-recurrent and rare. The rare event assumption is valid if non-recurrent with a cumulative incidence below 0.1 per individual over the total observation period. [84,85]

One key assumption of the SCCS design is that the occurrence of an event should not affect subsequent exposures. For example, this assumption would be violated if people who experienced an event of interest may less likely, or delayed, in receiving the vaccination. This can be visually examined by plotting the number of occurrences of an event by time before or since vaccination for each outcome to assess the possibility that the event might affect subsequent vaccination. Including a pre-exposure period can potentially address this [86]. Another assumption is that an event does not influence the time of

subsequent follow-up. This assumption can be assessed by plotting the time from event to actual end of observation in patients who were censored and uncensored.

In the analytical chapter, this method was used with some modifications. First, to address the concern of misclassification in exposures, I restricted the study population to cases with one exposure (i.e., vaccination or SARS-CoV-2 infection). Second, due to the dosing schedule of COVID-19 in some of the study regions (for example, in Spain, the second dose of the Pfizer vaccine was recommended to be taken after 21 days following the first dose), I did not include the time after vaccination in the control window.

For statistical analysis of SCCS, the conditional Poisson regression model was used to compare the rates of adverse events in the risk and control window. The model assumes constant risk throughout the risk window. Incidence rate ratios, the relative rate of each outcome of interest in risk periods relative to baseline periods, and their 95% confidence intervals were estimated using each model.

3.2 Confounding and bias addressment

Bias can be defined as "as a systematic difference between the results of the non-randomised studies of the effects of interventions and the results expected from the target trial".[87–89] Confounding, selection bias, and measurement bias are the major types of systematic bias in observational study.[90] Confounding is the bias that arises when the exposure and the outcome share a common cause, usually because treatment was not randomly assigned. Selection bias may arise from the procedure by which individuals are selected into the analysis, such as loss to follow-up and differential inclusion criteria. Measurement bias refers to the misclassification of intervention status or errors in measurement of outcomes, which also called information bias.

In this thesis, I applied the following methods to reduce the risk of bias in my analyses.

3.2.1 Active-comparator and new-user design

The active comparator new user (ACNU) study design is one of the most influential methodological approaches to address biases in pharmacoepidemiology research, specifically confounding and selection bias.

In clinical practice, treatments are selectively allocated based on specific indications, which can lead to confounding by indication. This occurs when the indication for treatment is associated with both the exposure of interest and the risk of the outcome, but not as an intermediate step in the causal pathway. Another common confounding bias is the healthy initiator bias, where preventive treatments are selectively initiated among health-conscious individuals who have a lower risk of adverse health outcomes due to their healthy lifestyle. Additionally, treatments may be removed from frail individuals who are at an increased risk of adverse outcomes. These biases can result in spurious associations and exaggerate the apparent benefits of a drug.

The ACNU design aims to emulate the intervention phase of a randomized controlled trial (RCT). It assembles cohorts of new drug users, including those prescribed an index drug A and a comparator drug B, and follows them over time to assess relevant health outcomes. Only patients who are naive to both the study and comparator drugs are included. An ideal active comparator should be interchangeable with the therapy under investigation and should represent the counterfactual risk associated with the alternative treatment. This requires the active comparator to be indicated for the same disease and disease severity, with comparable exclusion criteria in terms of absolute or relative measures. Moreover, the active comparator should be known to have no impact on the event(s) of interest or competing events. By incorporating an active comparator, the study population is restricted to individuals with an indication for treatment and without contraindications, effectively reducing bias. The new user component ensures

that individuals are aligned at a uniform point in time to initiate treatment, establishing the correct temporality between covariate and exposure assessment.[91,92]

In ACNU design, a specific therapy is compared to a therapeutic alternative, which is often a more meaningful clinical question in accessing the safety or effectiveness.[93] Identifying an appropriate active comparator should involve clinician input and adherence to relevant guidelines. Additionally, it is essential to verify the acceptability of its use within the selected data source and review the balance in patient characteristics.[92]

While observational cohort studies with active comparators have been widely used in research investigating the effectiveness and safety of medications and biologics, there have been fewer studies focusing on vaccines. For example, studies have compared the effectiveness and safety of high-dose versus standard-dose influenza vaccinations.[94–96] However, unlike COVID-19 vaccines, other commonly used vaccines were primarily developed using a single vaccine platform with a standard manufacturing process, making it challenging to identify an appropriate active comparator. The rapid development of COVID-19 vaccines using different platforms, such as viral vector and mRNA, made it possible to conduct the ACNU study. However, residual confounding could still exists. For example, residents and health care workers in care homes were in the first prioritised vaccinate group. However, due to the high demanding cold chain requirements for the BNT162b2 vaccines, while care home staffs were asked to travel to close hospital where the BNT162b2 vaccine can be stored, care home residents preferentially receiving the Chadox1 vaccine. [97]

3.2.2 Propensity score

The propensity score (PS) is the probability of treatment assignment conditional on observed baseline covariates.[98] The propensity score is considered as a balancing score, which means that, by conditioning

on the score, the distribution of measured baseline covariates is similar between exposed and unexposed subjects. Propensity scores have been widely used to reduce confounding by indication in observational studies.[99]

The common procedure of using the propensity score includes the following steps: variable selection, model selection for the propensity score, calculation of the score, use of the propensity score by matching, stratification (or subclassification), inverse probability of treatment weighting (IPTW) using the propensity score, adjust for the score as a covariate in the outcome model, and diagnostics of the covariates' balance.

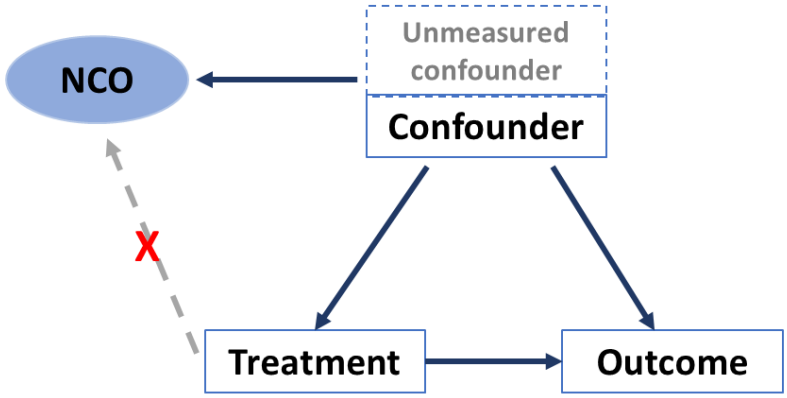
The variables used to build the propensity score can be identified in two ways. Traditionally, the variables in the propensity score are confounders picked based on clinical knowledge.[99] More recently, with the increasing volume and dimension of real-world data and computing power, data-driven approaches have been proposed as a better alternative. One such data-driven approach is large-scale PS, based on regularised regression: all potential confounders are included in a regression model of the outcome of treatment, usually, least absolute shrinkage and selection operator (LASSO) or ridge regression.[100,101] A simulation study by Tian et al. [101] found that this approach was preferable to a traditional propensity score calculated with a predefined set of covariates defined by experts.

3.3 Negative controls and empirical calibration

While some confounders can be addressed by study design or during modelling, there are confounders which are measured or captured with error in real-world data, causing residual confounding bias. Unmeasured confounding occurs when a confounder has not been measured, or when it is not controlled in the analysis. In theory, unmeasured confounding can be adjusted for only through randomisation. When randomisation is not possible, the potential impact of residual confounding on the results should be estimated and considered in the discussion.

In observational studies negative controls (NC) design is a method that aims to address unmeasured confounding, as well as information bias and selection bias.[78,102] It can be used to detect bias, reduce bias, correct bias, calibrate confidence interval or p-value.

Figure 3-1 Causal Directed Acyclic Graph for Negative Control outcome: Addressing Unmeasured Confounding



A negative control outcome (NCO) is a variable that shares the same potential source of bias with the outcome of interest but is not causally related to the exposure of interest. If an effect is observed between the exposure and NCO, it may indicate that unmeasured confounding or an unmeasurable source of bias is influencing the results.[102,103]

There are some key assumptions for NCO. First, there should be no causal effect between the exposure of interest and the NCO (i.e. no causal effect). Second, ideally, the NCO should share the same causal structure as the exposure – outcome association, which means that the NCO should be an outcome such that the set of common causes of exposure and outcome should be as identical as possible to the set of common causes of exposure and the NCO (share common bias structure). [102]

Beyond the detection of bias, negative controls can be used to correct unmeasured confounding through empirical calibration on p-values or confidence intervals.[79,80] By estimating the effect of exposure on outcomes across a collection of settings where the exposure is not believed to cause the NCO, one can estimate an empirical null distribution of the exposure effect and compute calibrated *p*-values and confidence intervals that take both random and systematic error into account.[79] While utilising NCOs may reduce the number of false positive results, the confidence interval may be inflated therefore reduce the ability to detect a true safety or efficacy signal, and is computationally expensive.[104,105]

In a recent scoping review targeting the usage of negative controls in pharmacoepidemiology, the author included 184 studies into the review. Among these, 115 were cohort studies, accounted for 62%, and about half of the studies used multiple negative controls. While 149 out of 184 studies used NCs for bias detection, only 16 studies used them for bias correction in effect estimate, either point estimate and/or calibrating the confidence interval, and 8 studies used NCs for p-value calibration.[106]

3.4 Covariates

3.4.1 Phenotyping and cohort diagnostics

For the phenotyping of both exposures and outcomes, especially events without established or validated definition algorithm, an iterative approach was used. I used the "CohortDiagnostics" R package, which is designed for the development and evaluation of phenotype algorithms for OMOP CDM compliant datasets. (https://ohdsi.github.io/CohortDiagnostics/)

Firstly, a list of appropriate codes (clinical coding in each original database, e.g. ICD-10 and READ codes for clinical conditions, NDC and dm+d codes for medications) was identified through literature review. Then, this list was linked to the standard vocabulary being used within the OMOP CDM (details in Chapter 4: Data source). This includes Standard Nomenclature of Medicine (SNOMED) or Logical Observation

Identifiers Names and Codes (LOINC). The standard vocabularies can be used in heterogenous data sources to identify the same clinical conditions. Initial exposure and outcome cohort definitions were developed by identifying people with records of these codes, together with other inclusion and exclusion criteria based upon demographic, disease and temporal factors, such as including only the first event for each individual, or exclude people with certain conditions prior to the index date.

The "CohortDiagnostics" package firstly generates cohorts of interest in each database based on the initial cohort definitions, and then run several pre-defined diagnostics, including the incidence rates with or without stratification by age, sex and calendar year, characteristics such as comorbidities and medication usage of the cohort members within a timeframe before and after entry of the cohort. This ensured that the cohort definitions accurately reflected the clinical population that I intended to study. It also enabled me to identify orphan concepts that were not initially included in the analysis but had similar names or high record counts in the data. Furthermore, this approach facilitates the identification of nuanced individual source codes mapped to the CDM, allowing us to refine the concept sets within the cohort definitions to better align with the data partners.

Aggregated cohort diagnostics results were exported and presented in an interactive Shiny App, and then reviewed and discussed by epidemiologists, clinicians, and other collaborators. Based on the discussions, the initial cohort definitions will be modified accordingly, and the diagnostics will be re-run, as an iterative process, till refined cohort definitions. In this thesis, all exposure and outcome cohort definitions went through this iterative process and were reviewed by expertise.

The following are examples of the interactive Shiny App of cohort diagnostics for the developed cohort definition used in the thesis.

Exposure of COVID-19 vaccines:

https://dpa-pde-oxford.shinyapps.io/diagCovVaxExposures_CPRD/

Outcomes:

https://livedataoxford.shinyapps.io/CovCoagOutcomesCohorts/

https://dpa-pde-oxford.shinyapps.io/diagCoagOutcome_USdata/

3.4.2 Exposure and outcome

3.4.2.1 Vaccines

Vaccines were identified by procedure, drug, or observation codes in each database identified by concept ids. (Appendix to chapter 3) First, the clinical codes for COVID-19 vaccinations were relatively new in both the source data coding and the standard vocabulary. Second, there were new codes added over time due to the approval of new available vaccines. I excluded codes where the vaccine type/ brand was not specified.

3.4.2.2 COVID-19 infection

SARS-CoV-2 cohort included people with a first positive reverse transcription polymerase chain reaction (RT-PCR) test result or antigen test, identified by SNOMED codes. (Appendix to chapter 3)

3.4.2.3 Events of interest

Thrombocytopenia: Thrombocytopenia was identified by a diagnostic code or measurement of <150 000 platelets per microliter, as proposed by the Brighton Collaboration.[38]

TTS: The definition of thrombosis with thrombocytopenia syndrome (supplementary B) was based on that proposed by the Brighton Collaboration and encompassed the occurrence of any thromboembolic event of interest with concurrent thrombocytopenia within 10 days before or after a thromboembolic event occurring within 28 days after vaccination.

All clinical codes being used to define study outcomes in this thesis were listed in Appendix to chapter 3.

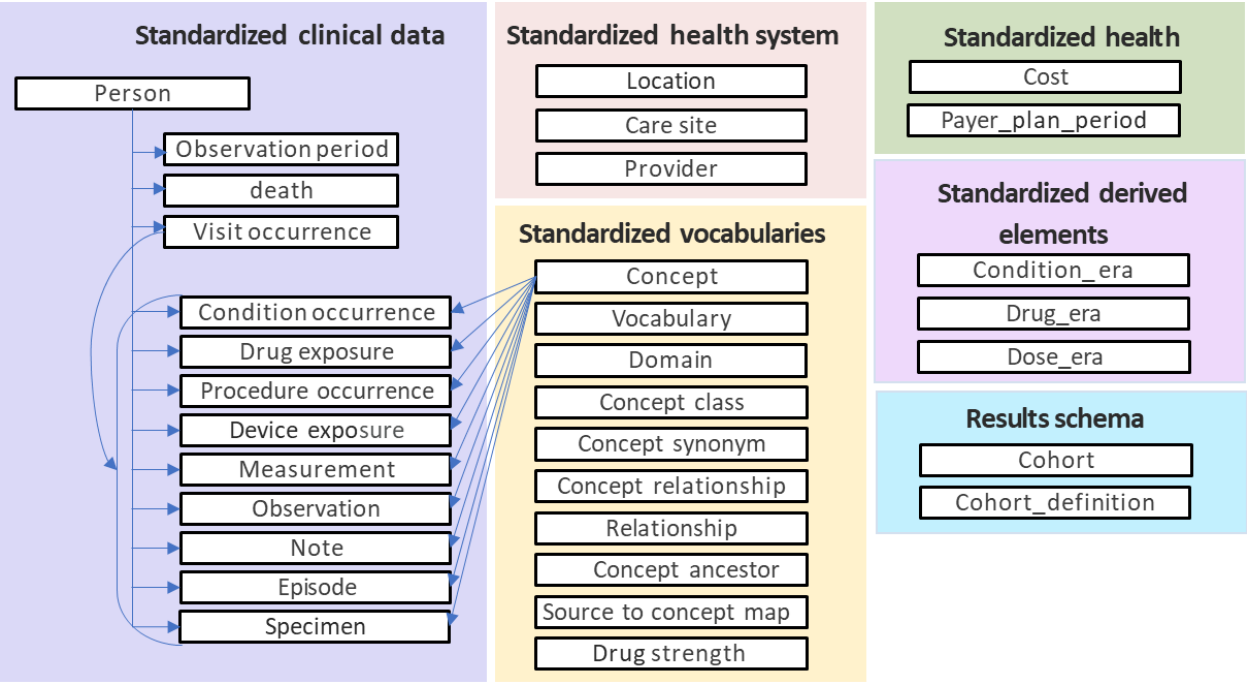
Chapter 4 Data sources

4.1 OMOP CDM and Federated Network analysis

The OMOP CDM is a person-centric database model that accommodates a wide range of data domains commonly found in observational data, which include demographics, visits, condition occurrences, drug exposures, procedures, and laboratory data. (Figure 4-1)

One core element of OMOP CDM is its standardised vocabularies, which allow for the harmonisation of clinical terminologies across different observational data sources. While each data source uses a different coding system, the OHDSI vocabularies allow organisation and standardisation of medical terms to be used across the various clinical domains of the OMOP common data model, therefore allowing for standardised analytics across different databases.

Figure 4-1 The OMOP CDM version 5.4



*Reproduced from The Book of OHDSI, under Creative Commons Zero v1.0 Universal license (https://ohdsi.github.io/CommonDataModel/)

Firstly, each data source is mapped to the OMOP common data model through an extraction, transform, and load (ETL) process. Once mapped, all datasets share the same format, the OMOP CDM, which allows standardised analytics to be conducted at each site without the need of sharing patient level records. This is called federated or distributed network analyses.

This approach ensures that research methods can be systematically applied to produce meaningfully comparable and reproducible results. (Figure 4-2) Such federated analyses also have the advantage of protecting the privacy of patient-level data by executing standardised analytic code or software in the local data environment, with only aggregated output shared across collaborating centres. Furthermore, this approach enables principal investigators to synthesise results from multiple data sources, gain a larger number of patients and draw conclusions with sufficient statistical power.

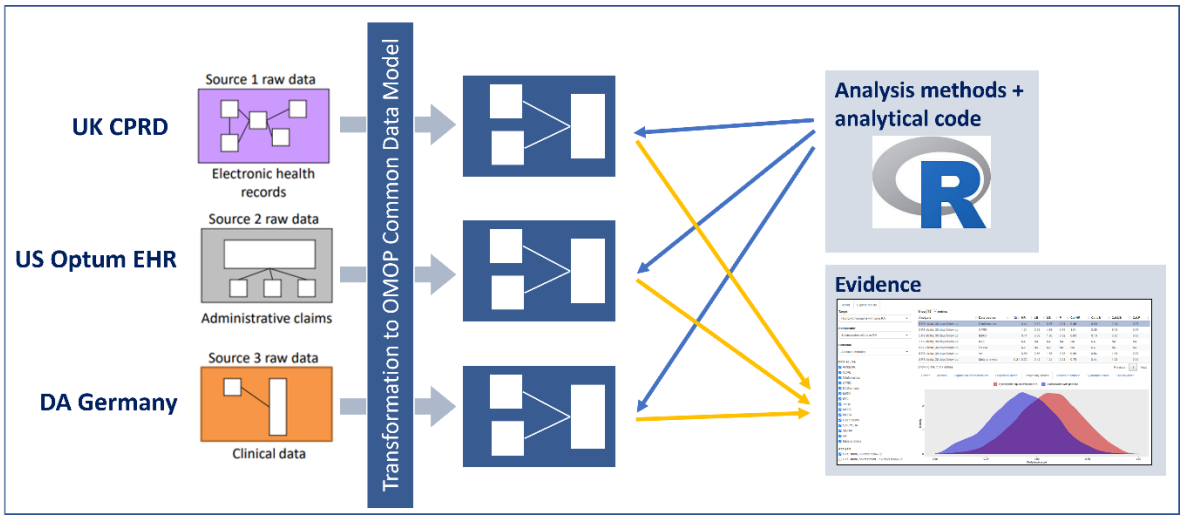


Figure 4-2 Common data model and federated network analyses

By utilising federated network analyses within the OHDSI network, researchers can harness the distinct advantage of externally validating local data analyses across a diverse range of healthcare systems within an international community. This approach provides a unique opportunity to benefit from the expertise

of individuals outside of my own institute, encompassing a wide array of perspectives from clinical, computer science, statistics, and epidemiology domains. The interdisciplinary collaboration improves research correctness, facilitates a deeper understanding of the analytical results and encourages meaningful discussions.

Furthermore, the transparency and reproducibility of research are enhanced through the availability of analytic packages and study codes, which are typically made publicly accessible. This allows other members within the network to review the code, ensuring accuracy of the research. Embracing such collaborative and open practices within the OHDSI community not only strengthens the validity of analyses but also fosters a culture of shared knowledge and continuous improvement.

Although federated network approaches offer advantages, they also come with drawbacks. One limitation is that the process of transforming data to the CDM is very time consuming and also may lead to a loss of granularity for certain conditions. Another limitation is the inability to identify duplicate patients across datasets, which may occur in situations where patients switch GP practices or insurers, or when there is an overlap between regional and national claims records. Administratively, in some situations, coordinating and communicating with numerous researchers across different regions and countries can be challenging in terms of time and effort.

4.2 Used datasets

In this Dphil thesis, all used datasets were mapped to the OMOP CDM. Here I provide a description of the datasets analysed for the studies that I will describe and report on in subsequent chapters, and discuss their respective strengths and limitations. All these datasets have been previously used in other peer-

reviewed studies, including COVID-19 related work.[107–110] A summary of all studied datasets is available in Table 4-1.

Table 4-1 Data sources used in the thesis

Data source	**Short name**	**Country**	**Type**	**Used in**		
				Chapter 5	Chapter 6	Chapter 7
Asia-pacific						
IQVIA Australia Electronic Medical Records (Australia EMR)	IQVIA_AUSTRALIA	Australia	GP records	Y		
Japan Medical Data Center (JMDC)	JMDC_JAPAN	Japan	Claims (insured general population 18-65y)	Y		
Europe						
IQVIA Longitudinal Patient Data France (LPD France)	IQVIA_FRANCE	France	GP records	Y		Y
IQVIA Disease Analyser Germany (DA Germany)	IQVIA_GERMANY	Germany	GP records	Y		Y
Integrated Primary Care Information (IPCI)	IPCI_NETHERLANDS	The Netherlands	GP records	Y		Y
Clinical Practice Research Datalink (CPRD) GOLD	CPRD_GOLD_UK	United Kingdom	GP records	Y		
Clinical Practice Research Datalink (CPRD) Aurum	CPRD_Aurum_UK	United Kingdom	GP records		Y	Y
Information System for Research in Primary Care – Hospitalisation Linked Data (SIDIAP-H)	SIDIAP_H_SPAIN	Spain	GP records linked to hospital care	Y	Y	Y
North America						
Optum® De-Identified Clinformatics® Data Mart Database – Socio-Economic Status (Optum SES)	OPTUM_SES_US	United States	Claims (commercially insured <65y)	Y		
Optum® de-identified Electronic Health Record Dataset (Optum I)	OPTUM_EHR_US	United StatIEHR (general population)		Y		

IBM MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR)	MDCR_US	United States	Claims (commercially insured >=65y)	Y		
Data source	**Short name**	**Country**	**Type**	**Used in**		
				Chapter 5	Chapter 6	Chapter 7
IBM MarketScan Multi-State Medicaid Database (MDCD)	MDCD_US	United States	Claims (Medicaid enrolees)	Y		
IBM MarketScan Commercial Claims and Encounters Database (CCAE)	CCAE_US	United States	Claims (commercially insured <65y)	Y		
Columbia University Irving Medical Center (CUIMC)	CUMC_US	United SIes	EHR (regional hospital)	Y		
IQVIA Medical and Institutional Claims (US Open Claims)	OPEN_Claims_US	United States	Claims (commercially insured)			Y
IQVIA Charge Data Master (US Hospital CDM)	Hospital_CDM_US	UniteItates	EHR (general population)			Y

Figure 4-3 Geographic countries/regions of data sources used.

4.2.1 European Datasets

4.2.1.1 UK: Clinical Practice Research Datalink (CPRD)

Two UK datasets, the Clinical Practice Research Datalink (CPRD) GOLD and Aurum, were used.

In the UK, the National Health System (NHS) provides universal access to comprehensive health services that are free at the point of use and financed by general taxation. Primary and secondary care are offered free of charge through general practitioners (GPs), with nearly the entire population registered to a local GP practice. GPs serve as the primary point of contact for the UK NHS, offering free consultations to registered patients and providing subsequent non-emergency healthcare, prescriptions, diagnostic tests, or referrals to secondary care when necessary. Following consultations, GPs are required to record visit

information, which may include patient symptoms, test requests/results, diagnoses, clinical measurements (such as BMI and smoking status), prescriptions, and vaccinations. This information started been recorded in the UK in digital format in the late eighties, providing a self-contained longitudinal medical record for each patient.[111]

The Clinical Practice Research Datalink (CPRD, https://cprd.com/) is a governmental, non-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency, part of the Department of Health of the United Kingdom. CPRD consists of anonymised electronic health records data collected from UK primary care practices. Since the patient electronic health records from GPs were collected using two different software systems, which have differences in structure and clinical coding, they are provided as two separate databases: CPRD GOLD [CPRD_GOLD_UK] and CPRD Aurum [CPRD_Aurum_UK].[112,113] These databases include symptoms, conditions, observations, referrals, measurements, and procedures that the GP is made aware of in addition to any prescriptions recorded. They also contain the patients' year of birth and (where applicable) a reliable record of date of death.[114,115]

In CPRD_GOLD_UK, which comes from practices using Vision® software, medical events are reported by Read coding.[116] On the other hand CPRD_Aurum_UK contains data from practices using EMIS Web® software, where clinical observations are recorded using a mixture of SNOMED and local EMIS® codes. In the OMOP CDM instance used for my research, GOLD and AURUM events were both mapped to SNOMED codes.[117]

As of 2022, there are over 3 million active individuals registered with CPRD GOLD, which represents 4.5% of the UK population. There are over 13 million current patients in CPRD Aurum, which account for over 20% of the UK population. Nonetheless, it is possible that the practices incorporated in CPRD do not accurately reflect the geographic and size-related characteristics of practices throughout the UK.[112]

CPRD GOLD was used in chapter 5, and CPRD Aurum was used for analyses in chapters 6 and 7.

4.2.1.2 Spain: Information System for Research in Primary Care – Hospitalisation Linked Data (SIDIAP-H)[SIDIAP_H_SPAIN]

Similar to the UK, in Catalonia, north-east Spain, healthcare is universal and taxpayer-funded and therefore primary care physicians serve as gatekeepers for all care and are responsible for repeat prescriptions.

The Sistema d'Informació pel Desenvolupament de la Investigació a l'Atenció Primària (The Information System for Research in Primary Care, SIDIAP; www.sidiap.org) is a plary care EHR database from Catalonia. SIDIAP includes data from over 300 primary care centres managed by the Catalan Health Institute in Catalonia, with records data for over 8 million people since 2006, with 5.8 million people active in June 2021. SIDIAP covers a representative 80% of the population of the Catalan region.[118,119] The database includes patient-level information on demographics, socioeconomic indicators, all-cause mortality, disease diagnoses, medication prescription and dispensation, routine measurements, laboratory results, therapeutic procedures including vaccinations, and lifestyle information as well.

The SIDIAP-H subset of the database includes around 2 million people who are registered in primary care practices with linked hospital data available as obtained from the Catalan Institute of Health hospitals.[120] Information on diagnoses and medical procedures at hospital, discharges and hospital medication for outpatient dispensation is available through the linkage.

The SIDIAP database was used in chapters 5, 6, and 7.

4.2.1.3 France: IQVIA Longitudinal Patient Data France (LPD France)[IQVIA_FRANCE]

The IQVIA Longitudinal Patient Data France is a anonymised electronic health record system in France that collects clinical records from patient management software utilized by general practitioners (GPs) and selected specialists during clinic visits. [121] In France, patients have the freedom to choose their

preferred physician when a medical need arises, although they are required to visit a GP before seeing a specialist in order to receive reimbursement, with the exception of paediatricians, gynaecologists, and ophthalmologists. Currently, the database includes contributions from over 1,200 GPs practicing at 400 different clinics, encompassing a total of 7.8 million patients across France. The database encompasses data spanning from 1994 up to the current date, with the observation period being defined by the initial and final consultation dates.

The IQVIA_FRANCE database was used in chapters 5 and 7.

4.2.1.4 Germany: IQVIA Disease Analyser Germany (DA Germany) [IQVIA_DA_GERMANY]

The IQVIA Germany Disease Analyzer Database comprises anonymised electronic medical records obtained from both general and specialized medical practices in Germany. These records encompass a wide range of demographic information, diagnoses, and prescription data pertaining to patients. It contains more than 34 million distinct current person records, which accounts for approximately 42.5% of the total population of 80 million in Germany. The data have been collected from 2,734 healthcare providers since 1992. Thorough assessme’ts of the data's validity and representativeness demonstrate its suitability for conducting pharmaco-epidemiological and pharmaco-economic studies. [122,123]

The IQVIA_DA_GERMANY database was used in the analyses in chapters 5 and 7.

4.2.1.5 The Netherlands: Integrated Primary Care Information (IPCI) [IPCI_NETHERLANDS]

The Integrated Primary Care Information (IPCI) database, accessible at https://www.ipci.nl/, is a comprehensive longitudinal observational database that houses the electronic health records of a carefully selected cohort of Dutch general practitioners. [124] In the Netherlands, every citizen is registered with a GP practice, which serves as a primary point of contact and facilitates a two-way flow of information with secondary care providers.

The IPCI database currently includes data from over 2.5 million patients, sourced from more than 600 GPs situated across various regions of the Netherlands. The median follow-up duration is 4.8 years. The database includes a diverse range of patient information, such as demographics, GP encounters, reported symptoms, diagnoses, laboratory and clinical measurements, prescription details, and records of secondary care utilisation.

The IPCI_NETHERLANDS database was used in the analyses in chapters 5 and 7.

4.2.2 Asia-Pacific data

4.2.2.1 Japan: Japan Medical Data Center (JMDC) [JMDC_JAPAN]

The Japan Medical Data Center (JMDC) database consists of data derived from 60 society-managed health insurance plans in Japan. These plans cover individuals aged 18 to 65, as well as their dependents, including children below 18 years of age and elderly individuals above 65 years of age. However, it should be noted that the representation of older people, particularly those aged 66 or older, may be less comprehensive compared to the entire population of the nation.

The JMDC database comprises two primary types of data: membership status information of the insured individuals and claims data provided by insurers with whom JMDC has contracts. Claims data are obtained from monthly claims issued by clinics, hospitals, and community pharmacies. The JMDC data includes a variety of patient-level demographic information, encompassing inpatient and outpatient data that consist of diagnoses and procedures, as well as information on prescriptions in the form of dispensed claims. The JMDC included about 9 million insured persons with approximately 6 million active persons.

The JMDC database was used in chapter 5.

4.2.2.2 Australia: IQVIA Australia Electronic Medical Records (Australia EMR) [IQVIA_AUSTRALIA]

The IQVIA_AUSTRALIA database comprises anonymized electronic health record system in Australia and gathers clinical records from patient management software utilised by general practitioners during clinic visits. These patient records are sourced from two distinct origins: longitudinal patient data and practice profiles, which are integrated into a unified data source. The data encompasses a time span from 2006 and information on over 3 million patients.

The IQVIA_AUSTRALIA database was used in chapter 5.

4.2.3 United States datasets

4.2.3.1 Optum® de-identified Electronic Health Record Dataset (Optum EHR) [OPTUM_EHR_US]

The OPTUM_EHR_US database is an anonymised electronic health record derived from numerous healthcare provider organisations in the United States, encompassing a broad spectrum of medical facilities. These facilities include over 700 hospitals and 7,000 clinics. This database also includes data from multi-specialty practices, small group practices, physician offices, and integrated delivery networks, which collectively serve a patient population exceeding 103 million individuals. The database contains demographic information, along with comprehensive clinical information, such as prescriptions, lab results, vital signs, body measurements, diagnoses, procedures, and data extracted from clinical notes through Natural Language Processing.

The OPTUM_EHR_US database was used in chapter 5.

4.2.3.2 Optum® De-Identified Clinformatics® Data Mart Database – Socio-Economic Status (Optum SES) [OPTUM_SES_US]

Optum® De-Identified Clinformatics® Data Mart Database (OPTUM_SES_US) is an adjudicated administrative health claims database for members with private health insurance, who are fully insured in commercial plans or Medicare Advantage. The database primarily represents individuals within the US

commercial claims patient population, ranging from 0 to 65 years old, with some inclusion of Medicare beneficiaries aged above 65. However, the age range is capped at 90 years. It includes data captured from administrative claims processed from inpatient and outpatient medical services and medications. Additionally, it incorporates outpatient lab test results obtained from major national laboratory vendors who engage in data exchange with Optum. This database also provides socio-economic status information for members possessing both medical and pharmacy coverage, alongside patient location data at the US Census Division level.

The OPTUM_SES_US dataset was used in chapter 5.

4.2.3.3 IBM MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR) [MDCR_US]

The IBM MarketScan Medicare Supplemental and Coordination of Benefits (COB) Database includes Medicare-eligible active and retired employees and their Medicare eligible dependents from employer-sponsored supplemental plans. The database contains adjudicated health insurance claims, including inpatient, outpatient, and outpatient pharmacy claims. Additionally, laboratory test results are available for a subset of people in the database.

This dataset was used in chapter 5.

4.2.3.4 IBM MarketScan Multi-State Medicaid Database (MDCD)[MDCD_US]

The IBM MarketScan Multi-State Medicaid Database (MDCD_US) contains adjudicated health insurance claims for Medicaid enrollees from multiple states in the United States. The database includes information on hospital discharge diagnoses, outpatient diagnoses and procedures, outpatient pharmacy claims, as well as information on long-term care. Data on eligibility, service and provider type are also included. In addition to standard demographic variables such as age and gender, the database includes variables of particular value to researchers investigating Medicaid populations, such as federal aid category (income

based, disability, Temporary Assistance for Needy Families) and race. The data does not contain laboratory results.

The MDCD_US dataset was used in chapter 5.

4.2.3.5 IBM MarketScan Commercial Claims and Encounters Database (CCAE) [CCAE_US]

The IBM® MarketScan® Research Databases (CCAE_US) are one of the largest proprietary collections of anonymised US patient data available for healthcare research, and includes patients with commercial insurance aged younger than 65 years old. It covered more than 264 million individuals, 37 billion service records, more than 160 contributing employers and 40 contributing private health plans.

The data include adjudicated health insurance claims covering various aspects such as inpatient, outpatient, and outpatient pharmacy services. Additionally, the database incorporates enrolment data obtained from large employers and health plans that offer private healthcare coverage to employees, their spouses, and dependents. Furthermore, laboratory test results are captured for a subset of the covered population. The IBM MarketScan Research Databases provide a comprehensive representation of diverse healthcare plans, including fee-for-service, preferred provider organisations, and Health Maintenance Organization.

The CCAE_US dataset was used in chapter 5.

4.2.3.6 Columbia University Irving Medical Center Clinical Data Warehouse r (CUIMC) [CUMC_US]

The Clinical Data Warehouse (CDW) contains longitudinal health records derived from patients receiving treatment at both Columbia University Irving Medical Center (CUIMC) and New York-Presbyterian Hospital. CDW comprises data sourced from the institutions' current and previous electronic health record systems, resulting in a comprehensive database spanning over 30 years. This database provides a wealth of clinical information for more than 6.4 million patients. It includes a wide range of data elements, such as demographic information, physician orders, laboratory results, radiology, pathology and other

pertinent reports. Furthermore, CDW comprises both structured and free-text clinical documents, as well as nurse flowsheets and task lists, offering a comprehensive overview of patients' medical records. (https://www.irvinginstitute.columbia.edu/services/trials-and-research-opportunities)

This dataset was used in chapter 5.

4.2.3.7 IQVIA Open Claims [OPEN _CLAIMS_US]

The IQVIA Open Claims database (OPEN_CLAIMS_US) includes claims of anonymised patients collected from office-based physicians and specialists, obtained through office management software and clearinghouse switch sources primarily for reimbursement purposes. The dataset includes both medical and institutional claims, encompassing a vast population of approximately 191 million patients across the entirety of the United States. The dataset has been steadily growing since 2006, accumulating a robust historical record. Claims include patient level diagnosis, procedures, and in-office treatments for visits to U.S. office-based professionals, ambulatory and general healthcare sites, skilled nursing facilities, and other institutions for outpatient and inpatient procedures and services.

The OPEN_CLAIMS_US dataset was used in chapter 7.

4.2.3.8 IQVIA hospital charge data master [HOSPITAL_CDM_US]

The IQVIA hospital charge data master (US Hospital CDM) database is a dataset comprised of records from US hospital charge data master files and records both inpatient and outpatient encounters. [125] The Hospital CDM information is collected from a panel of up to 400 non-federal hospitals and provides unique insights into what happens and how patients are treated during hospital visits. It tracks patients of all pay types with history beginning in 2009, and includes information on procedure or diagnosis codes, anonymised patient demographics, location of care, length of stay, and related hospital or care information.

US Hospital CDM dataset was used in chapter 7.

4.2.4 Ethical approval

Scientific approval for the research in this thesis was given by the CPRD Independent Scientific Advisory Committee (ISAC) for Medicine and Healthcare products Regulatory Agency database research (ISAC 20_000211, ISAC 21_000641), the IDIAPJGol (Fundacio Institut Universitari per a la recerca a l'Atencio Primaria de Salut Jordi Gol i Gurina) clinical research ethics committee (project code: 21/007-PCV, 21/052-PCV), the Integrated Primary Care Information governance board (application number 3/2021), and Columbia University institutional review board (AAAO7805). Informed consent of individual patients was not required as anonymised information was obtained from medical records.

4.3 Discussion on datasets

One key aspect of utilising routinely collected health data for distributed analyses is to select the appropriate databases to answer specific research questions. The "Guidelines for good database selection and use in pharmacoepidemiology research", endorsed by the International Society of Pharmacoepidemiology (ISPE)[126], highlight potential limitations of secondary data containing routinely collected healthcare information, such as EHRs (from either primary or secondary care) and claims databases, and recommend procedures for data analysis and interpretation.

The limitations and recommendations covered multiple perspective in RWE research, include completeness of data capture, bias in the assessment of exposure, outcome, and covariates, variability between data sources and the impact of changes in data over time (as seen in the pre- vs. post-COVID-19 period), and the healthcare system of the country or region covered by the study.

For both administrative claims and EHRs, information regarding medications is derived from GP prescriptions or dispensation, while drugs acquired by patients over-the-counter, are not documented.

For all commercial claims data in the US (Optum_SES_US, MarketScan_CCAE_US, and Open_Claims_US), one limitation is the relatively short continuous enrolment period, which may lead to incompleteness in patient history, as well as follow-up coverage. While requiring a longer visible time in the dataset can address this, as a compromise, the included study population may shrink, and the representativeness will be impacted as well. The enrolment period issue may be emphasised after the COVID-19 pandemic. In current standards of epidemiological studies, disenrollment is usually treated as loss-to-follow-up, where patients were censored at the time of lost enrolment. However, due to the impact of the pandemic, there was an increase in the loss of employment, which led to a loss in insurance coverage and possible missed outcomes. The risk of disenrollment is also different by patient and socioeconomic characteristics. Selecting the study population during the pandemic based on the enrolment period may lead to selection bias. [127]

Furthermore, while claim-based sources provide relatively complete data on inpatient, outpatient, and prescriptions and treatments, they lack measurement data and laboratory results. Other limitations of the US-based claims data include a lack of information on death, and the act that evidence generated from the data sources may not be generalizable to populations without health insurance. Misclassification of medication or treatment usage exists in the Open_Claims_US dataset because the claims are unadjudicated. Since CUMC_US is a regional electronic health record acquired from a local medical network, it might incompletely capture medical events that are recorded in other healthcare institutions.

The CPRD_GOLD_UK, IQVIA_DA_GERMANY, IQVIA_FRANCE, IQVIA_AUSTRALIA, and IPCI_Netherlands databases include primary care data, with patient-level linkage to hospitalisation data on hospital admissions. Therefore, inpatient information, including the treatments or procedures given in the hospital, could not be recorded.

The ability to capture vaccination exposure is especially important for COVID-19 vaccine studies. In countries with a universal healthcare system, including the UK and Spain, vaccines were and are delivered through the national immunisation programme. The data infrastructure of national and/or regional central health providers (NHS in the United Kingdom and Catalan Health Institute in Catalonia) facilitated the capture of COVID-19 and vaccination related events. The longstanding structure of GPs serving as the central node in healthcare service also enabled relatively complete COVID-19 vaccination records.

Those two data sources are also linked to the regional vaccination registry (SIDIAP) or central vaccination database (CPRD Aurum). In other databases, COVID-19 vaccines are captured incompletely. In the United States, for example, patients can receive vaccines outside of the insurance system and thus may not be covered by claims.[128] Identifying non-users of vaccines in US claims databases would therefore be impossible.

Table 4-2 Completeness of key data component for datasets used in comparative safety study.

Data source	**Short name**	**Key data component***			
		COVID-19 vaccines	Hospital treatments	Hospital outcomes	Outpatient treatments
IQVIA Longitudinal Patient Data France (LPD France)	IQVIA_FRANCE	Incomplete	No	Incomplete	Yes
IQVIA Disease Analyser Germany (DA Germany)	IQVIA_GERMANY	Incomplete	No	Incomplete	Yes
Integrated Primary Care Information (IPCI)	IPCI_NETHERLANDS	Incomplete	No	Incomplete	Yes
Clinical Practice Research Datalink (CPRD) Aurum	CPRD_Aurum_UK	Complete	No	Incomplete	Yes
Information System for Research in Primary Care – Hospitalisation Linked Data (SIDIAP-H)	SIDIAP_H_SPAIN	Complete	No	Linked	Yes

IQVIA Medical and Institutional Claims (US Open Claims)	OPEN_Claims_US	Incomplete	Incomplete	Incomplete	Yes
IQVIA Charge Data Master (US Hospital CDM)	Hospital_CDM_US	Incomplete	Yes	Yes	Incomplete

Chapter 5 Background incidence rates of adverse events of special interests

5.1 Chapter summary

In this chapter, I used 13 electronic health records and health claims databases from eight countries to estimate the background rates of 15 adverse events of special interests. (Objective 1) I found considerable heterogeneity both within-database and across-database on the estimated rates, and suggested within database comparisons as best practice in vaccine safety surveillance studies where background incidence rates are/should be used.

5.2 Background

In vaccine safety surveillance, the observed-to-expected analysis is commonly used to identify safety signals and to assess the relationship between vaccine exposure and an event, especially when new vaccines are launched and administered at a large scale.[129,130] As explained in chapter 3, this method largely relies on correct estimation of the background rates, which provide a baseline comparison for observed rates among the vaccine-eligible population.[39,130]

Based on the historical precedent set by prior vaccines and knowledge accumulated during the development of new vaccines, each new vaccine has potential adverse events of special interest (AESI) that need special focus during the post-authorisation monitoring. While identifying potential sources of background rates from the literature is often not too difficult, caution must be taken to make appropriate comparisons between rates. The background rate may be impacted by the year of measurement, age and sex distribution of the population, geographic location, co-morbidities, socioeconomic status, medication use, and study methodology. Therefore, Black et al. described two key steps for post- authorisation safety

evaluation of COVID-19 vaccines: the first is defining “a dynamic list of Adverse Events of Special Interest”, and the second is “establishing background rates for these AESI”.[131]

To better prepare for safety monitoring of COVID-19 vaccines, one of the objectives of my thesis was to estimate and characterise the background incidence rates of COVID-19 vaccine AESI. This study was designed and executed as an international open science study with observational data from Australia, France, Germany, Japan, the Netherlands, Spain, the United Kingdom, and the United States.

5.3 Methods

5.3.1 Data sources

A total of 13 databases from 8 countries participated in this study, 8 of which containing EHRs and 5 containing administrative claims.

The EHRs were: 1. IQVIA Australia Electronic Medical Records (IQVIA_AUSTRALIA); 2. Integrated Primary Care Information (IPCI_NETHERLANDS) from the Netherlands; 3. IQVIA Longitudinal Patient Data France (IQVIA_FRANCE); 4. IQVIA Disease Analyser Germany (IQVIA_GERMANY); 5. Information System for Research in Primary Care (SIDIAP_H_SPAIN), from Catalonia, Spain; 6. Clinical Practice Research Datalink GOLD from the UK (CPRD_GOLD_UK); 7. Columbia University Irving Medical Center (CUMC_US)from the US; and 8. Optum® de-identified Electronic Health Record Dataset (OPTUM_EHR_US) also from the US.

The claims-based databases were the Japan Medical Data Center (JMDC_JAPAN) and four US administrative claims databases: IBM MarketScan Commercial Claims and Encounters Database (CCAE_US), IBM MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR_US), IBM MarketScan Multi-State Medicaid Database (MDCD_US), Optum® De-Identified Clinformatics® Data Mart Database – Socio-Economic Status (OPTUM_SES_US).

The detailed description of each database has been discussed in Chapter Four. All the study databases were previously mapped to OMOP CDM.

5.3.2 Study period and population

The study period was set from 1 January 2017 to 31 December 2019. The study period was chosen to represent the pre-covid time in each participating database. No record earlier than 2017 was included in the study period to minimize the change in trends in clinical practice and coding, and to better represent the "unvaccinated" comparator population.

The study population was made by those who were observed on 1 January 2017, 2018, or 2019 and with at least 365 days visibility in the data prior to that. This was to make sure that an incident event can be captured in the database. The index date was January 1 of each calendar year, or the date after the clean window (as illustrated in Figure 5-1).

5.3.3 Events of interest

The events of interest in this study were AESIs that had been identified for COVID-19 safety assessment. A list of outcomes was chosen based on the FDA Center for Biologics Evaluation and Research's protocol, the Brighton Collaboration's prioritised COVID-19 vaccine AESI list, and previous research as well. [132,133] Fifteen event types were included: non-haemorrhagic and haemorrhagic stroke, acute myocardial infarction, deep vein thrombosis, pulmonary embolism, anaphylaxis, Bell's palsy, myocarditis or pericarditis, narcolepsy, appendicitis, immune thrombocytopenia, disseminated intravascular coagulation, encephalomyelitis (including acute disseminated encephalomyelitis), Guillain-Barré syndrome, and transverse myelitis.

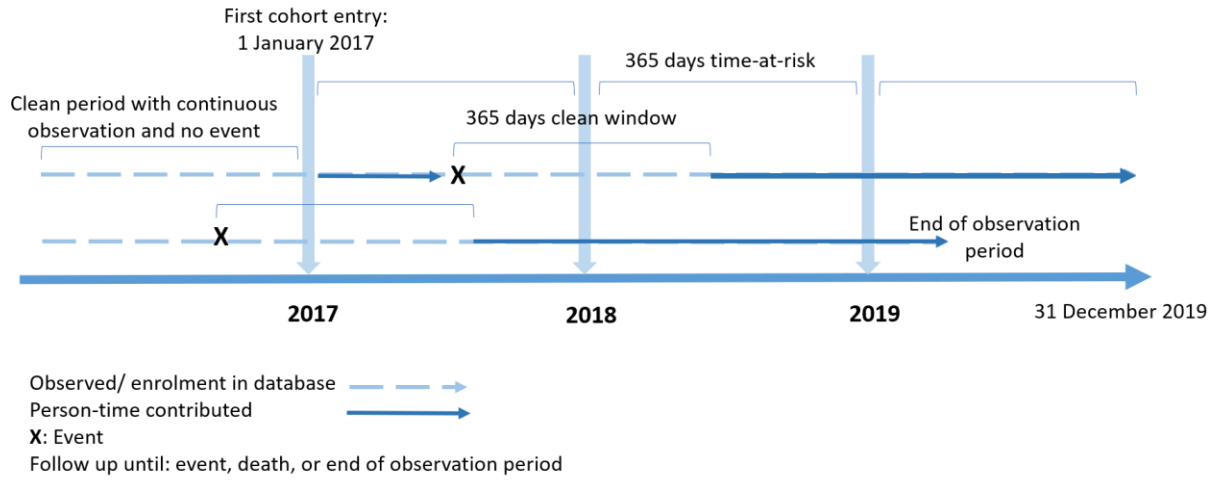
The events of interest were identified using SNOMED codes, which is the standard vocabulary in OMOP CDM mapped from diagnosis codes of the source data. The clinical codes included were based on previous

literatures. Encephalomyelitis, non-haemorrhagic and haemorrhagic stroke, and acute myocardial infarction all required hospital records in all diagnoses, whereas GBS required inpatient records in the primary code.[134–137] The SNOMED codes are listed in Appendix tables. As the CPRD_GOLD_UK, IQVIA_GERMANY, IQVIA_FRANCE, IQVIA_AUSTRALIA, and IPCI_Netherlands databases only included primary care data, they were not used for events whose definition required an inpatient diagnosis.

5.3.4 Statistical analysis

Incidence rates were estimated as the total number of incident events across the three yearly cohorts divided by the person-time-at-risk per 100,000 person years across the three yearly cohorts. Time-at-risk was defined as a 365-day period following the index date or until the end of observation if the person left the database during that 365-day period. People contributed time-at-risk from 1 January to 31 December for each qualifying year in 2017 to 2019.

Figure 5-1 Study design



One person could contribute with more than one event, with outcome-specific pre-specified clean window to avoid duplicate counts. A “clean window” was defined as the period before each index date

where qualifying events (AESIs) could not be observed. If an AESI was observed during this period, the participant did not enter the study cohort for that event. Figure 5-1 illustrates how two example cases contributed to three yearly cohorts with a clean window of 365 days for all events except anaphylaxis (30 days), facial nerve palsy and encephalomyelitis (183 days).[132] During the follow-up, if the participant had a qualified event, he/she would be temporary censored from that cohort, and re-join to contribute person-time once the clean window requirement was met.

Age and sex specific incidence rates were calculated and reported for each database separately when the event counts exceeded a minimum of 5. I categorized age into eight mutually exclusive age groups: 1-5, 6-17, 18-34, 35-54, 55-64, 65-74, 75-84, and 85 and older.

Firstly, the incidence rates were estimated in each database. Then, for each AESI, a pooled incidence rate across datasets for each age-sex group was estimated using random effects meta-analysis. The DerSimonian-Laird approach was used to estimate variance between databases. [138] The 95% prediction intervals were calculated, which represented 95% range of true incidence rates to be expected in similar studies using the "meta" R package. [139] The prediction interval represents the predicted uncertainty if a new study's estimation rate would be included.[140]

In order to have a standard illustration of the frequency of the AESI, the WHO Council for International Organisations of Medical Sciences thresholds were used to categorise meta-analytic age and sex specific rates: very common (≥1/10), common (<1/10 to ≥1/100), uncommon (<1/100 to ≥1/1,000), rare (<1/1,000 to ≥1/10,000), and very rare (<1/10,000). [141]

5.4 Results

5.4.1 Database characteristics

In this study, a total of 126,661,070 persons contributed 227,043,370 person years of follow-up from the 13 databases included. Optum's De-Identified Electronic Health Record Database (OPTUM_EHR_US) provided the majority of people (n=40,955,085), followed by IBM MarketScan's Commercial Claims and Encounters Database (CCAE_US) with over 25.5 million people. Among European data sources, the largest population was contributed by IQVIA Disease Analyser Germany, where 9.3 million people were included. The second largest cohort was from the UK Clinical Practice Research Datalink GOLD, including over 4.5 million people. The IQVIA Australia Electronic Medical Records data contributed least people, with 250,212 individuals included. Demographic data including age and sex groupings for study participants are shown in Table 5-1.

Except for JMDC_JAPAN, (45.0% female) most databases featured more females than males (range from 50.5% female in SIDIAP_H_SPAIN to 57.5% female in IQVIA_GERMANY). The CCAE US database had persons aged 0 to 74 years, whereas the MDCR_US database contained only persons aged 65 years or older. Individuals of all age groups were included in the other databases. In most databases, persons aged 35-54 years accounted for the largest proportion of the population (ranging from 23.5% in OPEUM_SES_US to 35.8% in JMDC_JAPAN). Persons aged 18-34 years, however, accounted for the largest proportion (22.3%) of the IBM MarketScan Multi-State Medicaid Database (MDCD_US) database. The proportion of persons aged 65 years and older ranged from 32.1% in the OPTUM_SES_US database to less than 10% in the CCAE_US, JMDC_JAPAN, and MDCD_US databases. Individuals younger than 18 years accounted for 45.8% of the MDCD_US database.

Table 5-1 Characteristics of included populations, stratified by database.

	CCAE_US	MDCD_US	MDCR_US	OPTUM_EHR_US	OPTUM_SES_US	CUMC_US	CPRD_GOLD_UK	IPCI_NETHERLANDS	SIDIAP_H_SPAIN	IQVIA_FRANCE	IQVIA_GERMANY	IQVIA_AUSTRALIA	JMDC_JAPAN
Full name	IBM MarketScan Commercial claims and encounters Database	IBM MarketScan Multi-State Medicaid Database	IBM MarketScan Medicare Supplemental and Coordination of Benefits Database	Optum deidentified Electronic Health Record Data	Optum De-Identified Clinformatics Data Mart Database–Socio-Economic Status	Columbia University Irving Medical Center	Clinical Practice Research Datalink	Integrated Primary Care Information	Information System for Research in Primary Care-Hospitalisation Linked Data	IQVIA Longitudinal Patient Data France	IQVIA Disease Analyser Germany	IQVIA Australia Electronic Medical Records	Japan Medical Data Center
Data type	Claims	Claims	Claims	EHR	Claims	EHR	EHR	EHR	EHR	EHR	EHR	EHR	Claims
Country	US	US	US	US	US	US	UK	Netherlands	Spain	France	Germany	Australia	Japan
Total No of patients	25,315,777	12,966,011	1,533,709	40,955,085	18,643,608	1,164,196	4,532,766	1,536,283	2,217,536	1,746,371	9,295,525	252,212	6,501,991
Person years	42,889,550	23,203,712	2,484,782	72,328,897	32,474,685	2,174,312	9,638,136	3,326,570	5,497,613	3,008,350	16,784,613	383,668	12,848,482
Age group (years):													
1-5	1,256,501 (4.96)	1,755,796 (13.54)	0	1,852,425 (4.52)	627,032 (3.36)	40,678 (3.49)	245,525 (5.42)	78,848 (5.13)	99,838 (4.5)	99,309 (5.69)	308,728 (3.32)	13,430 (5.32)	414,167 (6.37)
6-17	4,122,110 (16.28)	4,188,247 (32.3)	0	4,773,000 (11.65)	1,930,638 (10.36)	105,520 (9.06)	635,115 (14.01)	211,037 (13.74)	260,102 (11.73)	268,591 (15.38)	823,235 (8.86)	31,780 (12.6)	1,044,041 (16.06)
18-34	6,395,387 (25.26)	2,885,991 (22.26)	0	8,182,549 (19.98)	3,331,356 (17.87)	199,020 (17.1)	946,153 (20.87)	304,971 (19.85)	374,994 (16.91)	328,759 (18.83)	1,411,620 (15.19)	50,995 (20.22)	1,533,866 (23.59)
35-54	8,096,864 (31.98)	2,006,493 (15.48)	0	10,737,664 (26.22)	4,389,220 (23.54)	300,818 (25.84)	1,217,618 (26.86)	394,868 (25.7)	663,537 (29.92)	446,804 (25.58)	2,338,535 (25.16)	69,872 (27.7)	2,330,010 (35.84)

	CCAE_US	MDCD_US	MDCR_US	OPTUM_EHR_US	OPTUM_SES_US	CUMC_US	CPRD_GOLD_UK	IPCI_NETHERLANDS	SIDIAP_H_SPAIN	IQVIA_FRANCE	IQVIA_GERMANY	IQVIA_AUSTRALIA	JMDC_JAPAN
55-64	4,716,207 (18.63)	1,004,957 (7.75)	0	6,655,199 (16.25)	2,384,571 (12.79)	183,612 (15.77)	594,115 (13.11)	219,990 (14.32)	288,494 (13.01)	229,016 (13.11)	1,580,565 (17)	36,329 (14.4)	880,065 (13.54)
65-74	728,708 (2.88)	633,262 (4.88)	733,157 (47.8)	4,829,968 (11.79)	3,106,611 (16.66)	171,940 (14.77)	469,682 (10.36)	180,581 (11.75)	246,763 (11.13)	197,816 (11.33)	1,279,048 (13.76)	27,272 (10.81)	279,277 (4.3)
75-84	0	341,267 (2.63)	536,970 (35.01)	2,652,453 (6.48)	1,985,356 (10.65)	110,883 (9.52)	290,225 (6.4)	104,288 (6.79)	180,903 (8.16)	117,067 (6.7)	1,191,402 (12.82)	15,319 (6.07)	20,565 (0.32)
≥85	0	149,998 (1.16)	263,582 (17.19)	1,271,827 (3.11)	888,824 (4.77)	51,725 (4.44)	134,333 (2.96)	41,700 (2.71)	102,905 (4.64)	59,009 (3.38)	362,392 (3.9)	7215 (2.86)	0
Sex:													
Female	13,037,440 (51.5)	7,322,471 (56.47)	849,301 (55.38)	23,220,748 (56.7)	9,595,675 (51.47)	693,190 (59.54)	2,287,698 (50.47)	783,660 (51.01)	1,120,373 (50.52)	926,180 (53.03)	5,340,273 (57.45)	137,203 (54.4)	2,926,702 (45.01)
Male	12,278,337 (48.5)	5,643,540 (43.53)	684,408 (44.62)	17,734,337 (43.3)	9,047,933 (48.53)	471,006 (40.46)	2,245,068 (49.53)	752,623 (48.99)	1,097,163 (49.48)	820,191 (46.97)	3,955,252 (42.55)	115,009 (45.6)	3,575,289 (54.99)

5.4.2 Within database heterogeneity (individual level)

Across the 15 AESI analysed within all data sources, background rates varied considerably across demographic strata. There was individual level heterogeneity of incidence rates in the database-specific estimates of each event. While substantial variation by age and sex was observed, the age and sex patterns were similar in the majority of databases and pooled rates (Figure 5-2,Figure 5-3).

For example, deep vein thrombosis incidence rates increased with age. The incidence rates for men aged 6-17 increased from 20 (95%CI 19 to 22) per 100,000 person years to 2030 (95%CI 2009 to 2051) per 100,000 person years among 75-84 years old group in the OPTUM_SES_US data (

Figure 5-2). With increasing age, the incidence of haemorrhagic and non-haemorrhagic stroke, pulmonary embolism, Bell's palsy, immune thrombocytopenia, GBS, and disseminated intravascular coagulation also increased. For appendicitis, the incidence rates peaked at adolescent and young adults, then decreased by age. Incidence of anaphylaxis showed a decreased trend with age as well.

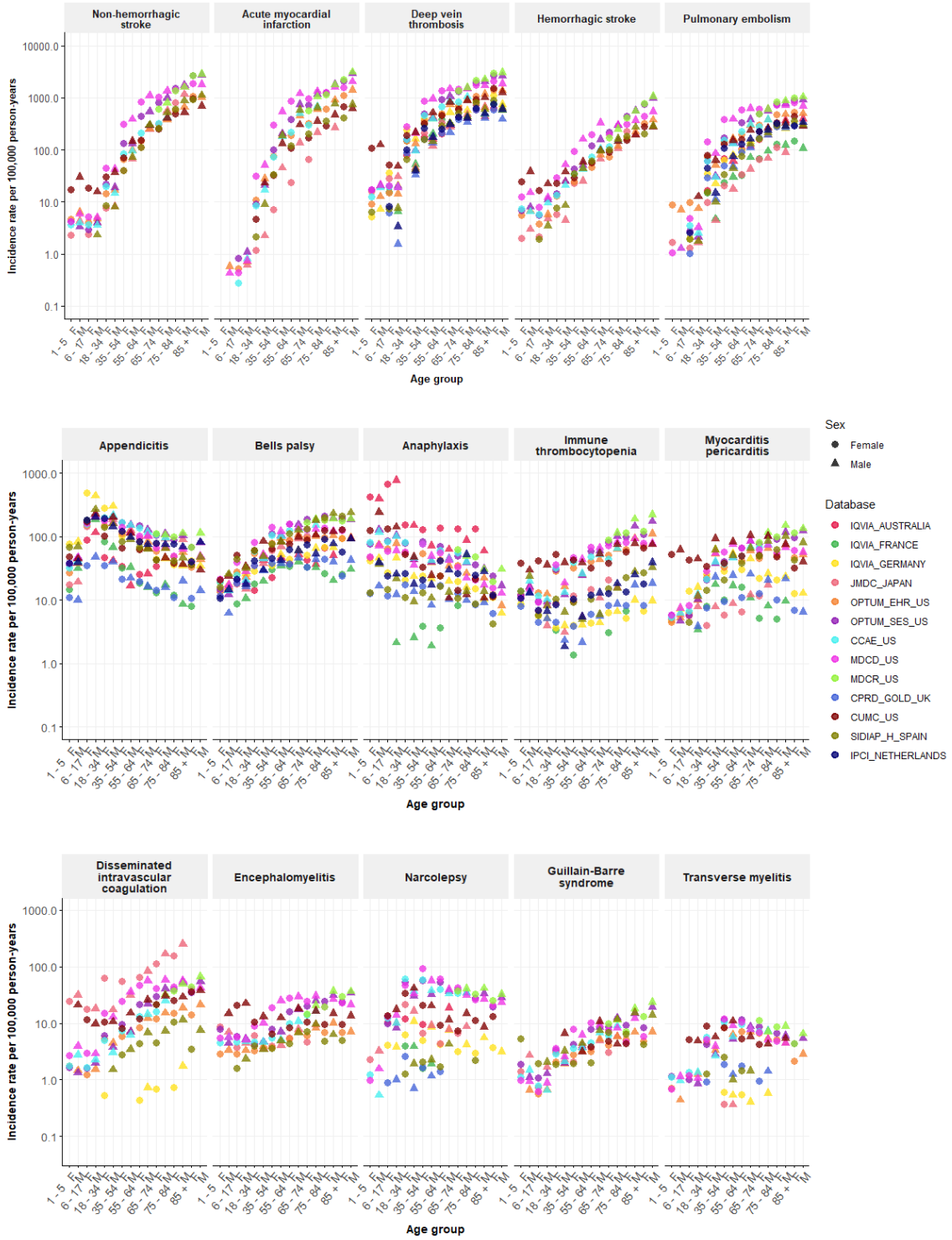
Higher incidence rates were observed in male than female for cardiovascular events in most age groups including stroke, myocardial infarction, and thromboembolism events. For example, the incidence rates among those aged 65-74 years in the MDCR_US database were 251 (95%CI 238 to 265) per 100,000 person years for male participants and 170 (95%CI 160 to 180) per 100,000 person years for female participants.

5.4.3 Heterogeneity between databases (population level)

Other than the individual level heterogeneity in age and sex, I also observed database level heterogeneity in the incidence rate for a given age-sex group. The rates of deep vein thrombosis, which were estimated in all thirteen databases, highlight population variance. Taking the 65-74 years old women group as an example, the rates ranged from 387 (95% CI: 370-404) per 100,000 person years in CPRD GOLD to 1,443 (1,416-1,470) in MDCD US. Eight databases reported rates of less than 650 per 100,000 person years (CPRD GOLD in the United Kingdom; CUIMC in the United States; IPCI in the Netherlands; IQVIA in Australia, France, and Germany; JMDC in Japan; and SIDIAP-H in Spain), whereas three databases reported rates of more than 1,300 per 100,000 person years (MDCD, MDCR, and OPTUM-SES in the US). For women aged 35-54 years, the incidence rates ranged from 159 (95% CI: 151 to 167) per 100,000 person years in Spain (SIDIAP) to 866 (95% CI: 854 to 878) per 100,000 person years in the US (MDCD). Among women aged 75-84 years, the lowest incidence rate was 585 (95% CI: 559 to 612) per 100,000 person years (CPRD GOLD in the UK) and the highest was 2,167 (95% CI: 2,126 to 2,210) per 100,000 person years (MDCR in the US).

Disseminated intravascular coagulation were identified in 8 databases, and I observed the variance in the incidence rates as well. For example, in the 65 – 74 years female group, a lowest incidence rate of 0.69 (95% CI: 0.32 - 1.32) per 100,000 person years was observed in the IQVIA_DA_GERMANY data, while the JMDC_JAPAN showed rate of 111.58 (95% CI: 97.23 - 127.46) per 100,000 person years. However, for disseminated intravascular coagulation, the IQVIA_GERMANY data always showed a lowest age-sex specific rate, and the highest rate was in the JMDC_JAPAN data for most age-sex groups. The rates for each database are depicted in Figure 5-2, and the full result data are available at the interactive web app (https://data.ohdsi.org/Covid19VaccineAesiIncidenceCharacterisation/).

Figure 5-2 Age and sex stratified incidence rates for 15 adverse events of special interest by database.



5.4.4 Pooled incidence rates across databases

The pooled incidence rates with prediction intervals of the 15 AESIs, stratified by age and sex, are summarised in Figure 5-3. The pooled incidence rates were then categorised based on the CIOMS definition, ranging from dark green (very rare, <1/10,000) to red (very common, ≥1/10) to better illustrate the pattern. I observed similar age and sex pattern of each event compared with the database-specific estimation. For example, acute myocardial infarction was very rare (<1/10,000) in women younger than 35 years, rare (<1/1000 to ≥1/10,000) in women 35-54 years, uncommon (<1/100 to ≥1/1000) in men and women aged 55-84 years, and common (<1/10 to ≥1/100) in men and women aged 85 years and older.

For non-haemorrhagic stroke, the pooled incidence rates were numerically higher in male among all age groups, but were classified into different CIOMS frequency group. Among female aged 35-54 years, the rate was classified as “rare” (pooled IR = 83, 95% prediction interval 11 to 67) per 100,000 person-years, and was classified as “uncommon” for male of the same age (pooled IR =119, 95% prediction interval 21 to 664). In the 75 – 84 years old group, pooled IR in female fell into the “uncommon” group, and “common” for male. Bell’s palsy, appendicitis, and immune thrombocytopenia were largely rare in all age groups. Guillain-Barré syndrome and transverse myelitis were very rare in nearly all subgroups.

Figure 5-3 Pooled estimated age and sex stratified incidence rates per 100,000 person years (95% prediction intervals), calculated from meta-analyses

	Incidence rate per 100,000 person to years (95% prediction interval)							
Outcome by sex	1 to 5	6 to 17	18 to 34	35 to 54	55 to 64	65 to 74	75 to 84	85+
Non-hemorrhagic stroke								
Female	4 (2 to 9)	4 (1 to 12)	18 (4 to 86)	83 (11 to 617)	217 (25 to 1882)	413 (77 to 2198)	874 (197 to 3884)	1523 (320 to 7239)
Male	6 (2 to 20)	5 (2 to 10)	17 (4 to 75)	119 (21 to 664)	370 (67 to 2046)	612 (145 to 2578)	1063 (242 to 4662)	1495 (260 to 8607)
Acute myocardial infarction								
Female	<1 (<1 to 1)	<1 (<1 to 1)	6 (1 to 49)	54 (7 to 430)	171 (24 to 1235)	312 (76 to 1280)	617 (184 to 2069)	1144 (313 to 4184)
Male	<1 (<1 to 1)	1 (1 to 1)	16 (4 to 72)	172 (40 to 740)	467 (135 to 1611)	653 (214 to 1994)	934 (290 to 3013)	1514 (356 to 6432)
Deep vein thrombosis								
Female	12 (3 to 50)	18 (8 to 40)	140 (66 to 298)	306 (117 to 797)	428 (150 to 1224)	683 (257 to 1820)	975 (360 to 2642)	1206 (407 to 3572)
Male	14 (4 to 55)	14 (6 to 32)	80 (28 to 228)	272 (88 to 836)	499 (194 to 1289)	695 (250 to 1931)	831 (254 to 2720)	1003 (278 to 3616)
Hemorrhagic stroke								
Female	7 (2 to 28)	5 (2 to 16)	13 (4 to 47)	36 (7 to 175)	77 (15 to 389)	124 (29 to 527)	249 (56 to 1108)	412 (85 to 1986)
Male	8 (2 to 43)	8 (3 to 24)	19 (5 to 76)	51 (10 to 268)	115 (23 to 562)	178 (49 to 650)	312 (73 to 1340)	506 (86 to 2961)
Pulmonary embolism								
Female	1 (<1 to 36)	3 (1 to 13)	38 (11 to 124)	81 (21 to 309)	125 (33 to 470)	217 (77 to 611)	358 (135 to 951)	427 (154 to 1184)
Male	1 (<1 to 24)	2 (<1 to 12)	20 (5 to 80)	80 (20 to 318)	171 (59 to 497)	256 (96 to 683)	349 (119 to 1030)	398 (124 to 1277)
Appendicitis								
Female	32 (12 to 84)	154 (55 to 430)	134 (69 to 260)	85 (42 to 172)	66 (28 to 156)	53 (20 to 143)	40 (13 to 124)	35 (12 to 98)
Male	38 (17 to 85)	194 (101 to 372)	146 (81 to 266)	88 (49 to 159)	65 (32 to 132)	57 (23 to 144)	47 (15 to 152)	45 (14 to 143)
Bell's palsy								
Female	15 (9 to 27)	25 (12 to 51)	44 (23 to 84)	61 (26 to 140)	76 (31 to 184)	86 (29 to 256)	101 (31 to 330)	92 (31 to 274)
Male	15 (10 to 24)	21 (13 to 34)	43 (29 to 64)	68 (37 to 125)	86 (43 to 172)	94 (35 to 252)	92 (29 to 291)	100 (34 to 292)
Anaphylaxis								
Female	49 (16 to 150)	50 (16 to 154)	39 (16 to 95)	34 (13 to 91)	35 (14 to 85)	29 (11 to 76)	23 (7 to 73)	12 (4 to 36)
Male	74 (26 to 209)	56 (18 to 175)	29 (14 to 63)	24 (11 to 53)	25 (11 to 53)	24 (9 to 68)	18 (7 to 49)	10 (2 to 50)
Immune thrombocytopenia								
Female	12 (8 to 19)	9 (4 to 21)	14 (6 to 36)	15 (5 to 43)	18 (6 to 53)	25 (8 to 82)	30 (8 to 110)	36 (11 to 118)
Male	17 (12 to 23)	8 (3 to 19)	8 (2 to 23)	10 (3 to 35)	19 (6 to 57)	30 (9 to 105)	41 (10 to 170)	56 (15 to 210)
Myocarditis pericarditis								
Female	6 (1 to 25)	7 (2 to 21)	16 (8 to 32)	22 (9 to 53)	31 (13 to 72)	35 (12 to 97)	39 (11 to 138)	34 (8 to 143)
Male	7 (1 to 32)	11 (5 to 24)	37 (16 to 88)	37 (16 to 87)	45 (20 to 102)	49 (17 to 139)	54 (15 to 193)	41 (9 to 193)
Disseminated intravascular coagulation								
Female	2 (<1 to 104)	2 (<1 to 48)	4 (<1 to 99)	5 (<1 to 75)	10 (1 to 89)	14 (2 to 97)	19 (4 to 94)	16 (3 to 82)
Male	3 (<1 to 137)	2 (<1 to 44)	4 (<1 to 31)	5 (1 to 56)	12 (1 to 120)	17 (2 to 154)	23 (4 to 152)	24 (5 to 126)
Encephalomyelitis								
Female	5 (2 to 15)	5 (2 to 16)	5 (2 to 19)	6 (1 to 44)	9 (1 to 61)	11 (2 to 62)	12 (2 to 77)	14 (2 to 100)
Male	5 (2 to 12)	5 (2 to 14)	5 (2 to 17)	7 (1 to 55)	12 (3 to 58)	16 (3 to 73)	18 (3 to 101)	16 (1 to 180)
Narcolepsy								
Female	1 (<1 to 5)	7 (3 to 17)	15 (4 to 52)	11 (2 to 55)	9 (2 to 42)	10 (2 to 46)	8 (1 to 49)	9 (2 to 42)
Male	1 (<1 to 5)	6 (2 to 18)	13 (4 to 40)	10 (2 to 47)	11 (3 to 44)	10 (2 to 50)	10 (2 to 68)	10 (2 to 60)
Guillain-Barre syndrome								
Female	1 (<1 to 8)	1 (<1 to 2)	3 (1 to 5)	3 (1 to 11)	5 (1 to 18)	6 (2 to 19)	6 (3 to 16)	7 (2 to 22)
Male	2 (<1 to 18)	1 (<1 to 3)	2 (1 to 4)	4 (2 to 7)	7 (4 to 14)	8 (3 to 25)	11 (3 to 40)	12 (2 to 68)
Transverse myelitis								
Female	1 (<1 to 3)	1 (<1 to 3)	3 (1 to 8)	4 (1 to 12)	4 (2 to 13)	4 (2 to 13)	4 (1 to 11)	2 (1 to 9)
Male	1 (<1 to 2)	1 (<1 to 3)	2 (1 to 6)	3 (1 to 10)	4 (1 to 10)	4 (1 to 11)	4 (1 to 13)	4 (1 to 11)

Council for International Organizations of Medical Sciences Frequency classification

Very rare: <1/10,000	Rare: ≥1/10,000 to <1/1,000	Uncommon: ≥1/1,000 to <1/100	Common: ≥1/100 to <1/10	Very common: ≥1/10

5.5 Discussion

5.5.1 Summary of study

In this chapter, I conducted a multinational network cohort study and assessed the incidence rates of 15 AESIs prioritised for post-marketing surveillance of COVID-19 vaccines. Specifically, I included deep vein thrombosis, pulmonary embolism, stroke, immune thrombocytopenia, myocarditis and pericarditis, on which safety concerns later raised during the global massive immunisation program around the world.[142–145]

I observed considerable variability with age and sex, emphasising the need for standardisation or stratification of the background rates used for vaccine surveillance. Notably, I found substantial heterogeneity between databases, suggesting that observed rates among COVID-19 vaccinees should be compared with background rates obtained from the same database where possible.

5.5.2 Research in context

With the advantage of improving statistical power for detecting rare events and enabling the detection of signals earlier than other methods,[130,146] the observed-to-expected method has been widely implemented in many countries and by many organisations, including the Vaccine Safety Datalink project in the US[146] and the Vaccine Adverse Event Surveillance and Communication project in Europe.[130] These background rates are often obtained from the literature or healthcare databases. However, there are limitations due to the methods used to estimate rates, case and outcome definitions, variations in clinical codes, and geographical and temporal variations. [146] While in this study, I presented estimates across a distributed network using a common data model, the same study design, standardised analysis procedures and phenotyping algorithms.

The majority of the studied outcomes had considerable within-source individual level heterogeneity that followed age and sex patterns. With increasing age, the prevalence of cardiovascular diseases, such as myocardial infarction, haemorrhagic stroke, deep vein thrombosis, and pulmonary embolism,

increased. The prevalence of GBS and Bell's palsy increased with age. Conversely, narcolepsy and appendicitis were more common in younger people. The patterns found in this study were generally comparable to those found in previous studies. [39,130,147–151] When comparing incidence rates across populations, age and sex stratification, as well as standardisation, are likely to be beneficial analytical strategies to reduce confounding. However, the extent of variation reported across sources within age and sex subgroups suggests that residual individual level differences will remain, including the differences in other risk factors' distributions such as comorbidities and medication utilisation.

Even after standardising outcome definitions, and age and sex stratification, there was a significant population-level heterogeneity across data sources for all events. For example, a threefold difference between the highest and lowest incidence rates for deep vein thrombosis measured in each database was observed. Previous studies using one or a small number of databases have also observed these variations. US based studies, for example, recorded incidence rates per 100,000 person years for idiopathic thrombocytopenia of around 3 among men and women aged 26-62 years,[151] 9 among those aged 25-44 years, and 12 among those aged 45-64 years.[148] In this study, the estimated incidence rates of transverse myelitis ranging from 1 to 4 per 100,000 person years in meta-analyses, depend on age and sex strata. Previous studies have reported overall incidence rates of transverse myelitis ranging from 0.4 to 4.6 per 100,000 person years.[39,152] However, direct comparison of published results can be challenging due to variations in study methods such as the definition of time-at-risk, study period, event definitions, population coverage, calendar year, and geographical location.[131] Additionally, different subgroup definitions can further complicate comparisons. For instance, in previous studies of Guillain-Barré syndrome, age strata were defined differently, leading to incomplete overlap between them.[39,147,151,153,154]

My study utilised consistent event definitions, common data models, and analyses across all studied databases. Therefore, any heterogeneity observed cannot be attributed to differences in analytical methods. This remaining heterogeneity may be attributable to differences in underlying

demographics, healthcare systems, and data collection techniques. While some variability may be due to systematic errors, selection bias, or database-specific differences, other factors such as socioeconomic status and comorbidities may also play a role in producing genuine demographic differences.

Given the observed differences in incidence rates by age, sex and database, caution should be exercised when comparing incidence rates across populations or over time. Incidence rates obtained from various sources might be subject of substantial systematic errors, thereby limiting the interpretability of the reported 95% confidence intervals for database-specific rates in this study. The notably wide prediction intervals for each age and sex subgroups also reflect the significant population-level heterogeneity observed across sources. Furthermore, I observed considerable variations in rates between electronic health records and claims data sources, despite using identical analysis and outcome definitions. The variability in rates from randomised trials or spontaneous reporting data could be even greater.

The WHO Council for International Organisations of Medical Sciences (CIOMS) guidance indicated that the most valid data for comparison in a particular area are the background rates from the local population.[41] The heterogeneity observed in this study also emphasize the importance of using local population. By the time of the planning and conducting of this project (manuscript first submitted in April 2021), most previous studies were outcome-specific, used one database, and none focused on COVID-19 AESI. By the time of writing the thesis, there are more studies on the background rates of COVID-19 AESIs available from different countries and/or regions with diverse data sources, including Sweden [155], the US [156,157], Australia [158], Canada[159], and the Europe[160].

Despite the differences in study design, case definition, and underlying population, the background rates reported here line up well with other published reports, with similar age and sex patterns observed.

5.5.3 Strength and limitations

Because of the large number of collaborative databases, geographic coverage, and research population, global comparisons of AESI background incidence rates across healthcare systems and geographies were possible. This study is an example of the collaborative projects, which become possible within the Observational Health Data Sciences and Informatics network. This study took advantage of the OMOP common data model and standard vocabularies as explained in chapter 3, which enabled me to use the same study design and analytical code in all databases and to gather results from participating data partners rapidly and without anybody transferring patient level data.[161] To maximise reproducibility and reuse, all outcome definitions, clinical codes, and phenotype algorithms are open source and available online.

As an observational study, the primary limitation of this study is that all events could have been subject to measurement errors. The outcome definitions were based on the presence of specific diagnostic codes and were not validated further, which could result in imperfect sensitivity or specificity. Additionally, this study used data from 2017 to 2019, a target population of individuals from each database with more than 365 days of observation indexed on 1 January, 365 days of time at risk, and outcome-specific clean windows that allowed for recurrent events. Further exploration is required to assess the impact of these design decisions.

In this study, I pooled the estimates across all databases using mixed-effects meta-analysis. In contrast to fixed-effect meta-analysis, which assumes a true estimate underlies all the studies and weights each study by the inverse variance of its estimate, the mixed-effect model assumes that the true effect could vary due to the heterogeneity among studies. The pooled estimate from the mixed effect model would be the mean or average effect, and the weight of each study is not predominantly decided by study size. [162] However, in some of the study outcomes, such as narcolepsy and DIC, the large heterogeneity between studies suggested that the pooled estimate could be close to the arithmetic mean rather than reflecting the underlying incidences. This study will benefit from identifying the

factors associated with heterogeneity and only including databases with low ascertainment bias in the meta-analysis.

Some limitations relate to the particularities of each individual database. This study included 13 data sources from 8 countries, which contained various type of data (EHRs and claims) that are either representative of the national population (e.g. CPRD in the UK, IPCI in the Netherlands), or have a regional scope (e.g. SIDIAP in Catalonia, Spain; CUMC in New York, the US). Electronic health records may include data collected exclusively at the primary care or hospital level, while others are collected or linked from both hospital and GP data. As information on hospital admissions was not available in most primary care datasets used (CPRD GOLD in the UK, IQVIA in France, Germany, and Australia, and IPCI in the Netherlands), events that happened during inpatient visits may not be competently recorded. The EHRs data sources were subject to incomplete capture of medical events recorded in other healthcare institutions. The bias of incomplete information was partially mitigated by including only those patients who had at least one year of continuous observation. Although the five administrative claims data sources offered reliable data capture, they lacked certain data elements, such as laboratory test results. Additionally, the US-based claims database had poor recording of death information. Nevertheless, comparing background rates within each database minimized bias related to these limitations, thereby mitigating their potential impact.

Chapter 6 COVID-19 vaccination and immune-mediated neurological events

6.1 Chapter summary

In this chapter, I assessed the association between COVID-19 vaccines, SARS-CoV-2 infection, and the risk of 4 immune mediated neurological events. I used data from the UK and Spain, and found no increased risk after vaccinations but after SARS-CoV-2 infection.

6.2 Background

At the preparedness stage of COVID-19 vaccines safety surveillance, several immune-mediated neurological disorders were recognized as adverse events of special interest by regulatory agencies such as the FDA in the US, EMA in Europe, and the Brighton Collaboration in the UK.[37,163,164]

Rare cases of severe neurological disorders were reported during the initial clinical trials of COVID-19 vaccines.[2,3,165–169] Cases of Guillain-Barré syndrome (GBS), a rare condition in which a person's immune system attacks the peripheral nerves, were reported during the clinical trial of the adenovirus-vectored vaccine (Ad.26-COV2.S), in both the placebo and active trial arms.[170] While it is important to note that these events may not have been directly caused by COVID-19 vaccines, the temporal association between the events and vaccination called for robust post-vaccination surveillance.

During immunisation campaigns, the British, European and US regulators issued warnings based on spontaneous reports of GBS following vaccination with adenovirus-vectored vaccines in mid-2021. The Food and Drug Administration in the US issued a warning on the risk of rare but serious cases of GBS after Ad26.COV2.S vaccine. The EMA had listed Guillain-Barré syndrome as a rare side effect associated with Ad26.COV2.S and ChAdOx1 nCoV-19 based on reported adverse events. Additionally,

a few cases of Guillain-Barré syndrome were associated with mRNA vaccines in cohort studies from Mexico and Israel.[171,172]

In the trials of the BNT162b2 and mRNA-1273 vaccines, there were indications of a possible association with Bell's palsy, with seven cases in the active arms compared to one case in the placebo arms. [173–178] Furthermore, subsequent reports described additional cases of Bell's palsy, encephalomyelitis, and transverse myelitis following COVID-19 vaccination using both viral vector and mRNA vaccines. [179–183] A disproportionality analysis on the USA Vaccine Adverse Event Reporting System, showed an increased risk of facial nerve palsy following COVID-19 mRNA vaccines both for BNT162b2 and mRNA-1273, which was similar to the risk associated with influenza vaccination.[184] However, this study is limited by the nature of self-reported passive surveillance, that events may be underreported, and no denominator value is available.

While a causal relationship had not been established, the recommendation for continued monitoring has led to the publication of numerous case reports worldwide. In this study, I aimed to investigate the potential association between COVID-19 vaccination and the risk of developing Guillain-Barré syndrome, Bell's palsy, encephalomyelitis, and transverse myelitis. Additionally, I examined the associations between SARS-CoV-2 infection and the risk of these immune-mediated neurological outcomes to provide counterfactual contextualisation for the findings.

6.3 Methods

6.3.1 Data sources

This study used two data sources: the Clinical Practice Research Datalink (CPRD) AURUM from the UK, and the Information System for Research in Primary Care (SIDIAP) from Spain. Both are routinely collected electronic health records and have been mapped to the OMOP CDM.

The SIDIAP data were linked at the individual-level to hospital inpatient data. These hospital data included information from all public and private hospitals in Catalonia (*Conjunt Mínim Bàsic de Dades d'Alta Hospitalària*, CMBD-AH).[21]

Details of both databases were given in Chapter 4.

6.3.2 Study participants

The populations of interest were individuals who had received at least one dose of a COVID-19 vaccine and people infected with SARS-CoV-2.

Individuals vaccinated were required to have received their first dose between the start of the vaccination campaign in each country (8 December 2020 in the UK and 27 December 2020 in Spain) and one week before the end of data availability of each database (9 May 2021 for CPRD AURUM and 30 June 2021 for SIDIAP), with the vaccination date used as index date.

Vaccination cohorts were constructed of people vaccinated according to the product (ChAdOx1, BNT162b2, mRNA-1273, or Ad.26.COV2.S) and the dose (first or second dose, with only a single dose cohort identified for Ad.26.COV2.S vaccine, which has a single dose regimen) administered. Ad26.COV2.S and mRNA-1273 cohorts were only available in SIDIAP.

For the second doses cohorts, I excluded persons who received a different brand of first dose COVID-19 vaccine. For the second dose cohorts, participants were required to have received their second dose in pre-specified intervals after first dose administration. For both databases, the interval allowed between doses of two-dose vaccines (all except Ad26.COV2.S) was 14 to 180 days. The SARS-Cov-2-infected cohort included people with a first positive RT-PCR or antigen test between 1st September 2020 and one week before the end of data availability of each database, with the test date used as index date. Data from both RT-PCR and antigen tests were available in SIDIAP, whereas only RT-PCR tests were available in CPRD AURUM. I excluded from this cohort individuals vaccinated against COVID-19 before the SARS-CoV-2 infection.

A background population cohort, which included all individuals registered in CPRD AURUM and SIDIAP as of the 1 January 2017 (index date), was also identified. All study cohorts were illustrated in Figure 6-1.

All study participants were required to be aged 18 years or older and to have at least 365 days of data availability before their index date. For each specific outcome, individuals with an occurrence of that outcome during the year prior to the index date were excluded.

For each cohort, individuals were followed from the index date until the earliest of the following: first occurrence of event of interest, 21 days (for people vaccinated) or 90 days (for those diagnosed with COVID-19) after index date, end of data availability, or until they were transferred out of the database or died. For first-dose vaccination cohorts, follow-up was also censored one day before the second dose, if a second dose was observed prior to 21 days. For the background general population cohort, follow-up was censored on 31 December 2019.

6.3.3 Events of interest

The events of interest were four immune-mediated neurological disorders pre-specified as potential adverse events of special interest for COVID-19 vaccine safety: Bell's palsy, encephalomyelitis, Guillain Barre Syndrome (including Miller Fisher syndrome) and transverse myelitis. Only the first outcome event during the study period was included. Details of the SNOMED codes used to define the outcomes are available in Appendix to chapter 3, table 3.

6.3.4 Risk window

The choice of risk window was defined to reflect a time period whereby events might be most likely to be related to the receipt of the vaccine. The 21-day window after vaccination was also set to accommodate the data from Spain, where most recipients of the BNT162b2 had their second dose 21 days after their first dose (unlike in the UK where second doses of BNT162b2 were administered 8 to

12 weeks after the first dose). Therefore, the 21-day period was used in both databases following both the first and second doses of vaccines to increase consistency.

For the SARS-CoV-2-infected cohort, a 90-day risk window was used, to capture events that could plausibly be associated with the exposure of interest. This difference in time periods was accounted for in the calculation of person-time used for calculating incidence rates. A sensitivity analysis was run considering an exposure window of 21 days for SARS-CoV-2 infected cohorts to place the same time at risk for vaccinated and SARS-CoV-2 infected cohorts.

6.3.5 Study design

At the time when this study was planned, there was only safety signals of the event of interest from case reports/series and spontaneous report system. As discussed in chapter 3, the observed-to-expected analysis is usually used for strengthening the signal that previously raised from passive surveillance, and the SCCS method for hypothesis testing. Therefore, I applied both designs in the presented study.

6.3.5.1 Observed-to-expected analysis

First, I used the observed-to-expected analysis method (Figure 6-1). Incidence rates of each outcome in the vaccinated cohorts and SARS-CoV-2-infected cohort were used as “observed” rates and compared with the “expected” background incidence rates estimated from the general population cohort. For the vaccinated cohorts, I estimated the rates during the 1 to 21 days after the first-dose vaccination (day 0). A 90-day post-test period was used for the SARS-Cov-2-infected cohort.

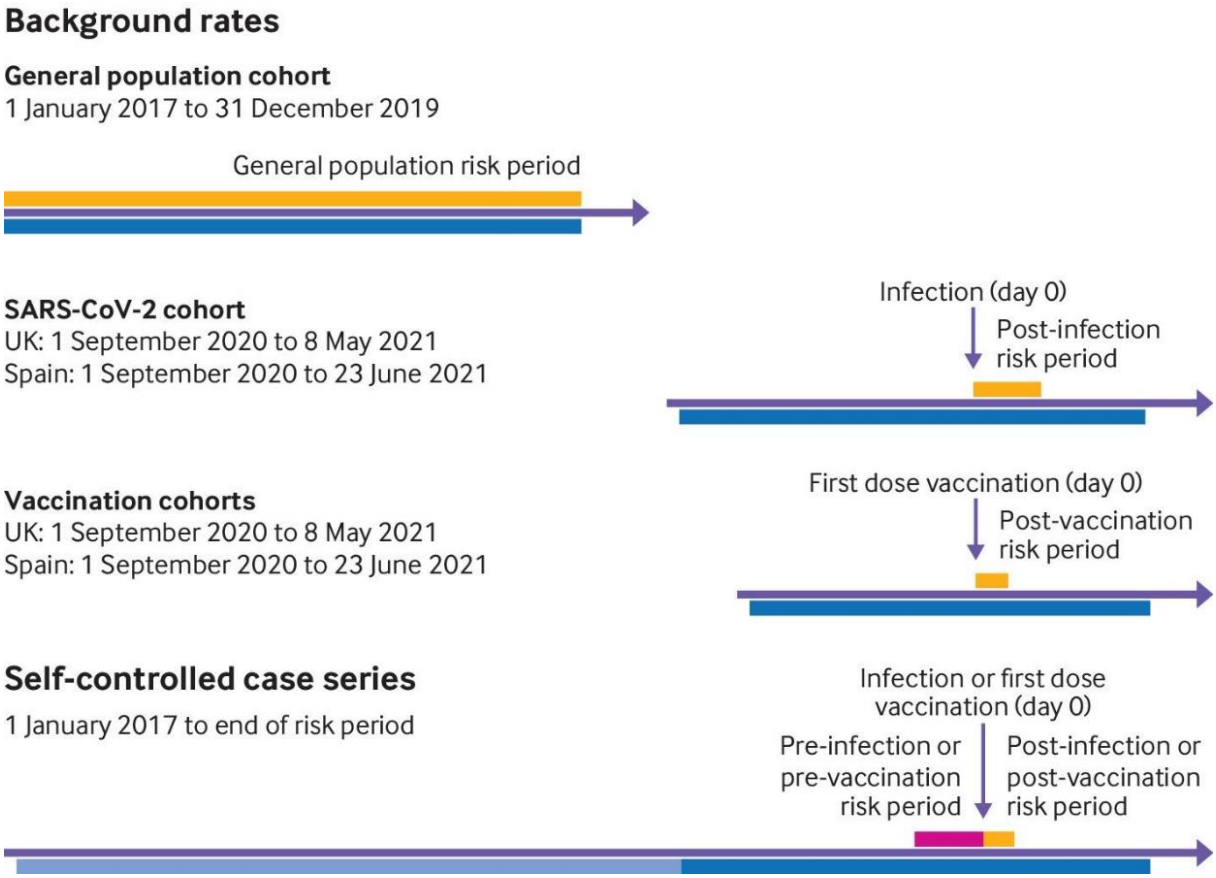
6.3.5.2 Self-controlled case series

Second, a modified self-controlled case series (SCCS) method was used. As explained in Chapter 3, the SCCS method includes only exposed individuals experiencing the outcome, and participants act as controls for themselves, therefore eliminating time-fixed confounding. Within-person comparisons of event rates were made between the pre-exposure (baseline) and the post-exposure risk period.

The SCCS was restricted to first dose vaccine administration, and only the first event for each individual was considered.

Persons were censored at 21 days after first dose vaccination or 90 days after a SARS-CoV-2 positive test, without including any time after the exposure window into the baseline period. I did not include a post exposure time in the baseline risk window for the vaccinated because of the receipt of second doses among the study participants (with the majority of those vaccinated in Spain getting their second dose on day 21). Time after 90 days for the COVID-19 cohort was not included firstly to be consistent with the analysis of the vaccinated, and secondly because of the potential long-term sequelae of COVID-19, which can be expected to last beyond 90 days for a substantial proportion of cases.

Figure 6-1 Study cohorts and study design.



* Study design. Potential risk period (dark blue) for vaccination cohorts was defined as time between the start of the vaccination campaign and one week before the end of data availability for each database (CPRD AURUM: 8 December 2020 to 2 May 2021; SIDIAP: 27 December 2020 to 23 June

2021). For the SARS-CoV-2 infected cohort, the potential risk period started on 1 September 2020. The baseline period for the self-controlled case series analysis (light blue) was defined from 1 January 2017 to 21 days before the day of vaccination or SARS-CoV-2 positive test result. The pre-risk period (pink) was defined as −21 to −1 days before vaccination or SARS-CoV-2 positive test result, and the risk period (orange) was defined as 1 to 21 days after vaccination and 1-90 days after a SARS-CoV-2 positive test result.

6.3.6 Statistical analysis

I characterised the study participants in each cohort in terms of socio-demographics (e.g., age, sex), comorbidities (any time before vaccination), and recent medication use during the 6 months before the index date. Detailed information of the definitions used for comorbidities and medications is reported in the Appendix to chapter 3, table 4.

6.3.6.1 Observed-to-expected analysis

I estimated the observed rates during the 21 days post-immunisation and 90 days post-test period for the vaccinated and SARS-CoV-2 infected cohorts, respectively. Similarly, background rates were estimated for the general population from 1st January 2017 to 31st December 2019. I calculated crude incidence rates as the total number of events divided by the person-time at risk per 100,000 person-years. I used indirect standardisation to account for differences between the age structure of the vaccinated or SARS-CoV-2 cohorts and the general population.[185] I calculated standardised incidence ratios (SIRs) and 95% confidence intervals (Cis) to compare observed and expected rates.

6.3.6.2 SCCS method

I fitted conditional Poisson regression models to estimate incidence rate ratios (IRRs) and 95% Cis for each outcome according to exposure window[186], comparing post- vs pre-exposure time. These models were estimated separately for each cohort of interest and were adjusted for age (in 5-year bands) and seasonality (four seasons). SCCS analyses were only conducted for comparisons where the minimum detectable relative rate (MDRR) was two or less (have the ability to detect 100% increased risk if present) [187].

The exposure window was divided in several pre-specified time periods: 0, 1-7, 8-14, 15-21, as well as collectively for the 1-21 days period. Day-zero events were considered separately, as the occurrence of these events could precipitate hospital admissions and subsequent COVID-19 testing, which could induce a positive association between SARS-CoV-2 infection and the studied events[188].

To account for the potential for events to temporarily decrease the probability of exposure (i.e., delay vaccination), a 21-day pre-risk period was removed from the baseline period and reported separately.

The assumptions of SCCS were assessed. Firstly, to assess potential for event-dependent observation periods, I plotted a histogram of the time between the event occurrence and the end of observation for individuals censored and uncensored. I also calculated the number of deaths that occurred following each event (Day 0 to day 7).

6.3.6.3 Sensitivity analyses

I applied sensitivity analyses to assess the impact of study design choices in both the observed-to-expected analysis and the SCCS method. First, I excluded individuals infected with SARS-CoV-2 prior to the index date for the vaccinated cohorts to exclude events potentially related to COVID-19 infection. Since RT-PCR tests were not routinely performed during the first wave of the pandemic, the COVID-19 definition was broadened to include positive SARS-CoV-2 test (RT-PCR or antigen test) or a compatible COVID-19 clinical code. Second, to include persons with little prior healthcare utilisation, I also replicated the analyses after removing the 1-year prior observation requirement for study participants.

Any subgroups with less than 5 people were blinded and reported as <5, following information governance requirements.

6.4 Results

Overall, 8,330,497 people who received at least one vaccine dose against COVID-19 (CPRD AURUM: 5,477,859; SIDIAP: 2,852,638), including 4,376,535 with ChAdOx1, 3,588,318 with BNT162b2, 244,913 with mRNA-1273, and 120,731 with Ad26.COV2.S were included in this study. A total of 3,745,017 completed the two-doses vaccination course (CPRD AURUM: 2,260,880; SIDIAP: 1,644,365), including 1,324,666 with ChAdOx1, 2,420,351 with BNT162b2, and 160,228 with mRNA-1273. Among those vaccinated, 594,407 participants had COVID-19 before the first-dose administration. In addition, I included 735,870 persons with a SARS-CoV-2 infection who were not previously vaccinated (CPRD AURUM: 447,233; SIDIAP: 288,637) and 14,330,080 general population participants (CPRD AURUM: 9,651,568; SIDIAP: 4,678,512).

The distribution of demographics, comorbidities and recent medication use for each database and cohort can be found in Table 6-1 and Table 6-2. Participants who received the first dose of any of the COVID-19 vaccines were older (median age from 56 to 64 years in CPRD AURUM, and 51 to 62 in SIDIAP) than the general population (median age 48 years old in CPRD AURUM and 47 in SIDIAP). In CPRD AURUM, participants infected with SARS-CoV-2 were younger than those vaccinated against COVID-19 (median age 41 years old). Vaccinated individuals also had more comorbidities than the general population, including autoimmune diseases, cancer, diabetes, obesity, heart diseases and renal impairment (Table 6-1, Table 6-2).

Table 6-1 Baseline characteristics of study participants in CPRD AURUM

	General population	ChAdOx1 first dose	ChAdOx1 second dose	BNT162b2 first dose	BNT162b2 second dose	SARS-CoV-2 infection
N	9,651,568	3,782,401	1,093,812	1,695,458	1,167,068	447,233
Age, median [IQR]	48 [33 to 63]	56 [47 to 66]	71 [59 to 76]	64 [49 to 76]	69 [52 to 78]	41 [30 to 54]
Sex: Male	4,825,624 (50.0%)	1,829,719 (48.4%)	482,007 (44.1%)	711,217 (41.9%)	462,061 (39.6%)	204,798 (45.8%)
Years of prior observation time, median [IQR]	12.8 [4.9 to 23.7]	16.1 [6.7 to 27.6]	20.4 [8.2 to 32.4]	18.0 [7.1 to 30.1]	19.1 [7.5 to 31.3]	11.6 [4.5 to 22.2]
Comorbidities						
Autoimmune disease	198,170 (2.1%)	105,128 (2.8%)	47,582 (4.4%)	62,384 (3.7%)	45,003 (3.9%)	8,988 (2.0%)
Malignant neoplastic disease	606,442 (6.3%)	319,601 (8.4%)	185,702 (17.0%)	243,369 (14.4%)	191,533 (16.4%)	17,519 (3.9%)
Diabetes mellitus	685,313 (7.1%)	383,250 (10.1%)	181,246 (16.6%)	263,365 (15.5%)	167,689 (14.4%)	32,705 (7.3%)
Obesity	349,425 (3.6%)	190,790 (5.0%)	67,163 (6.1%)	98,771 (5.8%)	61,431 (5.3%)	20,436 (4.6%)
Heart disease	827,039 (8.6%)	413,054 (10.9%)	232,267 (21.2%)	319,053 (18.8%)	236,329 (20.2%)	28,603 (6.4%)
Hypertensive disorder	1,761,421 (18.3%)	915,008 (24.2%)	449,684 (41.1%)	592,676 (35.0%)	445,346 (38.2%)	59,638 (13.3%)
Renal impairment	512,050 (5.3%)	242,170 (6.4%)	157,400 (14.4%)	208,112 (12.3%)	168,226 (14.4%)	15,982 (3.6%)
Medication use (183 days prior to four days prior)						
Systemic corticosteroids	545,205 (5.6%)	187,454 (5.0%)	77,668 (7.1%)	103,444 (6.1%)	73,838 (6.3%)	18,932 (4.2%)

Antithrombotic and anticoagulant therapies	190,145 (2.0%)	70,593 (1.9%)	43,189 (3.9%)	58,379 (3.4%)	44,663 (3.8%)	4,388 (1.0%)
Lipid modifying agents	290,710 (3.0%)	137,925 (3.6%)	71,958 (6.6%)	95,123 (5.6%)	69,719 (6.0%)	9,110 (2.0%)
Antineoplastic and immunomodulating agents	170,815 (1.8%)	40,343 (1.1%)	14,601 (1.3%)	25,199 (1.5%)	18,078 (1.5%)	9,041 (2.0%)

CPRD: Clinical Practice Research Datalink, IQR: Interquartile range

Table 6-2 Baseline characteristics of study participants in SIDIAP

	General population	ChAdOx1 first dose	ChAdOx1 second dose	BNT162b2 first dose	BNT162b2 second dose	mRNA-1273 first dose	mRNA-1273 second dose	Ad26.COV2.S	SARS-CoV-2 infection
N	4,678,512	594,134	230,854	1,892,860	1,253,283	244,913	160,228	120,731	288,637
Age, median [IQR]	47 [36 to 63]	62 [59 to 65]	60 [42 to 63]	56 [47 to 75]	72 [55 to 79]	53 [47 to 57]	54 [49 to 60]	51 [43 to 67]	46 [34 to 59]
Sex: Male	2,287,403 (48.9%)	270,830 (45.6%)	103,686 (44.9%)	859,915 (45.4%)	527,395 (42.1%)	117,409 (47.9%)	73,139 (45.6%)	64,435 (53.4%)	135,813 (47.1%)
Years of prior observation time, median [IQR]	11.0 [11.0 to 11.0]	15.3 [15.2 to 15.3]	15.5 [15.4 to 15.5]	15.3 [15.2 to 15.4]	15.3 [15.2 to 15.4]	15.4 [15.3 to 15.4]	15.4 [15.3 to 15.4]	15.4 [15.3 to 15.5]	14.9 [14.8 to 15.1]
Comorbidities									
Autoimmune disease	79,044 (1.7%)	14,501 (2.4%)	5,075 (2.2%)	49,029 (2.6%)	40,899 (3.3%)	13,012 (5.3%)	10,776 (6.7%)	2,242 (1.9%)	5,857 (2.0%)
Malignant neoplastic disease	339,004 (7.2%)	60,074 (10.1%)	18,504 (8.0%)	241,140 (12.7%)	220,040 (17.6%)	38,401 (15.7%)	32,964 (20.6%)	9,397 (7.8%)	22,242 (7.7%)
Diabetes mellitus	464,603 (9.9%)	80,492 (13.5%)	23,582 (10.2%)	280,876 (14.8%)	243,412 (19.4%)	25,669 (10.5%)	19,699 (12.3%)	15,049 (12.5%)	30,542 (10.6%)
Obesity	863,167 (18.4%)	161,047 (27.1%)	50,787 (22.0%)	502,689 (26.6%)	396,021 (31.6%)	55,920 (22.8%)	38,727 (24.2%)	30,154 (25.0%)	66,933 (23.2%)
Heart disease	568,547 (12.2%)	87,455 (14.7%)	26,640 (11.5%)	403,280 (21.3%)	367,372 (29.3%)	30,494 (12.5%)	24,061 (15.0%)	15,827 (13.1%)	37,800 (13.1%)

Hypertensive disorder	1,139,337 (24.4%)	200,957 (33.8%)	56,782 (24.6%)	701,276 (37.0%)	628,916 (50.2%)	60,731 (24.8%)	46,745 (29.2%)	33,102 (27.4%)	65,030 (22.5%)
Renal impairment	229,136 (4.9%)	21,359 (3.6%)	5,650 (2.4%)	194,837 (10.3%)	187,662 (15.0%)	14,622 (6.0%)	12,998 (8.1%)	4,923 (4.1%)	16,807 (5.8%)
Medication use (183 days prior to four days prior)									
Systemic corticosteroids	258,775 (5.5%)	32,229 (5.4%)	11,213 (4.9%)	123,581 (6.5%)	98,008 (7.8%)	20,231 (8.3%)	15,854 (9.9%)	7,296 (6.0%)	15,654 (5.4%)
Antithrombotic and anticoagulant therapies	110,809 (2.4%)	16,968 (2.9%)	4,851 (2.1%)	68,891 (3.6%)	59,083 (4.7%)	10,122 (4.1%)	7,838 (4.9%)	3,548 (2.9%)	7,618 (2.6%)
Lipid modifying agents	80,561 (1.7%)	18,339 (3.1%)	5,310 (2.3%)	45,250 (2.4%)	37,985 (3.0%)	6,292 (2.6%)	4,804 (3.0%)	2,934 (2.4%)	4,268 (1.5%)
Antineoplastic and immunomodulating agents	59,360 (1.3%)	7,097 (1.2%)	3,277 (1.4%)	24,460 (1.3%)	18,331 (1.5%)	7,861 (3.2%)	6,455 (4.0%)	1,310 (1.1%)	5,046 (1.7%)

SIDIAP: Information System for Research in Primary Care, IQR: Interquartile range

Table 6-3 Association between COVID-19 vaccination or infection and the occurrence of immune-mediated neurological disorders of special interest in CPRD AURUM

	N	Person-years	Observed events	Expected events	SIR (95% CI)
Bell's Palsy					
ChAdOx1 first-dose	3,776,803	384,250	117	164.5	0.71 (0.59 to 0.85)
ChAdOx1 second-dose	1,093,258	218,629	25	94.7	0.26 (0.18 to 0.39)
BNT162b2 first-dose	1,693,453	275,333	46	116.4	0.40 (0.30 to 0.53)
BNT162b2 second-dose	1,166,571	235,351	24	99.5	0.24 (0.16 to 0.36)
SARS-CoV-2 positive test	446,851	106,342	53	39.8	1.33 (1.02 to 1.74)
Encephalomyelitis					
ChAdOx1 first-dose	3,778,358	384,420	11	7.6	1.45 (0.80 to 2.62)
BNT162b2 first-dose	1,694,161	275,451	<5		
BNT162b2 second-dose	1,167,028	235,447	<5		
SARS-CoV-2 positive test	447,037	106,391	11	1.6	6.89 (3.82 to 12.44)
Guillain Barre Syndrome					
ChAdOx1 first-dose	3,778,430	384,427	11	14.9	0.74 (0.41 to 1.33)
ChAdOx1 second-dose	1,093,778	218,737	<5		
BNT162b2 first-dose	1,694,167	275,452	<5		

BNT162b2 second-dose	1,167,031	235,448	<5		
SARS-CoV-2 positive test	447,029	106,390	9	2.6	3.53 (1.83 to 6.77)
Transverse myelitis					
ChAdOx1 first-dose	3,778,434	384,430	<5		
BNT162b2 first-dose	1,694,190	275,456	<5		
SARS-CoV-2 positive test	447,041	106,393	<5		

CPRD: Clinical Practice Research Datalink; SIR: standardised incidence ratio.

Events with less than 5 occurrences have been omitted for privacy reasons.

Table 6-4 Association between COVID-19 vaccination or infection and the occurrence of immune-mediated neurological disorders of special interest in SIDIAP

	N	Person years	Observed events	Expected events	SIR (95% CI)
Bell's Palsy					
ChAdOx1 first-dose	592,365	89,055	27	82.6	0.33 (0.22 to 0.48)
ChAdOx1 second-dose	230,692	55,659	10	47.3	0.21 (0.11 to 0.39)
BNT162b2 first-dose	1,890,434	102,623	100	116.7	0.86 (0.70 to 1.04)
BNT162b2 second-dose	1,251,680	74,638	85	97.1	0.88 (0.71 to 1.08)
mRNA-1723 first-dose	244,467	17,648	14	15.2	0.92 (0.54 to 1.55)
mRNA-1723 second-dose	159,995	12,563	5	11.3	0.44 (0.18 to 1.06)
Ad26.COV2.S first-dose	120,470	5,228	6	5.2	1.15 (0.52 to 2.56)
SARS-CoV-2 positive test	288,030	66,603	93	54.7	1.70 (1.39 to 2.08)
Encephalomyelitis					
ChAdOx1 first-dose	592,843	89,121	5	6.1	0.82 (0.34 to 1.97)
ChAdOx1 second-dose	230,844	55,696	<5		
BNT162b2 first-dose	1,892,409	102,738	9	11.7	0.77 (0.40 to 1.48)
BNT162b2 second-dose	1,253,202	74,733	<5		
mRNA-1273 first-dose	244,744	17,669	<5		

Ad26.COV2.S first-dose	120,588	5,233	<5		
SARS-CoV-2 positive test	288,374	66,689	17	4.5	3.75 (2.33 to 6.02)
Guillain Barre Syndrome					
ChAdOx1 first-dose	592,860	89,123	<5		
BNT162b2 first-dose	1,892,423	102,739	5	6.3	0.79 (0.33 to 1.91)
BNT162b2 second-dose	1,253,201	74,733	<5		
mRNA-1273 second-dose	160,213	12,58	<5		
SARS-CoV-2 positive test	288,392	66,693	17	2.9	5.92 (3.68 to 9.53)
Transverse myelitis					
BNT162b2 first-dose	1,892,510	102,744	<5		
mRNA-1723 second-dose	160,222	12,581	<5		

SIDIAP: Information System for Research in Primary Care; SIR: standardised incidence ratio.
*Events with less than 5 occurrences have been omitted for privacy reasons.

Table 6-5 Incidence rate ratios (95% confidence intervals) for COVID-19 vaccination or infection and the occurrence of Bell's Palsy estimated from the SCCS analysis in CPRD AURUM

	ChAdOx1 first dose		BNT162b2 first dose		SARS-CoV-2 infection	
	Number of events	IRR	Number of events	IRR	Number of events	IRR
Baseline	7403	Ref	4564	Ref	2335	Ref
−21 to −1 days pre-exposure	83	0.68 (0.54 to 0.84)	41	0.74 (0.54 to 1.00)	14	1.03 (0.58 to 1.68)
Day 0	5	0.83 (0.30 to 1.79)	<5		<5	
Days 1 – 7 post exposure	25	0.59 (0.38 to 0.85)	14	0.75 (0.42 to 1.22)	6	1.30 (0.52 to 2.65)
Days 8 – 14 post exposure	34	0.81 (0.56 to 1.11)	17	0.91 (0.54 to 1.41)	<5	
Days 15 – 21 post exposure	43	1.03 (0.75 to 1.37)	16	0.84 (0.49 to 1.33)	7	1.51 (0.64 to 2.93)
Days 1 – 21 post exposure	102	0.81 (0.66 to 0.98)	47	0.83 (0.61 to 1.10)	*	0.94 (0.51 to 1.55)
Days 1 – 90 post exposure					48	0.76 (0.56 to 1.01)

SCCS: Self-controlled case series; CPRD: Clinical Practice Research Datalink. Events with less than 5 occurrences have been omitted for privacy reasons.
* Since event number below 5 was blinded, the total number of events during the time period was blinded as well.

Table 6-6 Incidence rate ratios (95% confidence intervals) for COVID-19 vaccination or infection and the occurrence of Bell's Palsy estimated using the SCCS methodology in SIDIAP

	ChAdOx1		BNT162b2		mRNA-1273		SARS-CoV-2 infection	
	Number of events	IRR	Number of events	IRR	Number of events	IRR	Number of events	IRR
Baseline		Ref		Ref		Ref		Ref
−21 to −1 days pre-exposure	12	0.44 (0.23 to 0.73)	74	0.69 (0.54 to 0.86)	11	0.80 (0.41 to 1.38)	44	2.93 (2.13 to 3.92)
Day 0	<5		<5		<5		8	11.13 (5.07 to 20.80)
Days post-vaccination/post-infection:								
Days 1 – 7 post exposure	7	0.77 (0.33 to 1.49)	22	0.62 (0.39 to 0.91)	<5		17	3.38 (2.01 to 5.28)
Day 8 – 14 post exposure	7	0.77 (0.33 to 1.50)	35	1.02 (0.72 to 1.40)	6	1.37 (0.54 to 2.79)	6	1.21 (0.48 to 2.47)
Days 15 – 21 post exposure	11	1.22 (0.63 to 2.10)	27	0.85 (0.57 to 1.22)	<5		<5	
Days 1 – 21 post exposure	25	0.92 (0.60 to 1.33)	84	0.83 (0.66 to 1.02)	*	0.99 (0.54 to 1.64)	*	1.82 (1.21 to 2.61)
Days 1 – 90 post exposure							7	1.31 (1.03 to 1.64)

SCCS: Self-controlled case series; SIDIAP: Information System for Research in Primary Care. Events with less than 5 occurrences have been omitted for privacy reasons.

* Since event number below 5 was blinded, the total number of events during the time period was blinded as well.

6.4.1 Bell's palsy

In the first 21 days following a first dose of ChAdOx1, a total of 144 cases of Bell's Palsy were observed. Of them, 117 were registered in CPRD AURUM, where 164.5 cases would have been expected given background rates in the general population (SIR: 0.71 [95%CI 0.59-0.85]) (Table 6-3). In SIDIAP, equivalent figures were 27 observed compared to 82.6 expected cases (SIR: 0.33 [0.22 to 0.48]) (Table 6-4). I observed 35 cases (CPRD AURUM: 25; SIDIAP: 10) following the second dose of ChAdOx1, which resulted in SIR of 0.26 [0.18 to 0.39] in CPRD AURUM and 0.21 [0.11 to 0.39] in SIDIAP (Figure 6-2). I observed 46 and 24 cases of Bell's Palsy following first and second dose vaccination with BNT162b2 in CPRD AURUM, which compared to 116.4 and 99.5 expected cases (0.40 [0.30 to 0.53] and 0.24 [0.16 to 0.36]). In SIDIAP, rates of Bell's palsy following BNT162b2 (first and second doses) and mRNA-1273 and Ad26.COV2.S first doses were similar to those expected, with equivalent SIRs of 0.86 [0.70 to 1.04], 0.88 [0.71 to 1.08], 0.92 [0.54 to 1.55] and 1.15 [0.52 to 2.56].

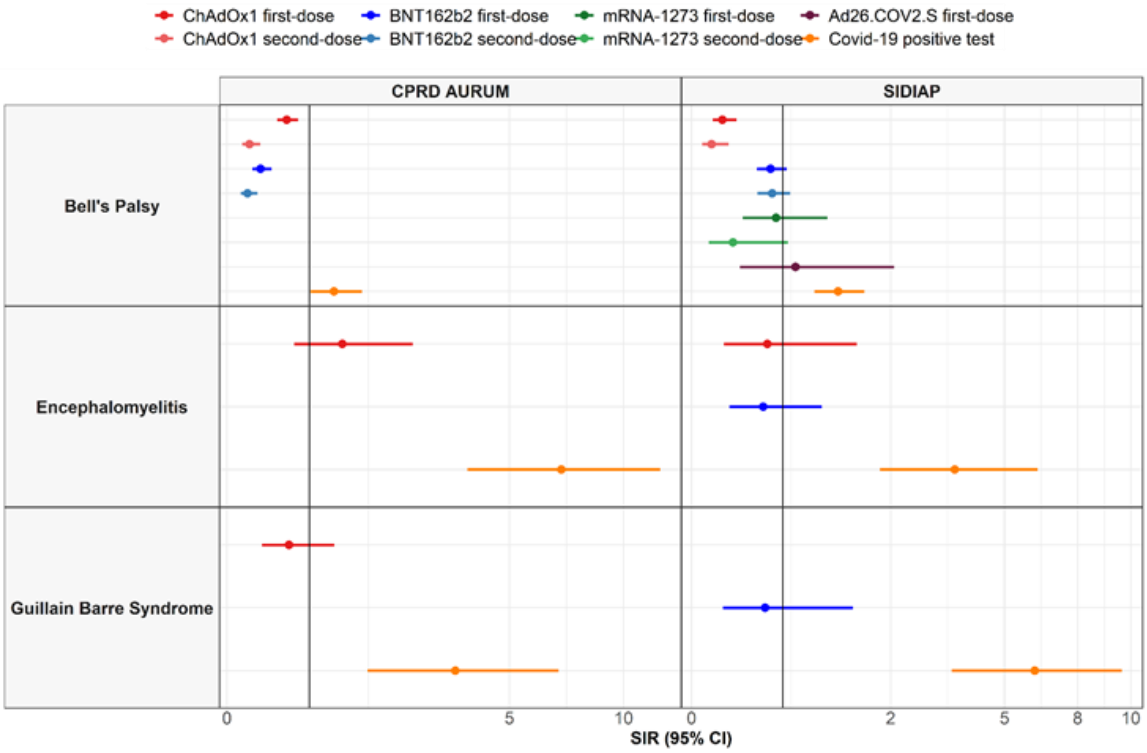
Age and sex stratified IRRs were only calculated for Bell's palsy among some vaccine cohorts, as it was the only event in which some of the strata had more than five occurrences. (Figure 6-3) No significantly higher rates were observed after vaccination in either CPRD AURUM or SIDIAP.

Rates of Bell's Palsy were higher than expected among those infected with SARS-CoV-2, who experienced 53 and 93 events in the 90 days following an infection registered in CPRD AURUM and in SIDIAP, which compared to 39.8 and 54.7 expected events, respectively. Equivalent SIRs were 1.33 [1.02 to 1.74] in the CPRD AURUM and 1.70 [1.39 to 2.08] in SIDIAP.

The SCCS analysis was only sufficiently powered to study individuals vaccinated with a first dose of BNT162b2, ChAdOx1 and mRNA-1273 (in SIDIAP only), and those infected with SARS-CoV-2 (Appendix table 1 and Appendix table 1). In the SCCS analysis, I observed an adjusted IRR of 0.81 (0.66 to 0.98) for ChAdOx1 and 0.83 (0.61 to 1.10) for BNT162b2 between 1 to 21 days in CPRD

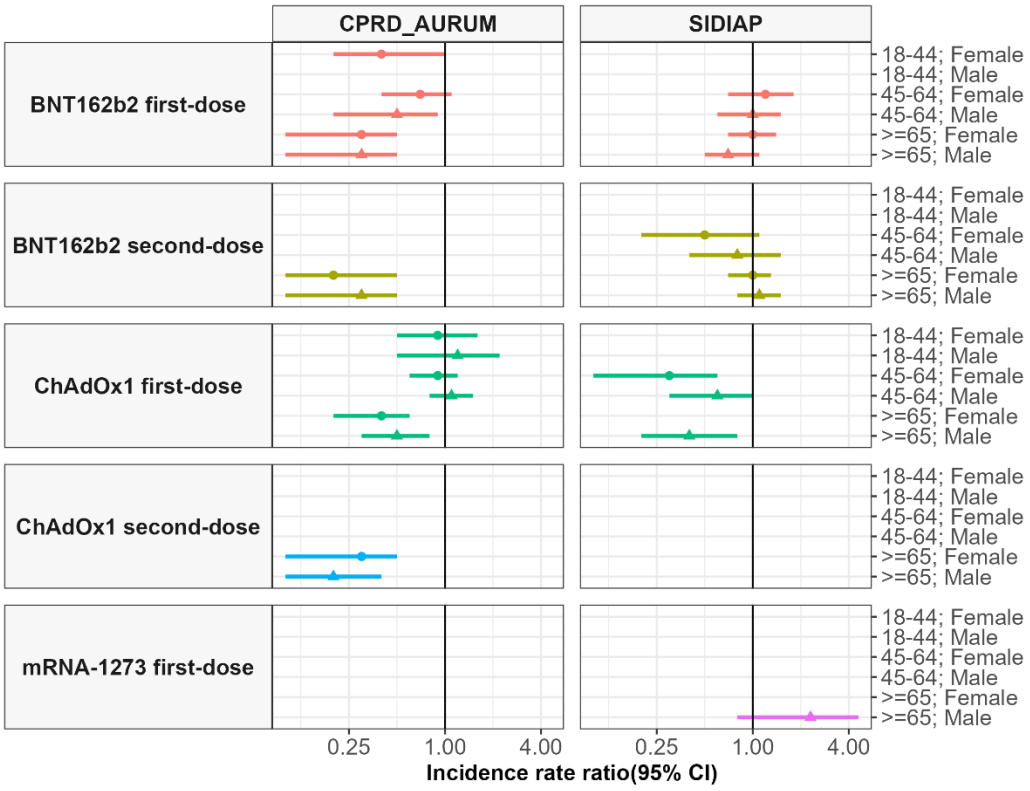
AURUM. Equivalent figures in SIDIAP were 0.92 (0.60 to 1.33) and 0.83 (0.66 to 1.02), respectively. The adjusted IRR for mRNA-1273 was 0.99 (0.54 to 1.64), and 1.82 (1.21 to 2.61) for SARS-CoV-2. I observed an increased risk for this outcome on the day of a SARS-CoV-2 positive test, and the 1-21 days before and 1-7 days after the positive test (Table 6-5, Table 6-6).

Figure 6-2 Standardised incidence ratios of immune mediated neurological disorders from observed-to-expected analysis.



In the visual check of the event-dependent observation period, there was no evidence of event-dependent censoring shortly after the bells' palsy in the CPRD data. (Appendix figure 1) There was a spike close to zero among the censored in the SIDIAP data, which indicated the presence of event-dependent observation periods. (Appendix figure 2). The percentage of outcome events observed in SIDIAP that were associated with hospitalisation was 35.9% for Bell's palsy (Appendix table 3). Deaths in the first week after an event were seen for fewer than 0.5% across both databases (Appendix table 4).

Figure 6-3 Incidence Rate Ratios (IRR) of Bell's palsy by age group and sex



* CPRD: Clinical Practice Research Datalink; SIDIAP: Information System for Research in Primary Care**.** Events with less than 5 occurrences have been omitted for privacy reasons. No events were observed for the omitted cohorts. Study period: CPRD AURUM: 8 December 2020 to 2 May 2021; SIDIAP: 27 December 2020 to 23 June 2021.

6.4.2 Encephalomyelitis

I observed 16 cases of encephalomyelitis following first-dose of ChAdOx1, resulting in SIRs of 1.45 [0.80 to 2.62] in CPRD AURUM and 0.82 [0.34 to 1.97] in SIDIAP. I observed 9 cases of encephalomyelitis following the first dose of BNT162b2 in SIDIAP (0.77 [0.40 to 1.48]). No cases following mRNA-1273 second dose were reported, and fewer than five events occurred among the remaining vaccination cohorts. Conversely, I observed 28 events of encephalomyelitis following SARS-CoV-2 infection (CPRD AURUM: 11; SIDIAP: 17), which compared with 1.6 expected in CPRD AURUM (6.89 [3.82 to 12.44]) and 4.5 expected in SIDIAP (3.75 [2.33 to 6.02]).

6.4.3 Guillain-Barré syndrome (GBS)

In CPRD AURUM, 11 and less than five cases of GBS were observed after first and second doses of ChAdOx1, respectively. Equivalent figures in SIDIAP were less than five and zero events, respectively. Five or fewer events occurred following BNT162b2 first and second dose and mRNA-1723 second dose, with no cases following first doses of mRNA or Ad26.COV2.S. On the other hand, I observed 26 events of GBS following SARS-CoV-2 infection, which compared with 2.6 (3.53 [1.83 to 6.77]) and 2.9 (5.92 [3.68 to 9.53]) expected cases in CPRD AURUM and SIDIAP, respectively.

6.4.4 Transverse myelitis

I only observed cases of transverse myelitis following ChAdOx1 first dose (in CPRD AURUM), BNT162b2 first dose (in CPRD AURUM and SIDIAP) and mRNA-1723 second dose (in SIDIAP), with fewer than five cases after the exposure. I also observed less than five cases following SARS-CoV-2 infection in CPRD AURUM.

6.4.5 Sensitivity analyses

I found similar results in sensitivity analyses after 1) excluding individuals with a prior COVID-19 infection among those vaccinated; and 2) removing the requirement for at least one year of prior history in all cohorts (Table 6-7, Table 6-8). In CPRD Aurum, there was an increased risk of encephalomyelitis among female aged 45-64 after first dose of ChAdOx1 when the one-year prior observation requirement was removed, with an IRR of 3.6, 95% CI 1.2 to 8. The incidence rate ratio of this age-sex group was not reported in the main analysis or in SIDIAP since the subgroup were smaller than five. I obtained similar results after reducing the follow-up period from 90 to 21 days in the SARS-CoV-2-infected cohorts, although SIRs were higher and had wider confidence intervals than those obtained in the main SARS-CoV-2 cohort.(Table 6-7)

Table 6-7 Sensitivity analyses for the O/E analyses in the Clinical Practice Research Datalink (CPRD AURUM)

	N	Person-years	Observed events	Expected events	SIR (95% CI)
Results reducing time at risk to 21 days among persons in the SARS-CoV-2-infected cohort					
Bell's Palsy	446,851	25,54	15	9.6	1.57 (0.95 to 2.60)
Encephalomyelitis	447,037	25,551	<5		
Guillain Barré Syndrome	447,029	25,551	<5		
Results excluding vaccinated persons with a history of COVID-19					
Bell's Palsy					
ChAdOx1 first-dose	3,522,487	361,239	108	155.0	0.70 (0.58 to 0.84)
ChAdOx1 second-dose	1,029,876	206,264	24	89.5	0.27 (0.18 to 0.40)
BNT162b2 first-dose	1,599,332	261,317	43	110.7	0.39 (0.29 to 0.52)
BNT162b2 second-dose	1,099,149	221,961	22	94.0	0.23 (0.15 to 0.36)
Encephalomyelitis					
ChAdOx1 first-dose	3,523,932	361,399	9	7.2	1.25 (0.65 to 2.41)
BNT162b2 first-dose	1,599,992	261,429	<5		
BNT162b2 second-dose	1,099,577	222,051	<5		
Guillain Barré Syndrome					
ChAdOx1 first-dose	3,523,992	361,405	11	14.1	0.78 (0.43 to 1.41)

ChAdOx1 second-dose	1,030,356	206,365	<5		
BNT162b2 first-dose	1,600,002	261,43	<5		
BNT162b2 second-dose	1,099,583	222,052	<5		
Transverse Myelitis					
ChAdOx1 first-dose	3,523,987	361,407	<5		
BNT162b2 first-dose	1,600,021	261,434	<5		
Results without requiring year of prior history					
Bell's Palsy					
ChAdOx1 first-dose	4,045,751	409,047	124	174.2	0.71 (0.60 to 0.85)
ChAdOx1 second-dose	1,148,737	229,423	27	99.0	0.27 (0.19 to 0.40)
BNT162b2 first-dose	1,805,347	292,343	49	122.8	0.40 (0.30 to 0.53)
BNT162b2 second-dose	1,226,529	247,211	24	103.9	0.23 (0.15 to 0.34)
SARS-CoV-2 positive test	491,257	116,681	57	43.0	1.33 (1.02 to 1.72)
Encephalomyelitis					
ChAdOx1 first-dose	4,047,348	409,22	13	8.1	1.60 (0.93 to 2.76)
BNT162b2 first-dose	1,806,088	292,465	<5		
BNT162b2 second-dose	1,226,997	247,31	<5		
SARS-CoV-2 positive test	491,453	116,733	11	1.7	6.35 (3.52 to 11.47)

Guillain Barré Syndrome					
ChAdOx1 first-dose	4,047,416	409,228	12	15.5	0.77 (0.44 to 1.36)
BNT162b2 first-dose	1,806,095	292,466	<5		
BNT162b2 second-dose	1,227,001	247,31	<5		
SARS-CoV-2 positive test	491,447	116,732	11	2.7	4.01 (2.22 to 7.24)
Transverse Myelitis					
ChAdOx1 first-dose	4,047,425	409,231	<5		
BNT162b2 first-dose	1,806,120	292,47	<5		
SARS-CoV-2 positive test	491,459	116,735	<5		

*SIR: Standardised incidence ratio

Table 6-8 Sensitivity analyses for the O/E analyses in the Information System for Research in Primary Care (SIDIAP)

	N	Person-years	Observed events	Expected events	SIR (95% CI)
Results reducing time at risk to 21 days among persons in the SARS-CoV-2-infected cohort					
Bell's Palsy	288,03	16,39	36	13.5	2.66 (1.92 to 3.69)
Encephalomyelitis	288,374	16,411	10	1.1	8.84 (4.76 to 16.43)
Guillain Barré Syndrome	288,392	16,412	8	0.7	11.33 (5.66 to 22.65)
Results excluding vaccinated persons with a history of COVID-19					
Bell's Palsy					
ChAdOx1 first-dose	556,581	85,327	23	79.2	0.29 (0.19 to 0.44)
ChAdOx1 second-dose	223,631	53,893	10	45.9	0.22 (0.12 to 0.41)
BNT162b2 first-dose	1,718,765	93,23	85	106.4	0.80 (0.65 to 0.99)
BNT162b2 second-dose	1,143,065	67,988	78	88.7	0.88 (0.70 to 1.10)
mRNA-1723 first-dose	219,459	15,971	12	13.9	0.87 (0.49 to 1.52)
mRNA-1723 second-dose	147,538	11,573	<5		
Ad26.COV2.S first-dose	108,031	4,692	5	4.7	1.06 (0.44 to 2.56)
Encephalomyelitis					
ChAdOx1 first-dose	557,017	85,389	5	5.8	0.86 (0.36 to 2.06)
ChAdOx1 second-dose	223,777	53,929	<5	3.3	0.30 (0.04 to 2.14)

BNT162b2 first-dose	1,720,466	93,33	7	10.6	0.66 (0.31 to 1.38)
BNT162b2 second-dose	1,144,394	68,071	<5		
mRNA-1723 first-dose	219,706	15,99	<5		
Ad26.COV2.S first-dose	108,13	4,696	<5		
Guillain Barré Syndrome					
ChAdOx1 first-dose	557,037	85,392	<5		
BNT162b2 first-dose	1,720,473	93,33	5	5.8	0.87 (0.36 to 2.09)
BNT162b2 second-dose	1,144,392	68,071	<5		
mRNA-1723 second-dose	147,734	11,589	<5		
Transverse Myelitis					
BNT162b2 first-dose	1,720,543	93,334	<5		
Results without requiring year of prior history					
Bell's Palsy					
ChAdOx1 first-dose	594,588	89,329	27	82.7	0.33 (0.22 to 0.48)
ChAdOx1 second-dose	231,228	55,786	10	47.3	0.21 (0.11 to 0.39)
BNT162b2 first-dose	1,900,497	103,128	100	117.1	0.85 (0.70 to 1.04)
BNT162b2 second-dose	1,256,600	74,934	85	97.4	0.87 (0.71 to 1.08)
mRNA-1723 first-dose	245,672	17,731	14	15.3	0.92 (0.54 to 1.55)

mRNA-1723 second-dose	160,667	12,615	5	11.4	0.44 (0.18 to 1.06)
Ad26.COV2.S first-dose	122,497	5,319	6	5.3	1.14 (0.51 to 2.53)
SARS-CoV-2 positive test	289,417	66,921	93	54.9	1.69 (1.38 to 2.07)
Encephalomyelitis					
ChAdOx1 first-dose	595,068	89,396	5	6.1	0.82 (0.34 to 1.97)
BNT162b2 first-dose	1,902,480	103,244	9	11.7	0.77 (0.40 to 1.48)
BNT162b2 second-dose	1,258,126	75,03	4	10.3	0.39 (0.15 to 1.03)
mRNA-1723 first-dose	245,949	17,752	<5		
Ad26.COV2.S first-dose	122,615	5,324	<5		
SARS-CoV-2 positive test	289,761	67,006	17	4.5	3.75 (2.33 to 6.04)
Guillain Barré Syndrome					
ChAdOx1 first-dose	595,085	89,397	<5		
BNT162b2 first-dose	1,902,494	103,245	5	6.3	0.79 (0.33 to 1.90)
BNT162b2 second-dose	1,258,125	75,03	<5		
mRNA-1723 second-dose	160,885	12,633	<5		
SARS-CoV-2 positive test	289,78	67,011	17	2.9	5.93 (3.69 to 9.54)
Transverse Myelitis					
BNT162b2 first-dose	1,902,581	103,25	<5		

mRNA-1723 second-dose	160,894	12,634	<5		

*SIR: Standardised incidence ratio

Table 6-9 Sensitivity analyses for the self-controlled case series excluding individuals infected with SARS-CoV-2 prior to vaccination.

	CPRD AURUM				**SIDIAP**					
	ChAdOx1		**BNT162b2**		**ChAdOx1**		**BNT162b2**		**mRNA-1273**	
	Number of events	IRR	Number of events	IRR	Number of events	IRR	Number of events	IRR	Number of events	IRR
Baseline	6935	Ref	4375	Ref	4455	Ref	8438	Ref	3695	Ref
−21 to −1 days pre-exposure	74	0.66 (0.52 to 0.83)	39	0.76 (0.54 to 1.03)	11	0.44 (0.23 to 0.75)	64	0.67 (0.52 to 0.85)	11	0.91 (0.47 to 1.56)
Day 0	5	0.91 (0.32 to 1.95)	<5		<5		<5		<5	
Days 1 - 7 post exposure	24	0.61 (0.40 to 0.89)	13	0.75 (0.41 to 1.23)	5	0.60 (0.21 to 1.29)	20	0.63 (0.39 to 0.95)	<5	
Day 8 - 14 post exposure	31	0.80 (0.55 to 1.12)	16	0.92 (0.54 to 1.45)	5	0.60 (0.22 to 1.30)	27	0.88 (0.59 to 1.26)	<5	
Days 15 - 21 post exposure	40	1.05 (0.75 to 1.41)	15	0.85 (0.48 to 1.36)	11	1.33 (0.69 to 2.29)	22	0.78 (0.50 to 1.16)	<5	
Days 1- 21 post exposure	95	0.82 (0.66 to 1.00)	44	0.84 (0.61 to 1.12)	21	0.84 (0.53 to 1.26)	69	0.76 (0.60 to 0.96)	*	0.95 (0.49 to 1.63)

Incidence rate ratios (95% confidence intervals) for COVID-19 vaccination or infection and the occurrence of Bell's Palsy.
CPRD: Clinical Practice Research Datalink; SIDIAP: Information System for Research in Primary Care. Events with less than 5 occurrences have been omitted for privacy reasons.

6.5 Discussion

6.5.1 Summary of findings

In this study, I examined the potential association between four COVID-19 vaccines (ChAdOx1 nCoV-19, BNT162b2, mRNA-1273 and Ad26.COV2.S) and the short-term risk of developing Bell's palsy, encephalomyelitis, Guillain-Barré syndrome and transverse myelitis. I also assessed the association between SARS-CoV-2 infection and the risk of these immune-mediated neurological events.

Overall, post-vaccine rates were consistent with expected background rates in a cohort of the general population for Bell's palsy, encephalomyelitis and Guillain-Barré syndrome. Furthermore, the self-controlled case series analysis conducted for Bell's palsy showed no safety signal for those vaccinated. In contrast, an increased risk of Bell's palsy, encephalomyelitis and Guillain-Barré syndrome was observed in people with SARS-CoV-2 infection.

6.5.2 Findings in context

Research findings from observational studies on the association between COVID-19 vaccines and immune-mediated neurological events have though been mixed.

In a disproportionality analysis used the World Health Organisation pharmacovigilance database, no signal of facial paralysis has been detected for mRNA COVID-19 vaccines.[189] Another disproportionality analysis using data from the USA Vaccine Adverse Event Reporting System found an increased risk of facial nerve palsy following both BNT162b2 mRNA-1273.[184] However, the nature of self-reported passive surveillance hampered the generalizability of these studies.

Studies using routinely collected health data showed mixed results as well. For example, a case series and nested case-control study using territory-wide electronic medical records from Hong Kong reported an increased risk of Bell's palsy for recipients of an inactivated vaccine CoronaVac

but not after BNT162b2.[188] Further, the author updated their previous analysis with extended study period and found increased risk of Bell's palsy following 0 - 13 days after second dose BNT162b2 in SCCS analysis, but not after the first dose.[190] Another cohort study using the same data source found no such increased risk of immune-related neurological disorders for either BNT162b2 or CoronaVac.[189] Further, a retrospective cohort study from Israel showed a nonsignificant increased risk of Bell's palsy within 42 days after BNT162b2 vaccination (risk ratio, 1.32 (0.92 to 1.86)). [191] However, they acknowledged that the results may be limited by confounding and selection bias. A pre-print systematic review and meta-analysis looked at Bell's palsy following SARS-CoV-2 vaccines, it pooled results from 9 observational studies and showed no significant increased risk for Bell's palsy compared to the unvaccinated group, yet the estimates from different studies showed substantial heterogeneity (odds ratio of 0.70 (95% CI 0.42-1.16), I2=94%).[192]

Self-controlled case series analysis has been conducted using similar data sources. A large self-controlled case series analysis in England and Scotland investigated the risk of hospital admissions from neurological events during the 28 days following vaccination.[193] The study included over 20 and 12 million people vaccinated with ChAdOx1 and BNT162b2, and reported increased risk of hospital admission with Bell's palsy following a first dose of ChAdOx1 vaccine, with an IRR of 1.3 [1.1–1.6] at 15–21,days. However, the overall IRR during the 1-28 days showed no evidence of increased risk (IRR 1.07 (0.94–1.21)). A more recent SCCS analysis from England also investigated the potential association of COVID-19 vaccination with three acute neurological event types, and found increased rate of Bell's palsy following a first dose of ChAdOx1 vaccine.[194] They observed no clear evidence of an association of ChAdOx1 vaccination with transverse myelitis, or between mRNA vaccines and these events. In my study, a much smaller vaccinated cohort was included (3.7 million ChAdOx1 and 1.7million BNT162b2 in CPRD), which limited the statistical power to identify small effects. Both the mentioned UK based studies used data sources with primary care, linked to

hospital admissions records, while in my study, only primary care data were available in CPRD. Although the SIDIAP data were linked to hospital data, the data size was much smaller. Besides, these two studies had longer study periods which allowed more vaccinees to be included, while both of my data sources were only available till June 2021 by the time this study was conducted. While both mentioned UK-based studies used a 28-days risk window after vaccination, I used a 21-day window to accommodate the different vaccine schedule in Spain.

In another study comparing events reported in a vaccine surveillance system to background rates, Ad26.COV2.S was associated with increased risks of GBS (observed/expected rate ratio of 4.2 [3.5-5.0] in the 42 days following vaccination). [195] In the study conducted by Patone et al., the researchers reported an association between hospital admission for GBS and first dose of ChAdOx1 vaccine, but not BNT162b2 vaccine, with IRR from their SCCS analysis of 2.90; 95 %CI 2.15–3.92 at 15–21 days after vaccination. [193] Walker et al. also found an IRR of 2.85; 95% CI2·33–3·47 during the 4–42 days following first dose of ChAdOx1 vaccinees.[194]

In this study, no safety signal of GBS was observed for ChAdOx1 or Ad26.COV2.S, but the latter was only available for the data from Spain and used in the smallest of the vaccine cohorts. I didn't perform the SCCS for GBS due to limited statistical power, and I found no increased risk for either GBS after ChAdOx1 from the observed-to-expected analysis. In line with these findings, a number of other studies have also reported increased risks of neuro-immune events following SARS-CoV-2 infection [196,197]. For example, Patone et al. found increased risks of Bell's palsy, encephalitis, meningitis and myelitis, and GBS following SARS-CoV-2 infection. The increase in risks following SARS-CoV-2 infection much exceed any of aforementioned associations reported after vaccination.[193]

Different risk periods for the events of interest have been used in previous studies. For example, a 42-day risk period had been used after the influenza A (H1N1) pandemic 2009 vaccination for Guillain Barre syndrome, Bell's Palsy and narcolepsy.[198–200]. This risk period was also used in

the “Draft Master Protocol Assessment of Risk of Safety Outcomes Following COVID-19 Vaccination” from the CBER Surveillance Program by the US FDA, and studies focused on COVID-19 vaccine[70,191].While other published studies on autoimmune conditions after COVID-19 vaccine using routinely collected health data used a shorter risk period of 28 days.[70,193] In my study, the time at risk of 21 days following vaccination was defined according to the shortest recommended interval between first and second doses. Although the second dose is not a consideration for Ad26.COV2.S, I used the same time at risk to ensure comparability between the results. Therefore, further evidence is required to understand the long-term adverse events of vaccination and SARS-CoV-2 infection. Larger cohorts are also needed to study the effect of vaccination on different age groups, particularly among younger populations. Moreover, while reassuring that a few outcomes of transverse myelitis were seen in this study, additional data may allow for this particular event to be studied in greater detail.

The SCCS model makes certain assumptions that should be met to provide valid and unbiased estimates. If vaccination is deferred because of an event before vaccination, this may violate the assumption that an outcome event should not alter the probability of subsequent exposure to vaccination. In my analysis, I applied a three-week pre-vaccination period and noticed the incidence rate ratios were lower during this period, indicating that the occurrence of an event might have reduced the likelihood of vaccination. However, the probability of receiving one vaccine may be affected by events other than the study outcomes. For example, individuals with myocarditis were less likely to receive an mRNA vaccine, and these people may have a different risk of the study outcomes after vaccination. The SCCS results from CPRD showed a lower risk of Bell’s palsy during the 1-7 days and 1-21 days after the first dose ChAdOx1 vaccine. This suggested that some outcomes may be under-ascertained, particularly if healthcare access was impacted by the COVID-19 pandemic, and there may be under-recording of recent events due to data lags. These potential violations of the assumptions for SCCS may underestimate the true risk in this analysis.

6.5.3 Study limitations

This study has limitations.

First, the CPRD Aurum data only included primary care data from the UK. Therefore, diagnoses from inpatient settings may not be captured and the absolute risk may be underestimated. However, a previous study has shown that CPRD primary care data captures immune-mediated neurological disorders like GBS relatively accurately, even in the absence of linked hospital data.[201] Meanwhile, the SIDIAP database did have patient-level hospital linkage, and results were consistent in both databases.

Second, the historical comparator method may be affected by confounding by indication. Although I account for differences in age, individuals vaccinated also had more comorbidities than the comparator cohort and may well differ in other, unobserved characteristics.

Third, comparisons between the vaccinated and historical cohorts might be limited due to changes over time, including seasonal variations[202] in outcomes and changes in exposure to other viral infections during the pandemic which might have reduced the incidence of immune-related neurological events.[203]

In addition, given public concerns over vaccine safety and SARS-CoV-2 complications, individuals vaccinated against or infected with SARS-CoV-2 might have been more prone to seek care when showing symptoms. Although seasonality was not addressed in the historical comparison method, I adjusted the SCCS for age and seasonality to control for time-varying confounding.

Fourth, since I excluded individuals with a recent history of the same immune-related neurological events, the results are not generalizable to individuals with chronic relapsing conditions such as multiple sclerosis or Devic's disease. This exclusion criteria would have also led to a depletion-of-susceptible bias in the second dose vaccination cohorts (because those who had an event following the first dose were excluded from the second dose cohort).

6.5.4 Conclusion

I found no safety signal for any of the studied neuro-immune events after vaccination against COVID-19. Infection with SARS-CoV-2 was, however, associated with an increased risk of Bell's palsy, encephalomyelitis and GBS.

Chapter 7 The comparative risk of different COVID-19 vaccines

7.1 Chapter summary

In this chapter, I examined the comparative risk of developing thrombosis with thrombocytopenia syndrome or thromboembolic events after different COVID-19 vaccines (Objective 4). I used datasets from the Europe and the US, and found an increased risk of thrombocytopenia after a first dose of the ChAdOx1 vaccine compared with BNT162b2 vaccine.

7.2 Background

After millions of vaccine doses were administered in large-scale immunisation campaigns, spontaneous case reports of thrombosis with thrombocytopenia syndrome (TTS) usually following the first dose of adenovirus vector-based vaccines emerged [204–206]. Although fewer concerns have been raised about similar safety signals for BNT162b2 (mRNA COVID-19 vaccine), instances of immune thrombocytopenia have also been observed among recipients of this vaccine[207]. Initial investigation by the EMA's Pharmacovigilance Risk Assessment Committee (PRAC) concluded that there might be causal relationship between vaccines and TTS. Therefore, the EMA updated the corresponding product information to include TTS as a very rare side effect.[208] As these unusual blood clots in combination with thrombocytopenia were reported predominantly among women aged younger than 60 years, several European countries restricted the use of adenovirus vector-based vaccines among the younger age groups as a precautionary measure, while some countries such as Denmark paused the use of the AstraZeneca COVID-19 vaccine.[209] While the pathogenesis was not yet fully understood, an immune response leading to the development of pathologic platelet-activating antibodies has been suggested, and named "vaccine-induced immune thrombotic thrombocytopenia" (VITT).[206,210] Although these events were very rare, based on the worldwide roll-out of these vaccines absolute numbers of affected patients may accumulate fast.

Given the importance of TTS risk for COVID-19 vaccines in various countries, insights into the magnitude of TTS risk associated with COVID-19 vaccination on a large scale were urgently needed. While some observational studies have been conducted in the European area to investigate the risk of TTS events following COVID-19 vaccination[50,51,211,212], there was currently no clear evidence on the comparative safety profile of different vaccines. In this study, the aim was to quantify the comparative risk of TTS or thromboembolic events (TE) associated with vaccination with adenovirus-based compared to mRNA-based COVID-19 vaccines. Additionally, I aimed to examine the comparative risk between the two mRNA vaccines, the BNT162b2 and mRNA-1273.

7.3 Method

7.3.1 Study design

This was a comparative cohort study. Federated network analysis was conducted with multiple datasets that were mapped to the OMOP CDM. Details of the OMOP CDM and federated network analysis were explained in Chapter Four.

The cohort design allowed me to compare the risk of events of interest for two populations with similar characteristics, as discussed in chapter Three: study design. It has been widely used in the hypothesis testing stage of vaccine safety surveillance. In this study, I used alternative COVID-19 vaccine as active comparator and included only new users of either vaccine.

7.3.2 Data sources

Five European countries (France, Germany, Netherlands, Spain, and the United Kingdom) and two databases from the United States informed the analyses.

- IQVIA Longitudinal Patient Data France (LPD France)
- IQVIA Disease Analyser Germany (DA Germany)
- The Integrated Primary Care Information (IPCI)

- The Information System for Research in Primary Care (SIDIAP)
- The Clinical Practice Research Datalink (CPRD) AURUM
- The IQVIA hospital charge data master (IQVIA Hospital CDM)
- The US Open Claims

Details of the contributing databases are available in chapter four.

The study period to identify exposures and outcomes covered from December 2020 (first vaccines administered) until mid-2021, depending on the latest data release available in each of the contributing databases.

7.3.3 Study participants

The study population was made of adults registered in any of the contributing databases and exposed to at least one dose of a brand-specific COVID-19 vaccine during the study period. Participants were required to have at least one year of history in the database prior to the index vaccination date. I excluded individuals who did not have a vaccine brand specified (unspecific vaccine codes) during the study period. I also excluded people who received their second dose within 14 days after the first dose, as these were likely errors in vaccination records.

Four brands of COVID-19 vaccines were included for analyses: ChAdOx1, Ad26.COV2.S, BNT162b2 and mRNA-1273. Vaccines were identified by procedure, drug, or observation codes in each database. I built the first- and second-dose cohorts for ChAdOx1, BNT162b2 and mRNA-1273. A single-dose cohort was built for Ad26.COV2.S as it was approved for a single-dose schedule. The primary comparisons were made between the adenovirus-based vaccines (ChAdOx1 or Ad26.COV2.S, the “target”) and mRNA vaccines (BNT162b2 or mRNA-1273x, the “comparator”).

In the secondary analysis, patients in vaccinated with first and second dose cohorts for BNT162b2, the “target”, and mRNA-1273, the “comparator”, were compared as well. This was only conducted in the two US databases: the US Hospital CDM and the US Open Claims.

The index date for the first dose vaccination cohorts was defined as the date of receiving the first COVID -19 vaccination for a specific brand. The index date for the second dose vaccination cohorts was defined as the date of the second COVID-19 vaccination. Individuals were followed from their index date to 28 days after vaccination, death, or loss of visibility in the database (e.g., person leaving the practice in electronic health records data, or end of continuous enrolment in claims data), whichever came first.

In the US Open claims data, given that the vaccinated cohort size was large and therefore highly computing expensive, I derived a 20% random sample for each cohort. All comparisons were limited to those with sufficient statistical power, defined as having a minimal detective relative risk (MDRR) less than five.

7.3.4 Primary outcomes

The primary outcomes of interest were thromboembolic events and thrombosis with thrombocytopenia syndrome.

Thromboembolic events of interest included: deep-vein thrombosis, pulmonary embolism, venous thrombo-embolism as a composite of deep-vein thrombosis or pulmonary embolism, cerebral venous sinus thrombosis (CVST), splanchnic and visceral vein thrombosis (SVT); ischemic stroke, myocardial infarction, and arterial thromboembolism as a composite of ischemic stroke, myocardial infarction and other rare arterial thromboembolisms, such as intestinal infarction. More detail on how the study outcomes were defined can be found in Appendix to chapter three.

The definition of TTS was based on the case definition proposed by the Brighton Collaboration: the occurrence of any thromboembolic event of interest with concurrent thrombocytopenia within 10 days before or after a thromboembolic event. [213] Thrombocytopenia was identified by a diagnostic code or measurement of <150,000 platelets per microliter of blood, as proposed by the Brighton collaboration. This definition has been used in previous OMOP CDM-based studies.[214]

I included two sensitivity analyses by using alternative TTS definitions: 1. Requiring the concurrent thrombocytopenia happened within 5 days before or after the thromboembolic event after vaccination. 2. Reducing the threshold to <100,000 platelets per microliter for the definition of thrombocytopenia based on laboratory data.

7.3.5 Negative control outcomes

As detailed in chapter 3, negative controls are outcomes that are not expected to have casual effect with the exposure, which in this case was COVID-19 vaccination. I used 92 negative control outcomes previously used for vaccine safety research[215]. These were pre-specified based on clinical knowledge and previous literature, validated by two clinicians, and tested in previous work on other vaccine safety projects[214]. The codes used to identify each of these negative control outcomes were available at appendix to chapter three, table 5.

7.3.6 Covariates

Baseline patient characteristics based on information before the index date were reported, including: demographics (age, sex, index year, and index month); medical history any time before cohort index; composite comorbidity (Charlson Index-Romano adaptation[216]) and thrombosis score (CHA2DS2-VASc, congestive heart failure hypertension – vascular disease[217]); records and total number of medicines, procedures, and measurement records observed in the 6 months before cohort index.

7.3.7 Propensity score

I used the propensity-score matching method to reduce observed confounding. Firstly, large-scale propensity scores were calculated for each pair of comparisons (target and comparator) using large-scale L1 regularized logistic regression,[101,186] which included all available baseline patient

characteristics in the databases. The covariates included: Demographics (Age group in 10-year bands, Sex), Index year and month; Condition group record for the concept or any of its descendants observed any time prior to cohort index; Drug exposure group record for the concept or any of its descendants observed during 180 days to 4 days prior to cohort index; Procedure occurrence record for the concept or any its descendants observed during 180 days to 4 days prior to cohort index; Measurement record for the verbatim concept observed during 180 days to 4 days prior to cohort index; Comorbidities index including the Charlson Index,[216] and the CHA2DS2-VASc (Congestive heart failure hypertension- vascular disease)[217]using conditions all time on or prior to cohort index. I also included some summary counts of drugs/procedures/visits/observations/measurements as a proxy of health care utilization, these included: Number of distinct drugs observed in 180 days to 4 days prior to cohort index (defined as unique RxNorm ingredient concepts); Number of distinct procedures observed in 180 days to 4 days prior to cohort index (defined as unique SNOMED concepts); Number of distinct observations observed in 180 days to 4 days prior to cohort index; Number of distinct measurements observed in 180 days to 4 days prior to cohort index (defined as unique SNOMED concepts); Number of visits observed in 180 days to 4 days prior to cohort index.

The propensity score models were also reviewed by epidemiologist and clinician to ensure that the selected variables are potential confounders. Due to the computing cost, a random sample of 250,000 participants were used to fit the propensity score model if cohort size was larger than this number, the models were then used to compute the propensity scores for all subjects.

Each person in the target cohort was matched to a different person from the comparator cohort without replacement based on greedy PS with a maximum caliper width of 0.2 standard deviations of the logit of the derived propensity score at a ratio of up to 1:4. If the target cohort was larger than the comparator cohort, reverse matching was allowed, and a ratio of 4:1 was used.

The propensity score estimating and matching was conducted for each target-comparator. In the analysis, individuals who experienced the outcomes prior to index date were excluded from the analysis of that outcome.

7.3.8 Statistical analysis

Descriptive statistics were used to report the baseline characteristics for each cohort before and after the matching. I reported the 28-day database-specific incidence rate and corresponding 95% confidence intervals for each event in the original and matched cohorts as well.

I used three diagnostics to evaluate measured confounding, statistical power, and unmeasured confounding. If the analysis for a dataset-target-comparator failed the measured confounding and statistical power diagnostics, the analysis was stopped to avoid biased estimates. First, to control for measured confounding, I considered satisfactory only target-comparator vaccine pairs with all covariates showing a standardized mean difference (SMD) lower than 0.1 after propensity-score matching. Second, I calculated the minimum detectable rate ratio (MDRR) for each outcome of interest and excluded analyses with inadequate statistical power, defined as MDRR>5[218]. Third, I used negative control outcomes to assess residual or unmeasured confounding. If there was less than 20% of the negative control outcomes showing association with the vaccine exposure, the analysis was less likely impacted by systematic errors. Results for those that failed the unmeasured confounding diagnostic are reported, but only empirically calibrated estimates should be used for clinical explanation or implementation.

I then used Poisson regression to calculate the incidence rate ratio (IRR) of outcomes according to each target-comparator vaccine pair. I used empirical calibration to account for residual systematic error due to potential unobserved confounding [219]. This approach has been used in many previous studies in different clinical areas, including COVID-19 repurposed therapies [109,220,221],

and was acknowledged in the latest version of the EnCePP guide on methodological standards in pharmacoepidemiology.[222]

Only treatment effect estimates that passed the covariate balance diagnostic were meta-analysed. Meta-analysis was performed using random effects models with the heterogeneity reported by I^2.

All analyses were conducted in R 3.6.0 using the open-source OHDSI tool-stack, including the Cyclops, CohortMethod, and EvidenceSynthesis packages. (https://cran.r-project.org/web/packages/Cyclops/, https://cran.r-project.org/web/packages/EvidenceSynthesis/ , https://ohdsi.github.io/CohortMethod/)

7.4 Results

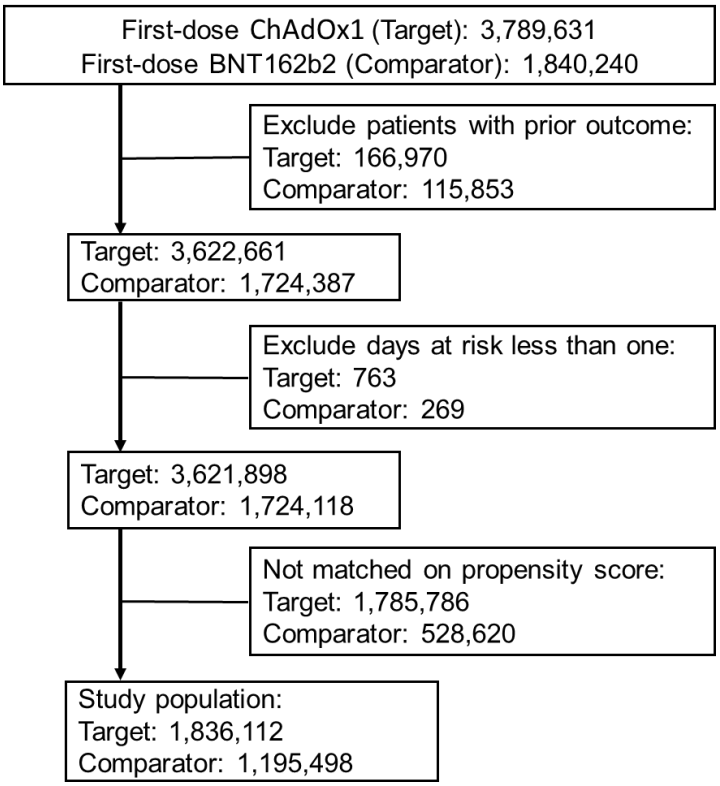
7.4.1 Study population

I identified a total of 4.6 million people vaccinated with a first dose of ChAdOx1 (3,789,631 UK CPRD, 606,399 Spain SIDIAP, 98,562 Germany DA, 27,698 France LPD, and 71,083 the Netherlands IPCI) and 1.6 million people vaccinated with a second dose of ChAdOx1 (1,195,626 UK CPRD Aurum, 307,344 Spain SIDIAP, 31,200 Germany DA, 15,067 France LPD, and 38,884 the Netherlands IPCI). I also identified 1.1 million people vaccinated with single dose Ad26.COV2.S in three databases (37,723 Germany DA, 138,351 Spain SIDIAP, and 939,748 US Open Claims). Similarly, I identified 10.6 million people vaccinated with a first dose of BNT162b2 (1,840,240 UK CPRD Aurum, 391,063 Germany DA, 6,055,754 US Open Claims, and 2,027,950 Spain SIDIAP), and 7.7 million people vaccinated with a second dose (1,369,238 UK CPRD Aurum, 321,099 Germany DA, 4,450,735 US Open Claims, and 1,357,509 Spain SIDIAP). Finally, I identified 4,261,016 people vaccinated with a first dose of mRNA-1273 in US Open Claims, and 2,938,023 people vaccinated with a second dose in US Open Claims.

Cohort characteristics stratified by vaccine and dose are summarised in Appendix tables 1 (European datasets), and 2 (US datasets).

When comparing first dose ChAdOx1 vaccinees to first dose BNT162b2 vaccinees in UK CPRD Aurum data, there were noticeable differences in baseline patient characteristics before propensity score matching (Table 7-2). BNT162b2 vaccinees were more likely to be female (58.2% vs. 51.5%) and older in age, and had a higher prevalence of comorbidities of interest. They were also more likely to use common medications such as hypertension and diabetes treatments. When compared to BNT162b2 users, mRNA-1273 users were older, had a higher prevalence of comorbidities, and used more medications.

Figure 7-1 Illustrative population flowchart using CPRD Aurum, first dose ChAdOx1 (target) vs BNT162b2 (comparator), and outcome of thrombocytopenia as an example.



*In practice, patients with prior outcomes were excluded after the propensity score matching. The propensity scores were calculated and matched once for each targe-comparator combination.

To reduce measured confounding, I estimated the propensity score for each vaccine pair per database (database-target-comparator combination). The propensity score matching led to a final cohort of 1.8 million ChAdOx1 vaccinees and 1.2 million vaccinees BNT162b2 vaccinees (Figure 7-1, Table 7-2). The pre- and post-matching baseline patient characteristics for each database-target-comparator combination are described in Table 7-2, Table 7-3,Table 7-4. Among combinations that passed the observed and unobserved confounding diagnostic tests, which is explained in the next paragraph, differences between pre and corrected post-matching characteristics were similar.

7.4.2 Confounding and analysis feasibility

To assess the robustness of these analyses, I used three diagnostic tests based on measured confounding, statistical power, and unmeasured confounding as explained in the previous section. Table 7-1 summarises the results of these diagnostics.

Firstly, to avoid bias due to confounding, I did not analyse cohorts with substantial differences after matching: 14 analyses passed this diagnostic, where no patient characteristic had a standardised mean difference of ≥0.1 after propensity score matching. The measured confounding requirements were met by all available comparisons in the UK CPRD Aurum, the Netherlands IPCI, and the US Open Claims. In Spain SIDIAP, only Ad26.COV2.S compared with BNT162b2 showed covariate balance after matching. Other combinations of database, target, and comparator that passed the covariate balance test included: first and second dose Germany DA ChAdOx1 compared with BNT162b2, Germany DA Ad26.COV2.S compared with BNT162b2, France LPD ChAdOx1 compared with first dose BNT162b2, and France LPD ChAdOx1 compared with first dose mRNA-1273. Conversely, no analysis was conducted in the US Hospital CDM, because residual confounding was noted (standardised mean difference >0.1 for ≥1 variables).

Table 7-1 Summary of diagnostics at the database-target-comparator level.

DIAGNOSTIC	Definition	*1st dose ChAdOx1 vs BNT162b2*	*1st dose ChAdOx1 vs mRNA-1273*	*2nd dose ChAdOx1 vs BNT162b2*	*2nddose ChAdOx1 vs mRNA-1273*	*Ad26.COV2.S vs BNT162b2*	*Ad26.COV2.S vs mRNA-1273*
UK CPRD Aurum							
Covariate balance	*SMD<0.1 for all*	√	n/a	√	n/a	n/a	n/a
Power	*MDRR<5 for 1+ outcomes*	10	n/a	9	n/a	n/a	n/a
Systematic error	*% NCOs not associated w exposure*	98.0%	n/a	94.4%	n/a	n/a	n/a
Spain SIDIAP							
Covariate balance	*SMD<0.1 for all*	x	x	x	x	√	n/a
Power	*4<5 for 1+ outcomes*					8	n/a

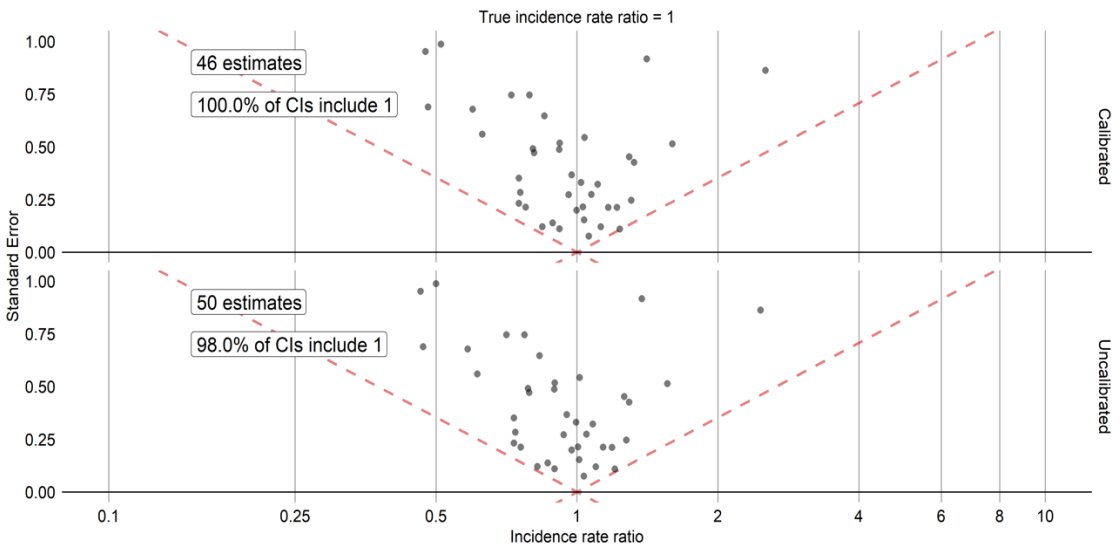
Systematic error	*% NCOs not associated w exposure*					96.4%	n/a
GERMANY DA							
Covariate balance	*SMD<0.1 for all*	√	x	√	x	√	n/a
Power	*MDRR<5 for 1+ outcomes*	7	0	0	0	1	n/a
		1st dose ChAdOx1 vs BNT162b2	*1st dose ChAdOx1 vs mRNA-1273*	*2nd dose ChAdOx1 vs BNT162b2*	*2nddose ChAdOx1 vs mRNA-1273*	*Ad26.COV2.S vs BNT162b2*	*Ad26.COV2.S vs mRNA-1273*
Systematic error	*% NCOs not associated w exposure*	84.2%		$		70.0%	n/a
FRANCE LPD							
Covariate balance	*SMD<0.1 for all*	√	√	x	x	n/a	n/a
Power	*MDRR<5 for 1+ outcomes*	0	0	0	0	n/a	n/a
Systematic error	*% NCOs not associated w exposure*	83.3%	$			n/a	n/a
Netherlands IPCI							
Covariate balance	*SMD<0.1 for all*	√	n/a	√	n/a	√	√
Power	*MDRR<5 for 1+ outcomes*	3	n/a	0	n/a	0	0
Systematic error	*% NCOs not associated w exposure*	$	n/a	$	n/a	$	$
IQVIA US OpenClaims							
Covariate balance	*SMD<0.1 for all*	n/a	n/a	n/a	n/a	√	√
Power	*MDRR<5 for 1+ outcomes*	n/a	n/a	n/a	n/a	12	13
Systematic error	*% NCOs not associated w exposure*	n/a	n/a	n/a	n/a	72.2%	67.9%
IQVIA US HOSPITAL CDM							
Covariate balance	*SMD<0.1 for all*	n/a	n/a	n/a	n/a	x	x
Power	*MDRR<5 for 1+ outcomes*	n/a	n/a	n/a	n/a	0	0
Systematic error	*% NCOs not associated w exposure*	n/a	n/a	n/a	n/a		

SMD: standardised mean difference. MDRR: minimal detective relative risk. NCOs: negative control outcomes.
Grey = Excluded from further analyses; Orange = Included for meta-analysis, underpowered for database-specific effect estimation; Green = Systematic error detected, only calibrated IRRs should be interpreted. $: empirical calibration was not conducted due to insufficient NCOs estimated.

Secondly, eight analyses had sufficient statistical power for at least one outcome, as indicated by a minimum detectable rate ratio <5. However, France LPD failed the power diagnostics for all study outcomes. Therefore, no database specific estimates were reported for France LPD, although this database contributed to meta-analyses (see below).

Thirdly, I identified residual confounding due to unobserved covariates and systematic error using negative control outcomes. Of seven combinations of database, target, and comparator with sufficient negative control outcomes, three had >20% associated with vaccine use (Germany DA Ad26.COV2.S v BNT162b2, US Open Claims Ad26.COV2.S v BNT162b2, and US Open Claims Ad26.COV2.S v mRNA-1273), suggesting the presence of substantial systematic error. For example, Figure 7-3 showed the incidence rate ratios for the negative controls outcomes when comparing the 1st dose ChAdOx1 cohort to the BNT162b2 in the UK CPRD data. A total of 50 negative control outcomes were included and IRRs were estimated, and 98.0% of the 95% confidence intervals included 1, suggested that there very low chance of systematic error in such analysis.

Figure 7-2 Passed systematic error diagnostic: <20% negative control outcomes associated with exposure in the uncalibrated analyses (CPRD Aurum, 1st dose ChAdOx1 vs BNT162b2)



*Estimates below the diagonal dashed lines are statistically significant (alpha = 0.05) different from the true effect size (expected IRR=1). A well-calibrated estimator should have the true effect size within the 95 percent confidence interval 95 percent of times.

On contrast, I observed potential systematic error in the comparison between Ad26.COV2.S to BNT162b2 or mRNA-1273 in the US Open Claims dataset (Figure 7-3). There were only 72.2% (figure a) and 67.9% (figure b) of the confidence intervals included 1, for BNT162b2 and mRNA-1273, respectively. The incidence rate ratio for the majority of these negative control outcomes was greater than one, indicating that the uncalibrated results overestimated risks, and only calibrated results should be considered and interpreted as adequate.

Figure 7-3 Systematic error in the comparison between Ad26.COV2.S and mRNA vaccines in US Open Claims

Figure a. Systematic error in the analysis of Ad26.COV2.S vs 1-dose BNT162b2

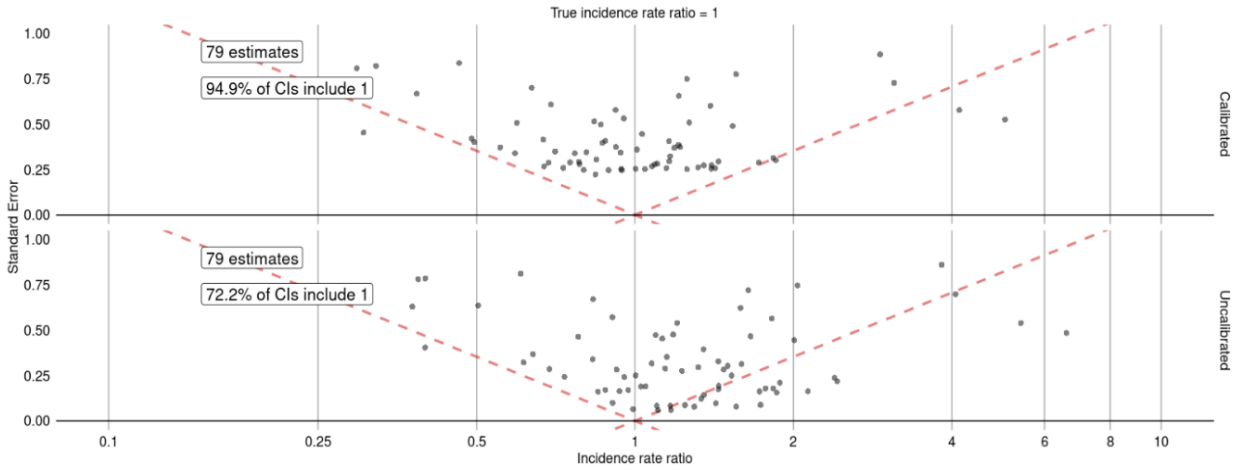
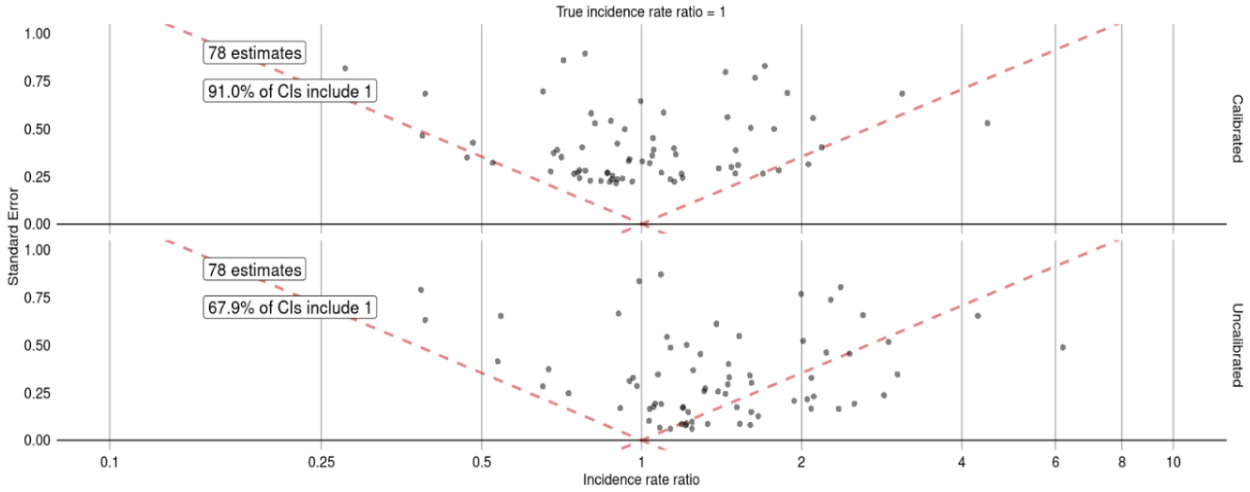


Figure b. Systematic error in the analysis of Ad26.COV2.S vs 1-dose mRNA-1273



* Estimates below the diagonal dashed lines are statistically significant (alpha = 0.05) different from the true effect size (expected IRR=1). A well-calibrated estimator should have the true effect size within the 95 percent confidence interval 95 percent of times.

For Germany DA ChAdOx1 compared with second dose BNT162b2, France DA ChAdOx1 compared with first dose mRNA-1273, and all comparisons within the Netherlands IPCI, I observed too few negative control outcomes, which precluded the use of empirical calibration.

In the secondary analysis of comparisons between the mRNA vaccines, all covariates were balanced, with no covariates showed standardized mean difference over 0.1 after propensity scores matching. However, 31.2% of the negative controls outcomes in Open Claims were

associated with exposure, suggesting residual (unobserved) confounding (Appendix to chapter seven). Most of the estimates for these NCOs had an incidence rate ratio >1, suggesting that the uncalibrated results overestimated risks, and that only calibrated results should be considered adequate. In the US Hospital CDM data, there was not enough power to analyse TTS events.

7.4.3 Comparative safety

In the matched cohorts from Germany and the UK databases, there were 862 thrombocytopenia events among the first dose ChAdOx1 recipients, and 520 events after a first dose of BNT162b2. Meta-analyses showed an increased risk of thrombocytopenia after first dose ChAdOx1 compared with BNT162b2, with a pooled calibrated incidence rate ratio of 1.33 (95% confidence interval 1.18 to 1.50 (Figure 7-4), a calibrated incidence rate difference of 1.18 (0.57 to 1.8) per 1,000 person years, and an absolute risk difference of 8.21 (3.59 to 12.82) per 100,000 recipients. In UK CPRD Aurum data, 827 and 442 thrombocytopenia events occurred after first dose ChAdOx1 and BNT162b2, respectively. The incidence rates were 6.06 per 1,000 person years (95% confidence interval 5.65 to 6.48) and 4.89 (4.45 to 5.37), respectively, with a calibrated incidence rate ratio of 1.31 (1.16 to 1.49). This finding was not replicated in the Germany DA data, where the calibrated incidence rate ratio was 1.01 (0.63 to 1.62). There was no difference in the risk of thrombocytopenia between the second doses of ChAdOx1 and BNT162b2 (meta-analytical calibrated incidence rate ratio 0.93 (0.78 to 1.11)). Similarly, I discovered no increased risk of thrombocytopenia after Ad26.COV2.S versus first dose BNT162b2 (meta-analytically calibrated incidence rate ratio 1.08 (0.58 to 1.99).

For venous thromboembolism and deep vein thrombosis, the meta-analysis was unreliable because of heterogeneity (I^2 values of 65% and 86%, respectively). After the first dose of ChAdOx1 versus BNT162b2, there was no evidence of increased risk of venous thromboembolism in either the German DA (calibrated incidence rate ratio 1.61 (95% confidence interval 0.92 to 2.83)) or the UK CPRD Aurum (0.91 (0.78 to 1.06); Table 7-5). An increased risk of deep vein thrombosis was seen

after first dose ChAdOx1 compared with BNT162b2 in Germany DA (2.62, 1.34 to 5.13), but not replicated in UK CPRD Aurum data (0.89, 0.71 to 1.11; Table 7-5). There was no increased risk of pulmonary embolism seen in either database, with calibrated incidence rate ratio 0.93 (0.77 to 1.12) and 0.69 (0.26 to 1.83) in UK and German data respectively.

In pooled meta-analysis or database specific analyses, no differential risks of venous thromboembolism, deep vein thrombosis, or pulmonary embolism were found when comparing second dose ChAdOx1 with BNT162b2(Figure 7-4). In line with this, no association was seen between vaccination with Ad26.COV2.S and any venous thromboembolic event in database specific (Table 7-5) or pooled meta-analysis. Regarding rare thrombosis, the meta-analysis showed a lower risk of intestinal infarction for the single dose Ad26.COV2.S users compared with first dose BNT162b2, with a pooled calibrated incidence rate ratio of 0.37 (0.15 to 0.89), an incidence rate difference of −0.41 (−1.17 to 0.35) per 1,000 person years, and an absolute risk difference of −3.34 (−9.77 to 3.09) per 100,000 vaccinations (Figure 7-4). I did not find any difference in risk between cohorts for any other rare thrombotic events.

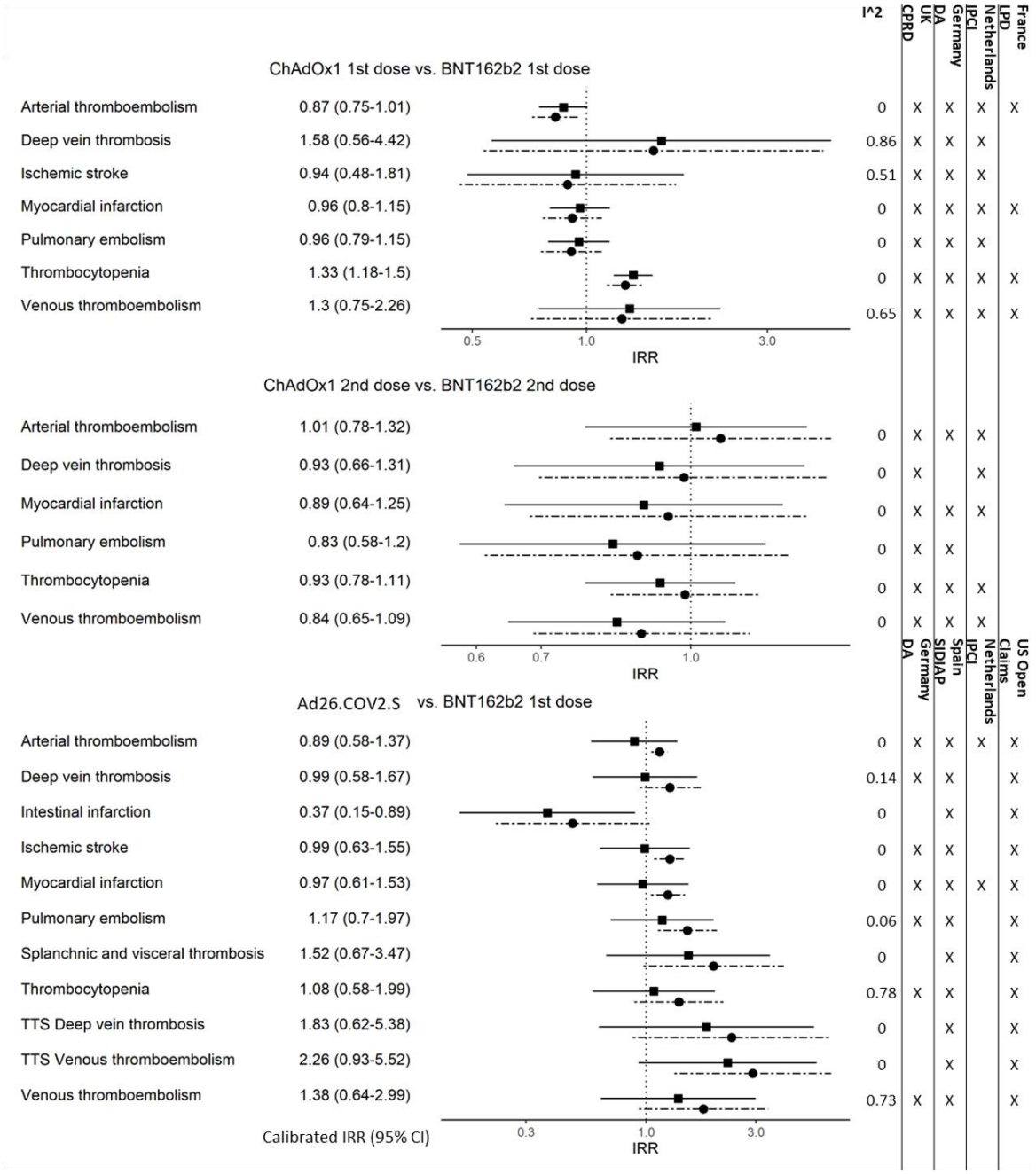
For composite arterial thromboembolism, the pooled calibrated incidence rate ratio for first dose ChAdOx1 compared with first dose BNT162b2 was 0.87 (95% confidence interval 0.75 to 1.01). The two reliable database specific analyses in Table 7-5 showed consistent findings—the calibrated incidence rate ratio was 0.85 (0.73 to 0.99) in CPRD UK and 0.76 (0.41 to 1.39) in Germany DA. In line with this, there were no differences in the risk of arterial thromboembolism, ischaemic stroke, or myocardial infarction after the second dose of ChAdOx1 versus the second dose of BNT162b2 or after Ad26.COV2.S versus the first dose of BNT162b2. When ischaemic stroke and myocardial infarction were separately examined, similar results were obtained.

Thrombosis with thrombocytopenia syndrome was very rare, and could only be analysed in UK data for ChAdOx1, and in US and Spanish data for Ad26.COV2.S. There was no evidence of increased risk of thrombosis with thrombocytopenia syndrome in UK CPRD after first dose ChAdOx1 compared

with first dose BNT162b2 (calibrated incidence rate ratio 1.29 (95% confidence interval 0.94 to 1.77)). The calibrated incidence rate ratio after second dose was 1.16 (0.71 to 1.89). For comparing Ad26.COV2.S with BNT162b2, meta-analyses were possible for venous thromboembolism with thrombocytopenia syndrome and deep vein thrombosis with thrombocytopenia syndrome. In a meta-analysis of US and Spanish data, the pooled calibrated incidence rate ratio was 2.26 (0.93 to 5.52) for venous thromboembolism with thrombocytopenia syndrome and 1.83 (0.62 to 5.38) for deep vein thrombosis with thrombocytopenia syndrome (

Figure 7-4). Database specific estimates from US Open Claims were consistent with the pooled results (Table 7-5).

Figure 7-4 Meta-analytic estimates of calibrated incidence rate ratios.



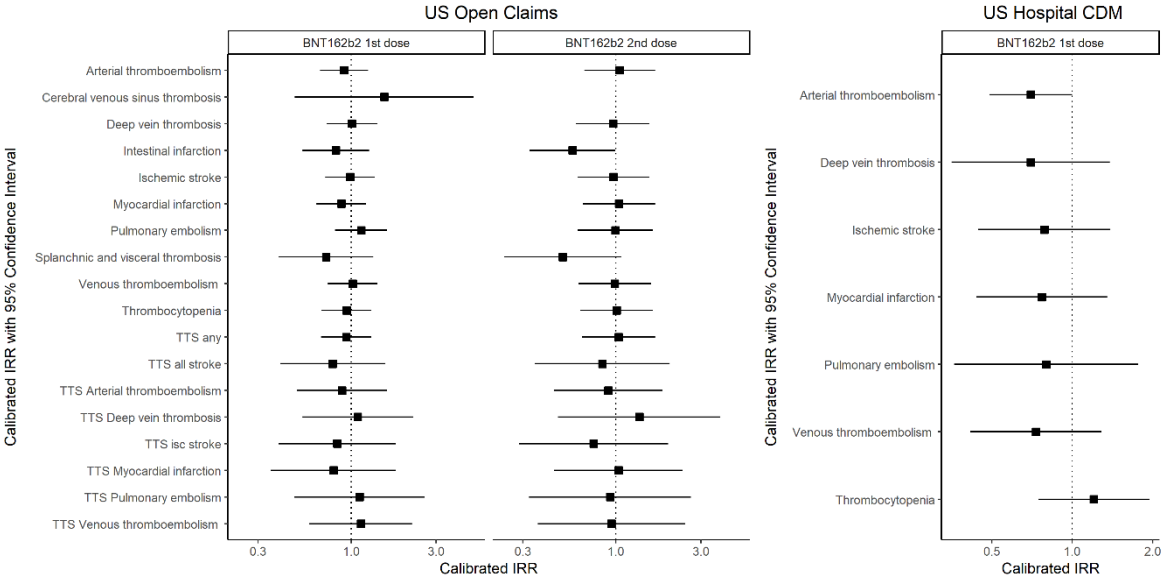
Square and solid line: calibrated estimates and 95% confidence intervals; Dot and dashed line: uncalibrated calibrated estimates and 95% confidence intervals; TTS-VTE: Venous thrombo-embolism with -thrombocytopenia syndrome; TTS-DVT: Deep vein thrombosis with thrombocytopenia syndrome.

In the secondary analysis between mRNA vaccines, I observed a slightly lower risk of arterial thromboembolism(ATE) with BNT162b2 vs mRNA-1273 in the Hospital CDM data, with calibrated

IRR 0.70 (95% confidence interval 0.49 to 1.00). This was not replicated in Open Claims: calibrated IRR 0.91 (0.67 to 1.24). Similarly, a lower risk of intestinal infarction BNT162b2 (calibrated IRR 0.57 (0.33 to 1.00)) and a marginal reduction in risk of splanchnic-visceral thrombosis (calibrated IRR 0.50 (0.24 to 1.07)) was seen after a second dose of BNT162b2 (vs 2-dose mRNA-1273) in Open Claims. No other differences in risk were detected between BNT162b2 and mRNA-1273. (Figure 7-5)

Figure 7-5 Incidence rate ratio of TTS or TE events comparing BNT162b2 to mRNA-1273 vaccinees.

Legend: TTS: Thrombosis with thrombocytopenia syndrome; IRR: incidence rate ratio. Solid line: calibrated IRR, dashed line: uncalibrated IRR.



I conducted two sensitivity analyses, the first restricting the window of thrombocytopenia to within 5 days before or after the thrombosis event, and the second reducing the threshold of platelet count from lower than 150,000 platelets per microliter to lower than 100,000 platelets per microliter. For both sensitivity analyses, calibrated IRRs were only estimated for a few TTS outcomes, with results consistent with the main analysis (Table 7-6).

7.5 Discussion

7.5.1 Summary of key results

In this cohort study, I used routinely collected health data from five European countries and the USA, and produced risk estimates after minimising confounding and systematic errors. I observed a 30% increased risk of thrombocytopenia and TTS after the first dose ChAdOx1 versus first dose BNT162b2 vaccination. I found no consistent differences in risk of TE or TTS between BNT162b2 and mRNA-1273 when comparing different mRNA-based vaccines in two large US databases.

7.5.2 Research in context

Head-to-head comparison on safety between vaccines have been lacking. By the time of this study, there was no other real-world cohort study comparing the risks of TTS and thromboembolic events after adenovirus-based or mRNA-based COVID-19 vaccination to my best knowledge. Instead, risk of TTE and/or TE events have been investigated in other study designs, and the conclusion remained unclear. For example, risk of potential adverse events was assessed between vaccinated and unvaccinated people using observed-to-expected analysis [223,224], in cohort study using EHRs [225], with non-risk period after vaccination [226], with SCCS design [211], and were studied in systematic review as well [227,228].

In this study, TTS was very rare, and I did not find any statistically significant increase in risk with either adenovirus-based vaccine compared with any mRNA-based vaccine. A US case series using the Vaccine Adverse Event Reporting System estimated rates of TTS were 3.83 per million vaccine doses of the Ad26.COV2.S and 0.00855 per million vaccine doses of mRNA-based COVID-19 vaccines [229]. However, the authors stated that TTS cases reported after mRNA vaccines have different demographic characteristics and medical histories than cases reported after Ad26.COV2.S. Passive surveillance tends to suffer from underreporting. In comparison, this study

used routinely collected health data to estimate the comparative risks of vaccines, thereby reducing surveillance bias.

Thrombosis with concomitant thrombocytopenia was first reported after administration of the ChAdOx1 vaccine in early 2021 [204,205]. A disproportionality analysis using the World Health Organisation's VigiBase database reported a safety signal for CVST and ischemic stroke for the ChAdOx1, BNT162b2, and mRNA-1273 vaccines [230]. The authors called for well-designed comparative safety studies on adverse events of all three vaccines. According to a study based on Danish and Norwegian data, there is an increase in venous thromboembolism, pulmonary embolism, and CVST after vaccination [210]. Although these studies gave valuable information about the frequency of adverse events reported after vaccination, they did not quantify the connection between COVID-19 vaccination and the occurrence of these events when taking into account potential risk factors.

Association between VTE and vaccination have been examined with different study designs. A nested case-control study from Scotland suggested no increased risk of venous thrombo-embolism with either the first dose ChAdOx1 or BNT162b2 vaccine [212]. Case-control studies, however, have recently come under fire for selection bias and their inability to account for confounding [231,232]. Hippisley-Cox et al. conducted a self-controlled case series analysis of approximately 30 million vaccinated people in England [211]. They provided epidemiological evidence of an increased risk of hospital admission with thrombocytopenia and venous thrombo-embolism after ChAdOx1, as well as an increased risk of cerebral venous sinus thrombosis after ChAdOx1 and BNT162b2. They also found a higher risk of arterial thromboembolism after BNT162b2, but not after ChAdOx1.

Study outcomes of cerebral venous sinus thrombosis, and splanchnic and visceral thrombosis were also very rare. Cerebral venous sinus thrombosis was observed in approximately 16.34 per million doses of ChAdOx1 and 12.60 per million doses of BNT162b2, according to Kerr et al. [222] Previous study using the same data source (UK CPRD Aurum) found that the rates of cerebral venous sinus

thrombosis after first dose of ChAdOx1 were higher than expected rates among the background population.[224] In a self-controlled case series analysis using data from England, Scotland, and Wales, ChAdOx1 was associated with an elevated risk of cerebral venous sinus thrombosis in the 28 days after ChAdOx1 vaccination (incidence rate ratio 1.93 (95% confidence interval 1.20 to 3.11)) but not after BNT162b2 vaccination.[231] Similarly, a large record linkage study of hospital admissions in England showed an increased risk of cerebral venous sinus thrombosis after first dose ChAdOx1, seen only in adults aged under 65 years, and not after BNT162b2.[233] In my study, only 13 cases of cerebral venous sinus thrombosis were identified in the pre-matching 1st dose ChAdOx1 cohort, and less than five cases after matching. I was not able to estimate the incidence rate ratio because the very low statistical power.

I used random-effect meta-analysis to pool the estimates from different databases. In the previous study included in chapter Five, I reported that background incidence rates varied across data sources, and suggested the use of analyses within databases for historical rate comparisons.[163] While the incidence rates of events of interest were similar across databases in this study, the relative rates showed substantial heterogeneity for some of the study outcomes. For example, in the analysis of comparing first-dose ChAdOx1 and BNT162b2, there was no evidence of increased risk of DVT (calibrated IRR 0.89 [0.71-1.11]). Whilst in the Germany DA data, a calibrated IRR of 2.62[1.34-5.13] was estimated. This heterogeneity resulted in the very wide confidence interval in the meta-analysis, where a pooled IRR of 1.58 [0.56-4.42] was estimated, with I^2 of 0.86. The pooled results with high heterogeneity showed be explained with caution. Further studies to identify the potential sources of between databases heterogeneity will benefit future research.

For the comparative safety of BNT162b2 and mRNA-1273 vaccines, only few other studies was found. Dickerman et al. conducted a cohort study using the US Veterans' EHR data to estimate the risk of set of potential adverse events up to 38 weeks after vaccination.[234] At 38 weeks, by comparing with mRNA-1273, the researchers found an excess risk of 10.9 events for ischemic stroke

and 14.8 events of myocardial infarction , and 11.3 events of other thromboembolic per 10 000 persons among BNT162b2 vaccinees, corresponding to 17%, 32%, and 20% increased risk, respectively. While looking at risk at 14 days and 42 days post-vaccination, there was no significant difference in risk. This may be partially explained by the difference in effectiveness of preventing post-covid sequelae for these two vaccines. [235] However, due to the Veteran population with median age of 69 years and 93% of male, and different estimated risk window, comparison with the presented study is hard.

7.5.3 Strengths and limitations

The findings of this study should be interpreted in light of its limitations. Due to heterogeneity across data sources, misclassification of exposures and outcomes may have remained. The UK and Spanish data sources captured vaccine information most reliably through their primary care systems and the latter also through linkage to official vaccination registries. In contrast, the German and French outpatient records and two US datasets are expected to include incomplete vaccine records. Inpatient data were only available for the Spanish data sources through linkage and the US hospital database. Some severe study events that cause hospitalisation might not have been captured in other databases, such as the German and French data sources.

I acknowledge the possibility of information bias in this study's outcome ascertainment. Robust methods were used for the creation and transportation of algorithms for the identification of all of the study events. However, previous research by the research group I worked with demonstrates that TTS as identified in these data does not necessarily match the reports of post-vaccination immune-induced TTS in the literature [214]. As a result, it is possible that these findings do not accurately reflect the potential impact of vaccines on uncommon and difficult-to-diagnose events such as vaccine-induced immune thrombotic thrombocytopenia.

Each country had its own immunisation schedule, and the studied vaccines were not all approved at the same time. For example, many countries began vaccinating their most vulnerable populations when the first vaccine, BNT162b2, was approved. By the time they were vaccinating younger, less vulnerable people, more vaccines were available. The vaccine cohorts being compared may therefore have had different background rates of the studied adverse events. Despite the use of propensity score matching and balance diagnostic tests, caution is still needed when interpreting and generalising the results.

I used the propensity score matching method to get comparable targets and comparator groups. However, while the propensity score allows one to remove the average effects of confounding due to the measured baseline covariate, it may mask some important factors that affect the treatment received. For example, in early 2021, when vaccines were prioritised for healthcare workers, the elderly, and clinically vulnerable groups, individuals aged below 40 years old could receive either BNT162b2 or ChAdOx1. With concerns about thromboembolic events, this age group has been offered mRNA vaccines since May 2021. This suggests that among people younger than 40, those who received mRNA vaccines may not be comparable with those who received ChAdOx1. A head-to-head comparison may not be appropriate. As the study data covered until mid-2021, only the first and second waves of the pandemic were represented. However, the proportion of included people with a history of COVID-19 infection before vaccination was balanced in all eligible comparisons, both before and after matching.

I presented the database-specific incidence rates of outcomes for both the original full cohorts and the propensity score matched cohorts in this study. The incidence rates from the full cohorts were crude without any adjustment. While they reflected real-world incidence, they were highly influenced by population characteristics and thus were not directly comparable across cohorts. The incidence rates from matched cohorts, on the other hand, can be compared since the propensity

score matching accounted for the measured confounding. Caution is needed when interpreting these incidence rates as the generalizability of the rates is limited.

While other epidemiological methods have been used in vaccine safety studies, a cohort study with active comparators enabled me to directly estimate the relative risk of developing thromboembolic events or TTS after different COVID-19 vaccines, which is not feasible in self-controlled designs or observed-to-expected analyses. The OMOP CDM allowed me to conduct the same study across different databases, which increased database comparability and allowed me to synthesize the results using meta-analysis.

To reduce bias and confounding and ensure the reported results are reliable, I included several diagnostics in the study design and statistical analysis plan. I used large-scale propensity score (PS) modelling based on an L1 regularized logistic regression to minimise observed confounding. This data-driven approach has been shown to improve the performance of the resulting PS compared to clinically-driven PS estimation, using rich real-world data [82,235]. I examined residual imbalances after matching and did not perform analyses where relevant confounding was observed. I used negative control outcomes to assess the residual systematic error of each comparison. These outcomes were reviewed and tested in previous studies on vaccine safety.[82,236] When systematic error was detected, identified by >20% of the proposed negative control outcomes being associated with the exposure of interest, I conducted empirical calibration to minimize the systematic error. In the OpenClaims data, for example, 72.2% of CIs included 1 in the analysis of Ad26.COV2.S vs one dose BNT162b2. The other outcomes were to the right of the null effect. After empirical calibration, this increased to 94.9%, showing that the calibrated estimates were more reliable. IRRs without calibration should be interpreted with care as they may overestimate risk (Figure 7-3).

Table 7-2 Baseline characteristics of eligible ChAdOx1 cohorts of vaccinated people identified EU based databases.

			UK CPRD Aurum								Germany DA					
	Before PS matching				After PS matching				Before PS matching				After PS matching			
	ChAdOx1	BNT162b2			ChAdOx1	BNT162b2			ChAdOx1	BNT162 b2			ChAdOx1	BNT162b 2		
Characteristics	%	%	SMD		%	%	SMD		%	%	SMD		%	%	SMD	
Cohort Count	3,741,359	1,804,763			1,912,752	1,247,556			88,463	344,916			85,163	211,587		
Age group																
<20	0.4	0.6	-0.03		0.5	0.6	-0.01		0.5	2	-0.14		0.4	0.4	0.01	
20 - 29	4.4	5.4	-0.05		4.7	4.7	0		4.3	7.4	-0.14		3.8	3.5	0.02	
30 - 39	7.5	7.9	-0.02		8.3	6.9	0.05		5.9	9.9	-0.15		5.5	5.2	0.01	
40 - 49	18.3	10.9	0.21		10.4	11.1	-0.03		8.7	12.8	-0.14		8.2	7.8	0.01	
50 - 59	28.8	16	0.31		15.6	16.9	-0.03		19.1	22.4	-0.08		19	18.7	0.01	
60 - 69	21.3	17.8	0.09		23.4	22.2	0.03		37.2	22.5	0.33		37.5	37.3	0.01	
70 - 79	14.7	20.4	-0.15		27.2	25.8	0.03		18	16.1	0.05		18.8	19.9	-0.03	
80 - 89	3.6	17.7	-0.47		8.8	9.5	-0.02		5.4	5.9	-0.02		5.7	6	-0.02	
90 - 99	1.1	3.3	-0.15		2.1	2.1	0		1.1	1	0.01		1.1	1.2	-0.01	
100 – 109	0.1	0.1	-0.01		0.1	0.1	0.01									
Gender: female	51.5	58.2	-0.14		56.1	56.1	0		41.2	55	-0.28		42	42.3	0	
Index month																
12/2020		11														
1/2021	13.7	47.6	-0.79		37.9	41.6	-0.08									
2/2021	29	36.3	-0.16		54.8	50.2	0.09									
3/2021	45.7	3.6	1.12		4.4	5.2	-0.04		1.9	0.1	0.18		0.5	0.4	0.02	
4/2021	9	1.6	0.34		2.3	2.3	-0.01		22.2	30.7	-0.2		24.6	25.1	-0.01	
5/2021	2.5	0.4	0.17		0.6	0.6	-0.01		60.4	26.4	0.73		61.1	61.3	0	
6/2021									13.2	24.5	-0.29		12.1	11.9	0.01	
7/2021									2	12.7	-0.42		1.6	1.4	0.02	
8/2021									0.3	5.5	-0.32		0.1	0.1	0	

COVID-19 infection prior vaccination	5.5	4	0.07		4.6	4.4	0.01							
Charlson index – Romano adaptation	79.8	131.1	-0.33		123.6	121.7	0.01		158.8	157.2	0.01		164.1	169.1	-0.02
CHADS2VASc	128.3	200.6	-0.54		190.4	187.3	0.02		196.9	194	0.02		204	209.1	-0.03
Medical history: General															
Acute respiratory disease	41.4	44.1	-0.05		44	43.9	0		57.9	59.6	-0.03		58.1	57.7	0.01
Attention deficit hyperactivity disorder	0.3	0.2	0.01		0.2	0.2	0		0.6	0.4	0.02		0.6	0.5	0
Chronic obstructive lung disease	3.1	5.3	-0.11		5.5	5.2	0.01		11.5	11.3	0		12	12.2	-0.01
Crohn’s disease	0.5	0.6	-0.01		0.7	0.6	0.01		0.6	0.8	-0.03		0.6	0.6	0
Dementia	1.2	2	-0.07		1.8	1.7	0.02		2.4	2.2	0.01		2.5	2.6	-0.01
Depressive disorder	21.3	20.3	0.02		21.3	21.2	0		22	24.2	-0.05		22.8	23	-0.01
Diabetes mellitus	10	15.5	-0.17		16.1	15.8	0.01		21.4	20.1	0.03		22.4	23.1	-0.02
Gastroesophageal reflux disease	5.1	5.9	-0.04		5.9	5.9	0		5.5	5.8	-0.01		5.8	5.8	0
Gastrointestinal haemorrhage	8.6	9.7	-0.04		9.7	9.5	0.01		3.6	3.2	0.02		3.6	3.5	0
Hyperlipidaemia	9.2	13	-0.12		13.1	12.9	0.01		34.5	29.6	0.11		35.5	36.1	-0.01
Hypertensive disorder	23.8	35.8	-0.26		34.7	34.2	0.01		49.8	43.8	0.12		52.1	52.9	-0.02
Lesion of liver	0.9	1.2	-0.03		1.3	1.3	0		1.3	1.4	0		1.4	1.4	0
Obesity	4.7	5.3	-0.03		5.7	5.6	0.01		15.5	15.3	0.01		16	16.2	-0.01
Osteoarthritis	18.9	28	-0.21		26.9	26.5	0.01		33.1	29	0.09		33.9	34.2	-0.01
Pneumonia	2.8	3.6	-0.05		3.6	3.4	0.01		6.6	6.9	-0.01		6.7	6.8	-0.01

Psoriasis	4.4	4.7	-0.01		4.9	4.9	0		4.2	3.7	0.02		4.3	4.4	0
Renal impairment	6.3	13.3	-0.24		11.2	11.1	0		8.2	7.3	0.03		8.3	8.6	-0.01
Rheumatoid arthritis	1.1	1.6	-0.05		1.6	1.6	0		3.7	4	-0.01		3.9	4	-0.01
Ulcerative colitis	0.7	0.8	-0.01		0.9	0.9	0.01		0.8	1	-0.02		0.8	0.8	0
Urinary tract infectious disease	13.5	16.9	-0.1		16.1	16	0		15	17.3	-0.06		15.6	15.6	0
Viral hepatitis C	0.2	0.2	0		0.2	0.2	0		0.4	0.4	0		0.4	0.4	0
Visual system disorder	39.7	48.7	-0.18		47.4	46.9	0.01		26.4	25.4	0.02		26.8	27.1	-0.01
Medical history: Cardiovascular disease															
Atrial fibrillation	2.9	6.6	-0.17		5.5	5.4	0		3.3	3.3	0		3.5	3.7	-0.01
Cerebrovascular disease	2.8	5.5	-0.14		4.9	4.8	0.01		7.7	7.3	0.02		8.1	8.5	-0.02
Coronary arteriosclerosis	0.8	1.4	-0.07		1.4	1.3	0		5.7	5.4	0.01		6	6.2	-0.01
Heart disease	10.5	19.5	-0.25		17.7	17.5	0.01		31	29.4	0.04		32.3	33.3	-0.02
Heart failure	1.3	2.6	-0.1		2.2	2.2	0.01		7.3	6.8	0.02		7.5	7.9	-0.02
Ischemic heart disease	4.1	8.3	-0.17		7.6	7.4	0.01		13.8	12.7	0.03		14.5	14.9	-0.01
Peripheral vascular disease	0.7	1.3	-0.06		1.3	1.2	0.01		6.7	5.8	0.04		6.9	7.1	-0.01
Pulmonary embolism	1	1.5	-0.05		1.3	1.3	0		0.9	1.5	-0.05		0.8	0.9	0
Venous thrombosis	4.4	6.1	-0.08		4.2	4.1	0.01		5.1	5.7	-0.03		4	4.1	0
Medical history: Neoplasms															
Malignant tumour of breast	1.6	2.7	-0.07		2.7	2.7	0		1.7	2.3	-0.04		1.8	1.9	0

Hematologic neoplasm	0.4	0.7	-0.04		0.6	0.6	0		0.9	1.2	-0.03		1	1	-0.01
Malignant lymphoma	0.3	0.6	-0.04		0.6	0.6	0		0.7	0.9	-0.02		0.7	0.7	-0.01
Malignant neoplasm of anorectum	0.2	0.4	-0.04		0.4	0.4	0		0.4	0.4	0		0.4	0.5	-0.01
Malignant neoplastic disease	8.2	14.7	-0.21		13.4	13.4	0		12.1	12.8	-0.02		12.6	13	-0.01
Malignant tumour of colon	0.3	0.7	-0.05		0.6	0.6	0		0.8	0.8	0		0.8	0.9	-0.01
Malignant tumour of lung	0.2	0.3	-0.03		0.3	0.3	0		0.2	0.3	-0.02		0.2	0.3	0
Medication use															
Lipid modifying agents	9.4	15.7	-0.19		15.6	15.3	0.01		19.5	17.1	0.06		20.8	21.5	-0.02
Antithrombotic agents	4.4	8.9	-0.18		7.8	7.7	0		14.1	14.2	0		14.9	15.5	-0.02
Agents acting on the renin-angiotensin system	9.1	14	-0.16		13.7	13.6	0		35.9	30.6	0.11		38.3	39.1	-0.02
Beta blocking agents	4.5	7.7	-0.13		7.1	7.1	0		23.1	20.8	0.06		24.7	25.4	-0.02
Diuretics	3.4	6.2	-0.13		5.6	5.6	0		19.8	17.2	0.07		21.1	21.6	-0.01
Drugs for acid related disorders	11.1	15.3	-0.13		15	14.9	0		17.4	18.1	-0.02		18.6	19	-0.01
Calcium channel blockers	6.5	9.9	-0.12		9.8	9.7	0		16	13.7	0.07		17	17.3	-0.01
Antiinflammatory and antirheumatic products	6.5	9.5	-0.11		9.3	9.1	0.01		20.7	19	0.04		21.7	22	-0.01

Drugs used in diabetes	4.3	6.8	-0.11		7.1	7	0.01		11.4	10.4	0.03		12.1	12.5	-0.01
Drugs for obstructive airway diseases	9.5	11.9	-0.08		12.1	12	0.01		9.3	10.2	-0.03		9.9	10	-0.01
Antibacterials for systemic use	8.6	10.6	-0.07		10.5	10.3	0.01		6.4	7.9	-0.06		6.6	6.8	-0.01
Opioids	5.7	7.3	-0.06		7.3	7.2	0		4.9	4.8	0.01		5.1	5.2	-0.01
Antineoplastic agents	0.4	0.6	-0.03		0.6	0.6	0		0.9	1	-0.01		0.9	1	0
Immunosuppressants	0.4	0.6	-0.03		0.6	0.6	0		0.7	1	-0.03		0.8	0.8	0
Antiepileptics	2.5	2.9	-0.03		3	3	0		2.4	2.7	-0.02		2.6	2.7	-0.01
Psycholeptics	2.7	3.1	-0.02		3.2	3.2	0		4.3	4.6	-0.02		4.6	4.7	-0.01
Antipsoriatics	0.3	0.3	0		0.3	0.4	0		0.3	0.2	0		0.3	0.3	0

Table 7-3 Baseline characteristics of eligible Ad26.COV2.S cohorts of vaccinated people identified EU based databases.

	Germany_DA							Spain SIDIAP					
	Before PS matching			After PS matching				Before PS matching			After PS matching		
	Ad26.COV2.S	BNT162b2		Ad26.COV2.S	BNT162b2			Ad26.COV2.S	BNT162b2		Ad26.COV2.S	BNT162b2	
Characteristics	%	%	SMD	%	%	SMD		%	%	SMD	%	%	SMD
Cohort Count	18,485	344,916		18,428	67,016			126,189	1,973,028		116,087	421,532	
Age group													
<20								0.3	0.2	0.02	0.3	0.2	0.01
20 - 29	10.3	7.4	0.10	5	4.9	0.01		3	1.9	0.07	3	2.5	0.03
30 - 39	13.1	9.9	0.10	8.1	8	0.00		4.8	8.3	-0.14	4.8	5	-0.01
40 - 49	16.9	12.8	0.11	13.8	13.8	0.00		40.8	25.6	0.33	42.4	44.7	-0.05
50 - 59	28.1	22.4	0.13	29.3	29.5	-0.01		13.7	24.9	-0.29	14.9	14.2	0.02
60 - 69	21	22.5	-0.04	28.3	28.5	-0.01		18.8	2.7	0.54	12.5	10.4	0.07
70 - 79	6.1	16.1	-0.32	9.6	9.5	0.00		18.4	19.9	-0.04	21.7	22.7	-0.03
80 - 89	2.6	5.9	-0.16	4.5	4.5	0.00		0.2	12.8	-0.53	0.3	0.2	0.02
90 - 99	0.5	1	-0.05	0.9	0.8	0.01		0.1	3.7	-0.27	0.1	0.1	0.01
Gender: female	38.2	55	-0.34	41.6	41.2	0.01		46.1	54.1	-0.16	46.3	46.8	-0.01
Index month													
1/2021								0.1	6.6	-0.37	0.1	0.1	-0.01
2/2021								0.1	5.9	-0.35	0.1	0.1	-0.01
3/2021		0.1						0.1	8.5		0.1	0.2	
4/2021		30.7						9.4	18.8	-0.27	11.1	10.8	0.01
5/2021	20.2	26.4	-0.15	24.1	23.7	0.01		16.6	22.4	-0.15	19.3	19.6	-0.01
6/2021	63.1	24.5	0.84	65.8	66.4	-0.01		73.7	37.7	0.78	69.3	69.2	0.00
7/2021	11	12.7	-0.05	7.7	7.4	0.01							
8/2021	5.7	5.5	0.01	2.3	2.4	-0.01							
COVID-19 infection prior vaccination	1.7	2.1	-0.03	1.9	2	-0.01		8.2	6.6	0.06	7.4	7.7	-0.01

Charlson index - Romano adaptation	109.1	157.2	-0.24	157.3	158.5	-0.01		68.6	106.2	-0.22	69.7	70.5	-0.01
CHADS2VASc	132.1	194	-0.42	185.7	186.6	-0.01		123.4	185.5	-0.41	126.7	122.7	0.03
Medical history: General													
Acute respiratory disease	62.8	59.6	0.07	64.6	64.9	-0.01		53.8	56.2	-0.05	54.5	54.9	-0.01
Attention deficit hyperactivity disorder	0.6	0.4	0.02	0.7	0.7	0.00		0.3	0.2	0.02	0.3	0.3	0.00
Chronic liver disease	0.9	1	-0.01	1.3	1.3	0.00		1.8	1.4	0.03	1.7	1.8	0.00
Chronic obstructive lung disease	9.3	11.3	-0.07	12.8	13	-0.01		4.1	5.4	-0.06	4.1	3.9	0.01
Crohn's disease	0.6	0.8	-0.03	0.7	0.7	0.01		0.2	0.2	-0.01	0.2	0.2	-0.01
Dementia	1.4	2.2	-0.06	2.4	2.3	0.00		0.3	2.7	-0.19	0.4	0.4	0.00
Depressive disorder	22	24.2	-0.05	28.1	28.4	-0.01		13.7	15.2	-0.04	13.7	14	-0.01
Diabetes mellitus	15.8	20.1	-0.11	24.6	24.7	0.00		11.7	14.1	-0.07	11.6	11.5	0.00
Gastroesophageal reflux disease	4.5	5.8	-0.06	6.3	6.3	0.00		7	7.9	-0.04	7	6.9	0.00
Human immunodeficiency virus infection	0.2	0.3	-0.02	0.2	0.2	-0.01		0.6	0.3	0.05	0.6	0.6	0.00
Hypertensive disorder	36.2	43.8	-0.15	55.6	55.7	0.00		25.5	34.7	-0.20	25.6	25	0.01
Lesion of liver	1.1	1.4	-0.02	1.7	1.7	0.00		1.1	1	0.01	1.1	1.1	-0.01
Obesity	14.2	15.3	-0.03	19.3	19.4	0.00		24	25.5	-0.04	24	23.9	0.00
Osteoarthritis	26.1	29	-0.07	34.2	34.3	0.00		16.8	24.5	-0.19	17	16.3	0.02
Pneumonia	5.5	6.9	-0.06	7	6.8	0.01		6.6	8.1	-0.06	6.6	6.7	-0.01
Psoriasis	3.4	3.7	-0.02	4.4	4.4	0.00		2.8	3	-0.01	2.8	2.9	0.00
Renal impairment	4.6	7.3	-0.11	7.3	7.3	0.00		3.8	9.6	-0.24	3.9	3.8	0.00

Rheumatoid arthritis	2.6	4	-0.08	3.7	3.7	0.00		0.5	0.8	-0.04	0.5	0.6	-0.01
Schizophrenia	0.5	0.4	0.01	0.7	0.6	0.00		1.2	0.5	0.07	1.2	1.3	-0.01
Ulcerative colitis	0.7	1	-0.03	0.9	0.8	0.01		0.4	0.4	-0.01	0.4	0.4	-0.01
Urinary tract infectious disease	12.8	17.3	-0.13	15.9	15.7	0.01		19.3	24	-0.12	19.4	19.8	-0.01
Viral hepatitis C	0.6	0.4	0.03	0.8	0.7	0.01		1.3	0.9	0.04	1.3	1.3	0.00
Medical history: Cardiovascular disease													
Atrial fibrillation	1.7	3.3	-0.10	2.8	2.7	0.00		2.1	5.5	-0.18	2.2	2.2	0.00
Cerebrovascular disease	3.9	7.3	-0.15	6.5	6.6	0.00		1.8	3.6	-0.12	1.8	2	-0.01
Heart disease	21.3	29.4	-0.19	31.6	31.7	0.00		12.2	20.1	-0.22	12.4	12.4	0.00
Heart failure	3.9	6.8	-0.13	6.7	6.8	0.00		1.5	4.2	-0.16	1.6	1.6	-0.01
Peripheral vascular disease	4.1	5.8	-0.08	6.9	7	0.00		1.7	2.5	-0.06	1.7	1.7	-0.01
Pulmonary embolism	0.6	1.5	-0.09	0.8	0.8	0.00		0.4	0.7	-0.04	0.4	0.4	0.00
Venous thrombosis	4.2	5.7	-0.07	4.3	4.3	0.00		2.9	3.8	-0.05	2.9	2.8	0.00
Medical history: Neoplasms													
Malignant tumour of breast	0.9	2.3	-0.12	1.2	1.1	0.01		1.2	1.8	-0.05	1.3	1.3	0.00
Hematologic neoplasm	0.8	1.2	-0.04	1.1	1.1	0.00		0.6	1	-0.04	0.6	0.7	-0.01
Malignant lymphoma	0.4	0.9	-0.06	0.6	0.6	0.00		0.2	0.4	-0.04	0.2	0.2	0.00
Malignant neoplasm of anorectum	0.2	0.4	-0.04	0.3	0.3	0.00		0.2	0.4	-0.03	0.3	0.3	0.00
Malignant neoplastic disease	7.1	12.8	-0.19	10	10	0.00		7.1	11.9	-0.16	7.5	7.4	0.00

Malignant tumour of lung	0.1	0.3	-0.05	0.2	0.2	-0.01		0.2	0.2	-0.02	0.2	0.2	0.00
Malignant tumour of urinary bladder	0.2	0.4	-0.04	0.3	0.3	0.00		0.5	0.9	-0.05	0.6	0.6	0.00
Medication use													
Lipid modifying agents	10.9	17.1	-0.18	20.3	20.4	0.00		15.3	19.6	-0.11	15.5	15.4	0.00
Antithrombotic agents	8.1	14.2	-0.20	14.7	14.9	-0.01		9.6	16.3	-0.20	9.7	9.7	0.00
Agents acting on the renin-angiotensin system	23	30.6	-0.17	42.4	42.6	0.00		18.9	25.7	-0.17	19	18.6	0.01
Diuretics	11.5	17.2	-0.16	21.4	21.5	0.00		12	18.9	-0.19	12.2	12	0.01
Calcium channel blockers	10.1	13.7	-0.11	18.9	19	0.00		6	9.2	-0.12	6.1	5.9	0.01
Antiinflammatory and antirheumatic products	17.6	19	-0.03	26.8	27.1	-0.01		23.9	24.1	0.00	23.9	24.1	0.00
Drugs used in diabetes	7.5	10.4	-0.10	13.8	13.8	0.00		9	10.7	-0.06	8.9	8.9	0.00
Drugs for obstructive airway diseases	7.3	10.2	-0.10	11.4	11.5	0.00		12.6	14	-0.04	12.7	12.7	0.00
Antibacterials for systemic use	6.3	7.9	-0.06	8.6	8.8	-0.01		13.6	14.6	-0.03	13.5	13.7	0.00
Antineoplastic agents	0.5	1	-0.05	0.9	0.9	0.00		0.7	1	-0.04	0.7	0.8	-0.01
Immunosuppressants	0.4	1	-0.08	0.6	0.6	0.00		0.6	1	-0.04	0.7	0.9	-0.02
Antiepileptics	1.8	2.7	-0.06	3.2	3.3	-0.01		7.2	7.1	0.01	7.2	7.5	-0.01
Psycholeptics	3.3	4.6	-0.07	5.8	5.9	0.00		19.8	23.1	-0.08	19.9	20.2	-0.01
Antipsoriatics	0.2	0.2	-0.01	0.3	0.3	-0.01		0.9	1	-0.02	0.9	0.9	0.00
Antidepressants	5.1	6.5	-0.06	8.8	9	-0.01		13.5	15.8	-0.07	13.6	14	-0.01

Psychostimulants, agents used for adhd and nootropics	0.1	0.1	-0.01	0.1	0.2	-0.01		1.1	1.4	-0.03	1.1	1.1	0.00

Table 7-4 Baseline characteristics of eligible Ad26.COV2.S cohorts of vaccinated people identified US database.

	US Open Claims						US Open Claims					
	Before PS matching			After PS matching			Before PS matching			After PS matching		
	Ad26.COV2.S	BNT162b2		Ad26.COV2.S	BNT162b2		Ad26.COV2.S	mRNA-1273		Ad26.COV2.S	mRNA-1273	
Characteristic	%	%	SMD	%	%	SMD	%	%	SMD	%	%	SMD
Cohort Count	628,004	5,288,749		628,002	2,363,428		628,164	3,923,371		628,164	2,230,157	
Age group												
<20	2.2	4.7	-0.14	1.8	1.9	-0.01	2.2	1.6	0.04	1.8	1.9	0.00
20 - 29	13.4	16	-0.07	10.4	10.9	-0.02	13.4	11.5	0.06	10.4	10.9	-0.01
30 - 39	15.2	17	-0.05	12.7	13.1	-0.01	15.2	13.1	0.06	12.7	13	-0.01
40 - 49	17.3	16.9	0.01	16.3	16.5	-0.01	17.3	14.2	0.08	16.3	16.1	0.01
50 - 59	22.1	17.7	0.11	23.5	23.6	0.00	22.1	17.9	0.10	23.5	22.9	0.02
60 - 69	18.9	15.1	0.10	21.8	21.4	0.01	18.9	20.6	-0.04	21.8	21.8	0.00
70 - 79	7.6	8.2	-0.02	9.3	8.6	0.02	7.6	14	-0.21	9.3	9.3	0.00
80 - 89	3.3	4.5	-0.06	4.1	3.9	0.01	3.3	6.9	-0.16	4.1	4.3	-0.01
Gender: female	47	54.5	-0.15	52	52.8	-0.02	47	54	-0.14	52	53.4	-0.03
Index month												
12/2020		1						0.7				
1/2021		6						8.9				
2/2021		6.9						15.4				
3/2021	23	20	0.07	26.9	29.3	-0.06	23	29.9	-0.16	26.9	30.7	-0.09
4/2021	33.5	21.6	0.27	35.2	34.3	0.02	33.5	23.9	0.21	35.2	35	0.00
5/2021	21.3	11.8	0.26	20	18	0.05	21.3	11.7	0.26	20	18.6	0.04
6/2021	10.5	7.8	0.09	9.2	8.9	0.01	10.5	2.9	0.31	9.2	7.2	0.07
7/2021	5.9	9	-0.12	4.7	5	-0.01	5.9	3.8	0.10	4.7	4.7	0.00
8/2021	5.3	14.1	-0.30	3.7	4.1	-0.02	5.3	2.5	0.15	3.7	3.5	0.01
9/2021	0.5	1.9	-0.13	0.3	0.3	-0.01	0.5	0.2	0.05	0.3	0.3	0.00
COVID-19 infection prior vaccination	2.9	3.3	-0.02	3.4	3.6	-0.01	2.9	2.8	0.01	3.4	3.5	0.00
Charlson index - Romano adaptation	108.2	113.3	-0.03	137.7	136.2	0.01	108.2	145.6	-0.17	137.7	139.7	-0.01
CHADS2VASc	129	138.5	-0.07	155.7	153.2	0.02	129	173.1	-0.29	155.7	158.3	-0.02

Medical history: General												
Acute respiratory disease	45.2	45.3	0.00	50.8	52	-0.03	45.2	46.4	-0.02	50.8	51.4	-0.01
Attention deficit hyperactivity disorder	4.4	4.2	0.01	5	5.3	-0.01	4.4	3.5	0.05	5	5	0.00
Chronic obstructive lung disease	5.4	5	0.02	7	6.6	0.02	5.4	7.1	-0.07	7	7.2	-0.01
Crohn's disease	0.5	0.6	0.00	0.7	0.7	0.00	0.5	0.6	-0.01	0.7	0.7	0.00
Dementia	1.2	1.9	-0.06	1.5	1.5	0.00	1.2	1.9	-0.06	1.5	1.5	0.00
Diabetes mellitus	12.9	13	0.00	16.8	16.6	0.01	12.9	17	-0.12	16.8	17.1	-0.01
Gastrointestinal haemorrhage	5.9	5.9	0.00	7.2	7.4	-0.01	5.9	7.2	-0.05	7.2	7.4	-0.01
Human immunodeficiency virus infection	0.4	0.4	-0.01	0.5	0.5	-0.01	0.4	0.4	-0.01	0.5	0.5	0.00
Hyperlipidaemia	31.4	29.8	0.03	39.6	39.4	0.01	31.4	38.9	-0.16	39.6	39.7	0.00
Hypertensive disorder	30.7	29.3	0.03	39.4	38.7	0.02	30.7	38.1	-0.16	39.4	39.5	0.00
Lesion of liver	1	1	0.00	1.2	1.3	0.00	1	1.2	-0.02	1.2	1.2	0.00
Obesity	13.7	13.9	-0.01	16.9	17	0.00	13.7	15.4	-0.05	16.9	17	-0.01
Osteoarthritis	25.6	23.4	0.05	32	31.6	0.01	25.6	30.8	-0.12	32	31.9	0.00
Pneumonia	6.9	7.3	-0.02	8.1	8.1	0.00	6.9	8	-0.05	8.1	8.3	0.00
Psoriasis	1.9	1.8	0.00	2.4	2.4	0.00	1.9	2.2	-0.02	2.4	2.4	0.00
Renal impairment	5.4	6	-0.03	6.9	6.8	0.01	5.4	8.1	-0.11	6.9	7.1	-0.01
Rheumatoid arthritis	1.6	1.7	-0.01	2.1	2	0.01	1.6	2.2	-0.04	2.1	2.2	-0.01
Schizophrenia	0.6	0.7	-0.02	0.8	0.8	0.00	0.6	0.7	-0.01	0.8	0.8	0.00
Urinary tract infectious disease	14.3	16.2	-0.05	17.6	18	-0.01	14.3	17.8	-0.10	17.6	18.2	-0.02
Visual system disorder	35.9	36.5	-0.01	41.7	42.1	-0.01	35.9	41.3	-0.11	41.6	42	-0.01
Medical history: Cardiovascular disease												

Atrial fibrillation	2.9	3.3	-0.02	3.7	3.5	0.01	2.9	4.7	-0.09	3.7	3.8	0.00
Cerebrovascular disease	4.9	5.2	-0.01	6.4	6.2	0.01	4.9	7.4	-0.10	6.4	6.5	-0.01
Coronary arteriosclerosis	6.1	6.1	0.00	8	7.6	0.02	6.1	9.1	-0.11	8	8.1	0.00
Heart disease	17.9	18.1	0.00	22.7	22.3	0.01	17.9	23.9	-0.15	22.7	22.9	0.00
Heart failure	3.3	3.8	-0.03	4.2	4.1	0.01	3.3	5	-0.09	4.2	4.3	-0.01
Peripheral vascular disease	4	4.4	-0.02	5.1	5	0.01	4	6	-0.09	5.1	5.3	-0.01
Pulmonary embolism	0.8	0.9	-0.01	0.6	0.7	0.00	0.8	1.1	-0.04	0.6	0.7	-0.01
Venous thrombosis	2.1	2.2	-0.01	1.4	1.4	0.00	2.1	2.8	-0.05	1.4	1.4	0.00
Medical history: Neoplasms												
Malignant tumour of breast	1.3	1.5	-0.02	1.7	1.7	0.00	1.3	2	-0.06	1.7	1.7	0.00
Hematologic neoplasm	1.1	1.2	-0.01	1.4	1.4	0.00	1.1	1.5	-0.04	1.4	1.4	0.00
Malignant neoplasm of anorectum	0.2	0.2	0.00	0.3	0.3	0.01	0.2	0.3	-0.01	0.3	0.3	0.00
Malignant neoplastic disease	8.5	8.7	-0.01	10.8	10.4	0.01	8.5	12.2	-0.12	10.8	10.7	0.00
Malignant tumour of colon	0.4	0.4	0.00	0.5	0.5	0.00	0.4	0.6	-0.03	0.5	0.5	0.00
Malignant tumour of urinary bladder	0.3	0.3	0.00	0.3	0.3	0.00	0.3	0.4	-0.03	0.3	0.3	0.00
Primary malignant neoplasm of prostate	1	1	-0.01	1.3	1.2	0.01	1	1.6	-0.06	1.3	1.3	0.00
Medication use												
Agents acting on the renin-angiotensin system	17.5	15.5	0.05	25.1	24.3	0.02	17.5	21.5	-0.10	25.1	24.7	0.01
Beta blocking agents	10.4	9.8	0.02	14.9	14.3	0.02	10.4	13.8	-0.10	14.8	14.8	0.00
Diuretics	11.4	10.6	0.03	16.2	15.7	0.01	11.4	14.3	-0.09	16.2	16	0.01
Drugs for acid related disorders	10.8	10.5	0.01	15.2	14.8	0.01	10.8	13.4	-0.08	15.2	15.2	0.00

Calcium channel blockers	8.3	7.9	0.02	11.8	11.5	0.01	8.3	10.8	-0.09	11.8	11.8	0.00
Anti-inflammatory and antirheumatic products	12.9	12.8	0.00	17.2	17	0.00	12.8	14.2	-0.04	17.2	17.2	0.00
Drugs used in diabetes	8.7	8.2	0.02	12.5	12.2	0.01	8.7	11	-0.08	12.5	12.5	0.00
Antibacterials for systemic use	20.9	21.5	-0.01	26.7	26.9	0.00	20.9	23.1	-0.05	26.7	26.6	0.00
Opioids	7.7	7.1	0.02	10.3	10.1	0.01	7.7	8.3	-0.02	10.3	10.2	0.01
Antineoplastic agents	2.3	2.5	-0.01	3.2	3.2	0.00	2.3	3	-0.04	3.2	3.3	-0.01
Immunosuppressants							1.4	1.9	-0.04	2	2	0.00
Antiepileptics	8.7	7.9	0.03	12.3	12.1	0.01	8.7	9.9	-0.04	12.3	12.1	0.01
Psycholeptics	11.3	10.6	0.02	15.9	15.8	0.00	11.3	12.3	-0.03	15.8	15.5	0.01
Antipsoriatics	0.3	0.3	-0.01	0.4	0.4	0.00	0.3	0.4	-0.01	0.4	0.4	0.00
Antidepressants	17.3	16.2	0.03	24.5	24.7	0.00	17.3	18.7	-0.04	24.5	24.2	0.01
Psychostimulants, agents used for adhd and nootropics	3.4	3.2	0.01	4.9	5.1	-0.01	3.4	2.9	0.03	4.9	4.7	0.01

Table 7-5 Incidence rates (IRs) per 1,000 person-years and incidence rate ratios (IRRs) for adenovirus vs mRNA-based vaccination in analyses that passed diagnostic tests among the matched cohorts

Event	Database	N after PS matching	Person-years	Events	IR per 1,000 person-years (95% CI)	calibrated IRR (95% CI)	
First-dose ChAdOx1 vs. BNT162b2							
Arterial thromboembolism (ATE)	UK CPRD	1,227,495	92,807	331	3.57 (3.19-3.97)	Ref	
		1,886,308	140,256	416	2.97 (2.69-3.27)	0.85 (0.73-0.99)	
	Germany DA	204,702	15,530	44	2.83 (2.06-3.8)	Ref	
		82,643	6,261	19	3.03 (1.83-4.74)	0.76 (0.41-1.39)	
Deep vein thrombosis (DVT)	UK CPRD	1,247,556	94,341	150	1.59 (1.35-1.87)	Ref	
		1,912,752	142,268	193	1.36 (1.17-1.56)	0.89 (0.71-1.11)	
	Germany DA	211,587	16,056	21	1.31 (0.81-2)	Ref	
		85,163	6,454	21	3.25 (2.01-4.97)	2.62 (1.34-5.13)	
Intestinal infarction	UK CPRD	1,270,917	96,126	14	0.15 (0.08-0.24)	Ref	
		1,945,248	144,743	22	0.15 (0.1-0.23)	1.06 (0.53-2.13)	
Ischemic stroke	UK CPRD	1,264,894	95,666	76	0.79 (0.63-0.99)	Ref	
		1,936,816	144,104	75	0.52 (0.41-0.65)	0.66 (0.48-0.92)	
	Germany DA	210,616	15,982	15	0.94 (0.53-1.55)	Ref	
		84,835	6,429	11	1.71 (0.85-3.06)	1.34 (0.58-3.09)	
Myocardial infarction	UK CPRD	1,233,874	93,294	201	2.15 (1.87-2.47)	Ref	
		1,895,358	140,942	283	2.01 (1.78-2.26)	0.94 (0.78-1.14)	
	Germany DA	208,975	15,856	26	1.64 (1.07-2.4)	Ref	
		84,048	6,368	10	1.57 (0.75-2.89)	0.70 (0.31-1.57)	
Pulmonary embolism	UK CPRD	1,254,781	94,894	197	2.08 (1.8-2.39)	Ref	
		1,922,818	143,038	269	1.88 (1.66-2.12)	0.93 (0.77-1.12)	
	Germany DA	212,362	16,115	20	1.24 (0.76-1.92)	Ref	
		85,493	6,479	6	0.93 (0.34-2.02)	0.69 (0.26-1.83)	
Thrombocytopenia	UK CPRD	1,195,498	90,381	442	4.89 (4.45-5.37)	Ref	
		1,836,112	136,523	827	6.06 (5.65-6.48)	1.31 (1.16-1.49)	
	Germany DA	204,508	15,516	78	5.03 (3.97-6.27)	Ref	
		82,281	6,234	35	5.61 (3.91-7.81)	1.01 (0.63-1.62)	

Any thrombosis (venous thrombo-embolism or arterial thromboembolism) with thrombocytopenia syndrome (Any-TTS)	UK CPRD	1,263,613	95,571	64	0.67 (0.52-0.86)	Ref	
		1,934,651	143,950	121	0.84 (0.7-1)	1.29 (0.94-1.77)	
Venous thrombo-embolism (VTE)	UK CPRD	1,233,788	93,290	314	3.37 (3-3.76)	Ref	
		1,893,469	140,803	420	2.98 (2.7-3.28)	0.91 (0.78-1.06)	
	Germany DA	209,244	15,878	40	2.52 (1.8-3.43)	Ref	
		84,436	6,398	25	3.91 (2.53-5.77)	1.61 (0.92-2.83)	
Second-dose ChAdOx1 vs. BNT162b2							
Thrombocytopenia	UK CPRD	1,012,563	60,302	347	5.75 (5.16-6.39)	Ref	
		747,810	38,474	230	5.98 (5.23-6.8)	0.94 (0.76-1.16)	
Any thrombosis (venous thrombo-embolism or arterial thromboembolism) with thrombocytopenia syndrome (Any-TTS)	UK CPRD	1,076,722	64,277	42	0.65 (0.47-0.88)	Ref	
		795,629	41,080	38	0.93 (0.65-1.27)	1.16 (0.71-1.89)	
Deep vein thrombosis (DVT)	UK CPRD	1,063,064	63,456	96	1.51 (1.23-1.85)	Ref	
		784,878	40,506	61	1.51 (1.15-1.93)	0.93 (0.65-1.34)	
Pulmonary embolism (PE)	UK CPRD	1,069,375	63,835	92	1.44 (1.16-1.77)	Ref	
		789,797	40,767	53	1.3 (0.97-1.7)	0.86 (0.58-1.26)	
Venous thrombo-embolism (VTE)	UK CPRD	1,050,916	62,715	179	2.85 (2.45-3.3)	Ref	
		775,486	39,998	105	2.63 (2.15-3.18)	0.87 (0.66-1.16)	
Ischaemic stroke	UK CPRD	1,078,360	64,368	28	0.43 (0.29-0.63)	Ref	
		796,695	41,129	23	0.56 (0.35-0.84)	1.20 (0.66-2.18)	
Myocardial infarction	UK CPRD	1,050,018	62,656	109	1.74 (1.43-2.1)	Ref	
		774,713	39,952	61	1.53 (1.17-1.96)	0.91 (0.64-1.3)	
Arterial thromboembolism	UK CPRD	1,044,491	62,307	153	2.46 (2.08-2.88)	Ref	
		770,339	39,705	101	2.54 (2.07-3.09)	1.05 (0.78-1.4)	

Ad26.COV2.S vs. BNT162b2							
Thrombocytopenia	Germany DA	65,217	4,894	14	2.86 (1.56-4.8)	Ref	
		17,933	1,213	12	9.89 (5.11-17.28)	1.30 (0.57-2.93)	*
	Spain SIDIAP	386,334	19,944	197	9.88 (8.55-11.36)	Ref	
		106,217	5,037	49	9.73 (7.2-12.86)	0.77 (0.55-1.08)	
	US Open Claims	2,364,195	172,698	470	2.72 (2.48-2.98)	Ref	
		628,293	46,997	170	3.62 (3.09-4.2)	1.03 (0.63-1.7)	*
Venous thrombo-embolism with thrombocytopenia syndrome (TTS-VTE)	US Open Claims	2,404,904	175,752	13	0.07 (0.04-0.13)	Ref	
		639,269	47,828	11	0.23 (0.11-0.41)	2.45 (0.95-6.29)	*
Any thrombosis (venous thrombo-embolism or arterial thromboembolism) with thrombocytopenia syndrome (Any-TTS)	US Open Claims	2,365,254	172,778	378	2.19 (1.97-2.42)	Ref	
		628,571	47,019	146	3.11 (2.62-3.65)	1.11 (0.67-1.84)	*
Deep vein thrombosis (DVT)	Spain SIDIAP	421,532	22,028	33	1.5 (1.03-2.1)	Ref	
		116,087	5,582	10	1.79 (0.86-3.29)	0.94 (0.45-1.96)	
	US Open Claims	2,363,428	172,627	347	2.01 (1.8-2.23)	Ref	
		628,002	46,974	121	2.58 (2.14-3.08)	0.98 (0.59-1.63)	*
Pulmonary embolism (PE)	Spain SIDIAP	422,330	22,072	14	0.63 (0.35-1.06)	Ref	
		116,315	5,593	5	0.89 (0.29-2.09)	1.06 (0.37-3.07)	
	US Open Claims	2,380,869	173,941	250	1.44 (1.26-1.63)	Ref	
		632,834	47,339	105	2.22 (1.81-2.69)	1.18 (0.7-1.98)	*
Venous thrombo-embolism (VTE)	Spain SIDIAP	420,502	21,960	42	1.91 (1.38-2.59)		
		115,760	,5,562	14	2.52 (1.38-4.22)	1.03 (0.55-1.93)	
	US Open Claims	2,348,419	171,499	506	2.95 (2.7-3.22)	Ref	
		624,001	46,670	190	4.07 (3.51-4.69)	1.06 (0.64-1.74)	*
Ischaemic stroke	Spain SIDIAP	417,793	21,749	61	2.8 (2.15-3.6)	Ref	

		114,999	5,509	18	3.27 (1.94-5.16)	1.04 (0.59-1.81)	
	US Open Claims	2,348,140	171,471	540	3.15 (2.89-3.43)	Ref	
		623,396	46,622	193	4.14 (3.58-4.77)	1.02 (0.62-1.67)	*
Myocardial infarction (MI)	Spain SIDIAP	418,734	21,822	38	1.74 (1.23-2.39)	Ref	
		115,276	5,528	10	1.81 (0.87-3.33)	0.81 (0.38-1.71)	
	US Open Claims	2,356,142	172,074	472	2.74 (2.5-3)	Ref	
		625,168	46,757	168	3.59 (3.07-4.18)	1.02 (0.62-1.68)	*
Intestinal infarction	US Open Claims	2,401,293	175,480	53	0.3 (0.23-0.4)	Ref	
		638,257	47,752	7	0.15 (0.06-0.3)	0.35 (0.14-0.87)	
Arterial thromboembolism (ATE)	Spain SIDIAP	413,039	21,426	119	5.55 (4.6-6.65)	Ref	
		113,588	5,421	34	6.27 (4.34-8.76)	0.93 (0.62-1.39)	
	US Open Claims	2,304,844	168,208	2,231	13.26 (12.72-13.83)	Ref	
		610,895	45,673	720	15.76 (14.63-16.96)	0.92 (0.57-1.48)	*
Splanchnic and Visceral Thrombosis (SVT)	US Open Claims	2,404,366	175,711	19	0.11 (0.07-0.17)	Ref	
		639,111	47,816	10	0.21 (0.1-0.38)	1.46 (0.59-3.61)	*
Ad26.COV2.S vs. mRNA-1273							
Deep vein thrombosis with thrombocytopenia syndrome (TTS-DVT)	US Open Claims	2,271,774	172,851	12	0.07 (0.04-0.12)	Ref	
		639,496	47,843	6	0.13 (0.05-0.27)	1.35 (0.45-4.05)	*
Venous thrombo-embolism with thrombocytopenia syndrome (TTS-VTE)	US Open Claims	2,271,552	172,835	14	0.08 (0.04-0.14)	Ref	
		639,432	47,838	11	0.23 (0.11-0.41)	1.92 (0.77-4.8)	*
Any thrombosis (venous thrombo-embolism or arterial thromboembolism) with thrombocytopenia syndrome (Any-TTS)	US Open Claims	2,232,550	169,861	380	2.24 (2.02-2.47)	Ref	
		628,737	47,028	146	3.1 (2.62-3.65)	0.97 (0.61-1.55)	*
Deep vein thrombosis (DVT)	US Open Claims	2,230,157	169,676	336	1.98 (1.77-2.2)	Ref	
		628,164	46,983	121	2.58 (2.14-3.08)	0.92 (0.57-1.48)	*

Pulmonary embolism (PE)	US Open Claims	2,247,746	171,017	227	1.33 (1.16-1.51)	Ref	
		632,997	47,349	105	2.22 (1.81-2.68)	1.15 (0.71-1.87)	*
Venous thrombo-embolism (VTE)	US Open Claims	2,215,499	168,558	488	2.9 (2.64-3.16)	Ref	
		624,163	46,679	190	4.07 (3.51-4.69)	0.99 (0.62-1.56)	*
Ischaemic stroke	US Open Claims	2,214,613	168,485	533	3.16 (2.9-3.44)	Ref	
		623,557	46,632	193	4.14 (3.58-4.77)	0.93 (0.59-1.47)	*
Myocardial infarction	US Open Claims	2,222,711	169,104	513	3.03 (2.78-3.31)	Ref	
		625,329	46,766	168	3.59 (3.07-4.18)	0.86 (0.54-1.36)	*
Intestinal infarction	US Open Claims	2,267,972	172,560	54	0.31 (0.24-0.41)	Ref	
		638,418	47,761	7	0.15 (0.06-0.3)	0.29 (0.12-0.73)	*
Arterial thromboembolism (ATE)	US Open Claims	2,171,445	165,188	2,246	13.6 (13.04-14.17)	Ref	
		611,054	45,682	720	15.76 (14.63-16.96)	0.83 (0.54-1.28)	*
Splanchnic and Visceral Thrombosis (SVT)	US Open Claims	2,271,071	172,798	17	0.1 (0.06-0.16)	Ref	
		639,274	47,826	10	0.21 (0.1-0.38)	1.48 (0.60-3.65)	*

* Did not pass the systematic error diagnostic test of over 80% uncalibrated confidence intervals covering 1.

PS: propensity score

Table 7-6 Sensitivity analysis: Incidence rates (IR per 1,000 py) and IRR for adenovirus vs mRNA-based vaccination in analyses that passed diagnostics

			N after ps matching	person-year	event	IR per 1,000 py	IRR calibrated	
Sensitivity analysis 1: Thrombocytopenia window to 5 days before/after thrombosis post vaccination								
	Any TTS (VTE or ATE)							
UK CPRD	Comparator	BNT162b2 1st dose	1,934,829	95,580	63	0.66 (0.51-0.84)		
	Target	ChAdOx1 1st dose	1,934,829	143,963	120	0.83 (0.69-1)	1.3 (0.95-1.79)	
	Comparator	BNT162b2 2nd dose	1,076,870	64,286	38	0.59 (0.42-0.81)		
	Target	ChAdOx1 2nd dose	795,723	41,085	37	0.9 (0.63-1.24)	1.23 (0.74-2.04)	
US Open Claims	Comparator	BNT162b2 1st dose	2,365,342	172,785	376	2.18 (1.96-2.41)		
	Target	Ad26.COV2.S 1st dose	628,592	47,020	143	3.04 (2.56-3.58)	1.09 (0.66-1.81)	*
	Comparator	mRNA-1273 1st dose	2,232,627	169,867	378	2.23 (2.01-2.46)		
	Target	Ad26.COV2.S 1st dose	628,758	47,030	143	3.04 (2.56-3.58)	0.96 (0.6-1.53)	*
Sensitivity analysis 2: Thrombocytopenia threshold of <100,000 platelets per microliter								
	DVT-thrombocytopenia syndrome (TTS-DVT)							
US Open Claims	Comparator	mRNA-1273 1st dose	2,271,774	172,851	12	0.07 (0.04-0.12)		
	Target	Ad26.COV2.S 1st dose	639,496	47,843	6	0.13 (0.05-0.27)	1.35 (0.45-4.05)	*
	VTE-thrombocytopenia syndrome (TTS-VTE)							
US Open Claims	Comparator	BNT162b2 1st dose	2,404,904	175,752	13	0.07 (0.04-0.13)		
	Target	Ad26.COV2.S 1st dose	639,269	47,828	11	0.23 (0.11-0.41)	2.45 (0.95-6.29)	*
	Comparator	mRNA-1273 1st dose	2,271,552	172,835	14	0.08 (0.04-0.14)		
	Target	Ad26.COV2.S 1st dose	639,432	47,838	11	0.23 (0.11-0.41)	1.92 (0.77-4.8)	*
	Any TTS (VTE or ATE)							
UK CPRD	Comparator	BNT162b2 1st dose	1,263,960	95,597	63	0.66 (0.51-0.84)		

	Target	ChAdOx1 1st dose	1,935,138	143,986	119	0.83 (0.68-0.99)	1.29 (0.94-1.78)	
	Comparator	BNT162b2 2nd dose	1,077,077	64,299	39	0.61 (0.43-0.83)		
	Target	ChAdOx1 2nd dose	795,893	41,094	37	0.9 (0.63-1.24)	1.25 (0.76-2.06)	
US Open Claims	Comparator	BNT162b2 1st dose	2,365,254	172,778	378	2.19 (1.97-2.42)		
	Target	Ad26.COV2.S 1st dose	628,571	47,019	146	3.11 (2.62-3.65)	1.11 (0.67-1.84)	*
	Comparator	mRNA-1273 1st dose	2,232,550	169,861	380	2.24 (2.02-2.47)		
	Target	Ad26.COV2.S 1st dose	628,737	47,028	146	3.1 (2.62-3.65)	0.97 (0.61-1.55)	*

* did not pass the systematic error diagnostics of over 80% uncalibrated CIs cover 1.

Chapter 8 Discussion

This work aimed to utilize routinely collected health data to support the surveillance of COVID-19 vaccines. I demonstrated that real-world data could generate timely and reliable evidence in post-authorisation vaccine safety. In this discussion chapter I summarise key findings from each chapter of the thesis, and discuss the strengths and limitations of my research. I also discuss the implications of my findings for clinical practice, healthcare systems and policy, and for future research.

8.1 Summary of main findings

I conducted three studies to address the overarching aim of my thesis and my three objectives. Chapter 1 provided an overview of my thesis, and the structure and rationale for my work. In chapter 2, I provided a summary of existing studies on post-authorisation COVID-19 vaccine safety. I reviewed the literature, highlighted the research gaps and methodology rationale for my planned studies. In chapter 3, I described the study designs and methods used to control bias and confounding in my thesis. I also explained the covariates definition and cohort phenotyping. In chapter 4, I provided descriptions of the databases used in the thesis.

In Chapter 5, I reported on background incidence rates of 15 prespecified adverse events of special interest associated with COVID-19 vaccines. I used 13 electronic health records and health claims databases from eight countries. All the datasets were mapped to the OMOP CDM. I then conducted this multinational network study with collaborators from the OHDSI community. Incidence rates of AESIs were stratified by age, sex, and database. Rates were pooled across databases using random effects meta-analyses and classified according to the frequency categories of the Council for International Organisations of Medical Sciences. This study found considerable heterogeneity between geographies and databases, suggesting caution when interpreting the differences

between observed and expected rates. These findings highlighted that, for future vaccine surveillance, the same data source should be used to compare post-COVID-19 vaccine (observed) and background (expected) AESI rates whenever possible.

In Chapter 6, I aimed to assess the association between COVID-19 vaccines, SARS-CoV-2 infection, and the risk of immune mediated neurological events. I applied the observed-to-expected analysis and the self-controlled case series methods using primary care records from the UK and Catalonia, Spain. I included over 8.3 million individuals who received at least one dose of COVID-19 vaccines ChAdOx1 nCoV-19, BNT162b2, mRNA-1273, or Ad.26.COV2.S in the UK and Spain, and about 0.7 millions of SARS-CoV-2 infected people. The outcomes of interest included Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis after vaccination or infection. This study found no increased risk after COVID-19 vaccination for the included immune-mediated neurological events. However, the study found that SARS-CoV-2 infection was associated with increased risk of all study outcomes.

Chapter 7 examined the comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different COVID-19 vaccines. The primary analysis compared between the adenovirus based COVID-19 vaccines versus mRNA based COVID-19 vaccines, and the secondary analysis compared between two brand of mRNA vaccines. I conducted an international network cohort study using mapped databases from five European countries and the US. Incidence rate ratios were estimated after propensity scores matching and were calibrated using negative control outcomes. Estimates specific to the database were pooled by use of random effects meta-analyses. The results showed a pooled 30% increased risk of thrombocytopenia after a first dose of the ChAdOx1 vaccine compared with BNT162b2 vaccine.

Table 8-1 presents a brief overview of studies undertaken in my thesis. The next section will discuss overall strengths and limitations in more detail.

Table 8-1 Summary of studies undertaken in the thesis.

Aim	**Participants**	**Main outcomes and measures**	**Conclusions & implication**
To estimate the background incidence rates characteristics of Adverse Events of Special Interest (AESI) for COVID-19 vaccines in databases world-wide. (Chapter 5)	126.7 million people observed for at least 365 days before 1 January 2017, 2018, or 2019 from 13 databases.	Incidence rates of 15 AESIs estimated by Poisson model, stratified by age, sex, and database.	This study found large variations in the observed rates of AESIs by age group and sex, showing the need for stratification or standardisation before using background rates for safety surveillance. Considerable population level heterogeneity in AESI rates was found between databases.
To evaluate the risk of developing adverse events after receiving COVID-19 vaccines using different epidemiological study designs. (Chapter 6)	8.3 million people who received at least one dose of COVID-19 vaccines ChAdOx1 nCoV-19, BNT162b2, mRNA-1273, or Ad.26.COV2.S from the UK and Spain	immune mediated neurological events: Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis. Incidence rate ratios from observed-to-expected analysis and SCCS.	No safety signal was observed between COVID-19 vaccines and the immune mediated neurological events of Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis. An increased risk of Bell's palsy, encephalomyelitis, and Guillain-Barré syndrome was, however, observed for people with SARS-CoV-2 infection.
To compare the risk of developing adverse events between different types and brands of COVID-19 vaccines (Chapter 7)	4.6 million people vaccinated with a first dose of ChAdOx1 and 1.6 million people vaccinated with a second dose of ChAdOx1 from all participating databases. 1.1 million people vaccinated with single dose Ad26.COV2.S. 10.6 million people vaccinated with a first dose of BNT162b2, and 7.7 million people vaccinated with a second dose.	Thrombosis with thrombocytopenia syndrome or venous or arterial thromboembolic events within the 28 days after COVID-19 vaccination. Incidence rate ratios between different vaccines. Estimates specific to the database were pooled by use of random effects meta-analyses.	a pooled 30% increased risk of thrombocytopenia after a first dose of the ChAdOx1 vaccine was observed. Although rare, the observed risks after adenovirus-based vaccines should be considered when planning further immunisation campaigns and future vaccine development.

8.2 Strengths and limitations

8.2.1 Strengths

This thesis has several strengths. First, by using routinely collected health care data, the population included in each study dataset has shown to be representative of the general population in each country/region in term of age and sex. Furthermore, all three analytical studies were conducted using data that mapped to a common data model that allowed for network collaboration and federated analysis, with all code used openly available. This enhances the transparency and reproducibility of my research. Additionally, using multi-national/regional datasets enhances the representation of the study's population, therefore improves generalisability of study findings. These studies also benefit from the quick response of health care systems in the countries where the datasets came from, where that COVID-19 vaccines have been rolled out at speed and scale, enabled the feasibility of conducting vaccine related study.

I also applied multiple methods to design my studies and to address bias and systematic error. For example, in chapter 6, I used the SCCS design, which inherently controlled for time-fixed confounding. In Chapter 7, I applied large-scale propensity score methods to control for observed confounding and used negative control outcomes and empirical calibration to examine and correct for unmeasured confounding. Three pre-specified diagnostics were applied to ensure the validity of study results.

8.2.2 Measurement error in exposure and outcome ascertainment

8.2.2.1 Misclassification in exposures

COVID-19 infection

In the study in chapter 6, where I included a cohort of COVID-19 infected patients, the COVID-19 infection was defined by either having a positive RT-PCR test result, or a diagnosis code. However,

during the first surge of the pandemic, community testing for SARS-CoV-2 was very limited, and people who were not hospitalized with COVID-19 were not tested. Furthermore, antigen test positive results may not be routinely coded within primary care. For example, a study from England showed that 20-30% of SARS-CoV-2 test positive cases might be missing from the primary care records.[237,238] Besides the incomplete records in tests, I was not able to identify and include asymptomatic infections among those who didn't receive a test. While asymptomatic COVID-19 cases can be identified through prospective cohorts,[239] it is not feasible to include asymptomatic cases with routinely collected health care data.

Vaccine status

As the main exposure in the thesis, misclassification of vaccine use might be problematic owing to heterogeneity across data sources. In the UK and Spain where a universal healthcare system is available and the vaccination program led by the national health provider (e.g., NHS in the UK), the GP EHRs were linked to official vaccination registries and captured vaccine information more completely. However, data sources such as German and French records and US databases may contain incomplete vaccine records. Especially in the US, vaccines were covered by the government rather than individual's health insurance, and thus not captured by the US based claims data. Therefore, vaccinees identified through these datasets would have high specificity, but may lack of sensitivity.

In my study described in chapter 6, since I included only the UK and Spanish data sources, misclassification in vaccine status is less of concern. In chapter 7, the use of comparative safety analyses minimises the impact of this problem, because only vaccinated cohorts are included for analysis. While continuous efforts need to be taken to improve the data quality in recording vaccinations, it is also important to choose proper study design to minimise the impact of misclassification.

8.2.2.2 Misclassification in outcomes

Given the nature of observational studies using routinely-collected health data, study outcomes are all subject to under-ascertained, which is particularly concerning given the limited healthcare access caused by the pandemic. In the study where I estimated the background rates of AESI, for example, all 15 events of interest could potentially subject to measurement errors. As the outcome definitions were based on the presence of specific diagnostic codes and were not validated further, sensitivity or specificity could have been imperfect.

In Chapter 7, where TTS and TE were the events of interest, robust methods were employed to develop and implement algorithms for their identification. However, certain events typically managed in a hospital setting may be incompletely captured in some of the databases, such as the German and French data sources. Nevertheless, this limitation was mitigated by the availability of inpatient data through linkage in the Spanish database and reimbursement-based information in US claims. Furthermore, the choice of a matched cohort design should help to minimize the impact of misclassification, as I do not anticipate incompleteness to be dependent on the vaccine received.

8.2.3 Residual and unmeasured confounding

As in other observational studies, my analyses are susceptible to residual and unmeasured confounders.[240] In chapter 6 where COVID-19 infected patients were included, COVID-19 severity was an important factor that was outside the scope of this analysis but may have affected the risk of AESIs and probably varied by database.

In chapter 7, although the routinely collected health data and the use of large-scale propensity scores allowed me to control for many potential confounders, I observed evidence of systematic errors in some analyses, especially in the US Open Claims database. Factors such as health seeking behaviour or family history of study outcomes were unmeasured or partially measured. In my study, I used empirical calibration to account for the unmeasured confounding. While empirical

calibration can increase coverage of the 95% confidence interval and decreased the bias of the estimated treatment effect, caution must be taken to select suitable negative controls.[241]

8.2.4 Selection bias

8.2.4.1 Confounding by indication

Confounding by indication and healthy vaccinee bias may be present in studies where COVID-19 vaccines were exposure. Given the priority schedule during the early stage of immunisation campaign, people vaccinated earlier were more likely to be clinically vulnerable. Healthy vaccinee bias can also occur when healthier people are more likely to follow COVID-19 vaccine recommendations. Besides, individuals also self-select for vaccination regardless of policy. In the SCCS analysis in chapter 6, I applied the pre-exposure period to assess and control for confounding by indication. On the other hand, in chapter 7, I used propensity score methods and only included comparison cohorts that were balanced after matching to maximise the comparability between the target and comparator group and thus minimise the potential confounding by indication.

8.2.5 Study design

The study designs used in the thesis have their own limitations.

In Chapter 6, it is important to acknowledge that using a historical population as a comparator can introduce an increased risk of type 1 error, potentially biasing the results away from the null hypothesis. For instance, individuals with COVID-19 disease may receive more clinical attention or examination compared to individuals in the pre-pandemic background population, leading to potential bias in the results. When defining the 'pre-pandemic background population', I used data from 2017 to 2019 of all people in each database with more than 365 days of observation indexed on 1 January. The impact of these design decisions, in particular the index date (anchoring effect) has been shown to influence rate estimates, however the effect of season has been shown to be moderate.[242]

In chapter 7 where I applied propensity score matching, individuals with extreme scores cannot be matched, which may limit the generalizability of study findings. Also, while matching can make the target and comparator group more comparable, it also decreases study population, which may reduce the power in the analysis. Other propensity score methods should be considered in future studies, with acknowledge their advantages and disadvantages. For example, while inverse probability weighting can retain data from all study participants, it may become unstable when extreme weights occur. Recently, a new propensity score weighting method using overlap weights showed effective in addressing extreme scores.[243,244]

8.2.6 Data sources related limitation

8.2.6.1 Representativeness of data

While real-world studies aimed to include populations that are far more representative of the unselected population than those of RCTs, the representativeness is never perfect. For example, the commercial claims data in the US only represent the insured people, which may be of different socioeconomic backgrounds from the general population. For the SIDIAP data that been used in all three studies, although it is representative of the population living in Catalonia and regions with similar socio-demographics, it is not necessarily so of other regions of Spain.

The UK CPRD Aurum data covers mostly England, with little data from Northern Ireland and none from Scotland, and Wales. Although 98% of the UK population is registered at an NHS GP surgery, certain patients are not represented in my studies, including prisoners, private patients, those in some residential homes, and some people with no fixed address.

8.2.6.2 No inpatient data

As information on hospital admission was not available in the primary care datasets used (CPRD in the UK, IQVIA in France, Germany, and Australia, and IPCI in the Netherlands), events that happened during inpatient visits were not included.

8.2.6.3 Missing or incomplete data in EHRs

In the analysis utilising EHRs from general practice, it is assumed that the absence of a diagnostic code indicates the absence of the disease, both for study outcomes and comorbidity variables. However, this assumption can be problematic as the absence of a diagnosis code could be due to general practitioners not recording the information or recording it as free text, which may not be mapped. However, previous study showed that the validity of diagnoses in one of the data sources used in the thesis, the CPRD GOLD, has been well-established.[245] Additionally, over-the-counter medications are not included in the available data.

Furthermore, the electronic health records data sources, particularly those from hospitals, may not fully capture medical events recorded in other healthcare institutions, resulting in incomplete information. The bias of incomplete information was partially mitigated by including only those patients who had at least one year of continuous observation.

8.2.6.4 Claims data

The use of administrative claims data sources offered reliable data capture but lacked data elements such as laboratory test results. The US based claims database did not record death information well. Comparing the background rates within the database helps to reduce biases associated with the specific limitations of each database. Similarly to EHRs, information on over-the-counter medications is not available in claims data.

8.2.7 Other limitations

The pandemic introduced certain limitation to observational studies as well. These included the changes in healthcare access, such as the shift to telehealth, which can result in missing data on clinical measurements and routine laboratory results. Additionally restrained healthcare resources and stay-home orders had led to delayed or missed healthcare encounters, potentially leading to underdiagnoses. Moreover, individuals from minority backgrounds, low-income households, or rural locales might be disproportionally impacted by the pandemic.[245]

8.3 Implications for clinical practice

While COVID-19 vaccines had been developed in an unprecedented speed after the pandemic started, and mass immunisation campaign began, the concerns raised on vaccine safety contributed to vaccine hesitancy.

In my studies, the AESIs after vaccination were rare (frequency of <1/1000 to ≥1/10,000), or very rare (<1/10,000) based on the WHO Council for International Organisations of Medical Sciences thresholds.[246] In the study estimating the risk of immune mediated neurological events, I also found that the short-term risk of these complications from SARS-CoV-2 infection was substantially increased. These findings can be used by health professionals to provide more information to individuals to support them to make decisions about whether or not to get vaccinated.

For example, both the "COVID-19 vaccination programme: Information for healthcare practitioners", [247] and the "COVID-19: the green book, chapter 14a"[248] published by the UK Health Security Agency, have included information on the potential adverse conditions after vaccination, including TTS, thrombocytopenia, and GBS. These guidelines for healthcare practitioners provided the background rates of events, as well as evidence related to vaccine safety.

Conversely, although immune-mediated neurological complications resulting from either vaccination or infection were rare in my study, these disorders can lead to enduring disabilities requiring long-term care. This underscores the importance for healthcare practitioners to enhance recognition of associated signs and symptoms. This heightened awareness aids in ensuring accurate diagnosis and the exclusion of alternative causes, ultimately facilitating the timely supportive care and treatment.

Apart from complications following the acute phase of SARS-CoV-2 infection, recent studies also reported high prevalence of post-acute sequelae of COVID-19, which persisted six months or even a year after infection.[249]

8.4 Implications for healthcare systems and policy

These findings also have several implications for the design of healthcare systems and public health strategies.

8.4.1 Use of background rate in safety surveillance

Observed-to-expected analysis have been widely used in vaccine surveillance as a method to strengthen safety signals, especially when rapid conclusions about the safety of a vaccine are needed.

The incidence rate ratio between observed and expected are commonly reported as a single number, regardless the fact that the background incidence rates differ between sex, age groups, geographical regions, or calendar time, and the distribution of characteristics among the vaccinated population among is rarely known. In the analysis of background rates for AESIs, notable differences in incidence rates were observed across various factors such as age, sex, and the database used. It is important to note that incidence rates obtained from different sources may be subject to substantial systematic errors. Additionally, the wide prediction intervals observed from the meta-analysis for each subgroup based on age and sex indicate the substantial heterogeneity at the population level across these sources.

Furthermore, when comparing electronic health records and claims data sources from the same country using the same analysis and outcome definitions, substantial variations were observed. It is reasonable to expect even greater variability in rates derived from randomized trials or spontaneous reporting data as the included population may differ from the real-world user.

Therefore, when utilising observational data to estimate historical “expected” rates and comparing them with observed rates from another source, it is crucial to appropriately incorporate the uncertainty associated with the background rate. This is necessary to prevent drawing misleading conclusions. Whenever possible, it is advisable to use the same data source for comparing post-COVID-19 vaccine (observed) AESI rates with background (expected) rates for vaccine surveillance.

8.4.2 Vaccine hesitancy

The acceptance and public trust in a new vaccine are heavily influenced by the availability of reliable evidence regarding its safety. While RCTs have demonstrated the effectiveness of vaccines, uncertainties regarding their safety persist due to the rapid development and authorisation processes. During phase III trials of COVID-19 vaccines, the participant pool was limited in its diversity, and the follow-up duration was relatively short.

In the current era of widespread dissemination of rumours through social media platforms, it is crucial to promptly address any concerns related to vaccine safety in order to uphold public confidence in large-scale vaccination programmes. When faced with alarming media coverage of a severe incident following the introduction of a new vaccine during a public vaccination campaign, knowledge of and transparent communication about the anticipated background rate can help to reassure the public. The availability of background rates therefore plays a significant role in promptly evaluating a safety signal. It is also important to test hypothesis based on known signals using robust epidemiological designs such as a cohort study or SCCS method.

8.4.3 Research networks and standardized analytics

All of the three studies were conducted using databases mapped to the OMOP CDM, which allowed me to distribute analytical codes to data partners without the need of sharing patient-level data. This also allowed me to include into my studies more heterogeneous populations and increase the external validity of results. From health regulator perspective, the ability to generate evidence

among a larger population in a fast speed is curial for regulatory decision making. Furthermore, having a research network with many data partners can help choosing the proper datasets to answer specific questions based on available data elements and completeness.

Since 2022, the European Medicines Agency (EMA) and the European Medicines Regulatory Network established a coordination centre to provide timely and reliable evidence on the use, safety and effectiveness of medicines for human use, including vaccines, from real world healthcare databases across the European Union (EU). This capability is called the Data Analysis and Real-World Interrogation Network (DARWIN EU®). (https://www.darwin-eu.org/) One key component of DARWIN EU® is using the CDM to ensure the interoperability of data with respect to structure (syntactic interoperability) and coding systems (semantic interoperability).

The studies presented in the thesis demonstrated the ability of using OMOP CDM to conduct large-scale observational health research and provided valuable experience for the development of Standardised Analytics in DARWIN EU®.

8.4.4 Representativeness of real-world data

While in traditional clinical trials, representation of diverse populations is a recognized problem, RWD have the potential and ability to expand diversity and fulfil the unequal medical knowledge map.[250] However, current real-world studies are majorly leading by high income countries. As the literature review in Chapter 2 showed, most of the included studies were from high-income countries or regions, with limited coverage of lower- and middle-income countries (LMICs). In this thesis, I used databases from many countries, but these data sources are from wealthy countries, mostly from the US or Europe.

There are 11 vaccines granted emergency use listing by WHO, but only four vaccines were studied in this thesis, and they were either mRNA-based or viral-vector based.[251] Inactivated vaccines such as the CoronaVac (manufactured by Sinovac) and Covilo (manufactured by Sinopharm

(Beijing)) were approved in 56 and 93 countries, separately. However, none of the datasets used in this thesis approved these vaccines, and I am not able to include them in my studies.

The lack of representation may introduce bias and limit the scope for generalisation.[252] It is essential that regulator and policymakers make efforts to deploy more EHRs systems, build trust between stakeholders, and create laws and regulations for local generation of data that are assented for secondary use.[253]

8.5 Implications for future research

8.5.1 Vaccine safety in special population

As mentioned earlier, the population included clinical trials of COVID-19 vaccines are of less diversity. While in the mass immunisation campaigns, the vaccines are usually rolled out with certain schedules where some population groups are prioritized for vaccination. For example, in the UK, the COVID-19 vaccines were first administrated to the residents in care home, then moved to all people over 80 years, to 75 years, and then to 70 years and clinically extremely vulnerable individuals, all of which were of limited representative in the clinical trials. [254] In my studies, while I accounted for these variables (e.g., age stratified results, included comorbidities in propensity score), I didn't specifically focus on these special populations. It is important that future research to generate evidence among special groups, such as pregnancy women, immunosuppressed individuals, and people with other pre-exist conditions as well.

8.5.2 Booster dose

Due to the data availability and time of the studies conducted, I only looked at the primary vaccination course (first and second doses for ChAdOx1 nCoV-19, BNT162b2, mRNA-1273, and single dose for Ad.26.COV2.S). With the booster program and accumulated real-world data, studies focus on the safety of booster dose will provide useful information. Methods used in this thesis,

including cohort method and self-controlled case series method, can be used in booster dose studies as well.

8.5.3 COVID-19 complication and long COVID

Studies demonstrated that a proportion of patient suffered from prolonged COVID-19 symptoms for months after the infection. Since the cause of long COVID remained uncertain, concerns were raised about the potential impact of vaccination to long COVID symptoms, which may lead to hesitancy of vaccines among people infected with COVID-19. Therefore, evidence is needed to assess the safety of COVID-19 vaccines among previous infected people.

8.5.4 Heterogeneity in multi-databases study

Regulatory authorities often require the use of multiple data sources in a single study to enhance result generalizability and ensure an adequate sample size, particularly when the exposure and/or outcome are rare. While multi-database studies are increasingly performed, one challenge is the observed heterogeneity between the data sources. My studies also showed heterogeneous results even after applying the same methods and analysis across databases.

The heterogeneity in multi-database studies can stem from various factors, including differences in population characteristics, healthcare systems, diagnostic practices, and overall health profiles. Therefore, conducting research to better understand, and to address, and to interpret the heterogeneity in multi-database studies, is crucial not only for vaccine safety studies but also for observational pharmacoepidemiology studies in general. In the meantime, it is always important to present the database specific results. Careful consideration on when and how to pool results across datasets using meta-analysis need to be taken when planning the analysis.

8.6 Conclusions

The COVID-19 pandemic provided a unique opportunity to further demonstrate how real-world data can be effectively leveraged for rapid responses to a public health emergency. This thesis used routinely collected health data from multiple countries to conduct observational studies and generate evidence regarding the safety of COVID-19 vaccines.

First, I emphasized the important of within database comparison when using the background incidence rates of adverse events of special interest for COVID-19 vaccines in safety surveillance. This is particularly crucial because these rates can vary between different databases and among patients with varying characteristics. Furthermore, I found that there was no increased risk of immune-mediated neurological outcomes after COVID-19 vaccines, but infection of SARS-Cov-2 would increase the risk of these events. Lastly, my results showed that there were increased risk of thrombocytopenia after first dose ChAdOx1 vaccine compare with mRNA vaccine, the risk of thromboembolic events or TTS after different COVID-19 vaccines were low, and similar across vaccine types.

These findings have important implications for clinical practice, health policy, and future research on assessing the safety of not only to the safety of COVID-19 vaccines but also to other vaccines. Most importantly, given the well-established benefits of COVID-19 vaccination, my finding should encourage continued confidence in vaccination. Additionally, there is a call for increased efforts to build and improve information systems to fully leverage routinely collected health data for research purposes.

Appendix

Appendix to Chapter Two

This section contains additional tables related to Chapter Two: Background.

Searching strategy for literature review

Medline

Medline (Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE®) 1946 to present.

Date: 2022.03.11

1 exp Vaccines/ 254806
2 exp Immunization/ 194262
3 (immuniz* or immunis* or vaccin*).ti,ab. 441585
4 Immunization Programs/ 12234
5 inoculat*.ti,ab. 131594
6 exp Vaccination/ 97570
7 1 or 2 or 3 or 4 or 5 or 6 660669
8 exp Electronic Health Records/ 25283
9 ("health record*" or "medical record*" or datalink or "General Practice data" or "outpatient data" or "admission data" or "emergency department data").ti,ab. 154300
10 ((Admin* adj2 clai*) or (admin* adj2 data*) or "routine data" or "routinely collected" or "health administrative" or "International Classification of Diseases").ti,ab. 45097
11 "International Classification of Diseases"/ 8960
12 (ICD-9-CM or ICD-10).ti,ab. 16721
13 exp Database Management Systems/ 7925
14 exp Medical Records Systems, Computerized/ 45833
15 (CPT or "Current procedural terminology").ti,ab. 15655
16 Current Procedural Terminology/ 1193
17 ("insurance database*" or "data warehouse" or ICD-9 or "international statistical classification" or "drug surveillance").ti,ab. 16895
18 exp Databases, Factual/ 162077
19 Databases as Topic/ 9681
20 Medical Record Linkage/ or incidence/ 294330
21 (ICD-10-CM or database* or "medical records" or "population surveillance" or "data collection" or "automatic data processing" or "patient discharge" or "hospital records").ti,ab. 741186
22 ("healthcare data" or "health care data" or "vaccine safety datalink" or "post-licensure rapid immunization safety monitoring" or "paediatric active enhanced disease surveillance" or medicare or

medicaid or "accelerated development of vaccine" or "canadian immunisation research network" or VAESCO or "vaccine adverse events monitoring and communication").ti,ab. 68650
23 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 1262467
24 exp Product Surveillance, Postmarketing/ 17283
25 Adverse Drug Reaction Reporting Systems/ 8372
26 "Drug-Related Side Effects and Adverse Reactions"/ 35383
27 exp Population Surveillance/ 73815
28 risk/ 127510
29 Risk Assessment/ 297466
30 (safety or surveillance or monitor* or postauthorization or postauthorisation or post-authorisation or post-authorization or post-licensure or postlicensure or post-market or postmarket or "adverse even*" or "adverse effect*").ti,ab. 1860395
31 24 or 25 or 26 or 27 or 28 or 29 or 30 2287981
32 ("Sequential test*" or "sequential analy*" or "sequential metho*" or "sequential monitor*" or "Rapid cycle analys*" or "rapid risk assessment" or "active surveillance" or "real-time monito*" or "early detection" or "sequential probability ratio test*" or maxSPRT or SPRT or "cumulative sum chart*" or "Sequential case series" or "signal refinement" or "signal strengthening" or "signal identification" or "signal generation" or "observed-expected" or "observed-to-expected" or "observed vs. expected" or "observed vs expected" or "observed versus expected" or "observed-vs-expected" or "current-historical" or "standardiSed incidence ratio" or "standardized incidence ratio" or "vaccine safety datalink" or "Post-Licensure Rapid Immunization Safety Monitoring" or "Paediatric Active Enhanced Disease Surveillance" or "historical comparison" or "signal detection" or "historical rate" or "historical rates" or "concurrent comparison" or "real-time surveillance" or "Vaccine Adverse Event Surveillance and Communication" or "observed-over-expected" or "near real time").ti,ab. 110864
33 observational study/ 122812
34 exp case-control studies/ 1293864
35 exp Cohort Studies/ 2309204
36 (cohort* or "follow up" or longitudinal or retrospective* or prospective* or (observ* and stud*) or "adverse effect*" or observational or "non-randomi* stud*" or case-control or self-contro* or SCCS or "historical rate*" or "historically-control*" or case-crossover or "expected cases" or "expected rate*" or "background rate*").ti,ab. 4922674
37 32 or 33 or 34 or 35 or 36 5775573
38 SARS-CoV-2/ or COVID-19/ 147119
39 (corona* adj1 (virus* or viral*)).ti,ab,kw,kf. 4809
40 (CoV not (Coefficien* or "co-efficien*" or covalent* or Covington* or covariant* or covarianc* or "cut-off value*" or "cutoff value*" or "cut-off volume*" or "cutoff volume*" or "combined optimi?ation value*" or "central vessel trunk*" or CoVR or CoVS)).ti,ab,kw,kf. 80046
41 (coronavirus* or 2019nCoV* or 19nCoV* or "2019 novel*" or Ncov* or "n-cov" or "SARS-CoV-2*" or "SARSCoV-2*" or SARSCoV2* or "SARS-CoV2*" or "severe acute respiratory syndrome*" or COVID*2).ti,ab,kw,kf.242650
42 or/38-41 249312
43 7 and 23 and 31 and 37 and 42 278
44 limit 43 to yr="2020-Current" 273
45 limit 44 to (case reports or comment or editorial or letter or historical article or comment or editorial or news) 9
46 44 not 45 264
47 (46 and english.lg.) not (Animals/ not humans/) 260

EMBASE

Date: 2022/03/11

Embase 1974 to present

1 exp Vaccines/ 369350
2 exp Immunization/ 319313
3 (immuniz* or immunis* or vaccin*).ti,ab. 502938
4 Immunization Programs/ 27914
5 inoculat*.ti,ab. 137908
6 exp Vaccination/ 197178
7 1 or 2 or 3 or 4 or 5 or 6 783262
8 exp Electronic Health Records/ 28527
9 ("health record*" or "medical record*" or datalink or "General Practice data" or "outpatient data" or "admission data" or "emergency department data").ti,ab. 259353
10 ((Admin* adj2 clai*) or (admin* adj2 data*) or "routine data" or "routinely collected" or "health administrative" or "International Classification of Diseases").ti,ab. 65004
11 "International Classification of Diseases"/ 15408
12 (ICD-9-CM or ICD-10).ti,ab. 36304
13 exp Database Management Systems/ 574
14 exp Medical Records Systems, Computerized/ 1939
15 (CPT or "Current procedural terminology").ti,ab. 24113
16 Current Procedural Terminology/ 4985
17 ("insurance database*" or "data warehouse" or ICD-9 or "international statistical classification" or "drug surveillance").ti,ab. 41166
18 exp Databases, Factual/ 75922
19 Databases as Topic/ 245987
20 Medical Record Linkage/ or incidence/ 675777
21 (ICD-10-CM or database* or "medical records" or "population surveillance" or "data collection" or "automatic data processing" or "patient discharge" or "hospital records").ti,ab. 1094317
22 ("healthcare data" or "health care data" or "vaccine safety datalink" or "post-licensure rapid immunization safety monitoring" or "paediatric active enhanced disease surveillance" or medicare or medicaid or "accelerated development of vaccine" or "canadian immunisation research network" or VAESCO or "vaccine adverse events monitoring and communication").ti,ab. 98312
23 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 1912802
24 exp Product Surveillance, Postmarketing/ 37658
25 Adverse Drug Reaction Reporting Systems/ 3149
26 "Drug-Related Side Effects and Adverse Reactions"/ 188403
27 exp Population Surveillance/ 90
28 risk/ 504463
29 Risk Assessment/ 656654

30 (safety or surveillance or monitor* or postauthorization or postauthorisation or post-authorisation or post-authorization or post-licensure or postlicensure or post-market or postmarket or "adverse even*" or "adverse effect*").ti,ab. 2666975
31 24 or 25 or 26 or 27 or 28 or 29 or 30 3785065
32 ("Sequential test*" or "sequential analy*" or "sequential metho*" or "sequential monitor*" or "Rapid cycle analys*" or "rapid risk assessment" or "active surveillance" or "real-time monito*" or "early detection" or "sequential probability ratio test*" or maxSPRT or SPRT or "cumulative sum chart*" or "Sequential case series" or "signal refinement" or "signal strengthening" or "signal identification" or "signal generation" or "observed-expected" or "observed-to-expected" or "observed vs. expected" or "observed vs expected" or "observed versus expected" or "observed-vs-expected" or "current-historical" or "standardiSed incidence ratio" or "standardized incidence ratio" or "vaccine safety datalink" or "Post-Licensure Rapid Immunization Safety Monitoring" or "Paediatric Active Enhanced Disease Surveillance" or "historical comparison" or "signal detection" or "historical rate" or "historical rates" or "concurrent comparison" or "real-time surveillance" or "Vaccine Adverse Event Surveillance and Communication" or "observed-over-expected" or "near real time").ti,ab. 155192
33 observational study/ 264980
34 exp case-control studies/ 203210
35 exp Cohort Studies/ 816192
36 (cohort* or "follow up" or longitudinal or retrospective* or prospective* or (observ* and stud*) or "adverse effect*" or observational or "non-randomi* stud*" or case-control or self-contro* or SCCS or "historical rate*" or "historically-control*" or case-crossover or "expected cases" or "expected rate*" or "background rate*").ti,ab. 7097092
37 32 or 33 or 34 or 35 or 36 7358332
38 exp severe acute respiratory syndrome coronavirus 2/ or coronavirus disease 2019/ or experimental coronavirus disease 2019/ 208550
39 (corona* adj1 (virus* or viral*)).ti,ab,kw. 4524
40 (CoV not (Coefficien* or co-efficien* or covalent* or covington or covariant* or covarianc* or "cut-off value*" or "cutoff value*" or "cut-off volume*" or "cutoff volume*" or "combined optimi?ation value*" or "central vessel trunk" or CoVR or CoVS)).ti,ab,kw. 74597
41 (coronavirus* or 2019nCoV* or 19nCoV* or "2019 novel*" or Ncov* or "n-cov" or "SARS-CoV-2*" or "SARSCoV-2*" or SARSCoV2* or "SARS-CoV2*" or "severe acute respiratory syndrome*" or COVID*2).ti,ab,kw. 256289
42 or/38-41 274257
43 7 and 23 and 31 and 37 and 42 587
44 limit 43 to yr="2020-Current" 576
45 limit 44 to (case reports or comment or editorial or letter or historical article or comment or editorial or news) 12
46 44 not 45 564
47 (46 and english.lg.) not (Animals/ not humans/) 561

Scopus

Results: 338

((TITLE-ABS-KEY (("health record*" OR "medical record*" OR datalink OR "General Practice data" OR "outpatient data" OR "admission data" OR "emergency department data" OR (admin* W/2 clai*) OR (admin* W/2 data*) OR "routine data" OR "routinely collected" OR "health

administrative" OR "International Classification of Diseases" OR icd-9-cm OR icd-10 OR cpt OR "Current procedural terminology" OR "insurance database*" OR "data warehouse" OR icd-9 OR "international statistical classification" OR "drug surveillance" OR icd-10-cm OR database* OR "medical records" OR "population surveillance" OR "data collection" OR "automatic data processing" OR "patient discharge" OR "hospital records" OR "healthcare data" OR "health care data" OR "vaccine safety datalink" OR "post-licensure rapid immunization safety monitoring" OR "paediatric active enhanced disease surveillance" OR medicare OR medicaid OR "accelerated development of vaccine" OR "canadian immunisation research network" OR vaesco OR "vaccine adverse events monitoring and communication")) AND TITLE-ABS-KEY (immuniz* OR immunis* OR vaccin* OR inoculat*) AND TITLE-ABS-KEY (safety OR surveillance OR monitor* OR postauthorization OR postauthorisation OR post-authorisation OR post-authorization OR post-licensure OR postlicensure OR post-market OR postmarket OR "adverse even*" OR "adverse effect*") AND TITLE-ABS-KEY (("Sequential test*" OR "sequential analy*" OR "sequential metho*" OR "sequential monitor*" OR "Rapid cycle analys*" OR "rapid risk assessment" OR "active surveillance" OR "real-time monito*" OR "early detection" OR "sequential probability ratio test*" OR maxsprt OR sprt OR "cumulative sum chart*" OR "Sequential case series" OR "signal refinement" OR "signal strengthening" OR "signal identification" OR "signal generation" OR "observed-expected" OR "observed-to-expected" OR "observed vs. expected" OR "observed vs expected" OR "observed versus expected" OR "observed-vs-expected" OR "current-historical" OR "standardiSed incidence ratio" OR "standardized incidence ratio" OR "vaccine safety datalink" OR "Post-Licensure Rapid Immunization Safety Monitoring" OR "Paediatric Active Enhanced Disease Surveillance" OR "historical comparison" OR "signal detection" OR "historical rate" OR "historical rates" OR "concurrent comparison" OR "real-time surveillance" OR "Vaccine Adverse Event Surveillance and Communication" OR "observed-over-expected" OR "near real time" OR cohort* OR "follow up" OR longitudinal OR retrospective* OR prospective* OR (observ* AND stud*) OR "adverse effect*" OR observational OR "non-randomi* stud*" OR case-control OR self-contro* OR sccs OR "historical rate*" OR "historically-control*" OR case-crossover OR "expected cases" OR "expected rate*" OR "background rate*")))) AND TITLE-ABS-KEY ("Wuhan coronavirus" OR "Wuhan seafood market pneumonia virus" OR "COVID19*" OR "COVID-19*" OR "COVID-2019*" OR "coronavirus disease 2019" OR "SARS-CoV-2" OR sars2 OR "2019-nCoV" OR "2019 novel coronavirus" OR "severe acute respiratory syndrome coronavirus 2" OR "2019 novel coronavirus infection" OR "coronavirus disease 2019" OR "coronavirus disease-19" OR "novel coronavirus" OR coronavirus OR "SARS-CoV-2019" OR "SARS-CoV-19") AND NOT (TITLE ("case report" OR comment OR editorial)) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2022))

Cochrane COVID-19 Study Register

Filtered by:
immuniz* or immunis* or vaccin* or inoculat*
AND
safety or surveillance or monitor* or postauthorization or postauthorisation or post-authorisation or post-authorization or post-licensure or postlicensure or post-market or postmarket or "adverse even*" or "adverse effect*"

Appendix

Included studies in literature review

Study	Study period	Country /Region	Data source	data type	Study design	study population	Vaccine & dose	comparator	Outcomes & risk window	confounding/error control & main analysis
(Ip et al. 2022)	till May 17th 2021	UK	NHS Digital Trusted Research Environment	EHRs	cohort	12+	BNT162b2, ChAdOx1; 1st, 2nd	unvaccinated	hospitalised or fatal myocarditis/pericarditis; 0-13; 14+ d	adjust for covariates, IPTW ; adjusted Cox
(Lai, Li, Peng, et al. 2022)		Hong Kong	Hospital Authority (HA)	EHRs	case-control	12+, Inpatients, first diagnosed with carditis	BNT162b2, CoronaVac		Carditis	
(Xu et al. 2021)		USA	Vaccine Safety Datalink (VSD)	EHRs	cohort	12+	BNT162b2, mRNA-1273, Ad26.COV2.S	unvaccinated	AESIs	Standardized mortality rates, adjusted relative risk
(Lai, Chua, Chan, et al. 2022)		Hong Kong	Hospital Authority (HA)	EHRs	cohort, comparative	12-18y	BNT162b2; 1st, 2nd	unvaccinated	30 AESI; 28d	age-sex match, vaccinated date
(Massari et al. 2022)	7 Dec 2020 -	Italy	TheShinISS: multi databases	claims	SCCS	12-39y	BNT162b2, mRNA-1273 ; 1st, 2nd		myocarditis and pericarditis; 31d	Calendar period was included as time-varying confounder

	30 Sept 2021									
(Lloyd et al. 2022)	Dec 11, 2020, through Jan 22, 2022, Jan 7, 2022, and Dec 31, 2021.	USA	Optum, HealthCore, and CVS Health databases	claims	Near real-time surveillance	12–64 y	BNT162b2, mRNA-1273, Ad26.COV2.S		17 outcomes of interest	biweekly or monthly sequential testing and generated rate ratios (RR) of observed outcome rates compared to historical (or expected) rates prior to COVID-19 vaccination.
(Xue Li, Gao, et al. 2022)	Jan 1, 2018 and June 30, 2021	Hong Kong	Hospital Authority (HA)	EHRs	descriptive cohort study	16+	BNT162b2, CoronaVac	matched unvaccinated	Autoimmune conditions	standardized incidence rate ratios (IRRs)
(Shibli et al. 2021)		Israel	Computerized database of Clalit Health Services (CHS)	EHRs	OvE	16+	BNT162b2		Bell's palsy; 21 (dose 1), 30 (2nd dose)	Standardized incidence ratios (SIRs), attributable risks (ARs)
(Yanir et al. 2022)	Dec 20, 2020, to May 31, 2021	Israel		EHRs	OvE	16+	BNT162b2; 1st, 2nd	population 2018-19	sudden sensorineural hearing loss (SSNHL)	Standardized incidence ratios (SIRs)

(Hippisley-Cox et al. 2021)	1 Dec 2020 and 24 April 2021	UK	QResearch	EHRs	SCCS	16+ & with outcome of interest	BNT162b2, ChAdOx1; 1st		hospital admission or death associated with thrombocytopenia, venous thromboembolism, and arterial thromboembolism; 28d	
(Sing et al. 2022)	Jan 1, 2018 and July 31, 2021	Hong Kong	Hospital Authority (HA)	EHRs	SCCS and Nested case–control	16+	BNT162b2, CoronaVac		hematological abnormalities	Adjusted odds ratios (aORs), incidence rate ratios (IRRs)
(Wan, Chui, Mok, et al. 2022)	23 February 2021 and 31 Jan 2022	Hong Kong	Hospital Authority (HA)	EHRs	SCCS	16+, T2B patients with at least 1 dose.	BNT162b2, CoronaVac		29 AESIs and acute diabetic complications; 21d	
(Shasha, Heymann, and Zacay 2022)	19 Dec 2020 to 12 February 12 2021.	Israel	Meuhedet Health Maintenance Organization (MHMO)	EHRs	OvE, historical cohort	16+	BNT162b2	unvaccinated	Bell's palsy, GBS, herpes zoster or symptoms of numbness or tingling	matched to a non-vaccinated control by sex, age, population sector (general Jewish, Arab, ultra-orthodox Jewish) and comorbidities. OR from logistic regression

(McMurry et al. 2021)	Dec 1, 2020 and April 20, 2021	USA	Mayo electronic health record (EHR) database	EHRs	cohort	18+	BNT162b2, mRNA-1273	unvaccinated	fatigue, fever, chills, myalgia, arthralgia, headache, lymphadenopathy, erythema, diarrhea, vomiting, and local pain and swelling, anaphylaxis, facial paralysis (Bell's palsy), and cerebral venous sinus thrombosis (CVST); 1,7,14,21d	PS matching ; incidence rate ratio (IRR) of the vaccinated and unvaccinated cohorts
(C. K. H. Wong, Lui, et al. 2022)	23 February 2021 to 30 September 2021	Hong Kong	Hospital Authority (HA)	EHRs	SCCS	18+	BNT162b2, CoronaVac; 1st, 2nd		Thyroid dysfunction encompassed anti-thyroid drug (ATD)/levothyroxine (LT4) initiation, biochemical picture of hyperthyroidism/ hypothyroidism, incident Graves' disease (GD), and thyroiditis. 56d	

(Wan et al. 2021)		Hong Kong	Hospital Authority (HA)	EHRs	case series and nested case-control	18+	BNT162b2, CoronaVac	Non cases	Bell's palsy; 42d	age-standardised difference; nested case-control study was also done using conditional logistic regression to estimate the odds ratio (OR)
(Whiteley et al. 2022)		UK	GDPPR	EHRs	cohort	18+	BNT162b2, ChAdOx1		major venous, arterial, or thrombocytopenic events; 1 to 28 and >28 days	HR
(Simone et al. 2022)	12/14/2020 and 2/18/2022	USA	Kaiser Permanente Southern California (KPSC)	EHRs	cohort	18+	BNT162b2, mRNA-1273; booster (3rd)	background population	Myocarditis; 21d	IR, IRR
(Xiong et al. 2022)	February 23, 2021, and September 9, 2021	Hong Kong	Hospital Authority (HA)	EHRs	cohort	18+, levothyroxine (LT4) users	BNT162b2, CoronaVac	unvaccinated, between vaccines	dosage reduction or escalation in LT4, emergency department (ED) visit, unscheduled hospitalization, AESI, and all-cause mortality	IPTW weighting; Cox regression models were fitted to estimate the hazard ratios (HR)
(Niesen et al. 2022)	Dec 2020 to October	USA	Mayo Clinic Enterprise	EHRs	cohort	18+, with booster	BNT162b2, mRNA-1273; booster (3rd)	before first dose;	anaphylaxis, arthralgia, cerebral venous	risk difference (RD). The RD of an adverse event was defined as the difference between the

	2021, Data were analyzed from September through November 2021.							after second dose	sinus thrombosis, chills, diarrhea, erythema, facial paralysis, fatigue, fever, headache, local pain, local swelling, lymphadenopathy, myalgia, myocarditis, nausea, pericarditis, soreness, and vomiting. ; 14d	percentage of the cohort that reported the adverse event after dose 3 and the percentage of the cohort that reported the adverse event during the control period (ie, before dose 1 or after dose 2).
(Goddard et al. 2022)	through Jan 15, 2022	USA	Vaccine Safety Datalink (VSD)	EHRs	cohort, OvE	18-39 y	BNT162b2, mRNA-1273; 1st, 2nd		myocarditis and pericarditis	Rate ratios estimated by conditional Poisson regression, adjusted for age, sex, race and ethnicity, health plan, and calendar day. Head-to-head comparison directly assessed risk following mRNA-1273 versus BNT162b2 during 0–7 days post-vaccination.
(H.-L. Wong, Hu, et al. 2022)		USA	Optum, HealthCore, Blue Health Intelligence, and CVS Health	claims	OvE	18-64 y	BNT162b2, mRNA-1273		myocarditis and pericarditis; 1-7 d, (1–21 days and 1–42 days for sensitivity)	multivariate Poisson regression to estimate the adjusted incidence rates, by brand of vaccine, and incidence rate ratios (IRRs) comparing mRNA-1273 and BNT162b2. Meta-analyses to pool

(Pottegård et al. 2021)	vaccinated 9 February 2021 to 11 March 2021	Denmark and Norway	multi databases	registry	OvE	18-65 y	ChAdOx1	general populations of Denmark (2016-18) and Norway (2018-19)	cardiovascular and haemostatic events; 28d	
(Botton et al. 2022)	27 Dec 2020 to 20 July 2021	France	French National Health Data System (Système National des Données de Santé [SNDS])	EHRs	SCCS (adapted)	18-74 y	BNT162b2, mRNA-1273, Ad26.COV2.S, ChAdOx1; 1st, 2nd		hospitalization with Myocardial Infarction, Stroke, and Pulmonary Embolism; 21d	
(Wan, Wang, et al. 2022)	Feb 23, 2021, and Jan 31, 2022	Hong Kong	Hospital Authority (HA)	EHRs	SCCS (modified)	60+	CoronaVac		30 adverse events of special interest; 21d	
(Jabagi et al. 2022)	Dec 15, 2020, and April 30, 2021	French	French National Health Data System linked to the national COVID-19	EHRs	SCCS	75+	BNT162b2		acute myocardial infarction, hemorrhagic stroke, ischemic stroke, or pulmonary embolism; 14d	

			vaccination database							
(Kang et al. 2022)	Dx between Jan 1, 2018, and September 30, 2021,	Hong Kong	Hospital Authority (HA)	EHRs	cohort	active cancer and history of cancer; 18+	BNT162b2, CoronaVac	unvaccinated	AESIs; 28d	unvaccinated patients, the pseudo index date was selected from a corresponding vaccine recipient matched on age and sex. Cox proportional hazards regression
(Ye et al. 2022)		Hong Kong	Hospital Authority (HA)	EHRs	SCCS	CVD patient	BNT162b2, CoronaVac		MACE	
(Lipkind et al. 2022)	Dec 15, 2020–July 22, 2021	USA	Vaccine Safety Datalink (VSD)	EHRs	cohort	Females aged 16–49 years with pregnancy	BNT162b2, mRNA-1273, Ad.26.COV2.S	unvaccinated	Preterm or Small-for-Gestational-Age at Birth	Time-dependent COVID-19 vaccine and COVID-19 diagnosis Cox models with standardized inverse probability weighting were used to estimate the aHR
(Hviid et al. 2022)	27 Dec 2020 to 13 April 2021.	Norway and Germany	Danish linkable registers on vaccinations, hospitalizations, occupation, and other covariates	registry	cohort, comparative	Frontline Personnel	BNT162b2, ChAdOx1	unvaccinated	Thromboembolic and Thrombocytopenic; 28d	risk differences

(Lai, Huang, Peng, et al. 2022)	Jan 1, 2018, and July 31, 2021	Hong Kong	Hospital Authority (HA)	EHRs	cohort	general population	BNT162b2, CoronaVac	vaccinated without covid-19 infection	30 AESIs; 28d	ps weighting
(Tu et al. 2022)	Jan 23, 2020, to August 3, 2021	Singapore		EHRs	cohort	general population	BNT162b2, mRNA-1273	SARS-CoV-2 Infection	CVST; 6 weeks	incidence rate ratio (IRR)
(Klein et al. 2021)	Dec 14, 2020, through June 26, 2021	USA	Vaccine Safety Datalink (VSD)	EHRs	cohort	general population	BNT162b2, mRNA-1273; 1st, 2nd	22 to 42 days for similar individuals after vaccine dose 1 or 2	acute myocardial infarction, Bell palsy, cerebral venous sinus thrombosis, Guillain-Barré syndrome, myocarditis/pericarditis, pulmonary embolism, stroke, and thrombosis with thrombocytopenia syndrome. 21d	Rate ratios (RRs) were estimated by Poisson regression, adjusted for age, sex, race and ethnicity, health plan, and calendar day.
(Goldshtein et al. 2022)	till Oct31, 2021	Israel	comprehensive database of the Maccabi	EHRs	cohort	Maternal BNT162b2 mRNA vaccination during	BNT162b2	unexposed to maternal vaccination	adverse neonatal and early infant outcomes (preterm birth, small birth weight for gestational	Stabilized inverse probability weighting; Risk ratios (RR)

			Healthcare Services			pregnancy.			age (SGA), congenital malformations, all-cause hospitalizations, and infant death)	
(Hadi et al. 2021)		USA	TriNetX	EHRs	cohort	patient with IBD	BNT162b2, mRNA-1273, Ad26.COV2.S	non-IBD	A list of AESIs	1-to-1 matching based on selected variables.
(Theiler et al. 2021)	Dec 10, 2020, and April 19, 2021	USA	vaccine registry+Mayo Clinic delivery registry	registry	cohort	patients aged 16 to 55 years with a delivery	BNT162b2, mRNA-1273, Ad26.COV2.S	unvaccinated	Pregnancy and birth outcomes	
(Pakhchanian et al. 2022)		USA	TriNetX	EHRs	cohort	patients with dermatomyositis	BNT162b2, mRNA-1273, Ad26.COV2.S	non-DM	AESIs	1-to-1 matching based on selected variables.
(Xue Li, Tong, et al. 2022)	31-Jul-21	Hong Kong	Hospital Authority (HA)	EHRs	cohort	Patients with rheumatoid arthritis (RA)	BNT162b2, CoronaVac		arthritis flare	matched each vaccine recipient with non-vaccinated individuals by age and sex using maximum ratio matching and assigned the vaccination date as the pseudo index date for non-vaccinated individuals (controls). multi-group Inverse Probability. Poisson regression to estimate the adjusted incidence rate ratio (IRR)

(Fell et al. 2022)	All births between Dec 14, 2020, and September 30, 2021	Canada	Better Outcomes Registry & Network Ontario Information System	EHRs	cohort	Women with pregnancy	BNT162b2, mRNA-1273, ChAdOx1	vaccinated after pregnancy and those with no record of COVID-19 vaccination at any point	Postpartum hemorrhage, chorioamnionitis, cesarean delivery (overall and emergency cesarean delivery), admission to neonatal intensive care unit (NICU), and low newborn 5-minute Apgar score (<7).	adjusted risk differences (aRDs) and risk ratios (aRRs), cumulative incidence
(Kerr et al. 2022)	Dec 8, 2020, and the end date was June 30, 2021	UK	linked primary care, secondary care, mortality, and virological testing data in each of England, Scotland, and Wales	EHRs	SCCS (pooled)		BNT162b2, ChAdOx1; 1st		CVST; 4 weeks	SCCS, the pool across
(Hertel et al. 2022)		USA	TriNetX	EHRs	cohort	vaccinated, a visit of the	BNT162b2, mRNA-1273, Ad26.COV2.S	unvaccinated	Onset of oral lichenoid lesions (OLL) or oral	matched for age, gender and the frequency of use of non-steroidal anti-inflammatory

						HCO for evaluation and management services, and a most recent BMI value of 19–30 kg/m2.			lichen planus (OLP); 6d	drugs, beta blockers, and angiotensin-converting enzyme inhibitors; Risk ratio (RR) as well as odds ratio (OR)
(Dickerman et al. 2022)	Jan 4 and September 20, 2021	USA	Veterans Health Administration	EHRs	cohort	Veterans	BNT162b2, mRNA-1273		AESIs	
(L. L. Li, Zheng, et al. 2022)	12/11/2020 and 8/31/2021	USA	Veterans Health Administration	EHRs	nested case-control	Veterans	BNT162b2, mRNA-1273		Hospitalisation follow vaccine	
(Lai, Huang, Chui, et al. 2022)	1 Jan 2018 and 31 July 2021	Hong Kong	Hospital Authority (HA)	EHRs	cohort	with 20 chronic conditions	BNT162b2, CoronaVac	unvaccinated	30 AESIs; 28d	age-sex matching to unvaccinated, Multiple vaccination group weighting for cohort in outcome model ; Cox - PH hazard
(Dick, Rosenbloom, Gutman-Ido, et al. 2022)	Dec 2020 and July 2021	Israel		EHRs	cohort	women who delivered	BNT162b2, mRNA-1273	unvaccinated	incidence of preterm labor and of small for gestational age	Adjusted Odds Ratios and 95% Confidence Intervals for Vaccinated vs. Unvaccinated Parturients

(Rottenstreich et al. 2022)		Israel		EHRs	cohort	Women who gave birth at >24 weeks of gestation	BNT162b2; 2nd	unvaccinated	maternal and neonatal outcomes	aOR from multivariable logistic regression
(Dick, Rosenbloom, Karavani, et al. 2022)	Dec 2020 and July 2021	Israel		EHRs	cohort	women with pregnancy	BNT162b2, mRNA-1273; booster	not vaccinated and with those who only received 2 vaccination doses	incidence of preterm labor and of small for gestational age neonates	odds ratios
(Mevorach et al. 2021)	Dec 20, 2020, to May 31, 2021	Israel	medical records obtained from the Ministry of Health database	EHRs	cohort		BNT162b2 ; 1st, 2nd	first vs. second; historical cohort; concurrent unvaccinated cohort	Myocarditis; 21 (dose 1), 30 (2nd dose)	standardized incidence ratio of the observed-to-expected incidence
(Pawlowski et al. 2021)	01/01/2017 and 03/15/2021.	USA	Mayo Clinic Health System	EHRs	cohort		covid-19 vac, other vacs		CVST; 30d	relative risk: post-vaccination incidence in the (+1 to +30 days) divided by the pre-vaccination incidence (-30 to -1 days)

(C. K. H. Wong, Lau, et al. 2022)	vaccinated 23 Feb to 9 Sept 2021	Hong Kong	Hospital Authority (HA)	EHRs	cohort, comparative		BNT162b2, CoronaVac; 1st, 2nd	BNT162b2	AESIs and all-cause mortality; 2-21	IRR from Poisson
(Ohaeri et al. 2022)		UK	NHS Wales	EHRs	cohort, comparative		BNT162b2, ChAdOx1	SARS-CoV-2 infection	CVST; 28d	hazard ratio
(C. K. H. Wong, Mak, et al. 2022)		Hong Kong	Hospital Authority (HA)	EHRs	SCCS (modified)		BNT162b2, CoronaVac		acute liver injury; 56d	
(Chui et al. 2022)	23 Febr to 30 Sept 2021	Hong Kong	Hospital Authority (HA)	EHRs	SCCS (modified)		BNT162b2, CoronaVac		thromboembolic events and hemorrhagic stroke	
(Xintong Li, Raventós, et al. 2022)		UK, Spain	CPRD, SIDIAP	EHRs	OvE and SCCS		BNT162b2, mRNA-1273, Ad26.COV2.S, ChAdOx1		Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis; 28d	
(Hanson et al. 2022)	Dec 13 2020 - Nov 13, 2021, Nov 2021 to	USA	Vaccine Safety Datalink (VSD)	EHRs	OvE, litea ature		BNT162b2,mRNA-1273, Ad.26.COV2.S; 1st, 2nd	background rate from literature	Guillain-Barré syndrome (GBS)	rate ratios (RRs) compare with background rates

	Feb 2022.									
(Keh et al. 2023)		UK	National Immunoglobulin Database + National Immunisation Management System in England		rate compare		BNT162b2, mRNA-1273, ChAdOx1		Guillain-Barré syndrome (GBS); 6 weeks	
(Wan, Chui, Wang, et al. 2022)	Feb 23 to July 31, 2021	Hong Kong	Hospital Authority (HA)	EHRs	SCCS, nested case-control		BNT162b2, CoronaVac		Herpes zoster related hospitalisation. 0-13; 14-27	

References for literature review

Botton, Jérémie, Marie Joelle Jabagi, Marion Bertrand, Bérangère Baricault, Jérôme Drouin, Stéphane Le Vu, Alain Weill, Paddy Farrington, Mahmoud Zureik, and Rosemary Dray-Spira. 2022. "Risk for Myocardial Infarction, Stroke, and Pulmonary Embolism Following COVID-19 Vaccines in Adults Younger than 75 Years in France." *Annals of Internal Medicine* 175 (9): 1250–57.

Chui, Celine Sze Ling, Min Fan, Eric Yuk Fai Wan, Miriam Tim Yin Leung, Edmund Cheung, Vincent Ka Chun Yan, Le Gao, et al. 2022. "Thromboembolic Events and Hemorrhagic Stroke after MRNA (BNT162b2) and Inactivated (CoronaVac) Covid-19 Vaccination: A Self-Controlled Case Series Study." *EClinicalMedicine* 50 (101504): 101504.

Dick, Aharon, Joshua I. Rosenbloom, Einat Gutman-Ido, Naama Lessans, Adva Cahen-Peretz, and Henry H. Chill. 2022. "Safety of SARS-CoV-2 Vaccination during Pregnancy- Obstetric Outcomes from a Large Cohort Study." *BMC Pregnancy and Childbirth* 22 (1): 166.

Dick, Aharon, Joshua I. Rosenbloom, Gilad Karavani, Einat Gutman-Ido, Naama Lessans, and Henry H. Chill. 2022. "Safety of Third SARS-CoV-2 Vaccine (Booster Dose) during Pregnancy." *American Journal of Obstetrics & Gynecology MFM* 4 (4): 100637.

Dickerman, Barbra A., Arin L. Madenci, Hanna Gerlovin, Katherine E. Kurgansky, Jessica K. Wise, Michael J. Figueroa Muñiz, Brian R. Ferolito, et al. 2022. "Comparative Safety of BNT162b2 and MRNA-1273 Vaccines in a Nationwide Cohort of US Veterans." *JAMA Internal Medicine* 182 (7): 739–46.

Fell, Deshayne B., Tavleen Dhinsa, Gillian D. Alton, Eszter Török, Sheryll Dimanlig-Cruz, Annette K. Regan, Ann E. Sprague, et al. 2022. "Association of COVID-19 Vaccination in Pregnancy with Adverse Peripartum Outcomes." *JAMA: The Journal of the American Medical Association* 327 (15): 1478–87.

Goddard, Kristin, Ned Lewis, Bruce Fireman, Eric Weintraub, Tom Shimabukuro, Ousseny Zerbo, Thomas G. Boyce, et al. 2022. "Risk of Myocarditis and Pericarditis Following BNT162b2 and MRNA-1273 COVID-19 Vaccination." *Vaccine* 40 (35): 5153–59.

Goldshtein, Inbal, David M. Steinberg, Jacob Kuint, Gabriel Chodick, Yaakov Segal, Shirley Shapiro Ben David, and Amir Ben-Tov. 2022. "Association of BNT162b2 COVID-19 Vaccination during Pregnancy with Neonatal and Early Infant Outcomes." *JAMA Pediatrics* 176 (5): 470–77.

Hadi, Yousaf Bashir, Shyam Thakkar, Sardar Momin Shah-Khan, William Hutson, Arif Sarwari, and Shailendra Singh. 2021. "COVID-19 Vaccination Is Safe and Effective in Patients with Inflammatory Bowel Disease: Analysis of a Large Multi-Institutional Research Network in the United States." *Gastroenterology* 161 (4): 1336-1339.e3.

Hanson, Kayla E., Kristin Goddard, Ned Lewis, Bruce Fireman, Tanya R. Myers, Nandini Bakshi, Eric Weintraub, et al. 2022. "Incidence of Guillain-Barré Syndrome after COVID-19 Vaccination in the Vaccine Safety Datalink." *JAMA Network Open* 5 (4): e228879.

Hertel, Moritz, Andrea-Maria Schmidt-Westhausen, Stephanie Wendy, Max Heiland, Susanne Nahles, Robert Preissner, and Saskia Preissner. 2022. "Onset of Oral Lichenoid Lesions and Oral Lichen Planus Following COVID-19 Vaccination: A Retrospective Analysis of about 300,000 Vaccinated Patients." *Vaccines* 10 (3): 480.

Hippisley-Cox, Julia, Martina Patone, Xue W. Mei, Defne Saatci, Sharon Dixon, Kamlesh Khunti, Francesco Zaccardi, et al. 2021. "Risk of Thrombocytopenia and Thromboembolism after Covid-19 Vaccination and SARS-CoV-2 Positive Testing: Self-Controlled Case Series Study." *BMJ (Clinical Research Ed.)* 374 (August): n1931.

Hviid, Anders, Jørgen Vinsløv Hansen, Emilia Myrup Thiesson, and Jan Wohlfahrt. 2022. "Association of AZD1222 and BNT162b2 COVID-19 Vaccination with Thromboembolic and Thrombocytopenic Events in Frontline Personnel : A Retrospective Cohort Study." *Annals of Internal Medicine* 175 (4): 541–46.

Ip, Samantha, Fatemeh Torabi, Spiros Denaxas, Ashley Akbari, Hoda Abbasizanjani, Rochelle Knight, Jennifer Cooper, et al. 2022. "Risk of Myocarditis and Pericarditis Following BNT162b2 and ChAdOx1 COVID-19 Vaccinations." *BioRxiv*. https://doi.org/10.1101/2022.03.06.21267462.

Jabagi, Marie Joelle, Jérémie Botton, Marion Bertrand, Alain Weill, Paddy Farrington, Mahmoud Zureik, and Rosemary Dray-Spira. 2022. "Myocardial Infarction, Stroke, and Pulmonary Embolism after BNT162b2 MRNA COVID-19 Vaccine in People Aged 75 Years or Older." *JAMA: The Journal of the American Medical Association* 327 (1): 80–82.

Kang, Wei, Jessica J. P. Shami, Vincent K. C. Yan, Xuxiao Ye, Joseph E. Blais, Xue Li, Victor H. F. Lee, et al. 2022. “Safety of Two-Dose COVID-19 Vaccination (BNT162b2 and CoronaVac) in Adults with Cancer: A Territory-Wide Cohort Study.” *Journal of Hematology & Oncology* 15 (1): 66.

Keh, Ryan Y. S., Sophie Scanlon, Preeti Datta-Nemdharry, Katherine Donegan, Sally Cavanagh, Mark Foster, David Skelland, et al. 2023. “COVID-19 Vaccination and Guillain-Barré Syndrome: Analyses Using the National Immunoglobulin Database.” *Brain: A Journal of Neurology* 146 (2): 739–48.

Kerr, Steven, Mark Joy, Fatemeh Torabi, Stuart Bedston, Ashley Akbari, Utkarsh Agrawal, Jillian Beggs, et al. 2022. “First Dose ChAdOx1 and BNT162b2 COVID-19 Vaccinations and Cerebral Venous Sinus Thrombosis: A Pooled Self-Controlled Case Series Study of 11.6 Million Individuals in England, Scotland, and Wales.” *PLoS Medicine* 19 (2): e1003927.

Klein, Nicola P., Ned Lewis, Kristin Goddard, Bruce Fireman, Ousseny Zerbo, Kayla E. Hanson, James G. Donahue, et al. 2021. “Surveillance for Adverse Events after COVID-19 MRNA Vaccination.” *JAMA: The Journal of the American Medical Association* 326 (14): 1390–99.

Lai, Francisco Tsz Tsun, Gilbert T. Chua, Edward Wai Wa Chan, Lei Huang, Mike Yat Wah Kwan, Tiantian Ma, Xiwen Qin, et al. 2022. “Adverse Events of Special Interest Following the Use of BNT162b2 in Adolescents: A Population-Based Retrospective Cohort Study.” *Emerging Microbes & Infections* 11 (1): 885–93.

Lai, Francisco Tsz Tsun, Lei Huang, Celine Sze Ling Chui, Eric Yuk Fai Wan, Xue Li, Carlos King Ho Wong, Edward Wai Wa Chan, et al. 2022. “Multimorbidity and Adverse Events of Special Interest Associated with Covid-19 Vaccines in Hong Kong.” *Nature Communications* 13 (1): 411.

Lai, Francisco Tsz Tsun, Lei Huang, Kuan Peng, Xue Li, Celine Sze Ling Chui, Eric Yuk Fai Wan, Carlos King Ho Wong, Esther Wai Yin Chan, Ivan Fan Ngai Hung, and Ian Chi Kei Wong. 2022. “Post-Covid-19-Vaccination Adverse Events and Healthcare Utilization among Individuals with or without Previous SARS-CoV-2 Infection.” *Journal of Internal Medicine* 291 (6): 864–69.

Lai, Francisco Tsz Tsun, Xue Li, Kuan Peng, Lei Huang, Patrick Ip, Xinning Tong, Celine Sze Ling Chui, et al. 2022. “Carditis after COVID-19 Vaccination with a Messenger RNA Vaccine and an Inactivated Virus Vaccine : A Case-Control Study.” *Annals of Internal Medicine* 175 (3): 362–70.

Li, Lucy L., Chunlei Zheng, Jennifer La, Nhan V. Do, Paul A. Monach, Judith M. Strymish, Nathanael R. Fillmore, and Westyn Branch-Elliman. 2022. “Impact of Prior SARS-CoV-2 Infection on Incidence of Hospitalization and Adverse Events Following MRNA SARS-CoV-2 Vaccination: A Nationwide, Retrospective Cohort Study.” *Vaccine* 40 (8): 1082–89.

Li, Xintong, Berta Raventós, Elena Roel, Andrea Pistillo, Eugenia Martinez-Hernandez, Antonella Delmestri, Carlen Reyes, et al. 2022. “Association between Covid-19 Vaccination, SARS-CoV-2 Infection, and Risk of Immune Mediated Neurological Events: Population Based Cohort and Self-Controlled Case Series Analysis.” *BMJ (Clinical Research Ed.)* 376 (March): e068373.

Li, Xue, Le Gao, Xinning Tong, Vivien K. Y. Chan, Celine S. L. Chui, Francisco T. T. Lai, Carlos K. H. Wong, et al. 2022. “Autoimmune Conditions Following MRNA (BNT162b2) and Inactivated (CoronaVac) COVID-19 Vaccination: A Descriptive Cohort Study among 1.1 Million Vaccinated People in Hong Kong.” *Journal of Autoimmunity* 130 (102830): 102830.

Li, Xue, Xinning Tong, Winnie Wan Yin Yeung, Peng Kuan, Samson Hin Hei Yum, Celine Sze Ling Chui, Francisco Tsz Tsun Lai, et al. 2022. “Two-Dose COVID-19 Vaccination and Possible Arthritis Flare among Patients with Rheumatoid Arthritis in Hong Kong.” *Annals of the Rheumatic Diseases* 81 (4): 564–68.

Lipkind, Heather S., Gabriela Vazquez-Benitez, Malini DeSilva, Kimberly K. Vesco, Christina Ackerman-Banks, Jingyi Zhu, Thomas G. Boyce, et al. 2022. “Receipt of COVID-19 Vaccine during Pregnancy and Preterm or Small-for-Gestational-Age at Birth - Eight Integrated Health Care Organizations, United States, December 15, 2020-July 22, 2021.” *MMWR. Morbidity and Mortality Weekly Report* 71 (1): 26–30.

Lloyd, Patricia C., Mao Hu, Hui-Lee Wong, Azadeh Shoaibi, Cindy Ke Zhou, An-Chi Lo, Kandace Amend, et al. 2022. “Near Real-Time Surveillance of Safety Outcomes in US COVID-19 Vaccine Recipients Aged 12 to 64 Years.” *Vaccine* 40 (45): 6481–88.

Massari, Marco, Stefania Spila-Alegiani, Cristina Morciano, Matteo Spuri, Pasquale Marchione, Patrizia Felicetti, Valeria Belleudi, et al. 2022. “Post-Marketing Active Surveillance of Myocarditis and Pericarditis Following Vaccination with COVID-19 MRNA Vaccines in Persons Aged 12-39 Years in Italy: A Multi-Database, Self-Controlled Case Series Study.” *BioRxiv*. https://doi.org/10.1101/2022.02.07.22270020.

McMurry, Reid, Patrick Lenehan, Samir Awasthi, Eli Silvert, Arjun Puranik, Colin Pawlowski, A. J. Venkatakrishnan, et al. 2021. “Real-Time Analysis of a Mass Vaccination Effort Confirms the Safety of FDA-Authorized MRNA COVID-19 Vaccines.” *Med (New York, N.Y.)* 2 (8): 965-978.e5.

Mevorach, Dror, Emilia Anis, Noa Cedar, Michal Bromberg, Eric J. Haas, Eyal Nadir, Sharon Olsha-Castell, et al. 2021. “Myocarditis after BNT162b2 MRNA Vaccine against Covid-19 in Israel.” *The New England Journal of Medicine* 385 (23): 2140–49.

Niesen, Michiel J. M., Colin Pawlowski, John C. O’Horo, Doug W. Challener, Eli Silvert, Greg Donadio, Patrick J. Lenehan, et al. 2022. “Surveillance of Safety of 3 Doses of COVID-19 MRNA Vaccination Using Electronic Health Records.” *JAMA Network Open* 5 (4): e227038.

Ohaeri, Columbus, Daniel Rhys Thomas, Jane Salmon, Simon Cottrell, Jane Lyons, Ashley Akbari, Ronan A. Lyons, Fatemeh Torabi, Gareth Gi Davies, and Christopher Williams. 2022. “Comparative Risk of Cerebral Venous Sinus Thrombosis (CVST) Following COVID-19 Vaccination or Infection: A National Cohort Study Using Linked Electronic Health Records.” *Human Vaccines & Immunotherapeutics* 18 (6): 2127572.

Pakhchanian, Haig, Ahmad Saud, Rahul Raiker, Sinan Kardes, Rohit Aggarwal, and Latika Gupta. 2022. “COVID-19 Vaccination Outcomes among Patients with Dermatomyositis: A Multicentered Analysis.” *Clinical Rheumatology* 41 (7): 2257–60.

Pawlowski, Colin, John Rincón-Hekking, Samir Awasthi, Viral Pandey, Patrick Lenehan, A. J. Venkatakrishnan, Sairam Bade, et al. 2021. “Cerebral Venous Sinus Thrombosis (CVST) Is Not Significantly Linked to COVID-19 Vaccines or Non-COVID Vaccines in a Large Multi-State US Health System.” *BioRxiv*. https://doi.org/10.1101/2021.04.20.21255806.

Pottegård, Anton, Lars Christian Lund, Øystein Karlstad, Jesper Dahl, Morten Andersen, Jesper Hallas, Øjvind Lidegaard, et al. 2021. “Arterial Events, Venous Thromboembolism, Thrombocytopenia, and Bleeding after Vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: Population Based Cohort Study.” *BMJ (Clinical Research Ed.)* 373 (May): n1114.

Rottenstreich, M., H. Y. Sela, R. Rotem, E. Kadish, Y. Wiener-Well, and S. Grisaru-Granovsky. 2022. "Covid-19 Vaccination during the Third Trimester of Pregnancy: Rate of Vaccination and Maternal and Neonatal Outcomes, a Multicentre Retrospective Cohort Study." *BJOG: An International Journal of Obstetrics and Gynaecology* 129 (2): 248–55.

Shasha, David, Anthony D. Heymann, and Galia Zacay. 2022. "Re: Real-World Safety Data for the Pfizer BNT162b2 SARS-CoV-2 Vaccine, Historical Cohort Study: Authors' Response." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*. Elsevier BV.

Shibli, Rana, Ofra Barnett, Zomoroda Abu-Full, Naomi Gronich, Ronza Najjar-Debbiny, Ilana Doweck, Gad Rennert, and Walid Saliba. 2021. "Association between Vaccination with the BNT162b2 MRNA COVID-19 Vaccine and Bell's Palsy: A Population-Based Study." *The Lancet Regional Health. Europe* 11 (100236): 100236.

Simone, Anthony, John Herald, Aiyu Chen, Rohith Nayak, Yuh-Jer Albert Shen, and Ming-Sum Lee. 2022. "Acute Myocarditis Following a Third Dose of COVID-19 MRNA Vaccination in Adults." *International Journal of Cardiology* 365 (October): 41–43.

Sing, Chor-Wing, Casey Tze Lam Tang, Celine Sze Ling Chui, Min Fan, Francisco Tsz Tsun Lai, Xue Li, Eric Yuk Fai Wan, et al. 2022. "COVID-19 Vaccines and Risks of Hematological Abnormalities: Nested Case-Control and Self-Controlled Case Series Study." *American Journal of Hematology* 97 (4): 470–80.

Theiler, Regan N., Myra Wick, Ramila Mehta, Amy L. Weaver, Abinash Virk, and Melanie Swift. 2021. "Pregnancy and Birth Outcomes after SARS-CoV-2 Vaccination in Pregnancy." *American Journal of Obstetrics & Gynecology MFM* 3 (6): 100467.

Tu, Tian Ming, Shen Jia Yi, Jasmine Shimin Koh, Seyed Ehsan Saffari, Rebecca Hui Min Hoe, Geraldine Jiangyan Chen, Hui Jin Chiew, et al. 2022. "Incidence of Cerebral Venous Thrombosis Following SARS-CoV-2 Infection vs MRNA SARS-CoV-2 Vaccination in Singapore." *JAMA Network Open* 5 (3): e222940.

Wan, Eric Yuk Fai, Celine Sze Ling Chui, Francisco Tsz Tsun Lai, Esther Wai Yin Chan, Xue Li, Vincent Ka Chun Yan, Le Gao, et al. 2021. "Bell's Palsy Following Vaccination with MRNA (BNT162b2) and Inactivated (CoronaVac) SARS-CoV-2 Vaccines: A Case Series and Nested Case-Control Study." *The Lancet Infectious Diseases*, August. https://doi.org/10.1016/S1473-3099(21)00451-5.

Wan, Eric Yuk Fai, Celine Sze Ling Chui, Anna Hoi Ying Mok, Wanchun Xu, Vincent Ka Chun Yan, Francisco Tsz Tsun Lai, Xue Li, et al. 2022. "MRNA (BNT162b2) and Inactivated (CoronaVac) COVID-19 Vaccination and Risk of Adverse Events and Acute Diabetic Complications in Patients with Type 2 Diabetes Mellitus: A Population-Based Study." *Drug Safety: An International Journal of Medical Toxicology and Drug Experience* 45 (12): 1477–90.

Wan, Eric Yuk Fai, Celine Sze Ling Chui, Yuan Wang, Vanessa Wai Sei Ng, Vincent Ka Chun Yan, Francisco Tsz Tsun Lai, Xue Li, et al. 2022. "Herpes Zoster Related Hospitalization after Inactivated (CoronaVac) and MRNA (BNT162b2) SARS-CoV-2 Vaccination: A Self-Controlled Case Series and Nested Case-Control Study." *The Lancet Regional Health. Western Pacific* 21 (100393): 100393.

Wan, Eric Yuk Fai, Yuan Wang, Celine Sze Ling Chui, Anna Hoi Ying Mok, Wanchun Xu, Vincent Ka Chun Yan, Francisco Tsz Tsun Lai, et al. 2022. “Safety of an Inactivated, Whole-Virion COVID-19 Vaccine (CoronaVac) in People Aged 60 Years or Older in Hong Kong: A Modified Self-Controlled Case Series.” *The Lancet. Healthy Longevity* 3 (7): e491–500.

Whiteley, William N., Samantha Ip, Jennifer A. Cooper, Thomas Bolton, Spencer Keene, Venexia Walker, Rachel Denholm, et al. 2022. “Association of COVID-19 Vaccines ChAdOx1 and BNT162b2 with Major Venous, Arterial, or Thrombocytopenic Events: A Population-Based Cohort Study of 46 Million Adults in England.” *PLoS Medicine* 19 (2): e1003926.

Wong, Carlos King Ho, Kristy Tsz Kwan Lau, Xi Xiong, Ivan Chi Ho Au, Francisco Tsz Tsun Lai, Eric Yuk Fai Wan, Celine Sze Ling Chui, et al. 2022. “Adverse Events of Special Interest and Mortality Following Vaccination with MRNA (BNT162b2) and Inactivated (CoronaVac) SARS-CoV-2 Vaccines in Hong Kong: A Retrospective Study.” *PLoS Medicine* 19 (6): e1004018.

Wong, Carlos King Ho, David Tak Wai Lui, Xi Xiong, Celine Sze Ling Chui, Francisco Tsz Tsun Lai, Xue Li, Eric Yuk Fai Wan, et al. 2022. “Risk of Thyroid Dysfunction Associated with MRNA and Inactivated COVID-19 Vaccines: A Population-Based Study of 2.3 Million Vaccine Recipients.” *BMC Medicine* 20 (1): 339.

Wong, Carlos King Ho, Lung Yi Mak, Ivan Chi Ho Au, Francisco Tsz Tsun Lai, Xue Li, Eric Yuk Fai Wan, Celine Sze Ling Chui, et al. 2022. “Risk of Acute Liver Injury Following the MRNA (BNT162b2) and Inactivated (CoronaVac) COVID-19 Vaccines.” *Journal of Hepatology* 77 (5): 1339–48.

Wong, Hui-Lee, Mao Hu, Cindy Ke Zhou, Patricia C. Lloyd, Kandace L. Amend, Daniel C. Beachler, Alex Secora, et al. 2022. “Risk of Myocarditis and Pericarditis after the COVID-19 MRNA Vaccination in the USA: A Cohort Study in Claims Databases.” *Lancet* 399 (10342): 2191–99.

Xiong, Xi, Carlos King Ho Wong, Ivan Chi Ho Au, Francisco Tsz Tsun Lai, Xue Li, Eric Yuk Fai Wan, Celine Sze Ling Chui, et al. 2022. “Safety of Inactivated and MRNA COVID-19 Vaccination among Patients Treated for Hypothyroidism: A Population-Based Cohort Study.” *Thyroid: Official Journal of the American Thyroid Association* 32 (5): 505–14.

Xu, Stanley, Runxin Huang, Lina S. Sy, Sungching C. Glenn, Denison S. Ryan, Kerresa Morrissette, David K. Shay, et al. 2021. “COVID-19 Vaccination and Non-COVID-19 Mortality Risk - Seven Integrated Health Care Organizations, United States, December 14, 2020-July 31, 2021.” *MMWR. Morbidity and Mortality Weekly Report* 70 (43): 1520–24.

Yanir, Yoav, Ilana Doweck, Rana Shibli, Ronza Najjar-Debbiny, and Walid Saliba. 2022. “Association between the BNT162b2 Messenger RNA COVID-19 Vaccine and the Risk of Sudden Sensorineural Hearing Loss.” *JAMA Otolaryngology-- Head & Neck Surgery* 148 (4): 299–306.

Ye, Xuxiao, Tiantian Ma, Joseph E. Blais, Vincent K. C. Yan, Wei Kang, Celine S. L. Chui, Francisco T. T. Lai, et al. 2022. “Association between BNT162b2 or CoronaVac COVID-19 Vaccines and Major Adverse Cardiovascular Events among Individuals with Cardiovascular Disease.” *Cardiovascular Research* 118 (10): 2329–38.

Appendix to Chapter Three

This section contains additional tables related to Chapter Three: Methods.

SNOMED: Systematized Nomenclature of Medicine

Table 1. Medical codes for persons vaccinated against SARS-CoV-2

Concept ID	Name	Vocabulary
59267100003	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	NDC
59267100002	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	NDC
592671000	bnt162b2 .23mg/1.8mL INTRAMUSCULAR INJECTION, SUSPENSION	NDC
80777027310	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension	NDC
2470234	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension	RxNorm
2470233	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML	RxNorm
2470232	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273	RxNorm
2468235	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	RxNorm
2468234	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension	RxNorm
2468233	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Product	RxNorm
2468232	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML	RxNorm
2468231	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein	RxNorm
2468230	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2	RxNorm
80777027399	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension	NDC
807770273	cx-024414 .2mg/mL INTRAMUSCULAR INJECTION, SUSPENSION	NDC
39214411000001100	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) 1170 dose	dm+d
39326611000001100	Generic COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials 100 dose	dm+d

39326811000001100	Generic COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials	dm+d
39214511000001100	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) (Pfizer-BioNTech) 1170 dose 195 x 6 dose vials	dm+d
39327011000001100	COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials (Moderna, Inc) 100 dose 10 x 10 dose vials	dm+d
39326911000001100	COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials	dm+d
39115611000001100	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech)	dm+d
39115311000001100	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) 6 dose	dm+d
39115711000001100	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) (Pfizer-BioNTech) 6 dose	dm+d
39116111000001100	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech)	dm+d
59676058005	Janssen COVID-19 vaccine, DNA, spike protein, adenovirus type 26 (Ad26) vector, preservative free, 5x1010 viral particles/0.5mL dosage, for intramuscular use	NDC
310122210	AZD1222 Astrazeneca COVID-19 vaccine, DNA, spike protein, chimpanzee adenovirus Oxford 1 (ChAdOx1) vector, preservative free, 5x1010 viral particles/0.5mL dosage, for intramuscular use	NDC
310122215	azd1222 50000000000[VP]/.5mL INTRAMUSCULAR INJECTION, SUSPENSION	NDC
59267100001	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	NDC
91303	Janssen Covid-19 Vaccine	CPT4
91302	AstraZeneca Covid-19 Vaccine	CPT4
91301	Moderna Covid-19 Vaccine	CPT4
91300	Pfizer-Biontech Covid-19 Vaccine	CPT4
0022A	AstraZeneca Covid-19 Vaccine Administration - Second Dose	CPT4
0021A	AstraZeneca Covid-19 Vaccine Administration - First Dose	CPT4
0031A	Janssen Covid-19 Vaccine Administration	CPT4
0012A	Moderna Covid-19 Vaccine Administration - Second Dose	CPT4
0011A	Moderna Covid-19 Vaccine Administration - First Dose	CPT4
0002A	Pfizer-Biontech Covid-19 Vaccine Administration - Second Dose	CPT4

0001A	Pfizer-Biontech Covid-19 Vaccine Administration - First Dose	CPT4
208	SARS-COV-2 (COVID-19) vaccine, mRNA, spike protein, LNP, preservative free, 30 mcg/0.3mL dose	CVX
207	SARS-COV-2 (COVID-19) vaccine, mRNA, spike protein, LNP, preservative free, 100 mcg/0.5mL dose	CVX
210	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-ChAdOx1, preservative free, 0.5 mL	CVX
212	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-Ad26, preservative free, 0.5 mL	CVX

Table 2. Medical codes used for the definition of compatible COVID-19

Concept Id	Name	Vocabulary
45763724	Suspected coronavirus infection	SNOMED
44810277	Suspected coronavirus infection	SNOMED
44797466	[X]Coronavirus infection, unspecified	SNOMED
40479642	Pneumonia due to Severe acute respiratory syndrome coronavirus	SNOMED
40380828	Coronavirus as the cause of diseases classified to other chapters	SNOMED
37396171	Severe acute respiratory syndrome of upper respiratory tract	SNOMED
37311061	Disease caused by 2019-nCoV	SNOMED
37311060	Suspected disease caused by 2019-nCoV	SNOMED
37310287	Myocarditis caused by 2019 novel coronavirus	SNOMED
37310286	Infection of upper respiratory tract caused by 2019 novel coronavirus	SNOMED
37310284	Encephalopathy caused by 2019 novel coronavirus	SNOMED
37310283	Gastroenteritis caused by 2019 novel coronavirus	SNOMED
37310269	Disease caused by 2019 novel coronavirus	SNOMED
37310268	Suspected disease caused by 2019 novel coronavirus	SNOMED

37310254	Otitis media caused by 2019 novel coronavirus	SNOMED
37016927	Pneumonia caused by Human coronavirus	SNOMED
4100065	Disease due to Coronaviridae	SNOMED
4092694	Coronavirus as the cause of diseases classified to other chapters	SNOMED
703441	COVID-19 confirmed by laboratory test	SNOMED
703440	Probable COVID-19 confirmed using clinical diagnostic criteria	SNOMED
439676	Coronavirus infection	SNOMED
320651	Severe acute respiratory syndrome	SNOMED

Table 3. Codes used for the definition of study outcomes

Concept Id	Name	Vocabulary
Bell's palsy		
44784628	Facial palsy House-Brackmann grade VI	SNOMED
44782796	Facial palsy House-Brackmann grade V	SNOMED
44782795	Facial palsy House-Brackmann grade IV	SNOMED
44782794	Facial palsy House-Brackmann grade III	SNOMED
44782526	Facial palsy House-Brackmann grade II	SNOMED
42710044	Peripheral facial palsy	SNOMED
37396434	Sellars Beighton syndrome	SNOMED
37204553	Familial recurrent peripheral facial palsy	SNOMED
4091559	Facial palsy	SNOMED
4063208	O/E - cranial 7 -paralysis-UMN	SNOMED
4044386	Familial facial nerve palsy	SNOMED
4002030	Supranuclear facial nerve paralysis	SNOMED
3182576	Partial seventh nerve palsy	Nebraska Lexicon
760893	Bell's palsy of right side of face	SNOMED
760892	Bell's palsy of left side of face	SNOMED
374923	Bell's palsy	SNOMED
Transverse myelitis		
4047628	Acute viral transverse myelitis	SNOMED
4044077	Acute non-infective transverse myelitis	SNOMED
4042204	Varicella transverse myelitis	SNOMED
443904	Transverse myelopathy syndrome	SNOMED
139803	Acute transverse myelitis	SNOMED
134330	Idiopathic transverse myelitis	SNOMED
Guillain-Barré syndrome		
37396786	Acute inflammatory demyelinating polyradiculoneuropathy	SNOMED
37396785	Acute motor sensory axonal neuropathy	SNOMED
37396086	Acute motor axonal neuropathy	SNOMED
37204541	Acute pure sensory neuropathy	SNOMED
37204369	Pharyngeal-cervical-brachial variant of Guillain-Barré syndrome	SNOMED
35624130	Paraparetic variant of Guillain-Barré syndrome	SNOMED
35623634	Acute sensory ataxic neuropathy	SNOMED
4164770	Guillain-Barré syndrome	SNOMED
4070552	Fisher's syndrome	SNOMED
374925	Acute infective polyneuritis	SNOMED
Encephalomyelitis		
46285143	Bacterial meningoencephalomyelitis	SNOMED
43530701	Acute disseminated encephalomyelitis following infectious disease	SNOMED
40480147	Toxic encephalomyelitis	SNOMED
36715541	Infection causing encephalomyelitis	SNOMED
36676712	Chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids	SNOMED
35607947	Acute necrotizing encephalopathy of childhood	SNOMED
4330496	Inflammation of spinal cord due to toxin	SNOMED
4320933	Enteroviral encephalomyelitis	SNOMED
4310021	Paraneoplastic encephalomyelitis	SNOMED
4249574	Acute hemorrhagic leukoencephalitis	SNOMED
4242575	Granulomatous meningoencephalomyelitis	SNOMED
4233538	Van Bogaert's sclerosing leukoencephalitis	SNOMED
4228612	Encephalomyelitis associated with AIDS	SNOMED
4220014	Encephalomyelitis caused by lymphocytic choriomeningitis virus	SNOMED
4206037	Pyogranulomatous meningoencephalomyelitis	SNOMED
4190307	Inflammatory disease of the central nervous system	SNOMED
4176250	Allergic encephalomyelitis	SNOMED
4147498	Encephalitis, myelitis and encephalomyelitis	SNOMED
4120303	Progressive congenital rubella encephalomyelitis	SNOMED

4105201	Post mixed vaccination encephalitis	SNOMED
4105200	Post diphtheria vaccination encephalitis	SNOMED
4104690	Post hepatitis B vaccination encephalitis	SNOMED
4104546	Meningoencephalomyelitis	SNOMED
4103116	Post influenza vaccination encephalitis	SNOMED
4103115	Post mumps vaccination encephalitis	SNOMED
4103114	Post polio vaccination encephalitis	SNOMED
4103113	Post yellow fever vaccination encephalitis	SNOMED
4103112	Post rabies vaccination encephalitis	SNOMED
4103111	Post pertussis vaccination encephalitis	SNOMED
4103110	Post plague vaccination encephalitis	SNOMED
4103109	Post cholera vaccination encephalitis	SNOMED
4103108	Post paratyphoid vaccination encephalitis	SNOMED
4103107	Post typhoid vaccination encephalitis	SNOMED
4102199	Post measles vaccination encephalitis	SNOMED
4102198	Post BCG vaccination encephalitis	SNOMED
4100112	Post hepatitis A vaccination encephalitis	SNOMED
4100111	Post rubella vaccination encephalitis	SNOMED
4100110	Post typhus vaccination encephalitis	SNOMED
4100109	Post smallpox vaccination encephalitis	SNOMED
4100108	Post tetanus vaccination encephalitis	SNOMED
4098889	Experimental allergic encephalomyelitis	SNOMED
4042202	Post-influenza encephalitis	SNOMED
4035637	Simian B encephalomyelitis	SNOMED
3655306	Encephalomyelitis caused by Coxiella burnetii	SNOMED
3655305	Encephalomyelitis caused by bacterium	SNOMED
3655275	Encephalomyelitis caused by Neisseria meningitidis	SNOMED
3654623	Encephalomyelitis caused by Burkholderia	SNOMED
765827	Multiphasic acute disseminated encephalomyelitis	SNOMED
761989	Recurrent acute disseminated encephalomyelitis	SNOMED
443802	Postvaccinal encephalomyelitis	SNOMED
380322	Post Varicella encephalitis	SNOMED
379798	Post-infectious encephalomyelitis	SNOMED
379792	Post-immunization encephalitis	SNOMED
378143	Encephalitis	SNOMED
377487	Post measles encephalitis	SNOMED
375185	Rubella encephalomyelitis	SNOMED
374021	Acute disseminated encephalomyelitis	SNOMED
373189	Encephalomyelitis	SNOMED
372615	Post-infectious encephalitis	SNOMED
138965	Myelitis	SNOMED
Concept ID	**Concept name**	**Vocabulary**
Cerebral venous sinus thrombosis (CVST)		
4102202	Cerebral venous sinus thrombosis	SNOMED
4048786	Cerebral venous thrombosis of sigmoid sinus	SNOMED
4043735	Cerebral venous thrombosis of straight sinus	SNOMED
4111713	Non-pyogenic venous sinus thrombosis	SNOMED
314667	Nonpyogenic thrombosis of intracranial venous sinus	SNOMED
4116206	Septic thrombophlebitis of cavernous sinus	SNOMED
4121335	Septic thrombophlebitis of lateral sinus	SNOMED
4119136	Septic thrombophlebitis of sagittal sinus	SNOMED
4041680	Septic thrombophlebitis of sigmoid sinus	SNOMED
4100225	Thrombophlebitis lateral venous sinus	SNOMED
4217471	Thrombophlebitis of basilar sinus	SNOMED
4104695	Thrombophlebitis of cavernous sinus	SNOMED
4167985	Thrombophlebitis of inferior sagittal sinus	SNOMED
764714	Thrombophlebitis of sigmoid sinus	SNOMED
4100224	Thrombophlebitis of superior longitudinal venous sinus	SNOMED
4098706	Thrombophlebitis of superior sagittal sinus	SNOMED
4277833	Thrombophlebitis of torcular Herophili	SNOMED
764710	Thrombophlebitis of transverse sinus	SNOMED

4228209	Thrombosis of basilar sinus	SNOMED
4234264	Thrombosis of cavernous venous sinus	SNOMED
4048890	Thrombosis of inferior sagittal sinus	SNOMED
4057329	Thrombosis of lateral venous sinus	SNOMED
4102203	Thrombosis of superior longitudinal sinus	SNOMED
4290940	Thrombosis of superior sagittal sinus	SNOMED
4079905	Thrombosis of torcular Herophili	SNOMED
4105338	Thrombosis transverse sinus	SNOMED
Deep vein thrombosis		
Concept ID	**Concept name**	**Vocabulary**
762047	Acute bilateral thrombosis of subclavian veins	SNOMED
762148	Acute deep vein thrombosis of bilateral iliac veins	SNOMED
761444	Acute deep vein thrombosis of bilateral lower limbs following coronary artery bypass graft	SNOMED
35616028	Acute deep vein thrombosis of left iliac vein	SNOMED
35615035	Acute deep vein thrombosis of left lower limb following procedure	SNOMED
761416	Acute deep vein thrombosis of left upper limb following coronary artery bypass graft	SNOMED
35615031	Acute deep vein thrombosis of left upper limb following procedure	SNOMED
43531681	Acute deep vein thrombosis of lower limb	SNOMED
35616027	Acute deep vein thrombosis of right iliac vein	SNOMED
35615034	Acute deep vein thrombosis of right lower limb following procedure	SNOMED
761415	Acute deep vein thrombosis of right upper limb following coronary artery bypass graft	SNOMED
35615030	Acute deep vein thrombosis of right upper limb following procedure	SNOMED
44782746	Acute deep venous thrombosis	SNOMED
44782751	Acute deep venous thrombosis of axillary vein	SNOMED
762008	Acute deep venous thrombosis of bilateral axillary veins	SNOMED
760875	Acute deep venous thrombosis of bilateral calves	SNOMED
765155	Acute deep venous thrombosis of bilateral ileofemoral veins	SNOMED
762017	Acute deep venous thrombosis of bilateral internal jugular veins	SNOMED
762417	Acute deep venous thrombosis of bilateral legs	SNOMED
762020	Acute deep venous thrombosis of bilateral popliteal veins	SNOMED
765546	Acute deep venous thrombosis of bilateral tibial veins	SNOMED
762004	Acute deep venous thrombosis of both upper extremities	SNOMED
44782742	Acute deep venous thrombosis of calf	SNOMED
44782747	Acute deep venous thrombosis of femoral vein	SNOMED
762015	Acute deep venous thrombosis of ileofemoral vein of left leg	SNOMED
765541	Acute deep venous thrombosis of ileofemoral vein of right lower extremity	SNOMED
44782748	Acute deep venous thrombosis of iliofemoral vein	SNOMED
44782752	Acute deep venous thrombosis of internal jugular vein	SNOMED
762009	Acute deep venous thrombosis of left axillary vein	SNOMED
760876	Acute deep venous thrombosis of left calf	SNOMED
765540	Acute deep venous thrombosis of left femoral vein	SNOMED
765922	Acute deep venous thrombosis of left internal jugular vein	SNOMED
762418	Acute deep venous thrombosis of left lower extremity	SNOMED
765537	Acute deep venous thrombosis of left upper extremity	SNOMED
44782767	Acute deep venous thrombosis of lower extremity as complication of procedure	SNOMED
46270071	Acute deep venous thrombosis of lower limb due to coronary artery bypass grafting	SNOMED

762022	Acute deep venous thrombosis of politeal vein of right leg	SNOMED
44782743	Acute deep venous thrombosis of popliteal vein	SNOMED
762021	Acute deep venous thrombosis of popliteal vein of left leg	SNOMED
762010	Acute deep venous thrombosis of right axillary vein	SNOMED
760877	Acute deep venous thrombosis of right calf	SNOMED
762013	Acute deep venous thrombosis of right femoral vein	SNOMED
762018	Acute deep venous thrombosis of right internal jugular vein	SNOMED
762419	Acute deep venous thrombosis of right lower extremity	SNOMED
762005	Acute deep venous thrombosis of right upper extremity	SNOMED
44782745	Acute deep venous thrombosis of thigh	SNOMED
44782744	Acute deep venous thrombosis of tibial vein	SNOMED
762026	Acute deep venous thrombosis of tibial vein of left leg	SNOMED
765156	Acute deep venous thrombosis of tibial vein of right leg	SNOMED
44782421	Acute deep venous thrombosis of upper extremity	SNOMED
764016	Acute deep venous thrombosis of upper extremity after coronary artery bypass graft	SNOMED
44782766	Acute deep venous thrombosis of upper extremity as complication of procedure	SNOMED
762048	Acute thrombosis of left subclavian vein	SNOMED
45757410	Acute thrombosis of mesenteric vein	SNOMED
762049	Acute thrombosis of right subclavian vein	SNOMED
36712892	Acute thrombosis of splenic vein	SNOMED
44782762	Acute thrombosis of subclavian vein	SNOMED
37109253	Bilateral acute deep vein thrombosis of femoral veins	SNOMED
40478951	Bilateral deep vein thrombosis of lower extremities	SNOMED
4046884	Deep vein thrombosis of leg related to air travel	SNOMED
4133004	Deep venous thrombosis	SNOMED
4181315	Deep venous thrombosis associated with coronary artery bypass graft	SNOMED
45773536	Deep venous thrombosis of femoropopliteal vein	SNOMED
763942	Deep venous thrombosis of left lower extremity	SNOMED
761980	Deep venous thrombosis of left upper extremity	SNOMED
443537	Deep venous thrombosis of lower extremity	SNOMED
4133975	Deep venous thrombosis of pelvic vein	SNOMED
40480555	Deep venous thrombosis of peroneal vein	SNOMED
4322565	Deep venous thrombosis of profunda femoris vein	SNOMED
763941	Deep venous thrombosis of right lower extremity	SNOMED
761928	Deep venous thrombosis of right upper extremity	SNOMED
4207899	Deep venous thrombosis of tibial vein	SNOMED
4028057	Deep venous thrombosis of upper extremity	SNOMED
193512	Embolism and thrombosis of the renal vein	SNOMED
435565	Embolism and thrombosis of the vena cava	SNOMED
4119760	Iliofemoral deep vein thrombosis	SNOMED
4124856	Inferior mesenteric vein thrombosis	SNOMED
4281689	Phlegmasia alba dolens	SNOMED
4284538	Phlegmasia cerulea dolens	SNOMED
4309333	Postoperative deep vein thrombosis	SNOMED
46285905	Provoked deep vein thrombosis	SNOMED
4033521	Splenic vein thrombosis	SNOMED
4055089	Superior mesenteric vein thrombosis	SNOMED
42538533	Thrombosis of iliac vein	SNOMED
44811347	Thrombosis of internal jugular vein	SNOMED
765049	Thrombosis of left peroneal vein	SNOMED
4317289	Thrombosis of mesenteric vein	SNOMED
4203836	Thrombosis of subclavian vein	SNOMED
4175649	Thrombosis of the popliteal vein	SNOMED
4153353	Traumatic thrombosis of axillary vein	SNOMED
46285904	Unprovoked deep vein thrombosis	SNOMED
4221821	Thrombophlebitis of deep veins of lower extremity	SNOMED
46271900	Recurrent deep vein thrombosis	SNOMED

4189004	Deep vein thrombosis of leg related to intravenous drug use	SNOMED
Splanchnic Vein Thrombosis		
Concept ID	**Concept name**	**Vocabulary**
4033521	Splenic vein thrombosis	SNOMED
196715	Budd-Chiari syndrome	SNOMED
199837	Portal vein thrombosis	SNOMED
4317289	Thrombosis of mesenteric vein	SNOMED
4092406	Portal thrombophlebitis	SNOMED
36712892	Acute thrombosis of splenic vein	SNOMED
4173167	Mesenteric embolus	SNOMED
4144032	Mesenteric thrombus and/or embolus	SNOMED
45757410	Acute thrombosis of mesenteric vein	SNOMED
45757409	Chronic thrombosis of mesenteric vein	SNOMED
4318407	Thrombophlebitis of mesenteric vein	SNOMED
4124856	Inferior mesenteric vein thrombosis	SNOMED
4055089	Superior mesenteric vein thrombosis	SNOMED
Pulmonary embolism		
Concept ID	**Concept name**	**Vocabulary**
4120091	Acute massive pulmonary embolism	SNOMED
45768439	Acute pulmonary embolism	SNOMED
45768888	Acute pulmonary thromboembolism	SNOMED
4309039	Hemorrhagic pulmonary infarction	SNOMED
762808	Infarction of lung due to embolus	SNOMED
40480461	Infarction of lung due to iatrogenic pulmonary embolism	SNOMED
4108681	Postoperative pulmonary embolus	SNOMED
4091708	Pulmonary air embolism	SNOMED
440417	Pulmonary embolism	SNOMED
37109911	Pulmonary embolism due to and following acute myocardial infarction	SNOMED
37016922	Pulmonary embolism on long-term anticoagulation therapy	SNOMED
43530605	Pulmonary embolism with pulmonary infarction	SNOMED
4119608	Pulmonary fat embolism	SNOMED
254662	Pulmonary infarction	SNOMED
4253796	Pulmonary microemboli	SNOMED
45766471	Pulmonary oil microembolism	SNOMED
4121618	Pulmonary thromboembolism	SNOMED
4119610	Pulmonary tumor embolism	SNOMED
4119607	Subacute massive pulmonary embolism	SNOMED
4119609	Subacute pulmonary fat embolism	SNOMED
4236271	Recurrent pulmonary embolism	SNOMED
Thrombocytopenia		
Platelet measurement		
Concept ID	**Concept name**	**Vocabulary**
3007461	Platelets [#/volume] in Blood	LOINC
3031586	Platelets [#/volume] in Blood by Estimate	LOINC
3024929	Platelets [#/volume] in Blood by Automated count	LOINC
3039827	Platelets [#/volume] in Body fluid by Automated count	LOINC
3024386	Platelet mean volume [Entitic volume] in Blood by Rees-Ecker	LOINC
4267147	Platelet count	SNOMED
37393863	Platelet count	SNOMED

Thrombocytopenia diagnosis		
Concept ID	**Concept name**	**Vocabulary**
37397537	Beta thalassemia X-linked thrombocytopenia syndrome	SNOMED
432870	Thrombocytopenic disorder	SNOMED
46272950	Thrombocytopathy, asplenia and miosis	SNOMED
44782445	Thrombocytopenia due to alcohol	SNOMED
42536958	Pancytopenia caused by medication	SNOMED
40321716	Secondary thrombocytopenia	SNOMED
37312165	Atypical hemolytic uremic syndrome	SNOMED
37209558	Pancytopenia caused by immunosuppressant	SNOMED
37204551	Hereditary isolated aplastic anemia	SNOMED
37204548	Hereditary thrombocytopenia with normal platelets	SNOMED
37204520	Bleeding diathesis due to thromboxane synthesis deficiency	SNOMED
37204478	Pancytopenia due to IKZF1 mutations	SNOMED
37117164	Revesz syndrome	SNOMED
37116398	Thyrocerebrorenal syndrome	SNOMED
37110394	Isolated thrombocytopenia	SNOMED
37019055	Aplastic anemia co-occurrent with human immunodeficiency virus infection	SNOMED
37018663	Thrombocytopenia co-occurrent and due to alcoholism	SNOMED
37017607	Antibody mediated acquired pure red cell aplasia caused by erythropoiesis stimulating agent	SNOMED
37017165	GATA binding protein 1 related thrombocytopenia with dyserythropoiesis	SNOMED
37016797	MYH9 related disease	SNOMED
37016151	Aplastic anemia caused by antineoplastic agent	SNOMED
36717326	DK phocomelia syndrome	SNOMED
36716406	Severe fever with thrombocytopenia syndrome virus	SNOMED
36716047	Radioulnar synostosis with amegakaryocytic thrombocytopenia syndrome	SNOMED
36715586	Refractory thrombocytopenia	SNOMED
36715053	Autosomal dominant macrothrombocytopenia	SNOMED
36713970	WT limb blood syndrome	SNOMED
36713443	MYH9 macrothrombocytopenia syndrome	SNOMED
36713112	Pancytopenia due to antineoplastic chemotherapy	SNOMED
36674972	Macrothrombocytopenia with mitral valve insufficiency	SNOMED
36674474	Pancytopenia with developmental delay syndrome	SNOMED
35625536	Ataxia pancytopenia syndrome	SNOMED
35623407	Adult pure red cell aplasia	SNOMED
4345236	Parvoviral aplastic crisis	SNOMED
4338386	Thrombocytopenia due to non-immune destruction	SNOMED
4316372	HELLP syndrome	SNOMED
4314802	Kasabach-Merritt syndrome	SNOMED
4311682	Radial aplasia-thrombocytopenia syndrome	SNOMED
4305588	Doan-Wright syndrome	SNOMED
4301602	Thrombotic thrombocytopenic purpura	SNOMED
4301128	Thrombocytopenia due to diminished platelet production	SNOMED
4300464	Wiskott-Aldrich autosomal dominant variant syndrome	SNOMED
4299560	Thrombocytopenic purpura due to defective platelet production	SNOMED
4298690	Immunologic aplastic anemia	SNOMED
4292531	Thrombocytopenic purpura due to platelet consumption	SNOMED
4292425	Sex-linked thrombocytopenia	SNOMED
4272928	Thrombocytopenia due to hypersplenism	SNOMED
4264464	Mediterranean macrothrombocytopenia	SNOMED

4258261	Drug induced thrombotic thrombocytopenic purpura	SNOMED
4247776	Posttransfusion purpura	SNOMED
4239484	Acquired pancytopenia	SNOMED
4235220	Hereditary thrombocytopenic disorder	SNOMED
4234973	Chronic acquired pure red cell aplasia	SNOMED
4233407	Megakaryocytic aplasia	SNOMED
4230266	Autoimmune thrombotic thrombocytopenic purpura	SNOMED
4226905	Thrombocytopenia associated with AIDS	SNOMED
4225810	Aplastic anemia associated with AIDS	SNOMED
4219476	Thrombocytopenia due to defective platelet production	SNOMED
4218171	Uremic thrombocytopenia	SNOMED
4214947	Thrombocytopenic purpura associated with metabolic disorder	SNOMED
4211348	Aplastic anemia associated with pancreatitis	SNOMED
4204900	Acquired thrombotic thrombocytopenic purpura	SNOMED
4197574	Dilutional thrombocytopenia	SNOMED
4188208	Estren-Dameshek anemia	SNOMED
4186108	Aplastic anemia associated with metabolic alteration	SNOMED
4185078	Bernard Soulier syndrome	SNOMED
4184758	Acquired aplastic anemia	SNOMED
4184200	Secondary aplastic anemia	SNOMED
4177177	Cellular immunologic aplastic anemia	SNOMED
4173278	Thrombocytopenia due to blood loss	SNOMED
4172008	Cyclic thrombocytopenia	SNOMED
4166754	Perinatal thrombocytopenia	SNOMED
4159966	Upshaw-Schulman syndrome	SNOMED
4159749	Idiopathic maternal thrombocytopenia	SNOMED
4159736	Radiation thrombocytopenia	SNOMED
4156233	Thrombocytopenia due to sequestration	SNOMED
4148471	Fanconi’s anemia	SNOMED

4147049	Thrombocytopenia due to extracorporeal circulation	SNOMED
4146088	Aplastic anemia due to drugs	SNOMED
4146086	Constitutional aplastic anemia with malformation	SNOMED
4145458	Thrombocytopenia due to hypothermia	SNOMED
4140545	Post infectious thrombocytopenic purpura	SNOMED
4139555	Thrombocytopenia due to massive blood transfusion	SNOMED
4137430	Idiopathic thrombocytopenic purpura	SNOMED
4133984	Alloimmune thrombocytopenia	SNOMED
4133983	Secondary autoimmune thrombocytopenia	SNOMED
4133981	Benign gestational thrombocytopenia	SNOMED
4125496	Pure red cell aplasia, acquired	SNOMED
4125494	Pancytopenia with pancreatitis	SNOMED
4123076	Montreal platelet syndrome	SNOMED
4123075	May-Hegglin anomaly	SNOMED
4123074	Megakaryocytic thrombocytopenia	SNOMED
4121265	Mediterranean thrombocytopenia	SNOMED
4121264	Epstein syndrome	SNOMED
4120620	Amegakaryocytic thrombocytopenia	SNOMED
4119134	Thrombocytopenic purpura	SNOMED
4103532	Immune thrombocytopenia	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	SNOMED
4101603	Thrombocytopenia due to extracorporeal circulation of blood	SNOMED
4101583	Aplastic anemia due to infection	SNOMED
4101582	Aplastic anemia due to chronic disease	SNOMED
4100998	Aplastic anemia due to toxic cause	SNOMED
4098148	Thrombocytopenia due to drugs	SNOMED
4098145	Idiopathic aplastic anemia	SNOMED
4098028	Transient acquired pure red cell aplasia	SNOMED
4098027	Aplastic anemia due to radiation	SNOMED
4082738	Autoimmune pancytopenia	SNOMED
4077348	Pancytopenia-dysmelia	SNOMED

4031699	Humoral immunologic aplastic anemia	SNOMED
4028065	Autoimmune thrombocytopenia	SNOMED
4027374	Alloimmune platelet transfusion refractoriness	SNOMED
4009307	Heparin-induced thrombocytopenia with thrombosis	SNOMED
4000065	Drug-induced immune thrombocytopenia	SNOMED
441264	Primary thrombocytopenia	SNOMED
440982	Wiskott-Aldrich syndrome	SNOMED
440372	Acquired thrombocytopenia	SNOMED
436956	Evans syndrome	SNOMED
433749	Heparin-induced thrombocytopenia	SNOMED
432881	Pancytopenia	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	SNOMED
140681	Constitutional aplastic anemia	SNOMED
138723	Acquired red cell aplasia	SNOMED
137829	Aplastic anemia	SNOMED
Thrombocytopenic purpura		
Concept ID	**Concept name**	**Vocabulary**
4119134	Thrombocytopenic purpura	SNOMED
4301602	Thrombotic thrombocytopenic purpura	SNOMED
4299560	Thrombocytopenic purpura due to defective platelet production	SNOMED
4292531	Thrombocytopenic purpura due to platelet consumption	SNOMED
4258261	Drug induced thrombotic thrombocytopenic purpura	SNOMED
4247776	Posttransfusion purpura	SNOMED
4230266	Autoimmune thrombotic thrombocytopenic purpura	SNOMED
4214947	Thrombocytopenic purpura associated with metabolic disorder	SNOMED
4204900	Acquired thrombotic thrombocytopenic purpura	SNOMED
4159966	Upshaw-Schulman syndrome	SNOMED
4140545	Post infectious thrombocytopenic purpura	SNOMED
4137430	Idiopathic thrombocytopenic purpura	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	SNOMED
313800	Thrombotic microangiopathy	SNOMED
Immune thrombocytopenia		
Concept ID	**Concept name**	**Vocabulary**
4103532	Immune thrombocytopenia	SNOMED
4137430	Idiopathic thrombocytopenic purpura	SNOMED
4133984	Alloimmune thrombocytopenia	SNOMED
4133983	Secondary autoimmune thrombocytopenia	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	SNOMED
4028065	Autoimmune thrombocytopenia	SNOMED
4027374	Alloimmune platelet transfusion refractoriness	SNOMED
4009307	Heparin-induced thrombocytopenia with thrombosis	SNOMED
4000065	Drug-induced immune thrombocytopenia	SNOMED
436956	Evans syndrome	SNOMED
433749	Heparin-induced thrombocytopenia	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	SNOMED
Ischemic stroke		
Concept ID	**Concept name**	**Vocabulary**
4045735	Anterior cerebral circulation infarction	SNOMED
4031045	Anterior choroidal artery syndrome	SNOMED
761110	Bilateral cerebral infarction due to precererbral arterial occlusion	SNOMED
4110189	Cerebral infarct due to thrombosis of precerebral arteries	SNOMED

443454	Cerebral infarction	SNOMED
762951	Cerebral infarction due to anterior cerebral artery occlusion	SNOMED
765515	Cerebral infarction due to basilar artery stenosis	SNOMED
43530683	Cerebral infarction due to carotid artery occlusion	SNOMED
762933	Cerebral infarction due to cerebral artery occlusion	SNOMED
762937	Cerebral infarction due to cerebral venous thrombosis	SNOMED
4111714	Cerebral infarction due to cerebral venous thrombosis, non-pyogenic	SNOMED
4108356	Cerebral infarction due to embolism of cerebral arteries	SNOMED
45772786	Cerebral infarction due to embolism of middle cerebral artery	SNOMED
4110190	Cerebral infarction due to embolism of precerebral arteries	SNOMED
762935	Cerebral infarction due to internal carotid artery occlusion	SNOMED
763015	Cerebral infarction due to middle cerebral artery occlusion	SNOMED
46273649	Cerebral infarction due to occlusion of basilar artery	SNOMED
35610084	Cerebral infarction due to occlusion of cerebral artery	SNOMED
46270031	Cerebral infarction due to occlusion of precerebral artery	SNOMED
762934	Cerebral infarction due to posterior cerebral artery occlusion	SNOMED
43531607	Cerebral infarction due to stenosis of carotid artery	SNOMED
35610085	Cerebral infarction due to stenosis of cerebral artery	SNOMED
46270381	Cerebral infarction due to stenosis of precerebral artery	SNOMED
4110192	Cerebral infarction due to thrombosis of cerebral arteries	SNOMED
45767658	Cerebral infarction due to thrombosis of middle cerebral artery	SNOMED
44782773	Cerebral infarction due to vertebral artery occlusion	SNOMED
46270380	Cerebral infarction due to vertebral artery stenosis	SNOMED
37110678	Cerebral ischemic stroke due to occlusion of extracranial large artery	SNOMED
37110679	Cerebral ischemic stroke due to stenosis of extracranial large artery	SNOMED
4043731	Infarction - precerebral	SNOMED
4131383	Infarction of basal ganglia	SNOMED
4046237	Infarction of optic radiation	SNOMED
4119140	Infarction of visual cortex	SNOMED
4141405	Left sided cerebral infarction	SNOMED
37116473	Multifocal cerebral infarction due to and following procedure on cardiovascular system	SNOMED
4077086	Occipital cerebral infarction	SNOMED
4046359	Partial anterior cerebral circulation infarction	SNOMED
4319146	Pituitary infarction	SNOMED
4146185	Right sided cerebral infarction	SNOMED
36717605	Silent cerebral infarct	SNOMED
4142739	Thalamic infarction	SNOMED
4046358	Total anterior cerebral circulation infarction	SNOMED
372924	Cerebral artery occlusion	SNOMED
Myocardial infarction		
Concept ID	Concept name	Vocabulary
4119457	Acute Q wave infarction - anterolateral	SNOMED
4119943	Acute Q wave infarction - anteroseptal	SNOMED
4121464	Acute Q wave infarction - inferior	SNOMED
4121465	Acute Q wave infarction - inferolateral	SNOMED
4124684	Acute Q wave infarction - lateral	SNOMED

4119948	Acute Q wave infarction - widespread	SNOMED
4126801	Acute Q wave myocardial infarction	SNOMED
4296653	Acute ST segment elevation myocardial infarction	SNOMED
46270162	Acute ST segment elevation myocardial infarction due to left coronary artery occlusion	SNOMED
761737	Acute ST segment elevation myocardial infarction due to occlusion of circumflex coronary artery	SNOMED
46270163	Acute ST segment elevation myocardial infarction due to right coronary artery occlusion	SNOMED
43020460	Acute ST segment elevation myocardial infarction involving left anterior descending coronary artery	SNOMED
45766076	Acute ST segment elevation myocardial infarction of anterior wall involving right ventricle	SNOMED
761736	Acute ST segment elevation myocardial infarction of anteroapical wall	SNOMED
46270159	Acute ST segment elevation myocardial infarction of anterolateral wall	SNOMED
46270160	Acute ST segment elevation myocardial infarction of anteroseptal wall	SNOMED
45766116	Acute ST segment elevation myocardial infarction of inferior wall	SNOMED
45766151	Acute ST segment elevation myocardial infarction of inferior wall involving right ventricle	SNOMED
35611570	Acute ST segment elevation myocardial infarction of inferolateral wall	SNOMED
35611571	Acute ST segment elevation myocardial infarction of inferoposterior wall	SNOMED
46274044	Acute ST segment elevation myocardial infarction of lateral wall	SNOMED
46270161	Acute ST segment elevation myocardial infarction of posterior wall	SNOMED
46273495	Acute ST segment elevation myocardial infarction of posterobasal wall	SNOMED
46270158	Acute ST segment elevation myocardial infarction of posterolateral wall	SNOMED
46270164	Acute ST segment elevation myocardial infarction of septum	SNOMED
45766075	Acute anterior ST segment elevation myocardial infarction	SNOMED
4178129	Acute anteroapical myocardial infarction	SNOMED
4267568	Acute anteroseptal myocardial infarction	SNOMED
312327	Acute myocardial infarction	SNOMED
44782769	Acute myocardial infarction due to left coronary artery occlusion	SNOMED
44782712	Acute myocardial infarction due to right coronary artery occlusion	SNOMED
45766115	Acute myocardial infarction during procedure	SNOMED
434376	Acute myocardial infarction of anterior wall	SNOMED
45766150	Acute myocardial infarction of anterior wall involving right ventricle	SNOMED
438438	Acute myocardial infarction of anterolateral wall	SNOMED
4243372	Acute myocardial infarction of apical-lateral wall	SNOMED
4108669	Acute myocardial infarction of atrium	SNOMED
4151046	Acute myocardial infarction of basal-lateral wall	SNOMED
4275436	Acute myocardial infarction of high lateral wall	SNOMED
438170	Acute myocardial infarction of inferior wall	SNOMED
45771322	Acute myocardial infarction of inferior wall involving right ventricle	SNOMED
438447	Acute myocardial infarction of inferolateral wall	SNOMED
441579	Acute myocardial infarction of inferoposterior wall	SNOMED
436706	Acute myocardial infarction of lateral wall	SNOMED
4324413	Acute myocardial infarction of posterobasal wall	SNOMED
4051874	Acute myocardial infarction of posterolateral wall	SNOMED
4303359	Acute myocardial infarction of septum	SNOMED
4147223	Acute myocardial infarction with rupture of ventricle	SNOMED
4145721	Acute non-Q wave infarction	SNOMED
4119944	Acute non-Q wave infarction - anterolateral	SNOMED
4119456	Acute non-Q wave infarction - anteroseptal	SNOMED

4119945	Acute non-Q wave infarction - inferior	SNOMED
4119946	Acute non-Q wave infarction - inferolateral	SNOMED
4121466	Acute non-Q wave infarction - lateral	SNOMED
4124685	Acute non-Q wave infarction - widespread	SNOMED
4270024	Acute non-ST segment elevation myocardial infarction	SNOMED
35610091	Acute nontransmural myocardial infarction	SNOMED
319039	Acute posterior myocardial infarction	SNOMED
444406	Acute subendocardial infarction	SNOMED
35610093	Acute transmural myocardial infarction	SNOMED
4119947	Acute widespread myocardial infarction	SNOMED
37109912	Arrhythmia due to and following acute myocardial infarction	SNOMED
438172	Atrial septal defect due to and following acute myocardial infarction	SNOMED
4124687	Cardiac rupture due to and following acute myocardial infarction	SNOMED
4215259	First myocardial infarction	SNOMED
4108678	Hemopericardium due to and following acute myocardial infarction	SNOMED
4173632	Microinfarct of heart	SNOMED
45771327	Mitral valve regurgitation due to acute myocardial infarction with papillary muscle and chordal rupture	SNOMED
45766214	Mitral valve regurgitation due to acute myocardial infarction without papillary muscle and chordal rupture	SNOMED
45766212	Mitral valve regurgitation due to and following acute myocardial infarction	SNOMED
4323202	Mixed myocardial ischemia and infarction	SNOMED
4329847	Myocardial infarction	SNOMED
37309626	Myocardial infarction due to demand ischemia	SNOMED
4170094	Myocardial infarction in recovery phase	SNOMED
4200113	Non-Q wave myocardial infarction	SNOMED
4030582	Postoperative myocardial infarction	SNOMED
35610087	Postoperative nontransmural myocardial infarction	SNOMED
4206867	Postoperative subendocardial myocardial infarction	SNOMED
35610089	Postoperative transmural myocardial infarction	SNOMED
4207921	Postoperative transmural myocardial infarction of anterior wall	SNOMED
4209541	Postoperative transmural myocardial infarction of inferior wall	SNOMED
37109911	Pulmonary embolism due to and following acute myocardial infarction	SNOMED
4108679	Rupture of cardiac wall without hemopericardium as current complication following acute myocardial infarction	SNOMED
4108219	Rupture of chordae tendinae due to and following acute myocardial infarction	SNOMED
4124686	Silent myocardial infarction	SNOMED
765132	Subendocardial myocardial infarction	SNOMED
45766114	Subsequent ST segment elevation myocardial infarction	SNOMED
45766113	Subsequent ST segment elevation myocardial infarction of anterior wall	SNOMED
45773170	Subsequent ST segment elevation myocardial infarction of inferior wall	SNOMED
4108217	Subsequent myocardial infarction	SNOMED
4108677	Subsequent myocardial infarction of anterior wall	SNOMED
4108218	Subsequent myocardial infarction of inferior wall	SNOMED
45766241	Subsequent non-ST segment elevation myocardial infarction	SNOMED
4108680	Thrombosis of atrium, auricular appendage, and ventricle due to and following acute myocardial infarction	SNOMED
439693	True posterior myocardial infarction	SNOMED
37109910	Ventricular aneurysm due to and following acute myocardial infarction	SNOMED

Other arterial thromboembolism		
Concept ID	Concept name	Vocabulary
4195665	Gastrointestinal tract vascular insufficiency	SNOMED
4148299	Ischemic colitis	SNOMED
4173167	Mesenteric embolus	SNOMED
4317289	Thrombosis of mesenteric vein	SNOMED
4319280	Acute bowel infarction	SNOMED
4144032	Mesenteric thrombus and/or embolus	SNOMED
45757410	Acute thrombosis of mesenteric vein	SNOMED
45757409	Chronic thrombosis of mesenteric vein	SNOMED
44811741	Acute ischaemia of large intestine	SNOMED
44811740	Acute ischaemia of small intestine	SNOMED
37117790	Insufficiency of mesenteric artery	SNOMED
37016198	Epiploic appendagitis	SNOMED
35622081	Nongangrenous ischemic colitis	SNOMED
35622080	Gangrenous ischemic colitis	SNOMED
4345926	Abdominal angina	SNOMED
4342767	Transient ischemic colitis	SNOMED
4341648	Hemorrhagic infarction of intestine	SNOMED
4341646	Occlusive mesenteric ischemia	SNOMED
4340939	Non-occlusive mesenteric ischemia	SNOMED
4340378	Transmural infarction of intestine	SNOMED
4340375	Focal segmental ischemia of small intestine	SNOMED
4318537	Large bowel gangrene	SNOMED
4318407	Thrombophlebitis of mesenteric vein	SNOMED
4240850	Acute ischemic enterocolitis	SNOMED
4239942	Embolic mesenteric infarction	SNOMED
4237654	Ischemic enterocolitis	SNOMED
4215949	Nonocclusive intestinal infarction	SNOMED
4214720	Thrombotic mesenteric infarction	SNOMED
4192856	Acute ischemic colitis	SNOMED
4188336	Chronic ischemic enterocolitis	SNOMED
4174014	Inferior mesenteric artery embolus	SNOMED
4149013	Mesenteric infarction	SNOMED
4148257	Chronic gastrointestinal tract vascular insufficiency	SNOMED
4148256	Acute GIT vascular insufficiency	SNOMED
4124856	Inferior mesenteric vein thrombosis	SNOMED
4055089	Superior mesenteric vein thrombosis	SNOMED
4055025	Superior mesenteric artery embolus	SNOMED
4045408	Ischemic stricture of intestine	SNOMED
201894	Acute vascular insufficiency of intestine	SNOMED
192673	Vascular insufficiency of intestine	SNOMED

Table 4. Codes used for the definition of comorbidities and medication use.

Id	**Code**	**Name**	**Vocabulary**
434621	85828009	Autoimmune disease	SNOMED
443392	363346000	Malignant neoplastic disease	SNOMED
201820	73211009	Diabetes mellitus	SNOMED
433736	414916001	Obesity	SNOMED
321588	56265001	Heart disease	SNOMED
316866	38341003	Hypertensive disorder	SNOMED

Table 5. Negative control outcome concept list.

Concept 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
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
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
135215	Hashimoto thyroiditis
442190	Hemorrhage of colon
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
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

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
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
197494	Viral hepatitis C
4284533	Vitamin D-dependent rickets

Appendix to Chapter Six

This section contains additional tables related to Chapter Six: COVID-19 vaccination and immune-mediated neurological events.

Table 1. Power calculation for the self-controlled case series (SCCS)

	CPRD AURUM				SIDIAP			
	Proportion time exposed	**Total time**	**Events**	**MDRR**	**Proportion time exposed**	**Total time**	**Events**	**MDRR**
ChAdOx1 first-dose								
Bell's palsy	0.015	44230222	7403	1.29	0.0135	9086299	1925	1.62
Encephalomyelitis	0.0143	43035252	415	2.51	0.0145	8068323	111	4.31
Guillain Barre syndrome	0.0167	40030856	477	2.27	0.0136	9005332	101	4.67
Transverse myelitis	0.0142	59366705	304	2.85	0.0135	9045039	9	24.69
BNT162b2 first-dose								
Bell's palsy	0.012	30195714	4564	1.42	0.0134	31723854	7386	1.30
Encephalomyelitis	0.0121	23621329	260	3.25	0.0145	27732788	489	2.31
Guillain Barre syndrome	0.0125	28983007	313	2.96	0.0134	31797412	305	2.82
Transverse myelitis	0.0114	33884636	176	4.04	0.0128	33399967	50	7.22
mRNA-1723 first-dose								

Bell's palsy					0.0131	4684571	978	1.92
Encephalomyelitis					0.014	4275406	151	3.76
Guillain Barre syndrome					0.0132	4664538	71	5.75
Transverse myelitis					0.0133	4642457	51	6.98
Ad26.COV2.S first-dose								
Bell's palsy					0.0107	1922746	454	2.63
Encephalomyelitis					0.0116	1727694	22	13.34
Guillain Barre syndrome					0.0107	1910933	18	16.33
Transverse myelitis					0.0107	1917374	<5	
SARS-CoV-2 infection (21 days at risk)								
Bell's palsy	0.0046	10037690	2335	2.05	0.0135	4329031	1090	1.85
Encephalomyelitis	0.0049	9498944	127	7.90	0.0141	3954950	101	4.578
Guillain Barre syndrome	0.0046	9289326	147	7.43	0.0135	4328106	46	7.36
Transverse myelitis	0.0045	12357608	85	10.89	0.0134	4358694	6	35.99
SARS-CoV-2 infection (90 days at risk)								
Bell's palsy	0.0046	10037690	2335	2.05	0.0135	4329031	1090	1.85
Encephalomyelitis	0.0049	9498944	127	7.90	0.0141	3954950	101	4.58
Guillain Barre syndrome	0.0046	9289326	147	7.43	0.0135	4328106	46	7.36

Transverse myelitis	0.0045	12357608	85	10.89	0.0134	4358694	6	35.99

CPRD: Clinical Practice Research Datalink; SIDIAP: Information System for Research in Primary Care; MDRR: minimum detectable relative risks.

Table 2. Power calculation for the self-controlled case series (SCCS) sensitivity analyses

	CPRD AURUM				**SIDIAP**			
	Proportion time exposed	**Total time**	**Events**	**MDRR**	**Proportion time exposed**	**Total time**	**Events**	**MDRR**
ChAdOx1 first-dose								
Bell's palsy	0.0148	41838475	6935	1.30	0.0135	8383419	1761	1.65
Encephalomyelitis	0.0142	40517414	379	2.61	0.0145	7443524	106	4.41
Guillain Barre syndrome	0.0164	37867929	447	2.33	0.0136	8308627	89	4.50
Transverse myelitis	0.014	56213708	290	2.92	0.0135	8345559	9	24.68
BNT162b2 first-dose								
Bell's palsy	0.0118	29035280	4375	1.43	0.0134	28290593	6542	1.32
Encephalomyelitis	0.0119	22495388	244	3.37	0.0145	24707838	405	2.46
Guillain Barre syndrome	0.0123	27919905	289	3.08	0.0134	28348798	266	2.98

Transverse myelitis	0.0112	32515243	168	4.17	0.0128	29783708	45	7.71
mRNA-1723 first-dose								
Bell's palsy					0.0132	4162063	866	1.99
Encephalomyelitis					0.014	3794490	125	4.12
Guillain Barre syndrome					0.0132	4144053	62	6.22
Transverse myelitis					0.0133	4127669	41	8.01
Ad26.COV2.S first-dose								
Bell's palsy					0.0107	1696564	396	2.77
Encephalomyelitis					0.0116	1523541	19	14.86
Guillain Barre syndrome					0.0108	1686084	13	21.10
Transverse myelitis					0.0107	1692484	<5	

*CPRD: Clinical Practice Research Datalink; SIDIAP: Information System for Research in Primary Care; MDRR: minimum detectable relative risks.

Table 3. Cases of immune-mediated neurological events with hospitalisation in SIDIAP between 1 January 2017 to 30 June 2021

	Number of cases	**Cases with hospitalisation (Day 0) (row %)**	**Cases with hospitalisation (Day 0 to Day 3) (row %)**
Bell's palsy	19,938	6,687 (33.5 %)	7,166 (35.9 %)
Encephalomyelitis	1,603	1,295 (80.8 %)	1,384 (86.3%)
Guillain Barre syndrome	1,018	704 (69.2 %)	801 (78.7%)
Transverse myelitis	193	106 (54.9 %)	114 (59.1%)

SIDIAP: Information System for Research in Primary Care

Table 4. Cases of death following immune-mediated neurological events between 1 January 2017 to 30 June 2021

	Number of cases	**Cases of death (Day 0 to Day 7) (row %)**
CPRD AURUM		
Bell's palsy	107,652	45 (0.4%)
Encephalomyelitis	9,253	32 (0.3%)
Guillain Barre syndrome	7,474	20 (0.3%)
Transverse myelitis	4,465	<5
SIDIAP		

Bell's palsy	19.938	53 (0.3%)
Encephalomyelitis	1,603	6 (0.4%)
Guillain Barre syndrome	1,018	<5
Transverse myelitis	193	0

CPRD: Clinical Practice Research Datalink; SIDIAP: Information System for Research in Primary Care. These counts were calculated by taking all persons who experienced any of the studied immune-mediated neurological disorders from 1 January 2017 to end of data availability in the two databases.

Figure 1. Time between end of study or censor date and occurrence of Bell's palsy from 1 January 2017 to end of data availability in CPRD AURUM

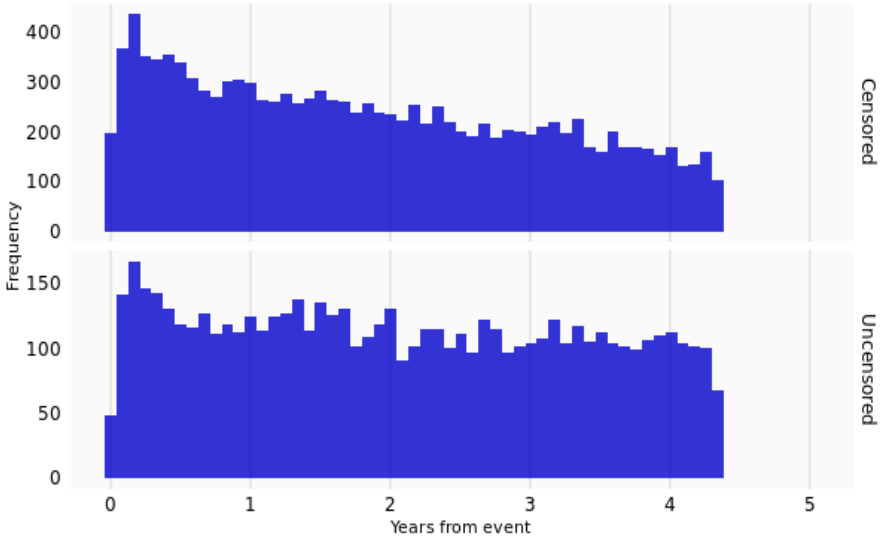
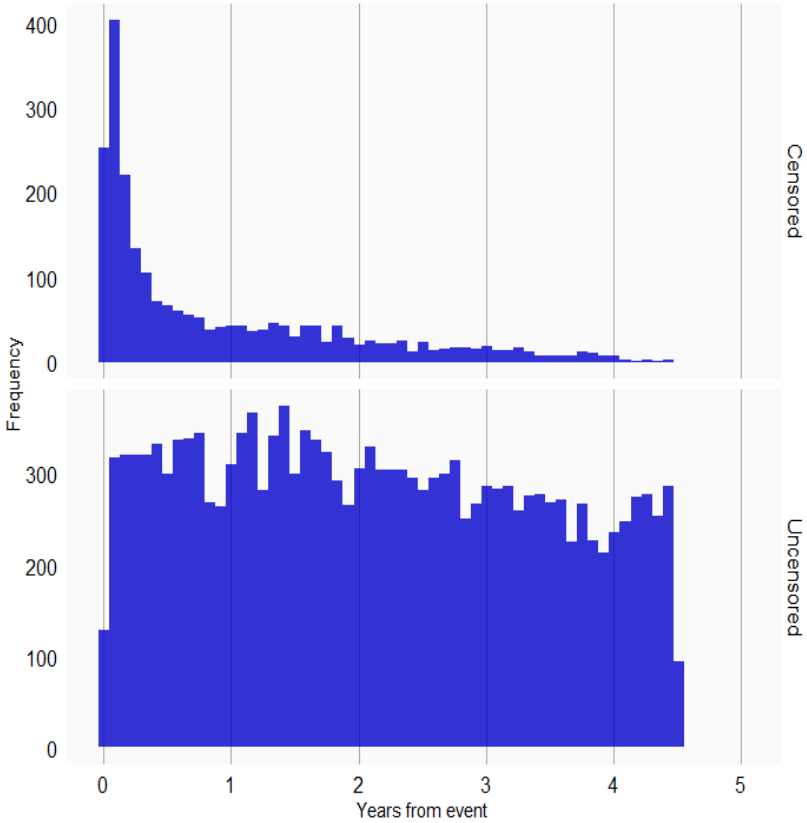


Figure 2. Time between end of study or censor date and occurrence of Bell's palsy from 1 January 2017 to end of data availability in SIDIAP



Appendix to Chapter Seven

Supplementary Table 1. Cohort characteristics: European data (by database and cohorts)

	CPRD_Aurum				Germany_DA_					SIDIAP			France_LPD				IPCI			
Covariate	**ChAdOx 1.1st.dose**	**ChAdOx 1.2nd.dose**	**Comirnaty.1st.dose**	**BNT 162b 2.2nd.dose**	**ChAdOx 1.1st.dose**	**ChAdOx 1.2nd.dose**	**BNT 162 b2.1 st.dose**	**BNT 162b 2.2nd.dose**	**Janssen. COVID.19**	**BNT 162 b2.1 st.dose**	**mRNA-1273 .1st. dose**	**Janssen. COVID.19**	**ChAdOx 1.1st.dose**	**ChAdOx 1.2nd.dose**	**BNT 162 b2.1 st.dose**	**BNT 162b 2.2nd.dose**	**ChAdOx 1.1st.dose**	**ChAdOx 1.2nd.dose**	**BNT 162 b2.1 st.dose**	**BNT 162b 2.2nd.dose**
Cohort Count	3,789,631	1,195,626	1,840,240	1,369,238	98,562	31,200	391,063	321,099	37,723	2,027,950	248,784	126,189	27,698	15,067	12,931	4,506	71,082	38,884	218,423	90,089
Covid-19 infection prior vaccination	5.5	4.5	4	4.3	1.2	0.6	2.1	0.6	1.7	6.6	7.6	8.2	4.7	4.2	6.1	3.8	6.9	4.8	5.4	3.1
gender: FEMALE	51.5	55.6	58.2	59.9	41.2	41.8	55	55.1	38.2	54.1	51.8	46.1	47.6	47.3	56.4	56.3	50.3	49.1	51.2	51.6
Index month																				
12										6.5	1.4									
1	13.7		47.6									0.1								
2	29	0.1	36.3	1.5						5.9	5	0.1	3.6		9		3		0.2	
3	45.7	10.6	3.6	32	1.9		0.1			8.5	8.2	0.1	50.3	0.1	20.5	12.5	18.5	0.4	7.6	0.2
4	9	73.1	1.6	53.8	22.2		30.7			18.8	18.2	9.4	26.5	0.9	22.3	29.9	42.4	0.5	24	3.1
5	2.5	16.3	0.4	7.3	60.4		26.4		20.2	22.5	48.6	16.6	9.7	34.2	17.8	15.9	29.4	17.5	30.7	42.1
6					13.2	46.7	24.5	41.7	63.1	37.7	18.7	73.7	7	47.8	10.6	16.9	6.8	81.6	37.5	54.5
7					2	42	12.7	24.7	11				2.6	15.9	6.1	8.5				
8					0.3	8.3	5.5	13.5	5.7				0.1	0.9	3.5	3.4				
9														0.1		0.7				
Age group																				
10 - 19	0.4	0.3	0.6	0.4	0.5	0.4	2			0.2	0.6	0.3					0.2	0.1	0.8	0.4
20 - 29	4.4	2.9	5.4	4.7	4.3	3.1	7.4	6.6	10.3	1.9	5.1	3.0	0.3	0.1	3.7	2.6	1.7	1.1	5.2	2.3
30 - 39	7.3	4.5	7.8	7.2	5.9	4.4	9.9	9	13.1	8.3	7.4	4.8	0.6	0.2	4.5	3.4	1.8	1.2	11.3	2.8
40 - 49	18.3	6.6	10.9	9.3	8.7	6.4	12.8	12.2	16.9	25.5	16.7	40.8	1.6	0.5	8.7	6.5	3	2.5	9.1	4.8

50 - 59	28.8	12	16	13.3	19.1	15.9	22.4	22.5	28.1	25	52.1	13.7	23.2	19.4	14.1	11.1	5.7	5.2	21.3	18.2
60 - 69	21.3	23	17.8	15.4	37.2	39	22.5	23.3	21	2.6	7.2	18.8	44.1	46.9	18.7	17	83.5	86.2	15	22.2
70 - 79	14.7	38.5	20.4	24.6	18	22.3	16.1	17.4	6.1	19.9	9.4	18.4	24.5	27.2	28.4	33.2	1.7	1.9	26.5	45.1
80 - 89	3.6	9.5	17.7	20.9	5.4	7.2	5.9	6.1				0.2	4.5	4.6	17.9	22.3	1.4	1.1	9.9	3.9
90 - 99	1.1	2.6	3.3	4.1	1.1	1.3	1	0.9	0.5	3.7	0.1	0.1	1.1	1	3.1	3.3	0.9	0.6	1	0.3
100 - 109	0.1	0.1	0.1	0.1																
Medical history																				
Charlson index - Romano adaptation	79.8	141.1	131.1	140	158.8	187.1	157.2	165.4	109.1	106	116	68.6	91.9	96.1	95.1	107.6	83.8	89	98.8	118
CHADS2VASc	128.3	216.2	200.6	219.9	196.9	225.6	194	201.6	132.1	185.2	116	123.4	191.4	201	232.1	259.1	136.6	145.8	184	218.4
Acute respiratory disease	41.4		44.1		57.9	58	59.6	59.3	62.8	56.3	55.7	53.8	44.2	45.3	45.5	45.9	44.3	44.7	44.3	44.3
Atrial fibrillation	2.9	6.6	6.6	7.4	3.3	4.2	3.3	3.5	1.7	5.4	2.2		0.8	0.9	1.5	2.1				
Attention deficit hyperactivity disorder	0.3	0.2	0.2	0.2	0.6		0.4		0.6	0.2	0.4	0.3					1	0.8	1.5	1
Cerebrovascular disease	2.8		5.5		7.7	8.9	7.3	7.8	3.9	3.6	1.7	1.8	4	4.2	5.2	6.8	4.7	5	5.7	6.9
Chronic liver disease		0.7		0.5	1	1.1	1	1.1	0.9	1.4	2.5	1.8	0.5	0.5	0.4	0.4	0.2		0.2	
Chronic obstructive lung disease	3.1	6.8	5.3	5.8	11.5	13	11.3	11.9	9.3	5.4	4	4.1	4.9		4.4		6	6.5	5.1	6.6
Coronary arteriosclerosis	0.8	1.6	1.4	1.5	5.7	7	5.4	5.7	3.2	2.4	1.7		2.2	2.4	2.3	2	0.8	0.9	0.8	1
Crohn's disease	0.5	0.8	0.6	0.6	0.6	0.6	0.8	0.9	0.6	0.2	1.5	0.2	0.2		0.4		0.4	0.4	0.5	0.6
Dementia	1.2	2.5	2	2.2	2.4	3.1	2.2	2.2	1.4	2.7	0.3	0.3	0.3		0.4		0.5	0.4	1	1
Depressive disorder	21.3		20.3		22	24.1	24.2	24.8	22	15.2	13.9	13.7	17.9	18.4	19	19.9	11.4	11.4	9.3	9.9
Diabetes mellitus	10	16.4	15.5	14.9	21.4	25.5	20.1	21.2	15.8	14	10.3	11.7	19.9	21.1	15.3	16.6	12.5	13.4	12.6	15.9

Gastroesophageal reflux disease	5.1	6.3	5.9	6.1	5.5	5.5	5.8	6	4.5	7.9	7	7.0	16.3	17.6	18.2	20.7	3.4	3.7	3.5	4.1
Gastrointestinal hemorrhage	8.6	10.3	9.7	9.9	3.6	3.6	3.2	3.4	3	5.3	5.2		3.3	3.4	3.8	4.5	6.1	6.5	5.8	6.7
Heart disease	10.5	20.2	19.5	20.8	31		29.4		21.3	20	12.3		14.7	16	18.1	21.6	17.1	18.4	19.6	24.1
Heart failure	1.3	2.8	2.6	3	7.3	9.5	6.8	7.3	3.9	4.2	2	1.5	1.2	1.4	2.1	2.6	1.8	1.8	2.4	2.6
Hematologic neoplasm		0.8		0.7	0.9	1.1	1.2	1.3	0.8	1	4	0.6	0.3	0.3	0.5	0.6	0.2	0.2	0.3	0.3
Human immunodeficiency virus infection					0.2	0.2	0.3	0.3	0.2	0.3	1.3	0.6	0.6	0.3	0.4	0.2	0.2	0.2	0.1	0.2
Hyperlipidemia	9.2	14.7	13	14.1	34.5	38.4	29.6	31	23.1	25.9	20.5	21.6	20.8	22.4	22.1	27.5	14.8		13.2	
Hypertensive disorder	23.8	39.8	35.8	38.6	49.8	56.6	43.8	45.8	36.2	34.6	24.4	25.5	44.1	46.6	44.1	49	32.7	34.6	31.3	39
Ischemic heart disease	4.1	8.7	8.3	9	13.8	16.7	12.7	13.4		5	3		4.9	5.2	4.8	6.1	8.1	8.7	9	11.2
Lesion of liver	0.9	1.5	1.2	1.2	1.3	1.5	1.4	1.4	1.1	1	2.6	1.1	0.6	0.6	0.5	0.5		0.1		0.1
Malignant lymphoma	0.3	0.7	0.6	0.6	0.7	0.7	0.9	0.9	0.4	0.4	1.9	0.2	0.3	0.3	0.6	0.8	0.2	0.2	0.3	0.3
Malignant neoplasm of anorectum	0.2	0.5	0.4	0.5	0.4	0.6	0.4	0.5		0.4	0.5	0.2	1.5	1.6	0.9	0.8				
Malignant neoplastic disease	8.2	15.9	14.7	16.4	12.1	14.2	12.8	13.6	7.1	12	15.4	7.1	7.8	8.6	9.2	11.1	14.8	15.6	17.3	21.2
Malignant tumor of breast		3.1		2.9	1.7	2	2.3	2.5	0.9	1.9	3.1	1.2		1.9		2.6	2.9	2.9	2.8	3.5
Malignant tumor of colon	0.3	0.8	0.7	0.8	0.8	1	0.8	0.9	0.5	1.1	1.2	0.6	1.5	1.7	0.9	0.8				
Malignant tumor of lung	0.2	0.4	0.3	0.4	0.2	0.3	0.3	0.4	0.1	0.3	1.2	0.2								
Malignant tumor of urinary bladder		0.6		0.7		0.6	0.4	0.4	0.2	0.9	0.8	0.5	0.2	0.3	0.4	0.6	0.5	0.5	0.7	0.8
Obesity	4.7		5.3		15.5	17.3	15.3	15.8	14.2	25.5	22.6	24	1.3	1.2	0.9	0.8	8	9.9	4.9	5.7

Osteoarthritis	18.9	31.7	28	30.6	33.1	37.7	29	30.2	26.1	24.4	14.5	16.8	22.3	24.1	22.7	28.8	18.4	19.8	19.7	24.3
Peripheral vascular disease	0.7	1.5	1.3	1.5	6.7		5.8		4.1	2.5	1.7	1.7	1.6	1.7	1.3	2	1.3	1.3	1.4	1.6
Pneumonia	2.8	4.4	3.6	3.8	6.6	7.5	6.9	7.1	5.5	8.1	8.5	6.6	4.5	4.5	4.8	5.5	11.1	11.8	11.4	13.1
Primary malignant neoplasm of prostate					1.6	1.8	1.4	1.5		1.6	0.9	0.8	1.7	2.1	1.7	2.2				
Psoriasis	4.4		4.7		4.2	4.8	3.7	3.9	3.4	3	4.7	2.8	4.4		4.2		4.7	4.9	4.3	5
Pulmonary embolism	1	1.7	1.5	1.6	0.9	1.1	1.5	1.6	0.6			0.4	0.7	0.7	1.1	1.4				
Renal impairment	6.3	13.6	13.3	15.1	8.2	10.3	7.3	7.7	4.6	9.5	5.9	3.8	1.4	1.6	2.5	3.1	3.7	4	6.8	7.5
Rheumatoid arthritis	1.1	2	1.6	1.7	3.7	4.4	4	4.2	2.6	0.8	2.1	0.5	0.7	0.7	1	1.2	2.7	2.8	2.9	3.5
Schizophrenia	0.5	0.5	0.4	0.3		0.4	0.4	0.4	0.5	0.5	0.7	1.2	0.1	0.2	0.1	-0.1	0.3		0.2	
Ulcerative colitis	0.7	1	0.8	0.9	0.8	0.9	1	1	0.7			0.4	0.3		0.3			1.2		1.7
Urinary tract infectious disease	13.5	17.1	16.9	17.7	15	15.8	17.3	17.7	12.8	24	21.1		7.1	7.4	8.9	9.7	24.1	24.2	26.2	27.8
Viral hepatitis C	0.2	0.2	0.2	0.1	0.4	0.5	0.4	0.4	0.6	0.9	1.6	1.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.1
Visual system disorder	39.7	50.5	48.7	50.7	26.4	29.5	25.4	26.3	21.7				21	22.5	23.7	27.1	48.5	49.6	51.9	55.3
Medication																				
AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	9.1	15.2	14	14.9	35.9	41.3	30.6	32.2	23	25.6	17.6	18.9	35.6	38.7	33.5	37.9	20.9	22.7	21.5	27.6
ANTIBACTERIALS FOR SYSTEMIC USE	8.6	11.3	10.6	10.6	6.4	6.6	7.9	7.4	6.3	14.5	15	13.6		14.9		15.2	12.9	12.8	14.5	15.6
ANTIDEPRESSANTS	9.7	10.5	10.3	9.9	6.1		6.5		5.1	15.9	14	13.5	10.3	10.7	11.4	12.7	9.1	9.2	7.9	8.8
ANTIEPILEPTICS	2.5	3.3	2.9	2.8	2.4	2.9	2.7	2.8	1.8	7.1	7.3	7.2	4	4.2	4.1	4.7	2.9	3	2.6	3.1

ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	6.5	10.1	9.5	9.9	20.7	23.2	19	19.2	17.6	24.2	24.3	23.9	34.5	37	34.7	37.9		16.3		18.2
ANTINEOPLASTIC AGENTS		0.7		0.6	0.9	1	1	1	0.5	1	3.1	0.7	0.9	1	0.9	1.2	1.7	1.8	2.4	3
ANTIPSORIATICS	0.3		0.3		0.3	0.3	0.2	0.3	0.2	1	1.8	0.9	0.5	0.5	0.6	0.6	0.5	0.5	0.5	0.6
ANTITHROMBOTIC AGENTS	4.4	9.4	8.9	9.8	14.1	17.3	14.2	14.9	8.1	16.2	10.7	9.6	22.8	25.4	25.9	31	13.7	14.8	17.5	21.4
BETA BLOCKING AGENTS	4.5	7.9	7.7	8.3	23.1	27.2	20.8	22	14	12.2	8.2		19.7	21.4	21.6	24.2	14.3	15.5	17.2	21.3
CALCIUM CHANNEL BLOCKERS	6.5	11.3	9.9	10.8	16	18.9	13.7	14.5	10.1	9.2	6.4	6.0	16.9	18.2	17.9	20.3	11	12.1	11.4	14.5
DIURETICS	3.4	6.7	6.2	6.9	19.8	23.6	17.2	18.2	11.5	18.8	11.1	12.0	19.5	21.2	20.4	22.7	13.3	14.4	13.7	17.4
DRUGS FOR ACID RELATED DISORDERS	11.1	16.8	15.3	16.2	17.4	19.4	18.1	18.4	13.5	27.7	25.5		28	29.8	30.1	32.2	25	26.1	27.3	32.4
DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	9.5	13.1	11.9	11.9	9.3		10.2		7.3			12.6	17.7	19.2	17.9	19.6	20.6	21.6	20.7	23.9
DRUGS USED IN DIABETES	4.3	7.3	6.8	6.4	11.4		10.4		7.5	10.7	7.8	9.0	17.2	18.5	12.9	14	8.6	9.4	8.5	11.1
IMMUNOSUPPRESSANTS	0.4	0.8	0.6	0.6	0.7	0.9	1	1	0.4	1	6.5	0.6	0.4	0.4	0.7	0.9	1.4	1.5	1.9	2.5
LIPID MODIFYING AGENTS	9.4	17.8	15.7	16.9	19.5	23.2	17.1	18.2	10.9	19.6	14.6	15.3	31.5	35.1	29.1	33.4	22.7	24.7	23.6	31
OPIOIDS	5.7		7.3		4.9	5.7	4.8	4.7	4	8.9	8.7	7.5	13.5	14	12.9	14.5	7	6.8	6.4	7.2
PSYCHOLEPTICS	2.7	3.5	3.1	3.1	4.3	4.9	4.6	4.7	3.3	23.2	20.2	19.8	17.3	18	18.4	19	9.9	9.9	9.8	11
PSYCHOSTIMULANTS, AGENTS USED FOR ADHD AND NOOTROPICS					0.1	0.1	0.1	0.1	0.1	1.4	1.1	1.1	3	3.2	3.3	3.7		0.4		0.5

Supplementary Table 2. Cohort characteristics: US data

	US_Hospital_CDM					US_Open_Claims				
Covariate	**BNT162b2.1st.dose**	**BNT162b2.2nd.dose**	**mRNA-1273.1st.dose**	**mRNA-1273.2nd.dose**	**Janssen.COVID.19**	**BNT162b2.1st.dose**	**BNT162b2.2nd.dose**	**mRNA-1273.1st.dose**	**mRNA-1273.2nd.dose**	**Janssen.COVID.19**
Cohort Count	281,743	230,036	70,957	58,688	6,190	6,055,337	4,450,735	4,260,380	2,938,023	939,748
Covid-19 infection prior vaccination	1.8	1.8	2.3	2.3	2	3.3	3.4	2.8	2.8	2.9
gender: FEMALE	59.1	58.8	58.2	57.9	54.3	54.5	54.4	54	54.1	47
Index month										
12	0.4		0.6			1		0.7		
1	17.2	1.5	17.3	0.3		6	2.4	8.9	1.1	
2	23.1	19.4	23.7	18.6		6.9	6.7	15.4	10.3	
3	33.7	32.3	37	24.3	31.1	20	10.2	29.9	20.4	23
4	20.6	31.6	13.5	37.4	62.6	21.6	26.8	23.9	36.2	33.5
5	3.4	12.6	6.4	12.8	3.1	11.8	19.7	11.7	22.8	21.3
6	1.2	2.1	1.1	6.3	2.1	7.8	12	2.9	6	10.5
7		0.5	0.3	0.4	1.1	9	7.7	3.8	1.7	5.9
8						14.1	12.6	2.5	1.3	5.3
9						1.9	1.9	0.2	0.2	0.5
Age group										
10 - 19	1.6	1.6	0.7	0.6	1.1	4.7	4.7	1.6	1.4	2.2
20 - 29	6.5	6.2	6.2	5.7	8.1	16	15.5	11.5	10.4	13.4
30 - 39	9.2	8.9	9.4	9.1	13	17	16.8	13.1	12.6	15.2
40 - 49	11.5	11.4	11.3	10.6	14.7	16.9	17.2	14.2	14.1	17.3
50 - 59	15.9	16.2	16.2	16	25.4	17.7	18.5	17.9	18.5	22.1
60 - 69	23.2	23.5	21.1	21.3	27.5	15.1	15.6	20.6	21.8	18.9
70 - 79	23.2	23	23.3	24.1	7.9	8.2	7.7	14	14.3	7.6
80 - 89	9	9.2	11.8	12.6	2.2	4.5	4	6.9	6.9	3.3
Medical hisory										

Charlson index - Romano adaptation	126.3	127.6	136.4	138.3	97.5	113.3	111.6	145.6	151.9	108.2
CHADS2VASc	196.5	197.5	202.4	206.6	138	138.5	136.5	173.1	177.1	129
Acute respiratory disease	13.5	13.6	16.2	16.4	14.2	45.3	45.7	46.4	47.5	45.2
Atrial fibrillation	5.6	5.6	6.1	6.3	2.6	3.3	3.1	4.7	4.9	2.9
Attention deficit hyperactivity disorder	1	1	1.1	1.1	1.1	4.2	4.2	3.5	3.4	4.4
Cerebrovascular disease	4.2		4.1		2.5	5.2	5.1	7.4	7.7	4.9
Chronic liver disease	2.3	2.4	2.7	2.8	2.9	2.3	2.3	2.9	3	
Chronic obstructive lung disease	5.4	5.5	5.6	5.7	5	5	5	7.1	7.4	5.4
Coronary arteriosclerosis	10.1	10.2	9.6	9.9	6	6.1	5.9	9.1	9.6	6.1
Crohn's disease	0.6	0.6	0.6	0.6	0.5	0.6	0.5		0.6	0.5
Dementia	1.1	1.1	1.1	1.2	0.8	1.9		1.9		1.2
Depressive disorder	11.1	11.1	12.6	12.7	10.8	16.8	16.9	18	18.3	17.1
Diabetes mellitus	16.9	17.2	17.7	17.9	13.8	13	13.2	17	17.8	12.9
Gastroesophageal reflux disease	18.3	18.4	17.5	17.8	14.9	15.8	15.9	19.2	20.1	
Gastrointestinal hemorrhage	3.5		3.6		3.6	5.9		7.2		5.9
Heart disease	22.5	22.8	22.2	22.7	14.3	18.1	17.9	23.9	24.9	17.9
Heart failure	4.6	4.7	4.7	4.7	3.5	3.8	3.7	5	5.2	3.3
Hematologic neoplasm	2	2	1.8	1.8	0.9		1.1		1.6	
Human immunodeficiency virus infection	0.8	0.8	1.7	1.6	1	0.4	0.4	0.4	0.4	0.4
Hyperlipidemia	36	36.6	34.1	34.9	24.7	29.8	30.3	38.9	40.8	31.4
Hypertensive disorder	40.8	41.4	40.4	41.1	34	29.3	29.4	38.1	39.6	30.7
Ischemic heart disease	6	6.1	6.6	6.8	4	5	4.9	7.1	7.5	
Lesion of liver	1.5	1.5	1.8	1.8	1.6	1	0.9	1.2	1.2	1
Malignant lymphoma	0.9	0.9	0.9	0.9	0.4	0.5	0.5	0.7	0.7	0.4
Malignant neoplasm of anorectum		0.3	0.3	0.3	0.2	0.2	0.2	0.3	0.3	0.2
Malignant neoplastic disease	11.6	11.7	11.2	11.6	6.5	8.7	8.3	12.2	12.8	8.5
Malignant tumor of breast	2.9	2.8	2.5	2.6	1.6	1.5	1.4	2	2.1	1.3
Malignant tumor of colon	0.6	0.6	0.5	0.5	0.3	0.4	0.4	0.6	0.6	0.4
Obesity	15.1	15.3	14.5	14.9	13.2	13.9	14.2	15.4	15.9	13.7
Osteoarthritis	21.3	21.6	23.3	24.3	15.2	23.4	23.3	30.8	32.2	25.6
Peripheral vascular disease	3.6	3.7	3.4	3.6	2.9	4.4	4.4	6	6.3	4
Pneumonia	4	4	4.3	4.3	3.7	7.3	7.2	8	8.2	6.9

Primary malignant neoplasm of prostate	2	2	1.7	1.8		1	1	1.6	1.7	1
Psoriasis	1	1.1	1.1	1.2	0.6	1.8	1.8	2.2	2.3	1.9
Pulmonary embolism	1	1.1	1	1		0.9		1.1		0.8
Renal impairment	8.7	8.8	9.6	9.7	6.5	6	5.8	8.1	8.4	5.4
Rheumatoid arthritis	1.7	1.8	1.8	1.9	1.1	1.7	1.6	2.2	2.3	1.6
Schizophrenia	0.3	0.3	0.5	0.4	0.8	0.7	0.7	0.7	0.6	0.6
Ulcerative colitis	0.5		0.5		0.3	0.6	0.6	0.7	0.8	0.6
Urinary tract infectious disease	9.2	9.4	8.6	8.7	8.4	16.2	16	17.8	18.2	14.3
Venous thrombosis	2		2.1		1.5	2.2	2.2	2.8	2.9	2.1
Viral hepatitis C	1	1	1.3	1.3	2.3	0.8	0.8	0.9	0.9	0.9
Visual system disorder	15	15.4	18.3	19.1	9.9	36.5	36.7	41.3	42.7	35.9
Medication										
AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	1.1	1.1	1.3	1.3		15.5	16	21.5	23	17.5
ANTIBACTERIALS FOR SYSTEMIC USE	4.2	3.9	5	4.8	5.3	21.5		23.1		20.9
ANTIDEPRESSANTS	1	0.9	1.3	1.2	1.5	16.2	16.6	18.7	19.6	17.3
ANTIEPILEPTICS	1.2	1.1	1.6	1.5	1.7	7.9	8.1		10.3	8.7
ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	3.4	3.3	3.9	3.9	5.1	12.8	12.9	14.2	14.6	12.9
ANTINEOPLASTIC AGENTS	1	1	1	1	0.6	2.5	2.4	3	3.2	2.3
ANTIPSORIATICS			0.2		-0.1	0.3	0.3	0.4	0.4	0.3
ANTITHROMBOTIC AGENTS	3.6	3.4	4.4	4.3	3.9	5.1	5.1	7.1	7.5	5
BETA BLOCKING AGENTS	1.6	1.5	2	1.9	2	9.8	10	13.8	14.8	10.4
CALCIUM CHANNEL BLOCKERS	1.2	1.1	1.4	1.4	1.4	7.9	8	10.8	11.5	8.3
DIURETICS	1.1	1	1.3	1.2	1.4	10.6	10.8	14.3	15.2	11.4
DRUGS FOR ACID RELATED DISORDERS	2.5	2.4	3	2.8	3	10.5	10.7	13.4	14.3	10.8
DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	3.7	3.6	4.6	4.7	4.1	13.8	14.1	16.1	17	14
DRUGS USED IN DIABETES	0.3	0.3	0.3	0.3	0.2	8.2	8.6	11	11.8	8.7
IMMUNOSUPPRESSANTS		0.3	0.3	0.3	0.1	1.7	1.6	1.9	2.1	1.4
LIPID MODIFYING AGENTS	1.6	1.5	1.8	1.8	1.8	16.4	16.9	23.5	25.4	17.6
OPIOIDS	5	4.7	5.9	5.8	5.5	7.1	7	8.3	8.3	7.7

PSYCHOLEPTICS	4.7	4.5	5.3	5.2	5.7	10.6	10.7	12.3	12.8	11.3
PSYCHOSTIMULANTS, AGENTS USED FOR ADHD AND NOOTROPICS	0.1	0.1	0.1	0.1	0.1	3.2	3.3	2.9	3	3.4

Supplementary Table 3. Baseline characteristics of eligible cohorts of vaccinated people identified in IPCI

Characteristic	Before PS matching			After PS matching		
	ChAdOx1	BNT162b2		ChAdOx1	BNT162b2	
	%	%	SMD	%	%	SMD
Age group						
10 - 19	0.2	0.8	-0.08	0.5	0.4	0.01
20 - 29	1.7	5.2	-0.19	4.1	3.6	0.03
30 - 39	1.8	11.3	-0.39	4.6	3.7	0.04
40 - 49	3.0	9.1	-0.26	6.6	5.7	0.04
50 - 59	5.7	21.3	-0.47	13.0	10.7	0.07
60 - 69	83.5	15.0	1.89	59.6	61.5	-0.04
70 - 79	1.7	26.5	-0.76	5.2	7.5	-0.09
80 - 89	1.4	9.9	-0.38	3.9	4.1	-0.01
90 - 99	0.9	1.0	-0.02	2.4	2.7	-0.02
Gender: female	50.3	51.2	-0.02	54.6	55.2	-0.01
Index month						
2	3.0	0.2	0.23	1.1	1.1	0.00
3	18.5	7.6	0.33	7.4	7.5	0.00
4	42.4	24.0	0.40	22.5	24.5	-0.05
5	29.4	30.7	-0.03	52.3	52.3	0.00
6	6.8	37.5	-0.80	16.6	14.5	0.06
Medical history: General						
Attention deficit hyperactivity disorder	1.0	1.5	-0.05	1.4	1.3	0.01
Chronic liver disease	0.2	0.2	0.01	0.2	0.2	0.00
Chronic obstructive lung disease	6.0	5.1	0.04	5.6	6.1	-0.02
Crohn's disease	0.4	0.5	-0.01	0.4	0.4	0.00
Dementia	0.5	1.0	-0.06	1.0	1.2	-0.02
Depressive disorder	11.4	9.3	0.07	11.1	11.2	0.00
Diabetes mellitus	12.5	12.6	0.00	12.4	13.0	-0.02
Gastroesophageal reflux disease	3.4	3.5	-0.01	3.3	3.5	-0.01
Gastrointestinal hemorrhage	6.1	5.8	0.01	5.9	6.1	-0.01
Human immunodeficiency virus infection	0.2	0.1	0.01	0.2	0.2	0.00
Hyperlipidemia	14.8	13.2	0.05	13.1	13.9	-0.02

Hypertensive disorder	32.7	31.3	0.03	30.6	32.6	-0.04
Obesity	8.0	4.9	0.13	7.7	7.7	0.00
Osteoarthritis	18.4	19.7	-0.03	18.3	19.8	-0.04
Pneumonia	11.1	11.4	-0.01	11.3	11.8	-0.02
Psoriasis	4.7	4.3	0.02	4.4	4.6	-0.01
Renal impairment	3.7	6.8	-0.14	5.0	5.6	-0.03
Rheumatoid arthritis	2.7	2.9	-0.01	2.8	3.0	-0.01
Schizophrenia	0.3	0.2	0.02	0.3	0.4	0.00
Urinary tract infectious disease	24.1	26.2	-0.05	26.6	27.2	-0.01
Viral hepatitis C	0.1	0.1	0.02	0.1	0.1	0.01
Visual system disorder	48.5	51.9	-0.07	49.1	50.3	-0.02
Medical history: Cardiovascular disease						
Cerebrovascular disease	4.7	5.7	-0.05	3.3	3.7	-0.03
Coronary arteriosclerosis	0.8	0.8	0.00	0.8	0.8	0.00
Heart disease	17.1	19.6	-0.06	17.5	18.9	-0.04
Heart failure	1.8	2.4	-0.04	2.5	2.8	-0.02
Ischemic heart disease	8.1	9.0	-0.04	8.2	8.8	-0.02
Peripheral vascular disease	1.3	1.4	-0.01	1.4	1.6	-0.01
Venous thrombosis	3.6	3.8	-0.01	3.9	4.3	-0.02
Medical history: Neoplasms						
Hematologic neoplasm	0.2	0.3	-0.01	0.2	0.2	0.00
Malignant lymphoma	0.2	0.3	-0.01	0.2	0.2	-0.01
Malignant neoplastic disease	14.8	17.3	-0.07	15.3	16.1	-0.02
Malignant tumor of breast	2.9	2.8	0.01	2.8	2.9	-0.01
Malignant tumor of urinary bladder	0.5	0.7	-0.03	0.5	0.5	0.00
Medication use						
Agents acting on the renin-angiotensin system	20.9	21.5	-0.01	19.9	21.3	-0.03
Antibacterials for systemic use	12.9	14.5	-0.05	14.6	15.3	-0.02
Antidepressants	9.1	7.9	0.04	8.9	9.3	-0.02
Antiepileptics	2.9	2.6	0.02	3.0	3.1	-0.01
Antineoplastic agents	1.7	2.4	-0.04	2.0	2.1	-0.01
Antipsoriatics	0.5	0.5	0.00	0.5	0.6	-0.01
Antithrombotic agents	13.7	17.5	-0.10	13.7	15.0	-0.04
Beta blocking agents	14.3	17.2	-0.08	14.9	16.1	-0.04
Calcium channel blockers	11.0	11.4	-0.01	10.2	11.1	-0.03

Diuretics	13.3	13.7	-0.01	13.0	14.3	-0.04
Drugs for acid related disorders	25.0	27.3	-0.05	25.6	27.2	-0.04
Drugs for obstructive airway diseases	20.6	20.7	0.00	21.3	21.8	-0.01
Drugs used in diabetes	8.6	8.5	0.00	8.6	9.0	-0.01
Immunosuppressants	1.4	1.9	-0.04	1.7	1.8	-0.01
Lipid modifying agents	22.7	23.6	-0.02	20.8	22.3	-0.04
Opioids	7.0	6.4	0.03	7.3	7.7	-0.01
Psycholeptics	9.9	9.8	0.00	10.6	11.0	-0.01
Psychostimulants, agents used for adhd and nootropics	0.5	0.7	-0.03	0.7	0.7	0.00

Supplementary Table 4. Crude incidence rate and corresponding 95% confidence intercals, European data

	UK CPRD				Germany DA				SIDIAP			
Vaccine	N	Person-year	Events	IR per 1,000 person-years (95% CI)	N	Person-year	Events	IR per 1,000 person-years (95% CI)	N	Person-year	Events	IR per 1,000 person-years (95% CI)
Thrombocytopenia												
ChAdOx1 1st dose	3,621,898	263,673	1,261	4.78 (4.52-5.05)	85,500	6,480	37	5.71 (4.02-7.87)				
ChAdOx1 2nd dose	1,115,370	54,189	347	6.4 (5.75-7.11)	17,426	1,219	<5					
BNT162b2 1st dose	1,724,118	130,845	641	4.9 (4.53-5.29)	334,413	25,186	116	4.61 (3.81-5.52)	1,754,222	109,785	1,516	13.81 (13.12-14.52)
BNT162b2 2nd dose	1,276,204	80,349	458	5.7 (5.19-6.25)	170,133	12,124	59	4.87 (3.7-6.28)	1,131,167		1,242	16.45 (15.54-17.39)
mRNA -1273 1st dose					614	46	<5		215,405	15,527	416	26.79 (24.28-29.49)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	158,659	8,191	310	37.85 (33.75-42.3)
Janssen COVID-19					17,988	1,217	13	10.68 (5.69-18.27)	115,521	5,434	57	10.49 (7.94-13.59)
Any thrombosis (venous thrombo-embolism or arterial thromboembolism) with thrombocytopenia syndrome (Any-TTS)												
ChAdOx1 1st dose	3,772,477	274,860	197	0.72 (0.62-0.82)								
ChAdOx1 2nd dose	1,186,267	57,864	51	0.88 (0.66-1.16)								
BNT162b2 1st dose	1,827,767	138,755	98	0.71 (0.57-0.86)					1,990,976	126,431	0	0 (0-0.03)
BNT162b2 2nd dose	1,358,901	85,737	68	0.79 (0.62-1.01)					1,328,942	89,897	0	0 (0-0.04)
mRNA -1273 1st dose									246,113	17,799	0	0 (0-0.21)
mRNA -1273 2nd dose									184,804	9,891	0	0 (0-0.37)
Janssen COVID-19									126,950	6,054	0	0 (0-0.61)
TTS_ATE												

ChAdOx1 1st dose	3,788,372	276,051	7	0.03 (0.01-0.05)	89,691	6,799	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,498	58,293	<5		18,765	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,463	139,647	6	0.04 (0.02-0.09)	350,881	26,434	<5		1,987,022	126,133	80	0.63 (0.5-0.79)
BNT162b2 2nd dose	1,368,269	86,343	<5		181,708	12,971	0	0 (0-0.28)	1,325,147	89,614	74	0.83 (0.65-1.04)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,696	17,768	10	0.56 (0.27-1.03)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,428	9,864	5	0.51 (0.16-1.18)
Janssen COVID-19					18,757	1,271	0	0 (0-2.9)	126,844	6,048	<5	
TTS_CVST												
ChAdOx1 1st dose	3,788,833	276,086	<5		89,718	6,802	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,815	58,310	0	0 (0-0.06)	18,774	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,964	139,685	0	0 (0-0.03)	351,008	26,444	<5		1,990,968	126,430	0	0 (0-0.03)
BNT162b2 2nd dose	1,368,704	86,372	0	0 (0-0.04)	181,818	12,979	0	0 (0-0.28)	1,328,935	89,897	0	0 (0-0.04)
mR -1273 1st dose					653	49	0	0 (0-74.7)	246,109	17,799	0	0 (0-0.21)
mR -1273 2nd dose					72	5	0	0 (0-694.52)	184,800	9,890	0	0 (0-0.37)
Janssen COVID-19					18,763	1,272	0	0 (0-2.9)	126,949	6,054	0	0 (0-0.61)
Deep vein thrombosis with thrombocytopenia syndrome (TTS-DVT)												
ChAdOx1 1st dose	3,788,419	276,054	10	0.04 (0.02-0.07)	89,706	6,801	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,560	58,298	<5		18,770	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,609	139,658	<5		350,943	26,439	<5		1,990,183	126,371	18	0.14 (0.08-0.23)
BNT162b2 2nd dose	1,368,413	86,353	7	0.08 (0.03-0.17)	181,768	12,976	<5		1,328,197	89,842	11	0.12 (0.06-0.22)
mRNA -1273 1st dose					652	49	0	0 (0-74.82)	245,907	17,784	5	0.28 (0.09-0.66)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,624	9,878	6	0.61 (0.22-1.32)
Janssen COVID-19					18,758	1,271	0	0 (0-2.9)	126,923	6,053	<5	
TTS_ATE												
ChAdOx1 1st dose	3,788,766	276,081	<5		89,704	6,800	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,766	58,307	<5		18,771	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,898	139,680	<5		350,954	26,440	<5		1,988,798	126,266	41	0.32 (0.23-0.44)
BNT162b2 2nd dose	1,368,650	86,368	<5		181,769	12,976	0	0 (0-0.28)	1,326,837	89,739	39	0.43 (0.31-0.59)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,927	17,786	5	0.28 (0.09-0.66)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,639	9,879	<5	
Janssen COVID-19					18,762	1,272	0	0 (0-2.9)	126,893	6,051	0	0 (0-0.61)
TTS_MI												
ChAdOx1 1st dose	3,788,454	276,057	<5		89,705	6,801	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,555	58,296	<5		18,768	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,542	139,653	<5		350,932	26,438	0	0 (0-0.14)	1,989,318	126,306	36	0.29 (0.2-0.39)
BNT162b2 2nd dose	1,368,336	86,347	<5		181,754	12,974	0	0 (0-0.28)	1,327,361	89,781	34	0.38 (0.26-0.53)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,917	17,785	<5	
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,621	9,878	<5	
Janssen COVID-19					18,758	1,271	0	0 (0-2.9)	126,902	6,051	<5	

TTS_PE												
ChAdOx1 1st dose	3,788,598	276,068	7	0.03 (0.01-0.05)	89,698	6,800	<5					
ChAdOx1 2nd dose	1,194,665	58,302	0	0 (0-0.06)	18,770	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,770	139,670	<5		350,891	26,435	<5		1,990,144	126,368	28	0.22 (0.15-0.32)
BNT162b2 2nd dose	1,368,545	86,362	<5		181,732	12,973	<5		1,328,162	89,839	17	0.19 (0.11-0.3)
mRNA -1273 1st dose					651	49	0	0 (0-74.94)	245,920	17,785	9	0.51 (0.23-0.96)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,642	9,879	<5	
Janssen COVID-19					18,759	1,272	<5		126,917	6,052	0	0 (0-0.61)
TTS_SVT												
ChAdOx1 1st dose	3,788,796	276,083	<5		89,714	6,801	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,797	58,309	0	0 (0-0.06)	18,772	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,935	139,683	0	0 (0-0.03)	350,997	26,443	0	0 (0-0.14)	1,990,776	126,416	<5	
BNT162b2 2nd dose	1,368,682	86,371	0	0 (0-0.04)	181,809	12,979	0	0 (0-0.28)	1,328,759	89,884	5	0.06 (0.02-0.13)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,964	17,788	<5	
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,680	9,882	0	0 (0-0.37)
Janssen COVID-19					18,763	1,272	0	0 (0-2.9)	126,942	6,054	0	0 (0-0.61)
Venous thrombo-embolism with thrombocytopenia syndrome (TTS-VTE)												
ChAdOx1 1st dose	3,788,220	276,039	14	0.05 (0.03-0.09)	89,697	6,800	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,431	58,291	<5		18,767	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,448	139,646	5	0.04 (0.01-0.08)	350,877	26,434	<5		1,989,653	126,332	39	0.31 (0.22-0.42)
BNT162b2 2nd dose	1,368,277	86,344	9	0.1 (0.05-0.2)	181,720	12,972	<5		1,327,705	89,805	25	0.28 (0.18-0.41)
mRNA -1273 1st dose					651	49	0	0 (0-74.94)	245,790	17,775	15	0.84 (0.47-1.39)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,523	9,870	9	0.91 (0.42-1.73)
Janssen COVID-19					18,755	1,271	<5		126,901	6,051	<5	
Cerebral venous sinus thrombosis												
ChAdOx1 1st dose	3,788,249	276,041	13	0.05 (0.03-0.08)	89,711	6,801	<5					
ChAdOx1 2nd dose	1,194,604	58,300	<5		18,770	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,631	139,660	0	0 (0-0.03)	350,965	26,441	<5		1,990,871	126,424	<5	
BNT162b2 2nd dose	1,368,495	86,361	0	0 (0-0.04)	181,790	12,977	0	0 (0-0.28)	1,328,866	89,892	0	0 (0-0.04)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	246,081	17,797	0	0 (0-0.21)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,775	9,889	<5	
Janssen COVID-19					18,762	1,272	<5		126,938	6,053	0	0 (0-0.61)
Splanchnic and Visceral Thrombosis												
ChAdOx1 1st dose	3,787,759	276,004	17	0.06 (0.04-0.1)	89,698	6,800	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,194,267	58,283	5	0.09 (0.03-0.2)	18,767	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,839,281	139,633	5	0.04 (0.01-0.08)	350,832	26,431	<5		1,989,938	126,356	15	0.12 (0.07-0.2)
BNT162b2 2nd dose	1,368,176	86,340	<5		181,688	12,970	<5		1,328,039	89,832	14	0.16 (0.09-0.26)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,579	17,760	7	0.39 (0.16-0.81)
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,343	9,859	6	0.61 (0.22-1.32)

Janssen COVID-19					18,759	1,271	0	0 (0-2.9)	126,887	6,051	<5	
Deep vein thrombosis (DVT)												
ChAdOx1 1st dose	3,741,359	272,472	335	1.23 (1.1-1.37)	88,463	6,706	21	3.13 (1.94-4.79)				
ChAdOx1 2nd dose	1,169,773	57,025	86	1.51 (1.21-1.86)	18,390	1,288	0	0 (0-2.86)				
BNT162b2 1st dose	1,804,763	136,994	236	1.72 (1.51-1.96)	344,916	25,982	33	1.27 (0.87-1.78)	1,973,028	125,127	219	1.75 (1.53-2)
BNT162b2 2nd dose	1,341,717	84,648	143	1.69 (1.42-1.99)	177,348	12,651	27	2.13 (1.41-3.11)	1,312,985	88,727	170	1.92 (1.64-2.23)
mRNA -1273 1st dose					634	47	<5		243,786	17,626	36	2.04 (1.43-2.83)
mRNA -1273 2nd dose					71	5	0	0 (0-704.69)	182,802	9,759	26	2.66 (1.74-3.9)
Janssen COVID-19					18,485	1,252	<5		126,189	6,012	10	1.66 (0.8-3.06)
Pulmonary embolism												
ChAdOx1 1st dose	3,757,176	273,676	387	1.41 (1.28-1.56)	88,838	6,734	6	0.89 (0.33-1.94)				
ChAdOx1 2nd dose	1,176,965	57,391	70	1.22 (0.95-1.54)	18,500	1,296	<5					
BNT162b2 1st dose	1,815,404	137,808	275	2 (1.77-2.25)	345,384	26,017	36	1.38 (0.97-1.92)	1,978,276	125,496	177	1.41 (1.21-1.63)
BNT162b2 2nd dose	1,349,726	85,156	122	1.43 (1.19-1.71)	177,597	12,667	20	1.58 (0.96-2.44)	1,317,392	89,043	131	1.47 (1.23-1.75)
mRNA -1273 1st dose					643	48	0	0 (0-75.88)	244,418	17,674	48	2.72 (2-3.6)
mRNA -1273 2nd dose					71	5	0	0 (0-704.69)	183,339	9,793	23	2.35 (1.49-3.52)
Janssen COVID-19					18,581	1,259	<5		126,442	6,024	5	0.83 (0.27-1.94)
Venous thrombo-embolism (VTE)												
ChAdOx1 1st dose	3,715,771	270,525	656	2.42 (2.24-2.62)	87,718	6,649	25	3.76 (2.43-5.55)				
ChAdOx1 2nd dose	1,155,369	56,282	143	2.54 (2.14-2.99)	18,163	1,272	<5					
BNT162b2 1st dose	1,784,864	135,473	462	3.41 (3.11-3.74)	340,391	25,639	59	2.3 (1.75-2.97)	1,964,216	124,479	358	2.88 (2.59-3.19)
BNT162b2 2nd dose	1,326,295	83,658	248	2.96 (2.61-3.36)	173,977	12,401	43	3.47 (2.51-4.67)	1,304,959	88,132	273	3.1 (2.74-3.49)
mRNA -1273 1st dose					629	47	<5		242,639	17,541	79	4.5 (3.57-5.61)
mRNA -1273 2nd dose					70	5	0	0 (0-715.16)	181,795	9,691	42	4.33 (3.12-5.86)
Janssen COVID-19					18,332	1,241	8	6.45 (2.78-12.7)	125,850	5,992	14	2.34 (1.28-3.92)
Ischaemic stroke												
ChAdOx1 1st dose	3,776,958	275,183	95	0.35 (0.28-0.42)	88,152	6,682	13	1.95 (1.04-3.33)				
ChAdOx1 2nd dose	1,187,113	57,901	33	0.57 (0.39-0.8)	18,246	1,278	0	0 (0-2.89)				
BNT162b2 1st dose	1,829,692	138,900	118	0.85 (0.7-1.02)	344,214	25,927	26	1 (0.66-1.47)	1,935,389	122,252	549	4.49 (4.12-4.88)
BNT162b2 2nd dose	1,360,495	85,825	52	0.61 (0.45-0.79)	176,358	12,574	8	0.64 (0.27-1.25)	1,276,052	85,949	523	6.08 (5.57-6.63)
mRNA -1273 1st dose					645	48	0	0 (0-75.64)	242,770	17,550	39	2.22 (1.58-3.04)
mRNA -1273 2nd dose					71	5	0	0 (0-704.69)	181,905	9,704	24	2.47 (1.58-3.68)
Janssen COVID-19					18,433	1,248	<5		125,015	5,935	18	3.03 (1.8-4.79)
Myocardial infarction												
ChAdOx1 1st dose	3,727,073	271,378	422	1.56 (1.41-1.71)	87,339	6,620	10	1.51 (0.72-2.78)				
ChAdOx1 2nd dose	1,155,608	56,294	85	1.51 (1.21-1.87)	17,915	1,253	<5					
BNT162b2 1st dose	1,783,357	135,356	268	1.98 (1.75-2.23)	341,848	25,748	34	1.32 (0.91-1.85)	1,956,402	123,866	289	2.33 (2.07-2.62)
BNT162b2 2nd dose	1,324,606	83,538	158	1.89 (1.61-2.21)	174,395	12,425	24	1.93 (1.24-2.87)	1,297,110	87,586	269	3.07 (2.72-3.46)
mRNA -1273 1st dose					645	48	0	0 (0-75.63)	242,726	17,546	30	1.71 (1.15-2.44)

mRNA -1273 2nd dose					71	5	0	0 (0-704.69)	181,881	9,716	21	2.16 (1.34-3.3)
Janssen COVID-19					18,277	1,236	<5		125,219	5,951	10	1.68 (0.81-3.09)
Intestinal infarction												
ChAdOx1 1st dose	3,787,035	275,950	28	0.1 (0.07-0.15)	89,687	6,799	0	0 (0-0.54)				
ChAdOx1 2nd dose	1,193,764	58,255	5	0.09 (0.03-0.2)	18,768	1,316	0	0 (0-2.8)				
BNT162b2 1st dose	1,838,646	139,584	19	0.14 (0.08-0.21)	350,850	26,432	0	0 (0-0.14)	1,988,290	126,231	24	0.19 (0.12-0.28)
BNT162b2 2nd dose	1,367,612	86,300	10	0.12 (0.06-0.21)	181,693	12,970	0	0 (0-0.28)	1,326,443	89,712	20	0.22 (0.14-0.34)
mRNA -1273 1st dose					653	49	0	0 (0-74.7)	245,853	17,780	<5	
mRNA -1273 2nd dose					72	5	0	0 (0-694.52)	184,577	9,875	5	0.51 (0.16-1.18)
Janssen COVID-19					18,758	1,271	0	0 (0-2.9)	126,855	6,048	<5	
Arterial thromboembolism (ATE)												
ChAdOx1 1st dose	3,716,038	270,539	629	2.32 (2.15-2.51)	85,890	6,509	21	3.23 (2-4.93)				
ChAdOx1 2nd dose	1,148,502	55,917	141	2.52 (2.12-2.97)	17,438	1,218	<5					
BNT162b2 1st dose	1,773,973	134,639	451	3.35 (3.05-3.67)	335,573	25,270	63	2.49 (1.92-3.19)	1,904,070	119,934	998	8.32 (7.81-8.85)
BNT162b2 2nd dose	1,317,134	83,041	232	2.79 (2.45-3.18)	169,391	12,055	31	2.57 (1.75-3.65)	1,247,356	83,871	918	10.95 (10.25-11.68)
mRNA -1273 1st dose					637	48	0	0 (0-76.59)	239,599	17,313	86	4.97 (3.97-6.13)
mRNA -1273 2nd dose					70	5	0	0 (0-715.16)	179,168	9,542	53	5.55 (4.16-7.27)
Janssen COVID-19					17,966	1,213	<5		123,408	5,839	34	5.82 (4.03-8.14)

Supplementary Table 5. Crude incidence rate and corresponding 95% confidence intercals, US data

	US Hospital CDM				US Open Claims			
Vaccine	N	Person-year	Events	IR per 1,000 person-years (95% CI)	N	Person-year	Events	IR per 1,000 person-years (95% CI)
thrombocytopenia								
BNT162b2 1st dose	240,105	15,552	135	8.68 (7.28-10.27)	5,285,326	376,318	1,038	2.76 (2.59-2.93)
BNT162b2 2nd dose	72,551	5,182	86	16.6 (13.27-20.5)	2,863,353	212,064	754	3.56 (3.31-3.82)
mRNA-1273 1st dose	61,081	4,603	25	5.43 (3.51-8.02)	3,923,483	298,953	946	3.16 (2.97-3.37)
mRNA-1273 2nd dose	19,561	1,382	25	18.08 (11.7-26.7)	2,189,283	165,463	642	3.88 (3.59-4.19)
Janssen COVID-19	1,645	114	6	52.28 (19.18-113.78)	628,294	46,997	170	3.62 (3.09-4.2)
TTS_any								

BNT162b2 1st dose					5,288,025	376,521	818	2.17 (2.03-2.33)
BNT162b2 2nd dose					2,865,177	212,203	595	2.8 (2.58-3.04)
mRNA-1273 1st dose					3,925,741	299,125	750	2.51 (2.33-2.69)
mRNA-1273 2nd dose					2,190,856	165,583	502	3.03 (2.77-3.31)
Janssen COVID-19					628,572	47,019	146	3.11 (2.62-3.65)
TTS_ATE								
BNT162b2 1st dose	247,434	16,066	12	0.75 (0.39-1.3)	5,374,657	382,956	36	0.09 (0.07-0.13)
BNT162b2 2nd dose	76,551	5,475	6	1.1 (0.4-2.39)	2,921,124	216,416	32	0.15 (0.1-0.21)
mRNA-1273 1st dose	62,720	4,726	<5		3,999,861	304,784	41	0.13 (0.1-0.18)
mRNA-1273 2nd dose	20,392	1,442	<5		2,240,829	169,389	33	0.19 (0.13-0.27)
Janssen COVID-19	1,712	119	<5		639,180	47,822	<5	
TTS_CVST								
BNT162b2 1st dose	247,714	16,085	0	0 (0-0.23)	5,377,723	383,187	0	0 (0-0.01)
BNT162b2 2nd dose	76,682	5,484	0	0 (0-0.67)	2,923,166	216,571	0	0 (0-0.02)
mRNA-1273 1st dose	62,781	4,731	0	0 (0-0.78)	4,002,519	304,987	0	0 (0-0.01)
mRNA-1273 2nd dose	20,431	1,445	0	0 (0-2.55)	2,242,605	169,524	0	0 (0-0.02)
Janssen COVID-19	1,715	119	0	0 (0-30.87)	639,538	47,849	<5	
TTS_DVT								
BNT162b2 1st dose	247,507	16,070	7	0.44 (0.18-0.9)	5,375,843	383,046	25	0.07 (0.04-0.1)
BNT162b2 2nd dose	76,552	5,475	<5		2,921,944	216,478	15	0.07 (0.04-0.11)
mRNA-1273 1st dose	62,737	4,727	<5		4,001,015	304,872	20	0.07 (0.04-0.1)
mRNA-1273 2nd dose	20,406	1,443	<5		2,241,586	169,446	12	0.07 (0.04-0.12)
Janssen COVID-19	1,715	119	0	0 (0-30.87)	639,334	47,833	6	0.13 (0.05-0.27)
TTS_ATE								
BNT162b2 1st dose	247,611	16,078	<5		5,376,277	383,078	18	0.05 (0.03-0.07)
BNT162b2 2nd dose	76,635	5,481	<5		2,922,248	216,501	14	0.06 (0.04-0.11)
mRNA-1273 1st dose	62,763	4,729	0	0 (0-0.78)	4,001,310	304,894	20	0.07 (0.04-0.1)
mRNA-1273 2nd dose	20,417	1,444	0	0 (0-2.55)	2,241,759	169,460	14	0.08 (0.05-0.14)
Janssen COVID-19	1,715	119	<5		639,374	47,836	<5	
TTS_MI								
BNT162b2 1st dose	247,553	16,074	9	0.56 (0.26-1.06)	5,376,066	383,062	16	0.04 (0.02-0.07)
BNT162b2 2nd dose	76,609	5,479	5	0.91 (0.3-2.13)	2,922,006	216,483	20	0.09 (0.06-0.14)
mRNA-1273 1st dose	62,740	4,727	<5		4,001,023	304,873	20	0.07 (0.04-0.1)
mRNA-1273 2nd dose	20,408	1,443	<5		2,241,633	169,450	19	0.11 (0.07-0.18)
Janssen COVID-19	1,712	119	<5		639,336	47,833	<5	

TTS_PE								
BNT162b2 1st dose	247,545	16,072	<5		5,376,230	383,074	19	0.05 (0.03-0.08)
BNT162b2 2nd dose	76,580	5,477	0	0 (0-0.67)	2,922,111	216,491	12	0.06 (0.03-0.1)
mRNA-1273 1st dose	62,749	4,728	0	0 (0-0.78)	4,001,176	304,884	13	0.04 (0.02-0.07)
mRNA-1273 2nd dose	20,417	1,444	<5		2,241,723	169,457	13	0.08 (0.04-0.13)
Janssen COVID-19	1,713	119	0	0 (0-30.9)	639,368	47,836	6	0.13 (0.05-0.27)
TTS_SVT								
BNT162b2 1st dose	247,668	16,081	<5		5,377,449	383,166	<5	
BNT162b2 2nd dose	76,653	5,482	<5		2,922,992	216,558	<5	
mRNA-1273 1st dose	62,772	4,730	0	0 (0-0.78)	4,002,311	304,971	<5	
mRNA-1273 2nd dose	20,427	1,445	<5		2,242,473	169,514	5	0.03 (0.01-0.07)
Janssen COVID-19	1,714	119	0	0 (0-30.89)	639,507	47,847	<5	
TTS_VTE								
BNT162b2 1st dose	247,444	16,065	8	0.5 (0.21-0.98)	5,375,270	383,003	32	0.08 (0.06-0.12)
BNT162b2 2nd dose	76,517	5,472	<5		2,921,557	216,448	16	0.07 (0.04-0.12)
mRNA-1273 1st dose	62,725	4,726	<5		4,000,511	304,834	23	0.08 (0.05-0.11)
mRNA-1273 2nd dose	20,402	1,443	<5		2,241,261	169,422	16	0.09 (0.05-0.15)
Janssen COVID-19	1,713	119	0	0 (0-30.9)	639,270	47,828	11	0.23 (0.11-0.41)
Cerebral venous sinus thrombosis								
BNT162b2 1st dose	247,697	16,083	0	0 (0-0.23)	5,377,097	383,140	10	0.03 (0.01-0.05)
BNT162b2 2nd dose	76,675	5,484	0	0 (0-0.67)	2,922,724	216,537	9	0.04 (0.02-0.08)
mRNA-1273 1st dose	62,776	4,730	0	0 (0-0.78)	4,002,042	304,950	6	0.02 (0.01-0.04)
mRNA-1273 2nd dose	20,428	1,445	<5		2,242,315	169,502	<5	
Janssen COVID-19	1,714	119	0	0 (0-30.89)	639,457	47,843	<5	
Splanchnic and Visceral Thrombosis								
BNT162b2 1st dose	247,470	16,067	<5		5,374,280	382,930	32	0.08 (0.06-0.12)
BNT162b2 2nd dose	76,533	5,473	5	0.91 (0.3-2.13)	2,920,875	216,398	21	0.1 (0.06-0.15)
mRNA-1273 1st dose	62,704	4,725	<5		3,999,804	304,780	42	0.14 (0.1-0.19)
mRNA-1273 2nd dose	20,394	1,442	<5		2,240,723	169,381	28	0.17 (0.11-0.24)
Janssen COVID-19	1,713	119	0	0 (0-30.91)	639,112	47,817	10	0.21 (0.1-0.38)
Deep vein thrombosis								
BNT162b2 1st dose	243,746	15,808	45	2.85 (2.08-3.81)	5,288,296	376,489	787	2.09 (1.95-2.24)
BNT162b2 2nd dose	74,589	5,332	41	7.69 (5.52-10.43)	2,862,601	211,990	508	2.4 (2.19-2.61)

mRNA-1273 1st dose	61,716	4,650	16	3.44 (1.97-5.59)	3,922,758	298,892	674	2.25 (2.09-2.43)
mRNA-1273 2nd dose	19,885	1,405	18	12.8 (7.59-20.23)	2,187,848	165,348	453	2.74 (2.49-3)
Janssen COVID-19	1,680	116	<5		628,004	46,974	121	2.58 (2.14-3.08)
Pulmonary embolism								
BNT162b2 1st dose	244,991	15,895	36	2.26 (1.59-3.14)	5,324,735	379,216	598	1.58 (1.45-1.71)
BNT162b2 2nd dose	75,211	5,378	22	4.09 (2.56-6.19)	2,887,007	213,839	351	1.64 (1.47-1.82)
mRNA-1273 1st dose	62,104	4,679	10	2.14 (1.02-3.93)	3,955,656	301,407	465	1.54 (1.41-1.69)
mRNA-1273 2nd dose	20,064	1,418	13	9.16 (4.88-15.67)	2,210,655	167,087	317	1.9 (1.69-2.12)
Janssen COVID-19	1,684	117	<5		632,835	47,339	105	2.22 (1.81-2.69)
Venous thrombo-embolism								
BNT162b2 1st dose	242,061	15,691	67	4.27 (3.31-5.42)	5,256,785	374,134	1,146	3.06 (2.89-3.25)
BNT162b2 2nd dose	73,682	5,266	52	9.87 (7.37-12.95)	2,841,124	210,370	726	3.45 (3.2-3.71)
mRNA-1273 1st dose	61,319	4,620	22	4.76 (2.98-7.21)	3,895,067	296,777	971	3.27 (3.07-3.48)
mRNA-1273 2nd dose	19,667	1,390	25	17.99 (11.64-26.55)	2,169,182	163,926	646	3.94 (3.64-4.26)
Janssen COVID-19	1,659	115	<5		624,003	46,670	190	4.07 (3.51-4.69)
Ischaemic stroke								
BNT162b2 1st dose	243,522	15,794	53	3.36 (2.51-4.39)	5,252,695	373,795	1,213	3.25 (3.07-3.43)
BNT162b2 2nd dose	74,521	5,327	45	8.45 (6.16-11.3)	2,837,126	210,050	841	4 (3.74-4.28)
mRNA-1273 1st dose	61,702	4,649	19	4.09 (2.46-6.38)	3,889,140	296,321	1,160	3.91 (3.69-4.15)
mRNA-1273 2nd dose	19,851	1,402	12	8.55 (4.42-14.94)	2,165,395	163,635	800	4.89 (4.56-5.24)
Janssen COVID-19	1,664	116	<5		623,397	46,622	193	4.14 (3.58-4.77)
Myocardial infarction								
BNT162b2 1st dose	243,360	15,786	61	3.86 (2.96-4.96)	5,281,985	375,988	1,003	2.67 (2.51-2.84)
BNT162b2 2nd dose	74,570	5,330	48	9 (6.64-11.94)	2,857,066	211,557	734	3.47 (3.22-3.73)
mRNA-1273 1st dose	61,614	4,643	23	4.95 (3.14-7.43)	3,909,897	297,907	1,024	3.44 (3.23-3.65)
mRNA-1273 2nd dose	19,816	1,400	22	15.7 (9.84-23.78)	2,178,281	164,615	691	4.2 (3.89-4.52)
Janssen COVID-19	1,672	116	<5		625,169	46,757	168	3.59 (3.07-4.18)
Intestinal infarction								
BNT162b2 1st dose	247,266	16,053	<5		5,368,328	382,481	99	0.26 (0.21-0.32)
BNT162b2 2nd dose	76,446	5,467	7	1.28 (0.51-2.64)	2,916,832	216,091	70	0.32 (0.25-0.41)
mRNA-1273 1st dose	62,682	4,723	<5		3,993,545	304,301	95	0.31 (0.25-0.38)
mRNA-1273 2nd dose	20,371	1,440	0	0 (0-2.56)	2,236,418	169,052	103	0.61 (0.5-0.74)
Janssen COVID-19	1,714	119	0	0 (0-30.89)	638,258	47,752	7	0.15 (0.06-0.3)

Arterial thromboembolism								
BNT162b2 1st dose	239,472	15,517	140	9.02 (7.59-10.65)	5,172,486	367,768	4,666	12.69 (12.33-13.06)
BNT162b2 2nd dose	72,596	5,187	110	21.2 (17.43-25.56)	2,781,792	205,853	3,211	15.6 (15.06-16.15)
mRNA-1273 1st dose	60,601	4,566	54	11.83 (8.88-15.43)	3,810,079	290,277	4,615	15.9 (15.44-16.36)
mRNA-1273 2nd dose	19,275	1,361	40	29.38 (20.99-40.01)	2,110,385	159,437	2,926	18.35 (17.69-19.03)
Janssen COVID-19	1,625	113	<5		610,896	45,673	720	15.76 (14.63-16.96)

Supplementary Table 6. Cohort selection and minimal detectable relative risk, database-target-outcome.

	Original cohort		Without prior outcomes		With at least 1 days at risk		Matched on propensity score		MDRR
Outcome	**Comparator**	**Target**	**Comparator**	**Target**	**Comparator**	**Target**	**Comparator**	**Target**	
CPRD Aurum	**ChAdOx1 1st dose vs. BNT162b2 1st dose**								
CVST	1840240	3789631	1839907	3789045	1839631	3788249	1946223	1271557	>12.97
DVT	1840240	3789631	1805035	3742139	1804763	3741359	1912752	1247556	1.36
Intestinal infarction	1840240	3789631	1838922	3787830	1838646	3787035	1945248	1270917	2.6
Ischemic stroke	1840240	3789631	1829966	3777750	1829692	3776958	1936816	1264894	1.59
ATE	1840240	3789631	1774242	3716815	1773973	3716038	1886308	1227495	1.23
MI	1840240	3789631	1783628	3727853	1783357	3727073	1895358	1233874	1.3
PE	1840240	3789631	1815678	3757964	1815404	3757176	1922818	1254781	1.3
SVT	1840240	3789631	1839557	3788555	1839281	3787759	1945829	1271295	>3.72
Thrombocytopenia	1840240	3789631	1724387	3622661	1724118	3621898	1836112	1195498	1.17
VTE	1840240	3789631	1785134	3716545	1784864	3715771	1893469	1233788	1.24
TTS_ATE	1840240	3789631	1839739	3789168	1839463	3788372	1946184	1271530	>6.12
TTS_CVST	1840240	3789631	1840240	3789629	1839964	3788833	1946593	1271821	Inf
TTS_DVT	1840240	3789631	1839885	3789215	1839609	3788419	1946277	1271591	>12.97
TTS_isc stroke	1840240	3789631	1840174	3789562	1839898	3788766	1946536	1271777	>12.97
TTS_MI	1840240	3789631	1839818	3789250	1839542	3788454	1946252	1271585	>6.12
TTS_PE	1840240	3789631	1840046	3789394	1839770	3788598	1946400	1271669	>6.12
TTS_SVT	1840240	3789631	1840211	3789592	1839935	3788796	1946566	1271800	>12.97
TTS_VTE	1840240	3789631	1839724	3789016	1839448	3788220	1946111	1271464	>5.23
TTS_all stroke	1840240	3789631	1839911	3789284	1839635	3788489	1946277	1271609	>6.12
TTS_any	1840240	3789631	1828042	3773269	1827767	3772477	1934651	1263613	1.52
CPRD Aurum	**ChAdOx1 2nd dose vs. BNT162b2 2nd dose**								
CVST	1369238	1195626	1369029	1195414	1368495	1194604	1084284	801323	Inf
DVT	1369238	1195626	1342243	1170572	1341717	1169773	1063064	784878	1.57
Intestinal infarction	1369238	1195626	1368146	1194574	1367612	1193764	1083567	800793	>4.82
Ischemic stroke	1369238	1195626	1361028	1187920	1360495	1187113	1078360	796695	2.21
ATE	1369238	1195626	1317653	1149297	1317134	1148502	1044491	770339	1.43
MI	1369238	1195626	1325126	1156405	1324606	1155608	1050018	774713	1.54
PE	1369238	1195626	1350256	1177764	1349726	1176965	1069375	789797	1.6
SVT	1369238	1195626	1368710	1195077	1368176	1194267	1084010	801100	>6.00

Thrombocytopenia	1369238	1195626	1276719	1116141	1276204	1115370	1012563	747810	1.27
VTE	1369238	1195626	1326817	1156159	1326295	1155369	1050916	775486	1.4
TTS_ATE	1369238	1195626	1368803	1195308	1368269	1194498	1084154	801245	>6.00
TTS_CVST	1369238	1195626	1369238	1195625	1368704	1194815	1084454	801467	Inf
TTS_DVT	1369238	1195626	1368947	1195370	1368413	1194560	1084231	801305	>6.00
TTS_isc stroke	1369238	1195626	1369184	1195576	1368650	1194766	1084417	801435	>6.00
TTS_MI	1369238	1195626	1368870	1195365	1368336	1194555	1084202	801289	>6.00
TTS_PE	1369238	1195626	1369079	1195475	1368545	1194665	1084319	801375	>12.61
TTS_SVT	1369238	1195626	1369216	1195607	1368682	1194797	1084432	801452	Inf
TTS_VTE	1369238	1195626	1368811	1195241	1368277	1194431	1084118	801225	>5.52
TTS_all stroke	1369238	1195626	1368976	1195381	1368442	1194571	1084262	801318	>6.00
TTS_any	1369238	1195626	1359432	1187076	1358901	1186267	1076722	795629	1.88
Germany_DA	**ChAdOx1 1st dose**		**vs. BNT162b2 1st dose**						
CVST	391063	98562	391019	98555	350965	89711	215180	86344	>7.10
DVT	391063	98562	384716	97278	344916	88463	211587	85163	2.6
Intestinal infarction	391063	98562	390901	98529	350850	89687	215106	86325	Inf
Ischemic stroke	391063	98562	384122	96966	344214	88152	210616	84835	3.37
ATE	391063	98562	375271	94660	335573	85890	204702	82643	2.18
MI	391063	98562	381689	96137	341848	87339	208975	84048	2.81
PE	391063	98562	385227	97653	345384	88838	212362	85493	3.37
SVT	391063	98562	390884	98542	350832	89698	215110	86330	>15.98
Thrombocytopenia	391063	98562	373590	94196	334413	85500	204508	82281	1.79
VTE	391063	98562	380017	96507	340391	87718	209244	84436	2.15
TTS_ATE	391063	98562	390936	98535	350881	89691	215092	86322	>15.98
TTS_CVST	391063	98562	391063	98562	351008	89718	215207	86351	Inf
TTS_DVT	391063	98562	390996	98550	350943	89706	215166	86340	Inf
TTS_isc stroke	391063	98562	391009	98548	350954	89704	215149	86337	>15.98
TTS_MI	391063	98562	390987	98549	350932	89705	215148	86337	Inf
TTS_PE	391063	98562	390944	98542	350891	89698	215135	86332	>15.98
TTS_SVT	391063	98562	391052	98558	350997	89714	215198	86347	Inf
TTS_VTE	391063	98562	390929	98541	350877	89697	215139	86332	>15.98
TTS_all stroke	391063	98562	390954	98543	350902	89699	215114	86330	>15.98
Germany_DA	**Janssen COVID-19**		**vs. BNT162b2 1st dose**						
CVST	391063	37723	391019	37721	350965	18762	68021	18706	>8.62
DVT	391063	37723	384716	37360	344916	18485	67016	18428	>8.62

Intestinal infarction	391063	37723	390901	37713	350850	18758	68007	18701	Inf
Ischemic stroke	391063	37723	384122	37359	344214	18433	66774	18379	>8.61
ATE	391063	37723	375271	36825	335573	17966	65173	17914	>4.98
MI	391063	37723	381689	37169	341848	18277	66336	18223	>7.15
PE	391063	37723	385227	37492	345384	18581	67323	18525	>8.62
SVT	391063	37723	390884	37718	350832	18759	68009	18703	Inf
Thrombocytopenia	391063	37723	373590	36591	334413	17988	65217	17933	3.8
VTE	391063	37723	380017	37166	340391	18332	66431	18274	6.17
TTS_ATE	391063	37723	390936	37716	350881	18757	68004	18701	Inf
TTS_CVST	391063	37723	391063	37723	351008	18763	68026	18707	Inf
TTS_DVT	391063	37723	390996	37718	350943	18758	68001	18702	Inf
TTS_isc stroke	391063	37723	391009	37722	350954	18762	68024	18706	Inf
TTS_MI	391063	37723	390987	37717	350932	18758	68006	18702	Inf
TTS_PE	391063	37723	390944	37718	350891	18759	68008	18703	>21.04
TTS_SVT	391063	37723	391052	37723	350997	18763	68026	18707	Inf
TTS_VTE	391063	37723	390929	37714	350877	18755	67989	18699	>21.03
IPCI	**ChAdOx1 1st dose**		**vs. BNT162b2 1st dose**						
CVST	218423	71083	218423	71083	210487	71001	49659	22239	Inf
DVT	218423	71083	216639	70509	208713	70427	49198	22039	>6.80
Intestinal infarction	218423	71083	218423	71083	210487	71001	49659	22239	Inf
Ischemic stroke	218423	71083	216518	70513	208584	70432	49162	22044	>6.80
ATE	218423	71083	211448	68697	203533	68619	47893	21489	3.54
MI	218423	71083	213239	69237	205322	69158	48370	21678	>5.05
PE	218423	71083	216686	70505	208768	70424	49225	22043	>6.22
SVT	218423	71083	218423	71083	210487	71001	49659	22239	Inf
Thrombocytopenia	218423	71083	214847	70030	206973	69952	48710	21854	>4.55
VTE	218423	71083	215067	69992	207158	69911	48804	21863	4.35
TTS_ATE	218423	71083	218410	71080	210474	70998	49649	22236	Inf
TTS_CVST	218423	71083	218423	71083	210487	71001	49659	22239	Inf
TTS_DVT	218423	71083	218417	71082	210481	71000	49655	22238	Inf
TTS_isc stroke	218423	71083	218421	71083	210485	71001	49659	22239	Inf
TTS_MI	218423	71083	218412	71080	210476	70998	49649	22236	Inf
TTS_PE	218423	71083	218410	71081	210474	70999	49653	22237	Inf
TTS_SVT	218423	71083	218423	71083	210487	71001	49659	22239	Inf
TTS_VTE	218423	71083	218407	71080	210471	70998	49650	22236	Inf

TTS_all stroke	218423	71083	218419	71082	210483	71000	49659	22239	Inf
SIDIAP	**Janssen COVID-19**		**vs. BNT162b2 1st dose**						
CVST	2027950	138351	2027844	138339	1990871	126938	423815	116789	Inf
DVT	2027950	138351	2009886	137540	1973028	126189	421532	116087	2.82
Intestinal infarction	2027950	138351	2025253	138252	1988290	126855	423558	116709	>8.61
Ischemic stroke	2027950	138351	1972241	136360	1935389	125015	417793	114999	2.15
ATE	2027950	138351	1940805	134695	1904070	123408	413039	113588	1.73
MI	2027950	138351	1993255	136561	1956402	125219	418734	115276	2.67
PE	2027950	138351	2015190	137815	1978276	126442	422330	116315	4.77
SVT	2027950	138351	2026906	138287	1989938	126887	423664	116747	>8.61
Thrombocytopenia	2027950	138351	1789028	126127	1754222	115521	386334	106217	1.54
VTE	2027950	138351	2001029	137188	1964216	125850	420502	115760	2.48
TTS_ATE	2027950	138351	2023991	138240	1987022	126844	423524	116696	>8.61
TTS_CVST	2027950	138351	2027942	138350	1990968	126949	423858	116802	Inf
TTS_DVT	2027950	138351	2027155	138321	1990183	126923	423781	116779	>8.61
TTS_isc stroke	2027950	138351	2025770	138290	1988798	126893	423667	116737	>21.00
TTS_MI	2027950	138351	2026290	138302	1989318	126902	423714	116762	>21.00
TTS_PE	2027950	138351	2027112	138316	1990144	126917	423767	116773	>21.00
TTS_SVT	2027950	138351	2027749	138343	1990776	126942	423830	116795	Inf
TTS_VTE	2027950	138351	2026620	138297	1989653	126901	423710	116760	>8.61
TTS_all stroke	2027950	138351	2025382	138280	1988414	126884	423637	116726	>21.00
US_Open_Claims	**Janssen COVID-19**		**vs. BNT162b2 1st dose**						
CVST	6055754	939748	6055085	939659	5377556	639457	2405656	639456	>7.96
DVT	6055754	939748	5963034	927070	5288749	628004	2363428	628002	1.37
Intestinal infarction	6055754	939748	6046033	938328	5368789	638258	2401293	638257	2.43
Ischemic stroke	6055754	939748	5926489	922291	5253145	623397	2348140	623396	1.29
ATE	6055754	939748	5843606	908728	5172933	610896	2304844	610895	1.14
MI	6055754	939748	5956952	924232	5282441	625169	2356142	625168	1.31
PE	6055754	939748	6000736	932417	5325183	632835	2380869	632834	1.44
SVT	6055754	939748	6052148	939279	5374738	639112	2404366	639111	3.59
Thrombocytopenia	6055754	939748	5958632	926904	5285775	628294	2364195	628293	1.31
VTE	6055754	939748	5930163	922661	5257227	624003	2348419	624001	1.3
TTS_ATE	6055754	939748	6052570	939364	5375116	639180	2404622	639179	>4.49
TTS_CVST	6055754	939748	6055739	939747	5378182	639538	2405961	639537	>21.67
TTS_DVT	6055754	939748	6053805	939521	5376301	639334	2405147	639333	5.58

TTS_isc stroke	6055754	939748	6054244	939571	5376736	639374	2405354	639373	>6.74
TTS_MI	6055754	939748	6054026	939529	5376525	639336	2405217	639335	>6.74
TTS_PE	6055754	939748	6054201	939563	5376688	639368	2405319	639367	6.29
TTS_SVT	6055754	939748	6055459	939714	5377908	639507	2405815	639506	>8.80
TTS_VTE	6055754	939748	6053209	939451	5375727	639270	2404904	639269	4.07
TTS_all stroke	6055754	939748	6053844	939527	5376349	639334	2405207	639333	7.28
TTS_any	6055754	939748	5961410	927210	5288474	628572	2365254	628571	1.35
US_Open_Claims	**Janssen COVID-19**		**vs. mRNA-1273 1st dose**						
CVST	4261016	939967	4260530	939878	4002663	639619	2272268	639619	>8.50
DVT	4261016	939967	4179884	927287	3923371	628164	2230157	628164	1.37
Intestinal infarction	4261016	939967	4251889	938545	3994166	638418	2267972	638418	2.38
Ischemic stroke	4261016	939967	4145567	922509	3889750	623557	2214613	623557	1.29
ATE	4261016	939967	4064922	908944	3810684	611054	2171445	611054	1.13
MI	4261016	939967	4166608	924449	3910513	625329	2222711	625329	1.3
PE	4261016	939967	4213302	932636	3956266	632997	2247746	632997	1.45
SVT	4261016	939967	4258228	939498	4000424	639274	2271071	639274	3.68
Thrombocytopenia	4261016	939967	4180086	927125	3924097	628459	2231498	628459	1.3
VTE	4261016	939967	4151615	922878	3895669	624163	2215499	624163	1.3
TTS_ATE	4261016	939967	4258299	939583	4000482	639342	2271303	639342	>3.51
TTS_CVST	4261016	939967	4261011	939966	4003140	639700	2272570	639700	>20.62
TTS_DVT	4261016	939967	4259478	939740	4001635	639496	2271774	639496	4.93
TTS_isc stroke	4261016	939967	4259779	939790	4001931	639536	2272003	639536	>5.43
TTS_MI	4261016	939967	4259484	939748	4001644	639498	2271842	639498	>4.72
TTS_PE	4261016	939967	4259644	939782	4001796	639530	2271912	639530	6.53
TTS_SVT	4261016	939967	4260797	939933	4002932	639669	2272454	639669	>8.50
TTS_VTE	4261016	939967	4258961	939670	4001130	639432	2271552	639432	3.87
TTS_all stroke	4261016	939967	4259463	939746	4001622	639496	2271845	639496	6.1
TTS_any	4261016	939967	4182368	927431	3926355	628737	2232550	628737	1.34

Supplementary Table 7. List of covariates with top 10 highest absolute value of propensity score model coefficient.

Covariate	Coefficient		Covariate	Coefficient
CPRD_Aurum - ChAdOx1 1st dose - BNT162b2 1st dose			**CPRD_Aurum - ChAdOx1 2nd dose - BNT162b2 2nd dose**	
index month: 12	-6.75		observation distinct concept count during day -180 through -4 concept_count relative to index	-7.05
observation distinct concept count during day -180 through -4 concept_count relative to index	-4.73		index month: 1	-6.61
visit_occurrence concept count during day -180 through -4 concept_count relative to index: Office Visit	4.51		index month: 2	-2.21
index month: 1	-4.46		observation during day -180 through -4 days relative to index: Housebound	1.95
measurement distinct concept count during day -180 through -4 concept_count relative to index	-4.08		procedure_occurrence during day -180 through -4 days relative to index: Administration of vaccine	-1.79
visit_occurrence concept count during day -180 through -4 concept_count relative to index	-4.07		index month: 12	-1.77
index year: 2020	-3.83		observation during day -180 through -4 days relative to index: Temporarily housebound	1.61
index month: 2	-3.33		measurement during day -180 through -4 days relative to index: Well person screening	1.57
procedure_occurrence during day -180 through -4 days relative to index: Administration of vaccine	-2.47		observation during day -180 through -4 days relative to index: Informed consent given for treatment	-1.56
drug_era distinct concept count during day -180 through -4 concept_count relative to index	-2.29		observation during day -180 through -4 days relative to index: Appointment cancelled by service	-1.44
France_LPD - ChAdOx1 1st dose - BNT162b2 1st dose			**France_LPD - ChAdOx1 1st dose - mRNA-1273 1st dose**	
index month: 1	-6.42		index month: 8	-6.64
CHADS2VASc	-6.00		index month: 9	-5.75
index month: 8	-4.69		index month: 1	-5.22
visit_occurrence concept count during day -180 through -4 concept_count relative to index	4.00		index month: 7	-4.10
age group: 20 - 29	-3.36		index month: 6	-3.69
index month: 9	-3.31		CHADS2VASc	-3.43
age group: 10 - 19	-3.08		age group: 20 - 29	-3.33
age group: 30 - 39	-3.04		age group: 10 - 19	-2.96

measurement during day -180 through -4 days relative to index: Potassium [Moles/time] in 24 hour Urine	-2.83		age group: 30 - 39	-2.92
age group: 40 - 49	-2.78		age group: 40 - 49	-2.88
Germany_DA - ChAdOx1 1st dose - BNT162b2 1st dose			Germany_DA - Janssen COVID-19 - BNT162b2 1st dose	
index month: 8	-3.76		index month: 6	5.92
index month: 7	-2.64		index month: 8	4.96
index month: 3	2.28		measurement distinct concept count during day -180 through -4 concept_count relative to index	4.77
age group: 10 - 19	-1.77		index month: 5	4.66
index month: 9	-1.68		index month: 7	4.52
gender = MALE	1.44		index month: 9	4.06
index month: 6	-1.42		Charlson index - Romano adaptation	2.25
drug_era distinct concept count during day -180 through -4 concept_count relative to index	-1.24		drug_era distinct concept count during day -180 through -4 concept_count relative to index	2.02
index month: 4	-1.22		gender = FEMALE	-1.75
age group: 20 - 29	-1.05		age group: 10 - 19	-1.28
Germany_DA - ChAdOx1 2nd dose - BNT162b2 2nd dose			**SIDIAP - Janssen COVID-19 - BNT162b2 1st dose**	
index month: 4	-4.10		CHADS2VASc	-10.04
index month: 5	-2.20		index month: 1	-6.91
age group: 10 - 19	-2.15		index month: 3	-6.24
drug_era distinct concept count during day -180 through -4 concept_count relative to index	-1.67		index month: 2	-6.00
age group: 30 - 39	-1.44		index month: 12	-4.02
drug_era group during day -180 through -4 days relative to index: ascorbic acid	1.39		visit_occurrence concept count during day -180 through -4 concept_count relative to index: Telehealth	-3.95
age group: 20 - 29	-1.37		age group: 60 - 69	3.75
condition_era group during day -9999 through -4 days relative to index: Attention deficit hyperactivity disorder	-1.25		age group: 70 - 79	2.75
age group: 40 - 49	-1.24		visit_occurrence concept count during day -180 through -4 concept_count relative to index: Outpatient Visit	2.56
visit_occurrence concept count during day -180 through -4 concept_count relative to index: Office Visit	1.14		condition_era group during day -9999 through -4 days relative to index: Transient cerebral ischemia	1.95

US_Open_Claims - Janssen COVID-19 - BNT162b2 1st dose			**US_Open_Claims - Janssen COVID-19 - mRNA-1273 1st dose**	
index month: 1	-9.36		index month: 1	-8.60
index month: 2	-7.67		index month: 2	-7.62
visit_occurrence concept count during day -180 through -4 concept_count relative to index	5.51		visit_occurrence concept count during day -180 through -4 concept_count relative to index	3.91
visit_occurrence concept count during day -180 through -4 concept_count relative to index: Off Campus-Outpatient Hospital	-5.05		index month: 12	-3.84
index month: 12	-4.54		drug_era distinct concept count during day -180 through -4 concept_count relative to index	3.80
drug_era distinct concept count during day -180 through -4 concept_count relative to index	3.79		index year: 2020	-2.50
visit_occurrence concept count during day -180 through -4 concept_count relative to index: Rural Health Clinic	3.05		index month: 6	1.51
index year: 2020	-2.91		Charlson index - Romano adaptation	1.28
index month: 9	-1.95		observation distinct concept count during day -180 through -4 concept_count relative to index	1.21
Charlson index - Romano adaptation	1.69		index month: 9	0.94

Figure 1. Systematic error in meta-analysis

Figure 1a. Systematic error in the analysis of first-dose ChAdOx1 vs 1-dose BNT162b2.

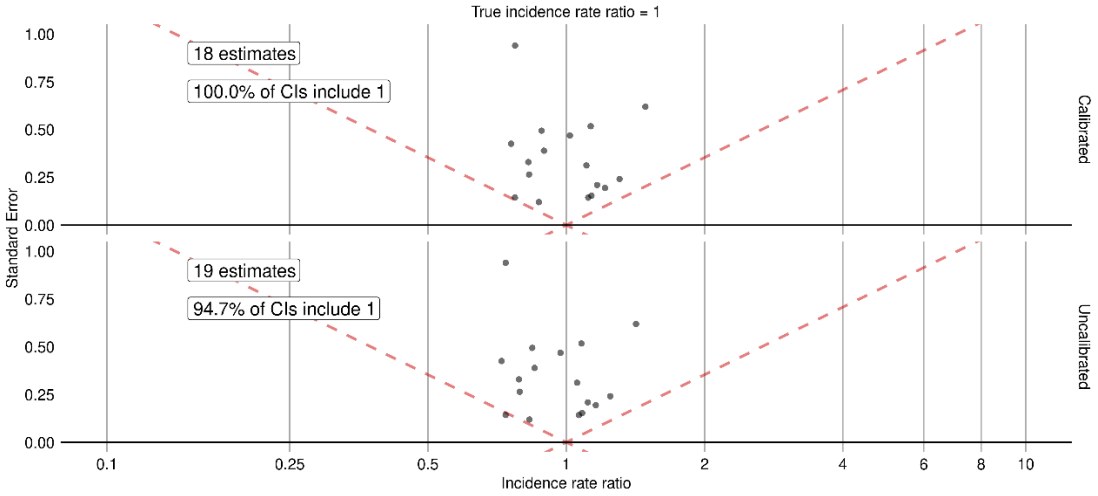


Figure 1b. Systematic error in the analysis of second-dose ChAdOx1 vs 1-dose BNT162b2.

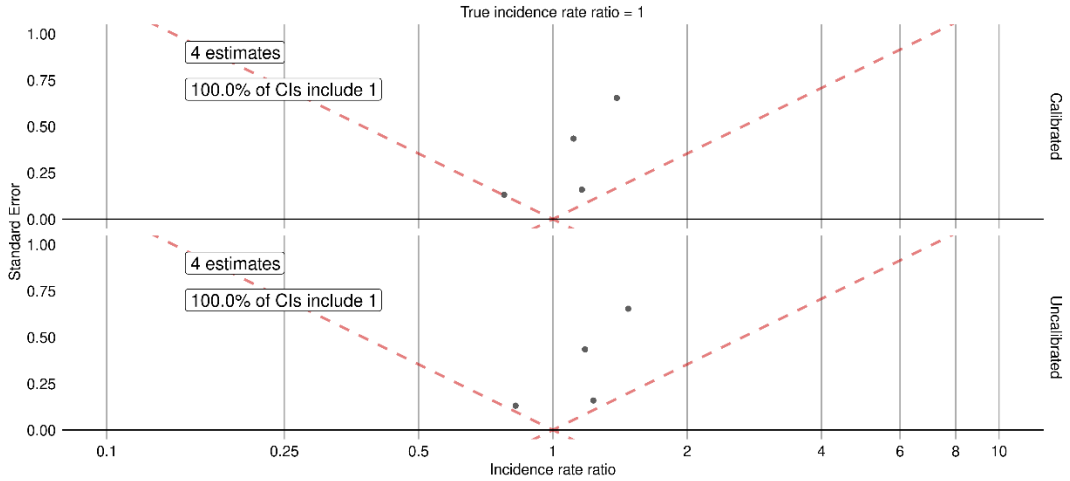
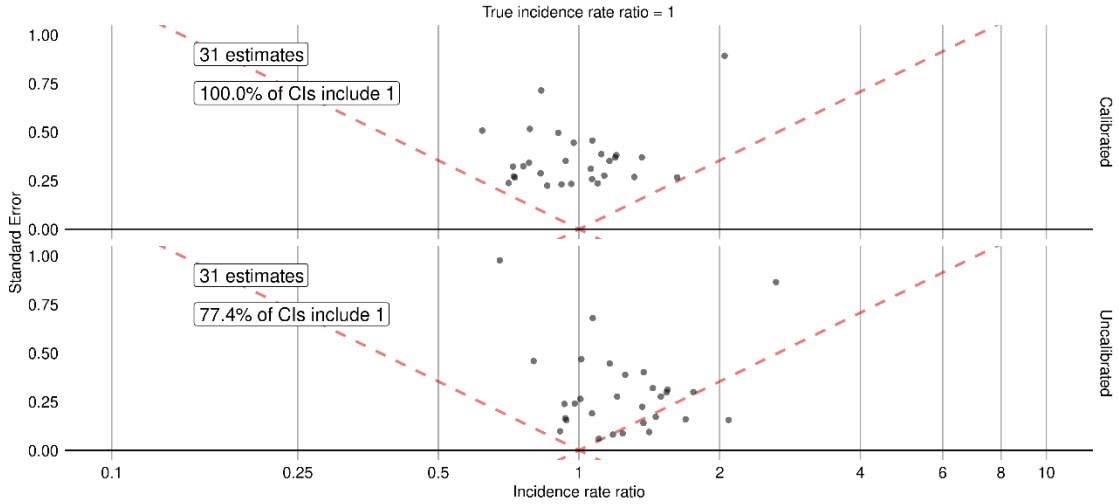


Figure 1c. Systematic error in the analysis of Ad26.COV2.S vs 1-dose BNT162



Online supplement

Chapter 5:

Aggregated data, analytical code, and detailed definitions of algorithms for identifying the events are available in a GitHub repository (https://github.com/ohdsi-studies/Covid19VaccineAesiIncidenceCharacterization). The results are available in an interactive web app (https://data.ohdsi.org/Covid19VaccineAesiIncidenceCharacterization).

Chapter 6:

All analytical code has been made available at https://github.com/SIDIAP/VaxAEsNeuroimmune

Chapter 7:

Patient level data cannot be shared without approval from data custodians owing to local information governance and data protection regulations. The analytical code is available at: https://github.com/oxford-pharmacoepi/ROC22_CovVaxComparativeSafety/tree/main/CovVaxComparativeSafety.

Shiny:

Patient level data cannot be shared without approval from data custodians owing to local information governance and data protection regulations. The analytical code is available at: https://github.com/oxford-pharmacoepi/ROC22_CovVaxComparativeSafety/tree/main/CovVaxComparativeSafety.

References

1 Polack FP, Thomas SJ, Kitchin N, *et al.* Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020;**383**:2603–15.

2 Baden LR, El Sahly HM, Essink B, *et al.* Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021;**384**:403–16.

3 Voysey M, Clemens SAC, Madhi SA, *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 2021;**397**:99–111.

4 Sadoff J, Gray G, Vandebosch A, *et al.* Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N Engl J Med* 2021;**384**:2187–201.

5 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;**12**:e1001885.

6 Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA* 2018;**320**:867–8.

7 Makady A, de Boer A, Hillege H, *et al.* What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health* 2017;**20**:858–65.

8 Berger ML, Sox H, Willke RJ, *et al.* Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE special Task Force on real-world evidence in health care decision making. *Value Health* 2017;**20**:1003–8.

9 Center for Devices, Radiological Health. Use of real-world evidence to support regulatory decision-making for medical devices. U.S. Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices (accessed 4 Oct 2022).

10 Li X, Lo Re V III, Toh S. Profiling real-world data sources for pharmacoepidemiologic research: A call for papers. *Pharmacoepidemiol Drug Saf* 2022;**31**:929–31.

11 Real-world evidence. U.S. Food and Drug Administration. 2023.https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence (accessed 10 Aug 2023).

12 Center for Drug Evaluation, Research. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. U.S. Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug (accessed 4 Oct 2022).

13 Franklin JM, Liaw K-L, Iyasu S, *et al.* Real-world evidence to support regulatory decision making: New or expanded medical product indications. *Pharmacoepidemiol Drug Saf* 2021;**30**:685–93.

14 Health Canada. Optimizing the Use of Real World Evidence to Inform Regulatory Decision-Making. 2019.https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html (accessed 13 Jul 2023).

15 MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions. GOV.UK. https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions (accessed 4 Oct 2022).

16 Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: Challenges and possible solutions for Europe. *Clin Pharmacol Ther* 2019;**106**:36–9.

17 Pottegård A, Kurz X, Moore N, *et al.* Considerations for pharmacoepidemiological analyses in the SARS-CoV-2 pandemic. *Pharmacoepidemiol Drug Saf* 2020;**29**:825–31.

18 Schneeweiss S, Brown JS, Bate A, *et al.* Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products. *Clin Pharmacol Ther* 2020;**107**:827–33.
19 Sentinel common data model. https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model (accessed 16 Oct 2023).
20 National Patient-Centered Clinical Research Network. The National Patient-Centered Clinical Research Network. 2020.https://pcornet.org/data (accessed 16 Oct 2023).
21 Murphy SN, Weber G, Mendis M, *et al.* Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;**17**:124–30.
22 Voss EA, Makadia R, Matcho A. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inf Assoc* 1093;**22**:553–64.
23 Observational Health Data Sciences, Informatics. The book of OHDSI. 2021.https://ohdsi.github.io/TheBookOfOhdsi/index.html (accessed 5 Oct 2022).
24 Hripcsak G, Duke JD, Shah NH, *et al.* Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. *Stud Health Technol Inform* 2015;**216**:574–8.
25 CDISC. https://www.cdisc.org (accessed 16 Oct 2023).
26 CNODES. https://www.cnodes.ca/cdm-overview (accessed 16 Oct 2023).
27 Garza M, Del Fiol G, Tenenbaum J, *et al.* Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform* 2016;**64**:333–41.
28 Ball P. The lightning-fast quest for COVID vaccines — and what it means for other diseases. *Nature* 2021;**589**:16–8.
29 Krammer F. SARS-CoV-2 vaccines in development. *Nature* 2020;**586**:516–27.
30 Sethi S, Kumar A, Mandal A, *et al.* The UPTAKE study: a cross-sectional survey examining the insights and beliefs of the UK population on COVID-19 vaccine uptake and hesitancy. *BMJ Open* 2021;**11**:e048856.
31 Watson OJ, Barnsley G, Toor J, *et al.* Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *Lancet Infect Dis* 2022;**22**:1293–302.
32 Schneider EC, Shah A, Sah P, *et al.* The U.s. covid-19 vaccination program at one year: How many deaths and hospitalizations were averted? 2021. doi:10.26099/3542-5N54
33 Lindquist M. VigiBase, the WHO global ICSR database system: Basic facts. *Drug Inf J* 2008;**42**:409–19.
34 EMA. EudraVigilance. European Medicines Agency. 2018.https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance (accessed 13 Sep 2023).
35 Center for Biologics Evaluation, Research. VAERS overview. U.S. Food and Drug Administration. 2021.https://www.fda.gov/vaccines-blood-biologics/vaccine-adverse-events/vaers-overview (accessed 13 Sep 2023).
36 Monitoring and responding to adverse events of special interest (aesis). https://www.who.int/docs/default-source/covid-19-vaccines-safety-surveillance-manual/covid19vaccines_manual_aesi.pdf (accessed 13 Sep 2023).
37 bc-coordinator. Priority list of adverse events of special interest: COVID-19. Brighton Collaboration. 2020.https://old.brightoncollaboration.us/priority-list-aesi-covid/ (accessed 13 Sep 2023).
38 Support C. COVID-19 AESI List 5th update – October 2022. Brighton Collaboration. 2022.https://brightoncollaboration.org/covid-19-aesi-list-5th-update-october-2022/ (accessed 27 Jan 2024).
39 Black S, Eskola J, Siegrist C-A, *et al.* Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. *Lancet* 2009;**374**:2115–22.

40 Mahaux O, Bauchau V, Van Holle L. Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines. *Pharmacoepidemiol Drug Saf* 2016;**25**:215–22.
41 Report of CIOMS Working Group on Vaccine Safety. CIOMS Guide to Active Vaccine Safety Surveillance. https://cioms.ch/wp-content/uploads/2020/04/240WEB-CIOMS-Guide-AVSS-20170202-protected.pdf (accessed 11 Mar 2021).
42 Center for Biologics Evaluation, Research. CBER biologics effectiveness and Safety (BEST) System. U.S. Food and Drug Administration. 2022.https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/cber-biologics-effectiveness-and-safety-best-system (accessed 13 Sep 2023).
43 vaccination coverage, vaccine safety, benefit monitoring. D. Report on appraisal of vaccine safety methods. https://vac4eu.org/wp-content/uploads/2019/03/ADVANCE_D4-2_appraisal-safety-methods_final_PU.pdf (accessed 13 Sep 2023).
44 Baker MA, Lieu TA, Li L, *et al.* A vaccine study design selection framework for the postlicensure rapid immunization safety monitoring program. *Am J Epidemiol* 2015;**181**:608–18.
45 Stone PW. Popping the (PICO) question in research and evidence-based practice. *Appl Nurs Res* 2002;**15**:197–8.
46 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;**335**:806–8.
47 Massari M, Spila-Alegiani S, Morciano C, *et al.* Post-marketing active surveillance of myocarditis and pericarditis following vaccination with COVID-19 mRNA vaccines in persons aged 12-39 years in Italy: a multi-database, self-controlled case series study. bioRxiv. 2022. doi:10.1101/2022.02.07.22270020
48 Lloyd PC, Hu M, Wong H-L, *et al.* Near real-time surveillance of safety outcomes in US COVID-19 vaccine recipients aged 12 to 64 years. *Vaccine* 2022;**40**:6481–8.
49 Wong H-L, Hu M, Zhou CK, *et al.* Risk of myocarditis and pericarditis after the COVID-19 mRNA vaccination in the USA: a cohort study in claims databases. *Lancet* 2022;**399**:2191–9.
50 Pottegård A, Lund LC, Karlstad Ø, *et al.* Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study. *BMJ* 2021;**373**:n1114.
51 Kerr S, Joy M, Torabi F, *et al.* First dose ChAdOx1 and BNT162b2 COVID-19 vaccinations and cerebral venous sinus thrombosis: A pooled self-controlled case series study of 11.6 million individuals in England, Scotland, and Wales. *PLoS Med* 2022;**19**:e1003927.
52 Li X, Raventós B, Roel E, *et al.* Association between covid-19 vaccination, SARS-CoV-2 infection, and risk of immune mediated neurological events: population based cohort and self-controlled case series analysis. *BMJ* 2022;**376**:e068373.
53 Hviid A, Hansen JV, Thiesson EM, *et al.* Association of AZD1222 and BNT162b2 COVID-19 vaccination with thromboembolic and thrombocytopenic events in frontline personnel : A retrospective cohort study. *Ann Intern Med* 2022;**175**:541–6.
54 Lai FTT, Li X, Peng K, *et al.* Carditis after COVID-19 vaccination with a messenger RNA vaccine and an inactivated virus vaccine : A case-control study. *Ann Intern Med* 2022;**175**:362–70.
55 Wan EYF, Chui CSL, Wang Y, *et al.* Herpes zoster related hospitalization after inactivated (CoronaVac) and mRNA (BNT162b2) SARS-CoV-2 vaccination: A self-controlled case series and nested case-control study. *Lancet Reg Health West Pac* 2022;**21**:100393.
56 Chui CSL, Fan M, Wan EYF, *et al.* Thromboembolic events and hemorrhagic stroke after mRNA (BNT162b2) and inactivated (CoronaVac) covid-19 vaccination: A self-controlled case series study. *EClinicalMedicine* 2022;**50**:101504.
57 Wong CKH, Mak LY, Au ICH, *et al.* Risk of acute liver injury following the mRNA (BNT162b2) and inactivated (CoronaVac) COVID-19 vaccines. *J Hepatol* 2022;**77**:1339–48.

58 Wong CKH, Lau KTK, Xiong X, *et al.* Adverse events of special interest and mortality following vaccination with mRNA (BNT162b2) and inactivated (CoronaVac) SARS-CoV-2 vaccines in Hong Kong: A retrospective study. *PLoS Med* 2022;**19**:e1004018.
59 Lai FTT, Huang L, Chui CSL, *et al.* Multimorbidity and adverse events of special interest associated with Covid-19 vaccines in Hong Kong. *Nat Commun* 2022;**13**:411.
60 Li X, Tong X, Yeung WWY, *et al.* Two-dose COVID-19 vaccination and possible arthritis flare among patients with rheumatoid arthritis in Hong Kong. *Ann Rheum Dis* 2022;**81**:564–8.
61 Lai FTT, Huang L, Peng K, *et al.* Post-Covid-19-vaccination adverse events and healthcare utilization among individuals with or without previous SARS-CoV-2 infection. *J Intern Med* 2022;**291**:864–9.
62 Ye X, Ma T, Blais JE, *et al.* Association between BNT162b2 or CoronaVac COVID-19 vaccines and major adverse cardiovascular events among individuals with cardiovascular disease. *Cardiovasc Res* 2022;**118**:2329–38.
63 Kang W, Shami JJP, Yan VKC, *et al.* Safety of two-dose COVID-19 vaccination (BNT162b2 and CoronaVac) in adults with cancer: a territory-wide cohort study. *J Hematol Oncol* 2022;**15**:66.
64 Wan EYF, Wang Y, Chui CSL, *et al.* Safety of an inactivated, whole-virion COVID-19 vaccine (CoronaVac) in people aged 60 years or older in Hong Kong: a modified self-controlled case series. *Lancet Healthy Longev* 2022;**3**:e491–500.
65 Xiong X, Wong CKH, Au ICH, *et al.* Safety of inactivated and mRNA COVID-19 vaccination among patients treated for hypothyroidism: A population-based cohort study. *Thyroid* 2022;**32**:505–14.
66 Wan EYF, Chui CSL, Lai FTT, *et al.* Bell's palsy following vaccination with mRNA (BNT162b2) and inactivated (CoronaVac) SARS-CoV-2 vaccines: a case series and nested case-control study. *Lancet Infect Dis* 2022;**22**:64–72.
67 Wong CKH, Lui DTW, Xiong X, *et al.* Risk of thyroid dysfunction associated with mRNA and inactivated COVID-19 vaccines: a population-based study of 2.3 million vaccine recipients. *BMC Med* 2022;**20**:339.
68 Wan EYF, Chui CSL, Mok AHY, *et al.* MRNA (BNT162b2) and inactivated (CoronaVac) COVID-19 vaccination and risk of adverse events and acute diabetic complications in patients with type 2 diabetes mellitus: A population-based study. *Drug Saf* 2022;**45**:1477–90.
69 Sing C-W, Tang CTL, Chui CSL, *et al.* COVID-19 vaccines and risks of hematological abnormalities: Nested case-control and self-controlled case series study. *Am J Hematol* 2022;**97**:470–80.
70 Li X, Gao L, Tong X, *et al.* Autoimmune conditions following mRNA (BNT162b2) and inactivated (CoronaVac) COVID-19 vaccination: A descriptive cohort study among 1.1 million vaccinated people in Hong Kong. *J Autoimmun* 2022;**130**:102830.
71 Lai FTT, Chua GT, Chan EWW, *et al.* Adverse events of special interest following the use of BNT162b2 in adolescents: a population-based retrospective cohort study. *Emerg Microbes Infect* 2022;**11**:885–93.
72 Lai FTT, Li X, Peng K, *et al.* Carditis after COVID-19 vaccination with a messenger RNA vaccine and an inactivated virus vaccine. *Ann Intern Med* 2022;**175**:362–70.
73 Baggs J, Gee J, Lewis E, *et al.* The Vaccine Safety Datalink: a model for monitoring immunization safety. *Pediatrics* 2011;**127 Suppl 1**:S45-53.
74 Dick A, Rosenbloom JI, Karavani G, *et al.* Safety of third SARS-CoV-2 vaccine (booster dose) during pregnancy. *Am J Obstet Gynecol MFM* 2022;**4**:100637.
75 Niesen MJM, Pawlowski C, O'Horo JC, *et al.* Surveillance of safety of 3 doses of COVID-19 mRNA vaccination using electronic health records. *JAMA Netw Open* 2022;**5**:e227038.
76 Simone A, Herald J, Chen A, *et al.* Acute myocarditis following a third dose of COVID-19 mRNA vaccination in adults. *Int J Cardiol* 2022;**365**:41–3.
77 McMurry R, Lenehan P, Awasthi S, *et al.* Real-time analysis of a mass vaccination effort confirms the safety of FDA-authorized mRNA COVID-19 vaccines. *Med (N Y)* 2021;**2**:965-978.e5.

78 Arnold BF, Ercumen A, Benjamin-Chung J, *et al.* Brief report: Negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology* 2016;**27**:637–41.
79 Schuemie MJ, Hripcsak G, Ryan PB, *et al.* Robust empirical calibration of p-values using observational data. Stat. Med. 2016;**35**:3883–8.
80 Schuemie MJ, Cepeda MS, Suchard MA, *et al.* How confident are we about observational findings in healthcare: A benchmark study. *Harv Data Sci Rev* 2020;**2**. doi:10.1162/99608f92.147cc28e
81 Lai LY, Arshad F, Areia C, *et al.* Current approaches to vaccine safety using observational data: A rationale for the EUMAEUS (Evaluating Use of methods for adverse events under surveillance-for vaccines) study design. *Front Pharmacol* 2022;**13**:837632.
82 Li X, Lai LY, Ostropolets A, *et al.* Bias, precision and timeliness of historical (background) rate comparison methods for vaccine safety monitoring: An empirical multi-database analysis. *Front Pharmacol* 2021;**12**:773875.
83 Schuemie MJ, Arshad F, Pratt N, *et al.* Vaccine safety surveillance using routinely collected healthcare data-an empirical evaluation of epidemiological designs. *Front Pharmacol* 2022;**13**:893484.
84 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016;**354**:i4515.
85 Whitaker HJ, Steer CD, Farrington CP. Self-controlled case series studies: Just how rare does a rare non-recurrent outcome need to be? *Biom J* 2018;**60**:1110–20.
86 Whitaker HJ, Farrington CP, Spiessens B, *et al.* Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;**25**:1768–97.
87 Sterne JAC, Hernán MA, Reeves BC, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.
88 Sterne JA, Higgins JP, Elbers RG, *et al.* Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance. 2016.https://www.bristol.ac.uk/media-library/sites/social-community-medicine/images/centres/cresyda/ROBINS-I_detailed_guidance.pdf (accessed 23 Jan 2024).
89 Chapter 25: Assessing risk of bias in a non-randomized study. https://training.cochrane.org/handbook/current/chapter-25 (accessed 23 Jan 2024).
90 Hernán MA, Robins JM. Causal Inference: What If. 2024.https://www.hsph.harvard.edu/miguel-hernan/wp-content/uploads/sites/1268/2024/01/hernanrobins_WhatIf_2jan24.pdf (accessed 23 Jan 2024).
91 Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep* 2015;**2**:221–8.
92 Sendor R, Stürmer T. Core concepts in pharmacoepidemiology: Confounding by indication and the role of active comparators. *Pharmacoepidemiol Drug Saf* 2022;**31**:261–9.
93 Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol* 2015;**11**:437–41.
94 Layton JB, McGrath LJ, Sahrmann JM, *et al.* Comparative safety of high-dose versus standard-dose influenza vaccination in patients with end-stage renal disease. *Vaccine* 2020;**38**:5178–86.
95 Richardson DM, Medvedeva EL, Roberts CB, *et al.* Comparative effectiveness of high-dose versus standard-dose influenza vaccination in community-dwelling veterans. *Clin Infect Dis* 2015;**61**:171–6.
96 Butler AM, Layton JB, Dharnidharka VR, *et al.* Comparative effectiveness of high-dose versus standard-dose influenza vaccine among patients receiving maintenance hemodialysis. *Am J Kidney Dis* 2020;**75**:72–83.

97 BBC News. Covid-19: Pfizer vaccine may not go to care home residents. BBC. 2020.https://www.bbc.co.uk/news/uk-england-northamptonshire-55158467 (accessed 23 Jan 2024).
98 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41–55.
99 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;**46**:399–424.
100 Franklin JM, Eddings W, Glynn RJ, *et al.* Regularized regression versus the high-dimensional propensity score for confounding adjustment in secondary database analyses. *Am J Epidemiol* 2015;**182**:651–9.
101 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018;**47**:2005–14.
102 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;**21**:383–8.
103 Desai JR, Hyde CL, Kabadi S, *et al.* Utilization of positive and negative controls to examine comorbid associations in observational database studies. *Med Care* 2017;**55**:244–51.
104 Gruber S, Tchetgen Tchetgen E. Limitations of empirical calibration of p-values using observational data. *Stat Med* 2016;**35**:3869–82.
105 Schuemie MJ, Hripcsak G, Ryan PB, *et al.* Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A* 2018;**115**:2571–7.
106 Dure Kim P, Nafie M, Yankah S. Understanding the use of negative controls to assess the validity of non-interventional studies of treatment using real-world evidence. Margolis Center for Health Policy. https://healthpolicy.duke.edu/events/understanding-use-negative-controls-assess-validity-non-interventional-studies-treatment (accessed 6 Jun 2023).
107 Vora P, Morgan Stewart H, Russell B, *et al.* Time trends and treatment pathways in prescribing individual oral anticoagulants in patients with nonvalvular atrial fibrillation: An observational study of more than three million patients from Europe and the United States. *Int J Clin Pract* 2022;**2022**:6707985.
108 Recalde M, Roel E, Pistillo A, *et al.* Characteristics and outcomes of 627 044 COVID-19 patients living with and without obesity in the United States, Spain, and the United Kingdom. *Int J Obes (Lond)* 2021;**45**:2347–57.
109 Lane JCE, Weaver J, Kostka K, *et al.* Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. *Lancet Rheumatol* 2020;**2**:e698–711.
110 Burn E, You SC, Sena AG, *et al.* Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. *Nat Commun* 2020;**11**:5009.
111 The NHS website. nhs.uk. https://www.nhs.uk/ (accessed 30 Mar 2023).
112 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36.
113 Wolf A, Dedman D, Campbell J, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019;**48**:1740–1740g.
114 Gallagher AM, Dedman D, Padmanabhan S, *et al.* The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf* 2019;**28**:563–9.
115 Delmestri A, Prieto-Alhambra D. CPRD GOLD and linked ONS mortality records: Reconciling guidelines. *Int J Med Inform* 2020;**136**:104038.
116 Benson T. The history of the Read Codes: the inaugural James Read Memorial Lecture 2011. *Inform Prim Care* 2011;**19**:173–82.
117 Home. SNOMED International. https://www.snomed.org/ (accessed 10 Aug 2023).

118 Recalde M, Rodríguez C, Burn E, *et al.* Data resource profile: The information system for research in primary care (SIDIAP). *Int J Epidemiol* 2022;**51**:e324–36.
119 García-Gil MDM, Hermosilla E, Prieto-Alhambra D, *et al.* Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). *Inform Prim Care* 2011;**19**:135–45.
120 Conjunt mínim bàsic de dades (CMBD). CatSalut. Servei Català de la Salut. https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/ (accessed 11 Apr 2023).
121 Jouaville SL, Miotti H, Coffin G, *et al.* Validity and limitations of the longitudinal patient database France for use in pharmacoepidemiological and pharmacoeconomics studies. *Value Health* 2015;**18**:A18.
122 Rathmann W, Bongaerts B, Carius H-J, *et al.* Basic characteristics and representativeness of the German Disease Analyzer database. *Int J Clin Pharmacol Ther* 2018;**56**:459–66.
123 Becher H, Kostev K, Schröder-Bernhardi D. Validity and representativeness of the "Disease Analyzer" patient database for use in pharmacoepidemiological and pharmacoeconomic studies. *Int J Clin Pharmacol Ther* 2009;**47**:617–26.
124 de Ridder MAJ, de Wilde M, de Ben C, *et al.* Data resource profile: The integrated primary care information (IPCI) database, the Netherlands. *Int J Epidemiol* 2022;**51**:e314–23.
125 Available IQVIA data. https://www.iqvia.com/insights/the-iqvia-institute/available-iqvia-data (accessed 26 Feb 2022).
126 Hall GC, Sauer B, Bourke A, *et al.* Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012;**21**:1–10.
127 Butler AM, Burcu M, Christian JB, *et al.* Noninterventional studies in the COVID-19 era: methodological considerations for study design and analysis. *J Clin Epidemiol* 2023;**153**:91–101.
128 *Assessing the Completeness of Medicare Claims Data for Measuring COVID-19 Vaccine Administration*.
129 Black S, Eskola J, Siegrist C-A, *et al.* Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. *Lancet* 2009;**374**:2115–22.
130 Wijnans L, Lecomte C, de Vries C, *et al.* The incidence of narcolepsy in Europe: before, during, and after the influenza A(H1N1)pdm09 pandemic and vaccination campaigns. *Vaccine* 2013;**31**:1246–54.
131 Black SB, Law B, Chen RT, *et al.* The critical role of background rates of possible adverse events in the assessment of COVID-19 vaccine safety. *Vaccine* 2021;**39**:2712–8.
132 Center for Biologics Evaluation and Research Office of Biostatistics and Epidemiology. CBER Surveillance Program Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring Protocol. https://www.bestinitiative.org/wp-content/uploads/2021/02/C19-Vaccine-Safety-AESI-Background-Rate-Protocol-FINAL-2020.pdf (accessed 11 Mar 2021).
133 bc-coordinator. Priority list of adverse events of special interest: COVID-19. 2020.https://brightoncollaboration.us/priority-list-aesi-covid/ (accessed 11 Mar 2021).
134 Arya DP, Said MA, Izurieta HS, *et al.* Surveillance for Guillain-Barré syndrome after 2015-2016 and 2016-2017 influenza vaccination of Medicare beneficiaries. *Vaccine* 2019;**37**:6543–9.
135 Samannodi M, Hansen M, Hasbun R. Lack of accuracy of the international classification of disease, ninth (ICD-9) codes in identifying patients with encephalitis. *J Neurol* 2019;**266**:1034–5.
136 Aboa-Eboulé C, Mengue D, Benzenine E, *et al.* How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol* 2013;**260**:605–13.
137 Chrischilles EA, Winiecki SK, Fireman B, *et al.* Mini-sentinel assessment protocol thromboembolic events after immunoglobulin administration. https://www.sentinelinitiative.org/sites/default/files/Drugs/Assessments/Mini-

Sentinel_Thromboembolic-Events-After-Immunoglobulin-Administration-Protocol_0.pdf (accessed 8 Jun 2023).
138 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.
139 Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* 2019;**22**:153–60.
140 Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d549.
141 The Council for International Organizations of Medical Sciences (CIOMS). *Guidelines for Preparing Core Clinical-Safety Information on Drugs Second Edition – Report of CIOMS Working Groups III and V*. Geneva: 1999.
142 Wise J. Covid-19: European countries suspend use of Oxford-AstraZeneca vaccine after reports of blood clots. *BMJ* 2021;**372**:n699.
143 *CoViD Vaccines and thrombotic events: Possibility of mRNA translation and spike protein synthesis by platelets?*
144 Mevorach D, Anis E, Cedar N, *et al.* Myocarditis after BNT162b2 mRNA vaccine against covid-19 in Israel. *N Engl J Med* 2021;**385**:2140–9.
145 Jabagi MJ, Botton J, Bertrand M, *et al.* Myocardial infarction, stroke, and pulmonary embolism after BNT162b2 mRNA COVID-19 vaccine in people aged 75 years or older. *JAMA* 2022;**327**:80–2.
146 Yih WK, Kulldorff M, Fireman BH, *et al.* Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. *Pediatrics* 2011;**127 Suppl 1**:S54-64.
147 Willame C, Dodd C, van der Aa L, *et al.* Incidence rates of autoimmune diseases in European healthcare databases: A contribution of the ADVANCE project. *Drug Saf* 2021;**44**:383–95.
148 Baker MA, Nguyen M, Cole DV, *et al.* Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. *Vaccine* 2013;**31 Suppl 10**:K98-112.
149 Hense S, Schink T, Kreisel SH, *et al.* Estimation of background incidence rates of Guillain-Barré syndrome in Germany - a retrospective cohort study with electronic healthcare data. *Neuroepidemiology* 2014;**43**:244–52.
150 Granieri E, Andreasi NG, De Martin P, *et al.* Incidence study of Guillain-Barré syndrome in the province of Ferrara, Northern Italy, between 2003 and 2017. A 40-year follow-up. *Neurol Sci* 2019;**40**:603–9.
151 Klein NP, Ray P, Carpenter D, *et al.* Rates of autoimmune diseases in Kaiser Permanente for use in vaccine adverse event safety studies. *Vaccine* 2010;**28**:1062–8.
152 Rasmussen TA, Jørgensen MRS, Bjerrum S, *et al.* Use of population based background rates of disease to assess vaccine safety in childhood and mass immunisation in Denmark: nationwide population based cohort study. *BMJ* 2012;**345**:e5823.
153 ACCESS: Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines. 2020.http://www.encepp.eu/encepp/viewResource.htm?id=37274 (accessed 11 Mar 2021).
154 Levison LS, Thomsen RW, Christensen DH, *et al.* Guillain-Barré syndrome in Denmark: validation of diagnostic codes and a population-based nationwide study of the incidence in a 30-year period. *Clin Epidemiol* 2019;**11**:275–83.
155 Nyberg F, Lindh M, Vanfleteren LEGW, *et al.* Adverse events of special interest for COVID-19 vaccines - background incidences vary by sex, age and time period and are affected by the pandemic. bioRxiv. 2021;:2021.10.04.21263507. doi:10.1101/2021.10.04.21263507
156 Gubernot D, Jazwa A, Niu M, *et al.* U.S. Population-Based background incidence rates of medical conditions for use in safety assessment of COVID-19 vaccines. *Vaccine* 2021;**39**:3666–77.
157 Tworkoski E, Zhou CK, Wong H-L, *et al.* Biologics effectiveness and safety (BEST) initiative background rates of adverse events of special interest for COVID-19 vaccine safety monitoring.

2021.https://bestinitiative.org/wp-content/uploads/2022/01/C19-Vax-Safety-AESI-Bkgd-Rate-Final-Report-2021.pdf (accessed 8 Jun 2023).
158 Pillsbury A, Phillips A, Deng L, *et al.* Background incidence rates of selected adverse events of special interest (AESI) to monitor the safety of COVID-19 vaccines. *Vaccine* 2023;**41**:3422–8.
159 Nasreen S, Calzavara A, Buchan SA, *et al.* Background incidence rates of adverse events of special interest related to COVID-19 vaccines in Ontario, Canada, 2015 to 2020, to inform COVID-19 vaccine safety surveillance. *Vaccine* 2022;**40**:3305–12.
160 Willame C, Dodd C, Gini R, *et al.* Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines. Zenodo 2021. doi:10.5281/ZENODO.5255870
161 Overhage JM, Ryan PB, Reich CG, *et al.* Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;**19**:54–60.
162 Dettori JR, Norvell DC, Chapman JR. Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine J* 2022;**12**:1624–6.
163 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**:n1435.
164 CBER Surveillance Program Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring Protocol. *Biostatistics and Epidemiology*
165 Knoll MD, Wonodi C. Oxford-AstraZeneca COVID-19 vaccine efficacy. Lancet. 2021;**397**:72–4.
166 Briefing Document. Pfizer-BioNTech COVID-19 Vaccine. In: *Vaccines and Related Biological Products Advisory Committee Meeting*.
167 Sadoff J, Le Gars M, Shukarev G, *et al.* Interim results of a phase 1-2a trial of Ad26.COV2.S Covid-19 vaccine. *N Engl J Med* 2021;**384**:1824–35.
168 Márquez Loza AM, Holroyd KB, Johnson SA, *et al.* Guillain- Barré syndrome in the placebo and active arms of a COVID-19 vaccine clinical trial: Temporal associations do not imply causality. *Neurology* 2021;:10.1212/WNL.0000000000011881.
169 *Phase III Double-blind, Placebo-controlled Study of AZD1222 for the Prevention of COVID-19 in Adults*.
170 Márquez Loza AM, Holroyd KB, Johnson SA, *et al.* Guillain-Barré syndrome in the placebo and active arms of a COVID-19 vaccine clinical trial. *Neurology* 2021;**96**:1052–4.
171 García-Grimshaw M, Michel-Chávez A, Vera-Zertuche JM, *et al.* Guillain-Barré syndrome is infrequent among recipients of the BNT162b2 mRNA COVID-19 vaccine. *Clin Immunol* 2021;**230**:108818.
172 Shasha D, Bareket R, Sikron FH, *et al.* Real-world safety data for the Pfizer BNT162b2 SARS-CoV-2 vaccine: historical cohort study. *Clin Microbiol Infect* 2022;**28**:130–4.
173 Burrows A, Bartholomew T, Rudd J, *et al.* Sequential contralateral facial nerve palsies following COVID-19 vaccination first and second doses. *BMJ Case Rep* 2021;**14**:e243829.
174 Obermann M, Krasniqi M, Ewers N, *et al.* Bell's palsy following COVID-19 vaccination with high CSF antibody response. *Neurol Sci* Published Online First: 29 July 2021. doi:10.1007/s10072-021-05496-5
175 Iftikhar H, Noor SMU, Masood M, *et al.* Bell's palsy after 24 hours of mRNA-1273 SARS-CoV-2 vaccine. *Cureus* 2021;**13**:e15935.
176 Colella G, Orlandi M, Cirillo N. Bell's palsy following COVID-19 vaccination. *J Neurol* Published Online First: 21 February 2021. doi:10.1007/s00415-021-10462-4
177 Gómez de Terreros Caro G, Gil Díaz S, Pérez Alé M, *et al.* Parálisis de Bell tras vacunación COVID-19: a propósito de un caso. *Neurologia* Published Online First: April 2021. doi:10.1016/j.nrl.2021.04.004

178 Martin-Villares C, Vazquez-Feito A, Gonzalez-Gimeno MJ, *et al.* Bell's palsy following a single dose of mRNA SARS-CoV-2 vaccine: a case report. *J Neurol* Published Online First: 25 May 2021. doi:10.1007/s00415-021-10617-3
179 Goss AL, Samudralwar RD, Das RR, *et al.* ANA investigates: Neurological complications of COVID-19 vaccines. *Ann Neurol* 2021;**89**:856–7.
180 Román GC, Gracia F, Torres A, *et al.* Acute transverse myelitis (ATM):Clinical review of 43 patients with COVID-19-associated ATM and 3 post-vaccination ATM serious adverse events with the ChAdOx1 nCoV-19 vaccine (AZD1222). *Front Immunol* 2021;**12**:653786.
181 Nishizawa Y, Hoshina Y, Baker V. Bell's palsy following the Ad26.COV2.S COVID-19 vaccination. *QJM* 2021;**114**:657–8.
182 Zuhorn F, Graf T, Klingebiel R, *et al.* Postvaccinal encephalitis after ChAdOx1 nCov-19. *Ann Neurol* 2021;**90**:506–11.
183 Vogrig A, Janes F, Gigli GL, *et al.* Acute disseminated encephalomyelitis after SARS-CoV-2 vaccination. *Clin Neurol Neurosurg* 2021;**208**:106839.
184 Sato K, Mano T, Niimi Y, *et al.* Facial nerve palsy following the administration of COVID-19 mRNA vaccines: analysis of a self-reporting database. *Int J Infect Dis* 2021;**111**:310–2.
185 Sharples LD, Kirkwood BR. Essentials of medical statistics. *J R Stat Soc Ser A Stat Soc* 1989;**152**:263.
186 Suchard MA, Simpson SE, Zorych I, *et al.* Massive parallelization of serial inference algorithms for a complex generalized linear model. *ACM Trans Model Comput Simul* 2013;**23**:1–17.
187 Musonda P, Farrington CP, Whitaker HJ. Sample sizes for self-controlled case series studies. *Stat Med* 2006;**25**:2618–31.
188 Fonseca-Rodríguez O, Fors Connolly A-M, Katsoularis I, *et al.* Avoiding bias in self-controlled case series studies of coronavirus disease 2019. *Stat Med* 2021;**40**:6197–208.
189 Renoud L, Khouri C, Revol B, *et al.* Association of facial paralysis with mRNA COVID-19 vaccines: A disproportionality analysis using the world health organization pharmacovigilance database. *JAMA Intern Med* 2021;**181**:1243–5.
190 Wan EYF, Chui CSL, Ng VWS, *et al.* Messenger RNA Coronavirus disease 2019 (COVID-19) vaccination with BNT162b2 increased risk of Bell's palsy: A nested case-control and self-controlled case series study. *Clin Infect Dis* 2023;**76**:e291–8.
191 Barda N, Dagan N, Ben-Shlomo Y, *et al.* Safety of the BNT162b2 mRNA covid-19 vaccine in a nationwide setting. *N Engl J Med* 2021;**385**:1078–90.
192 Rafati A, Pasebani Y, Jameie M, *et al.* Bell's palsy following SARS-CoV-2 vaccines: A systematic review and meta-analysis. bioRxiv. 2022;:2022.10.26.22281537. doi:10.1101/2022.10.26.22281537
193 Patone M, Handunnetthi L, Saatci D, *et al.* Neurological complications after first dose of COVID-19 vaccines and SARS-CoV-2 infection. *Nat Med* 2021;**27**:2144–53.
194 Walker JL, Schultze A, Tazare J, *et al.* Safety of COVID-19 vaccination and acute neurological events: A self-controlled case series in England using the OpenSAFELY platform. *Vaccine* 2022;**40**:4479–87.
195 Woo EJ, Mba-Jonas A, Dimova RB, *et al.* Association of receipt of the Ad26.COV2.S COVID-19 vaccine with presumptive Guillain-Barré syndrome, February-July 2021. *JAMA* 2021;**326**:1606–13.
196 Paterson RW, Brown RL, Benjamin L, *et al.* The emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory findings. *Brain* 2020;**143**:3104–20.
197 Varatharaj A, Thomas N, Ellul MA, *et al.* Neurological and neuropsychiatric complications of COVID-19 in 153 patients: a UK-wide surveillance study. *Lancet Psychiatry* 2020;**7**:875–82.
198 Wijnans L, Dodd CN, Weibel D, *et al.* Bell's palsy and influenza(H1N1)pdm09 containing vaccines: A self-controlled case series. *PLoS One* 2017;**12**:e0175539.

199 Montplaisir J, Petit D, Quinn M-J, *et al.* Risk of narcolepsy associated with inactivated adjuvanted (AS03) A/H1N1 (2009) pandemic influenza vaccine in Quebec. *PLoS One* 2014;**9**:e108489.
200 Dodd CN, Romio SA, Black S, *et al.* International collaboration to assess the risk of Guillain Barré Syndrome following Influenza A (H1N1) 2009 monovalent vaccines. *Vaccine* 2013;**31**:4448–58.
201 Stowe J, Andrews N, Wise L, *et al.* Investigation of the temporal association of Guillain-Barre syndrome with influenza vaccine and influenzalike illness using the United Kingdom General Practice Research Database. *Am J Epidemiol* 2009;**169**:382–8.
202 Webb AJS, Brain SAE, Wood R, *et al.* Seasonal variation in Guillain-Barré syndrome: a systematic review, meta-analysis and Oxfordshire cohort study. *J Neurol Neurosurg Psychiatry* 2015;**86**:1196–201.
203 Wouk J, Rechenchoski DZ, Rodrigues BCD, *et al.* Viral infections and their relationship to neurological disorders. *Arch Virol* 2021;**166**:733–53.
204 Schultz NH, Sørvoll IH, Michelsen AE, *et al.* Thrombosis and thrombocytopenia after ChAdOx1 nCoV-19 vaccination. *N Engl J Med* Published Online First: 9 April 2021. doi:10.1056/NEJMoa2104882
205 Greinacher A, Thiele T, Warkentin TE, *et al.* Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *N Engl J Med* Published Online First: 9 April 2021. doi:10.1056/NEJMoa2104840
206 Cines DB, Bussel JB. SARS-CoV-2 vaccine-induced immune thrombotic thrombocytopenia. N. Engl. J. Med. 2021;**384**:2254–6.
207 Lee E-J, Cines DB, Gernsheimer T, *et al.* Thrombocytopenia following Pfizer and Moderna SARS-CoV-2 vaccination. *Am J Hematol* 2021;**96**:534–7.
208 EMA. Press briefing on the conclusion of the assessment of the Pharmacovigilance Risk Assessment Committee (PRAC) of Vaxzevria (previously COVID-19 Vaccine AstraZeneca) and thromboembolic events. European Medicines Agency. 2021.https://www.ema.europa.eu/en/events/press-briefing-conclusion-assessment-pharmacovigilance-risk-assessment-committee-prac-vaxzevria (accessed 14 Aug 2023).
209 Denmark continues its vaccine rollout without the COVID-19 vaccine from AstraZeneca. https://www.sst.dk/en/english/news/2021/denmark-continues-its-vaccine-rollout-without-the-covid-19-vaccine-from-astrazeneca (accessed 9 Jun 2023).
210 Scully M, Singh D, Lown R, *et al.* Pathologic antibodies to platelet factor 4 after ChAdOx1 nCoV-19 vaccination. *N Engl J Med* 2021;**384**:2202–11.
211 Hippisley-Cox J, Patone M, Mei XW, *et al.* Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: self-controlled case series study. *BMJ* 2021;**374**:n1931.
212 Simpson CR, Shi T, Vasileiou E, *et al.* First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland. *Nat Med* 2021;:1–8.
213 Chen RT, Monash JB. November 11 v 2b DRAFT updated Brighton collaboration case definition for thrombosis with thrombocytopenia syndrome (TTS). https://brightoncollaboration.us/wp-content/uploads/2021/11/TTS-Updated-Brighton-Collaboration-Case-Defintion-Draft-Nov-11-2021.pdf (accessed 26 Feb 2022).
214 Burn E, Li X, Kostka K, *et al.* Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. *Pharmacoepidemiol Drug Saf* Published Online First: 22 February 2022. doi:10.1002/pds.5419
215 Schuemie MJ, Arshad F, Pratt N, *et al.* Vaccine safety surveillance using routinely collected healthcare data – An empirical evaluation of epidemiological designs. bioRxiv. 2021;:2021.08.09.21261780. doi:10.1101/2021.08.09.21261780

216 Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* 1993;**46**:1075–9; discussion 1081-90.
217 Lip GYH, Nieuwlaat R, Pisters R, *et al.* Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;**137**:263–72.
218 World Health Organization(WHO). *Sample size determination in health studies*. Genève, Switzerland: : World Health Organization 1991.
219 Schuemie MJ, Ryan PB, DuMouchel W, *et al.* Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014;**33**:209–18.
220 McGrath LJ, Spangler L, Curtis JR, *et al.* Using negative control outcomes to assess the comparability of treatment groups among women with osteoporosis in the United States. *Pharmacoepidemiol Drug Saf* 2020;**29**:854–63.
221 Morales DR, Conover MM, You SC, *et al.* Renin-angiotensin system blockers and susceptibility to COVID-19: an international, open science, cohort analysis. Lancet Digit Health. 2021;**3**:e98–114.
222 European Medicines Agency: Amsterdam, the Neherlands. ENCePP Guide on Methodological Standards in Pharmacoepidemiology (Version 9). https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml (accessed 27 Feb 2022).
223 Burn E, Roel E, Pistillo A, *et al.* Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain. *Nat Commun* 2022;**13**:7169.
224 Burn E, Li X, Delmestri A, *et al.* Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in the United Kingdom. *Nat Commun* 2022;**13**:7167.
225 Barda N, Dagan N, Ben-Shlomo Y, *et al.* Safety of the BNT162b2 mRNA covid-19 vaccine in a nationwide setting. *N Engl J Med* 2021;**385**:1078–90.
226 Klein NP, Lewis N, Goddard K, *et al.* Surveillance for adverse events after COVID-19 mRNA vaccination. *JAMA* 2021;**326**:1390–9.
227 Hafeez MU, Ikram M, Shafiq Z, *et al.* COVID-19 vaccine-associated thrombosis with thrombocytopenia syndrome (TTS): A systematic review and post hoc analysis. *Clin Appl Thromb Hemost* 2021;**27**:10760296211048816.
228 Waqar U, Ahmed S, Gardezi SMHA, *et al.* Thrombosis with thrombocytopenia syndrome after administration of AZD1222 or Ad26.COV2.S vaccine for COVID-19: A systematic review. *Clin Appl Thromb Hemost* 2021;**27**:10760296211068488.
229 See I, Lale A, Marquez P, *et al.* Case series of thrombosis with thrombocytopenia syndrome after COVID-19 vaccination-United States, December 2020 to August 2021. *Ann Intern Med* Published Online First: 18 January 2022. doi:10.7326/M21-4502
230 Sodhi M, Samii A, Etminan M. A comparative safety study of reported neurological adverse events with three COVID-19 vaccines. *J Neurol* Published Online First: 9 January 2022. doi:10.1007/s00415-021-10919-6
231 Schuemie MJ, Ryan PB, Man KKC, *et al.* A plea to stop using the case-control design in retrospective database studies. *Stat Med* 2019;**38**:4199–208.
232 Sedgwick P. Bias in observational study designs: case-control studies. *BMJ* 2015;**350**:h560.
233 Andrews NJ, Stowe J, Ramsay ME, *et al.* Risk of venous thrombotic events and thrombocytopenia in sequential time periods after ChAdOx1 and BNT162b2 COVID-19 vaccines: A national cohort study in England. *Lancet Reg Health Eur* 2022;**13**:100260.
234 Dickerman BA, Madenci AL, Gerlovin H, *et al.* Comparative safety of BNT162b2 and mRNA-1273 vaccines in a nationwide cohort of US veterans. *JAMA Intern Med* 2022;**182**:739–46.
235 Dickerman BA, Gerlovin H, Madenci AL, *et al.* Comparative effectiveness of BNT162b2 and mRNA-1273 vaccines in U.s. veterans. *N Engl J Med* 2022;**386**:105–15.
236 Schuemie MJ. *Vaccine safety surveillance using routinely collected healthcare data - An empirical evaluation of epidemiological designs. medRxiv*.

237 Wood A, Denholm R, Hollings S, *et al.* Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ* 2021;**373**:n826.
238 6. Results and findings. NDRS. https://digital.nhs.uk/data-and-information/publications/statistical/coronavirus-as-recorded-in-primary-care/march-2020-21/results-and-findings (accessed 24 Jun 2023).
239 Amit S, Beni SA, Biber A, *et al.* Postvaccination COVID-19 among Healthcare Workers, Israel. *Emerg Infect Dis* 2021;**27**:1220–2.
240 Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 2011;**22**:42–52.
241 Hwang H, Quiroz JC, Gallego B. Assessing the effectiveness of empirical calibration under different bias scenarios. *BMC Med Res Methodol* 2022;**22**. doi:10.1186/s12874-022-01687-6
242 Ostropolets A, Li X, Makadia R, *et al.* Factors influencing background incidence rate calculation: Systematic empirical evaluation across an international network of observational databases. *Front Pharmacol* 2022;**13**:814198.
243 Li F, Thomas LE. Addressing extreme propensity scores via the overlap weights. *Am J Epidemiol* 2018;**188**:250–7.
244 Thomas LE, Li F, Pencina MJ. Overlap weighting. *JAMA* 2020;**323**:2417.
245 Herrett E, Thomas SL, Schoonen WM, *et al.* Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;**69**:4–14.
246 https://cioms.ch/wp-content/uploads/2018/03/Guidelines-for-Preparing-Core-Clinical-Safety-Info-Drugs-Report-of-CIOMS-Working-Group-III-and-V.pdf (Accessed 14 Jun 2023).
247 COVID-19 vaccination: information for healthcare practitioners. GOV.UK. 2020.https://www.gov.uk/government/publications/covid-19-vaccination-programme-guidance-for-healthcare-practitioners (accessed 27 Jan 2024).
248 COVID-19: the green book, chapter 14a. GOV.UK. 2023.https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a (accessed 15 Jun 2023).
249 Peter RS, Nieters A, Kräusslich H-G, *et al.* Post-acute sequelae of covid-19 six to 12 months after infection: population based study. *BMJ* 2022;**379**:e071050.
250 Wyber R, Vaillancourt S, Perry W, *et al.* Big data in global health: improving health in low- and middle-income countries. *Bull World Health Organ* 2015;**93**:203–8.
251 WHO – COVID19 vaccine tracker. https://covid19.trackvaccines.org/agency/who/ (accessed 3 Aug 2023).
252 Zhang J, Symons J, Agapow P, *et al.* Best practices in the real-world data life cycle. *PLOS Digit Health* 2022;**1**:e0000003.
253 McNair D, Lumpkin M, Kern S, *et al.* Use of RWE to inform regulatory, public health policy, and intervention priorities for the developing world. *Clin Pharmacol Ther* 2022;**111**:44–51.
254 UK COVID-19 vaccines delivery plan. GOV.UK. https://www.gov.uk/government/publications/uk-covid-19-vaccines-delivery-plan/uk-covid-19-vaccines-delivery-plan (accessed 19 Jun 2023).