Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases

Voss, Erica A.; Shoaibi, Azza; Yin Hui Lai, Lana; Blacketer, Clair; Alshammari, Thamir

DOI:
10.1016/j.eclinm.2023.101932

Publication date:
2023

Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study

Erica A. Voss,[a,b,c,*] Azza Shoaibi,[a,c] Lana Yin Hui Lai,[a,d] Clair Blacketer,[a,b,c] Thamir Alshammari,[a,e] Rupa Makadia,[a,c] Kevin Haynes,[c] Anthony G. Sena,[a,b,c] Gowtham Rao,[a,c] Sebastiaan van Sandijk,[a,f] Clement Fraboulet,[g] Laurent Boyer,[g] Tanguy Le Carrour,[h] Scott Horban,[i] Daniel R. Morales,[j,k] Jordi Martínez Roldán,[l] Juan Manuel Ramírez-Anguita,[m,n] Miguel A. Mayer,[m,n] Marcel de Wilde,[a,b] Luis H. John,[a,b] Talita Duarte-Salles,[a,o] Elena Roel,[o] Andrea Pistillo,[o] Raivo Kolde,[p] Filip Maljković,[q] Spiros Denaxas,[r,s,t] Vaclav Papez,[r,s] Michael G. Kahn,[a,u] Karthik Natarajan,[a,v,w] Christian Reich,[a,x] Alex Secora,[x] Evan P. Minty,[a,y] Nigam H. Shah,[a,z] Jose D. Posada,[a,aa] Maria Teresa Garcia Morales,[ab] Diego Bosca,[ao] Honorio Cadenas Juanino,[ac] Antonio Diaz Holgado,[ac] Miguel Pedrera Jiménez,[ap] Pablo Serrano Balazote,[ap] Noelia García Barrio,[ap] Selçuk Şen,[ad] Ali Yağız Üresin,[ad] Baris Erdogan,[ae] Luc Belmans,[af] Geert Byttebier,[af] Manu L. N. G. Malbrain,[af,ag] Daniel J. Dedman,[ah] Zara Cuccu,[ah] Rohit Vashisht,[a,ai] Atul J. Butte,[a,ai,aj] Ayan Patel,[a,ai] Lisa Dahm,[a,aj] Cora Han,[a,aj] Fan Bu,[ak] Faaizah Arshad,[a,ak] Anna Ostropolets,[a,v] Fredrik Nyberg,[al] George Hripcsak,[a,v,w] Marc A. Suchard,[a,ak,am] Dani Prieto-Alhambra,[a,b,an] Peter R. Rijnbeek,[a,b] Martijn J. Schuemie,[a,c,ak] and Patrick B. Ryan[a,c,v]

[a]OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA
[b]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands
[c]Janssen Pharmaceutical Research and Development LLC, Titusville, NJ, USA
[d]Department of Informatics, Imaging & Data Sciences, University of Manchester, Manchester, UK
[e]College of Pharmacy, Prince Sattam Bin Abdulaziz University, Alkharj, Kingdom of Saudi Arabia
[f]Odysseus Data Services, Prague, Czechia
[g]Department of Medical Information, Assistance Publique - Hopitaux de Marseille, Marseille, France
[h]Easter Eggs, Paris, France
[i]Health Informatics Centre, University of Dundee, Dundee, UK
[j]Division of Population Health and Genomics, University of Dundee, Dundee, UK
[k]Department of Public Health, University of Southern Denmark, Odense, Denmark
[l]Hospital del Mar, Department of Innovation and Digital Transformation, Barcelona, Spain
[m]Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain
[n]Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[o]Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol I Gurina (IDIAPJGol), Barcelona, Spain
[p]University of Tartu, Tartu, Estonia
[q]Heliant, Belgrade, Serbia
[r]University College London, Institute of Health Informatics, London, UK
[s]Health Data Research UK, London, UK
[t]British Heart Foundation Data Science Centre, London, UK
[u]University of Colorado Anschutz Medical Campus, Denver, CO, USA
[v]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA
[w]New York-Presbyterian Hospital, New York, NY, USA
[x]IQVIA, Falls Church, VA, USA
[y]O'Brien Institute for Public Health, Faculty of Medicine, University of Calgary, Alberta, CA, USA
[z]Stanford School of Medicine, Stanford University, Palo Alto, CA, USA
[aa]Department of Systems Engineering, School of Engineering, Universidad del Norte, Barranquilla, Colombia
[ab]Instituto de Investigación, Hospital 12 de Octubre, Madrid, Spain
[ac]Fundación para la Investigación e Innovación Biosanitaria en Atención Primaria (FIIBAP), Madrid, Spain
[ad]Division of Clinical Pharmacology, Department of Medical Pharmacology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Türkiye
[ae]Clinerion, Basel, Switzerland
[af]Medaman BV, Geel, Flanders, Belgium
[ag]1st Department Anaesthesiology and Intensive Therapy, Professor Critical Care Research, Medical University Lublin, Lublin, Poland
[ah]Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK
[ai]Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, USA

Abbreviations: AESI, adverse events of special interest; OHDSI, Observational Health Data Sciences and Informatics; OMOP, Observational Medical Outcomes Partnership; CDM, Common Data Model

*Corresponding author. Janssen Research & Development, LLC, 920 Route 202, Raritan, NJ, 08869, USA.
E-mail address: evoss3@its.jnj.com (E.A. Voss).

[aj]Center for Data-driven Insights and Innovation, University of California Health, Oakland, CA, USA
[ak]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA
[al]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[am]VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, UT, USA
[an]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, UK
[ao]VeraTech for Health, Valencia, Spain
[ap]Hospital Universitario 12 de Octubre, Madrid, Spain

eClinicalMedicine 2023;58: 101932
Published Online xxx
https://doi.org/10.1016/j.eclinm.2023.101932

Summary

Background Adverse events of special interest (AESIs) were pre-specified to be monitored for the COVID-19 vaccines. Some AESIs are not only associated with the vaccines, but with COVID-19. Our aim was to characterise the incidence rates of AESIs following SARS-CoV-2 infection in patients and compare these to historical rates in the general population.

Methods A multi-national cohort study with data from primary care, electronic health records, and insurance claims mapped to a common data model. This study's evidence was collected between Jan 1, 2017 and the conclusion of each database (which ranged from Jul 2020 to May 2022). The 16 pre-specified prevalent AESIs were: acute myocardial infarction, anaphylaxis, appendicitis, Bell's palsy, deep vein thrombosis, disseminated intravascular coagulation, encephalomyelitis, Guillain-Barré syndrome, haemorrhagic stroke, non-haemorrhagic stroke, immune thrombocytopenia, myocarditis/pericarditis, narcolepsy, pulmonary embolism, transverse myelitis, and thrombosis with thrombocytopenia. Age-sex standardised incidence rate ratios (SIR) were estimated to compare post-COVID-19 to pre-pandemic rates in each of the databases.

Findings Substantial heterogeneity by age was seen for AESI rates, with some clearly increasing with age but others following the opposite trend. Similarly, differences were also observed across databases for same health outcome and age-sex strata. All studied AESIs appeared consistently more common in the post-COVID-19 compared to the historical cohorts, with related meta-analytic SIRs ranging from 1.32 (1.05 to 1.66) for narcolepsy to 11.70 (10.10 to 13.70) for pulmonary embolism.

Interpretation Our findings suggest all AESIs are more common after COVID-19 than in the general population. Thromboembolic events were particularly common, and over 10-fold more so. More research is needed to contextualise post-COVID-19 complications in the longer term.

Funding None.

Keywords: COVID-19; Observational research; OMOP CDM; Adverse events of special interest

Research in context

Evidence before this study

During the rollout of the COVID-19 vaccinations, regulatory authorities paid special attention to a pre-specified list of adverse events of particular concern (AESIs). Some of the pre-specified AESIs have since been related with vaccinations. Not only are some AESIs potentially related with COVID-19 vaccinations, but also with the virus itself. To comprehend the benefit-risk of COVID-19 vaccinations, we must be aware of the projected COVID-19 infection rates. A literature search was conducted up until March 31, 2022 in PubMed to determine what evidence has previously been published on the occurrence of AESI after COVID-19 infection. Included in the search criteria were "COVID-19" and each of the 16 AESIs evaluated (the original 15 AESIs plus thrombosis with thrombocytopenia). There were 63 publications discovered and evaluated. However, the investigations were conducted on a single database or a small number of databases, and the majority of the papers concerned AESIs after immunization, with no mention of post-COVID occurrences.

Added value of this study

We demonstrate that all 16 studied AESIs are more common after COVID-19 than expected in the general population. Thromboembolic events were particularly common, and over 10-fold more so. These findings highlight the need for more research to contextualize post-COVID-19 complications in the longer term.

Implications of all the available evidence

Our study shows considerable database heterogeneity across the AESIs, suggesting individual study estimates should be interpreted with caution. Comparing the 'Subjects with COVID-19' to the 'Pre-Pandemic Background Population' shows a fairly consistent elevated rate in experiencing an AESI within 90-days after index. The results of this work are of public health importance as they help to put into perspective the risk of the AESIs post vaccination versus post SARS-CoV-2 infection.

Introduction

Since coronavirus disease 2019 (COVID-19) was first reported until July 2022, over 554 million cases and 6.3 million deaths have been reported worldwide.[1] COVID-19 vaccines have demonstrated to reduce COVID-19 hospitalisations and deaths, both in randomised controlled trials and in real-world observational studies.[2,3] As of 30th May 2022, more than 5 billion people (approximately 67% of the global population) have received at least one dose of a COVID-19 vaccine.[4] At least 39 vaccines have been authorised or approved for use by at least one country, of which 11 have been granted emergency use by the World Health Organization (WHO). Although vaccine safety has been rigorously monitored in clinical trials, rare adverse events can go undetected during vaccine development due to limited number of trial participants, a relatively short follow-up duration, and insufficient generalisability to the broader population.

Throughout the roll-out of the COVID-19 vaccines, adverse events of special interest (AESIs) have been prespecified to be monitored by medicines regulatory authorities. To provide context for these safety investigations, Li et al.,[5] in a multinational network cohort study, reported heterogeneity in the background incidence rates of 15 AESIs across age and sex stratification as well as various administrative claims and electronic health records (EHR) databases. Against this backdrop, some vaccine-related adverse events have gone on to be associated with the vaccines; two within the originally pre-specified 15 AESIs (myocarditis and pericarditis[6] and Guillain-Barré syndrome (GBS)[7]), and one newly-identified AESI (thrombosis with thrombocytopenia syndrome[8]).

It is, however, important to note that some AESIs are not only potentially associated with the COVID-19 vaccines, but with COVID-19 itself. And in order to understand the benefit risk of COVID-19 vaccines it is important to contextualise this with expected rates following COVID-19. To address this problem, the Observational Health Data Sciences and Informatics (OHDSI)[9] community carried out a network study using observational data from 26 databases across 11 countries with the aim of estimating the incidence rates of 16 vaccine related AESIs among people who had COVID-19, as compared to the background population during the pre-pandemic phase.

Methods

Data sources

Rates were obtained from 26 databases, which included 8 administrative claims databases, 12 EHRs, 1 EHR with a registry, and 5 general practitioner (GP) databases. These databases represented 11 countries: Belgium, Estonia, France, Germany, Japan, the Netherlands, Serbia, Spain, Turkey, the United Kingdom (UK), and the United States of America (US). All of these databases represent subsets of the total population from which they originate.

The administrative claims databases were: IBM® MarketScan® Commercial Claims and Encounters Database (IBM_CCAE); IBM® MarketScan® Multi-State Medicaid Database (IBM_MDCD); IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database (IBM_MDCR); IQVIA LRxDX Open Claims (IQVIA_OPENCLAIMS); IQVIA PharMetrics Plus (IQVIA_PharMetrics); JMDC[10]; Optum De-Identified Clinformatics® Data Mart Database–Socio-Economic Status (SES) (OPTUM_SES); and University of Tartu (U_OF_TARTU). The EHR systems were: Health Data Warehouse of Assistance Publique–Hopitaux de Marseille (APHM); University of Colorado Anschutz Medical Campus- Health Data Compass (CU_AMC); Columbia University Irving Medical Center (CUIMC); Fundación para la Investigación e Innovación Biosanitaria en Atención Primaria COVID-19 (FIIBAP); Health Informatics Centre (HIC); Parc de Salut Mar Barcelona Information System (IMASIS); Istanbul Faculty of Medicine, Istanbul University (IU); Medaman Hospital Data (MHD); Optum® de-identified Electronic Health Record Dataset (OPTUM_EHR); STAnford medicine Research data Repository-OMOP (STARR)[11]; University Clinical Center of Serbia (UCCS); and the University of California Health Data Warehouse (UCHDW). The EHR paired with a registry was the UK Biobank (UK_BIOBANK).[12] Finally, the GP databases were: Clinical Practice Research Datalink AURUM (CPRD_AURUM); Integrated Primary Care Information (IPCI)[13]; IQVIA Disease Analyzer France (IQVIA_FRANCE_DA)[14]; IQVIA Disease Analyzer Germany (IQVIA_GERMAN_DA)[14]; and The Information System for Research on Primary Care (SIDIAP).[15]

Table 1 provides a high-level overview of the databases. All datasets were mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) V5.3.1 or higher[9] which is maintained by OHDSI. See Appendix 1 for database details.

The use of APHM was approved by the health data access commission (CADS). The use of CPRD_AURUM was approved by the Research Data Governance Process, protocol 20_000211. The use of IPCI database was approved by its governance board (2020-04). For UK_BIOBANK, ethical approval was provided from the UKB Access Review Board, reference 58,356 "Defining and redefining human disease at scale: an atlas of the human phenome." COVID-19 research using UK Biobank data has been outlined here: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/covid-19-hub. Participant data for this project are available directly from the UKB following a protocol review and contractual agreements, more information can be found on the UKB website and participants that had withdrawn consent were excluded from the study. The use of de-identified STARR-OMOP data and contribution of aggregate results has been approved (IRB Panel #8 RB-53248, registered to Lelan Stanford Junior University) under the Stanford Human

Name	Country	Data provenance	Dates covered YYYY/MM	Total persons	% F	Ages covered mean age (IQR)	Patients with COVID-19 #	Pre-pandemic background population #	Measurements with values
APHM	FR	EHR	1998/11–2021/07	2,465,265	51.9	40 (21–60)	11,431	673,031	N
CPRD_AURUM	UK	GP	1995/01–2021/03	39,879,547	51.8	31 (18–44)	587,886	14,094,032	Y
CU_AMC	US	EHR	2011/01–2022/04	4,795,392	54.0	39 (22–56)	72,648	830,579	Y
CUIMC	US	EHR	1985/03–2021/08	6,808,470	55.8	38 (21–56)	28,044	1,197,983	Y
FIIBAP[a]	ES	EHR	2001/03–2021/10	292,305	54.7	41 (23–58)	7138	78	N
HIC	SC	EHR	2005/01–2021/12	1,254,464	50.4	45 (28–66)	11,813	885,236	Y
IBM_CCAE	US	Claims	2000/01–2021/07	159,440,276	51.1	31 (17–46)	983,089	23,483,191	N
IBM_MDCD	US	Claims	2006/01–2020/12	32,806,887	56.2	24 (5–38)	196,997	11,810,505	N
IBM_MDCR	US	Claims	2000/01–2021/07	10,356,249	55.3	72 (65–77)	41,542	1,467,963	N
IMASIS	ES	EHR	1990/02–2021/07	976,524	47.4	38 (23–54)	9330	198,012	Y
IPCI	NL	GP	2006/01–2021/06	2,529,355	51.2	37 (18–55)	91,759	1,329,674	Y
IQVIA_FRANCE_DA	FR	GP	2016/07–2021/06	3,767,012	52.3	38 (18–56)	2859	1,394,912	N
IQVIA_GERMANY_DA	DE	GP	2011/04–2021/03	30,780,239	55.7	44 (25–62)	45,508	9,040,531	N
IQVIA_OPENCLAIMS	US	Claims	2000/01–2021/10	306,000,000	52.6	34 (14–52)	17,848,443	306,000,000	N
IQVIA_PHARMETRICS	US	Claims	2013/01–2021/09	166,422,594	50.6	38 (19–50)	1,593,578	46,947,246	N
IU	TR	EHR	2018/01–2021/10	899,515	53.0	35 (18–52)	6194	619	Y
JMDC	JP	Claims	2005/01–2021/03	12,541,088	48.6	32 (19–46)	17,564	6,680,196	N
MHD	BE	EHR	2015/07–2021/12	117,131	50.9	52 (29–70)	203	23,754	N
OPTUM_EHR	US	EHR	2007/01–2021/03	99,454,715	53.3	37 (19–56)	693,334	41,281,147	Y
OPTUM_SES	US	Claims	2000/05–2021/06	90,285,937	50.5	36 (19–52)	899,986	17,212,611	Y
SIDIAP	ES	GP	2003/01–2021/06	8,022,374	50.1	35 (17–51)	495,237	5,934,449	Y
STARR	US	EHR	2008/01–2022/04	3,475,673	53.6	36 (18–54)	31,928	1,118,549	Y
U_OF_TARTU[b]	EE	Claims	2021/01–2021/02	386,557	53.2	39 (21–57)	84,957	376,842	N
UCCS	RS	EHR	2018/10–2021/03	823,962	54.1	51 (35–67)	16,764	49,643	Y
UCHDW[a]	US	EHR	2012/01–2022/05	316,119	53.8	37 (12–54)	61,037	240,831	Y
UK_BIOBANK	UK	EHR + Registry	1970/02–2020/07	502,504	54.4	58 (40–69)	1717	458,889	Y
Total	–	–	–	*945,520,607*	–	–	*23,840,986*	*492,730,503*	–

APHM = Health Data Warehouse of Assistance Publique - Hopitaux de Marseille, BE = Belgium, CPRD_AURUM = Clinical Practice Research Datalink AURUM, CU_AMC = University of Colorado Anschutz Medical Campus- Health Data Compass, CUIMC = Columbia University Irving Medical Center, DE = Germany, EE = Estonia, EHR = Electronic Health Record, ES = Spain, F = Female, FIIBAP = Fundación para la Investigación e Innovación Biosanitaria en Atención Primaria COVID19, FR = France, GP = General Practitioner, HIC = Health Informatics Centre, IBM_CCAE = IBM® MarketScan® Commercial Claims and Encounters Database, IBM_MDCD = IBM® MarketScan® Multi-State Medicaid Database, IBM_MDCR = IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database, IMASIS = Parc de Salut Mar Barcelona Information System, IPCI = Integrated Primary Care Information, IQR = Interquartile Range, IQVIA_FRANCE_DA = IQVIA Disease Analyzer France, IQVIA_GERMAN_DA = IQVIA Disease Analyzer Germany, IQVIA_OPENCLAIMS = IQVIA LRxDX Open Claims, IQVIA_PHARMETRICS = IQVIA Pharmetrics, IU = Istanbul Faculty of Medicine, Istanbul University, JP = Japan, MHD = Medaman Hospital Data, N = No, NL = The Netherlands, OPTUM_EHR = Optum® de-identified Electronic Health Record Dataset, OPTUM_SES = Optum De-Identified Clinformatics® Data Mart Database - Socio-Economic Status, RS = Serbia, SC = Scotland, SIDIAP = The Information System for Research on Primary Care, STARR = STAnford medicine Research data Repository-OMOP, TR = Turkey, U_OF_TARTU = University of Tartu (U_OF_TARTU), UCCS = University Clinical Center of Serbia, UCHDW = University of California Health Data Warehouse, UK = United Kingdom, UK_BIOBANK = UK Biobank, US = United States, Y = Yes. [a]COVID-19 only subset. [b]COVID-19 + Controls.

Table 1: **Database characteristics.**

Research Protection Program (HRPP). The following databases had institutional review board or ethics committee approval: CU_AMC (IRB COMIRB #21-2869), CUIMC (IRB protocol AAAO7805), FIIBAP (through Hospital Universitario 12 de Octubre, project code: 22/295), HIC (through Caldicott Guardian(s), reference number IGTCAL9397), IMASIS (CEIC Parc de Salut Mar, ref. Code 2021/9975), IU (ref. No 2021/2190), MHD (approved 7 FEB 2022), SIDIAP (project code: 21/052-PCV), U_OF_TARTU (protocol 330/T-10), and UCCS (Decision 410-9).

The New England Institutional Review Board (IRB) has determined that studies conducted in IBM_CCAE, IBM_MDCD, IBM_MDCR, OPTUM_SES, and OPTUM_EHR are exempt from study-specific IRB review, as these studies do not qualify as human subjects research. Research activity associated with access to the UC-wide UCHDW COVID Research Data Set (CORDS) Limited Data Set does not involve human subjects as defined by the federal regulations summarized in 45 CFR 46.102(e), and hence does not require IRB review. Based on Ethical Guidelines for Epidemiological Research issued by the Japanese Ministry of Health, Labor and Welfare, ethics approval and informed consent were not applicable for use of JMDC. For use of IQVIA_FRANCE_DA, IQVIA_GERMAN_DA, IQVIA_OPENCLAIMS, and IQVIA_PAHRMETRICS, no patient permission was necessary because all patient data are deidentified for research purposes.

Study participants

We determined the incidence of AESI within two target populations: a 'Pre-Pandemic Background Population' from 2017 to 2019 and 'Patients with COVID-19' from 2020 to 2022. This work focuses on the 'Patients with COVID-19' using the 'Pre-Pandemic Background Population' to understand the occurrence of those conditions in the absence of either COVID-19 or the vaccines.

We defined the 'Pre-Pandemic Background Population' cohort as patients who were observed in a database on the index dates 1 January 2017, 1 January 2018, or 1 January 2019 (i.e., the patient was active in the database on one or any of those dates they would be included in this cohort). The entry event was the date observed and patients were only included if they had at least 365 days of observable time prior to the index date. We used the time prior to the pandemic, 2017 till 2019, as a way to identify patients that we are confident have not been exposed, had, or were experiencing COVID-19. Once the COVID-19 pandemic started it would be hard to be confident a patient did not or does not have the disease.

The 'Patients with COVID-19' cohort was defined as persons with a positive laboratory test for SARS-CoV-2 (any SARS-CoV-2 laboratory testing method) or an observed diagnosis code for COVID-19 (without a negative test on or within 3 days after index). The index date, or entry event, was set to the first of these and did not have to occur on a specific month or day. Additional inclusion rules included keeping patients that had at least 365 days of observable time prior to the index date and having no diagnosis of COVID-19 or SARS-CoV-2 prior to index. The calendar time for this cohort included all time available within each database after 1 December 2019, see Table 1 for specifics (e.g., CPRD_AURUM last date of capture is March 2021). An additional 'Patients with COVID-19' variation cohort was created for sensitivity analysis; it was a similar cohort except with the additional inclusion rule that cohort entry had to be prior to the COVID-19 vaccines becoming available (i.e., index before 1 January 2021); this cohort was called 'Patients with COVID-19 prior to 2021'. Patients were not excluded from the COVID-19 cohorts if they were also in the 'Pre-Pandemic Background Population'. The codes for this COVID-19 definition are listed in the protocol appendix[16] and have been used in prior studies.[17–20]

All databases contributed to both study cohorts ('Pre-Pandemic Background Population' and 'Patients with COVID-19'), except for FIIBAP and IU, which only contributed to the 'Patients with COVID-19' cohort as these databases only contained COVID-19 patients and not a general population.

This study uses coded data that already exist in an electronic database. Confidentiality of patient records was always maintained. All study reports contain only aggregate data and do not identify individual patients or physicians. At no point during the study was identifying information about the subjects used.

Adverse events of special interest (AESIs)

We studied 16 AESIs. This list includes the original 15 prespecified by regulatory authorities and the addition of thrombosis with thrombocytopenia syndrome (TTS) which was an adverse event discovered when COVID-19 vaccinations started. The 16 outcomes include: GBS, facial nerve (Bell's) palsy, anaphylaxis, encephalomyelitis, narcolepsy, appendicitis, non-haemorrhagic stroke, haemorrhagic stroke, acute myocardial infarction (AMI), myocarditis and pericarditis, deep vein thrombosis (DVT), pulmonary embolism, disseminated intravascular coagulation (DIC), immune thrombocytopenia (ITP), transverse myelitis, and the cooccurrence of thrombosis with thrombocytopenia (TWT) as a proxy to TTS. All AESI definitions were used in previous published studies[5] and were based on the FDA's Center for Biologics Evaluation and Research protocol.[21] The AESIs were reviewed within each data partner to ensure the outcome could be reliably captured and if not, that AESI would be excluded for that particular data partner (Appendix 2 provides further detail). The definitions are fully specified in the protocol.[16]

Statistical analysis

We defined time at risk as 1–90 days after the study participant index date. Patients contributed time-at-risk from the index date until the earliest of 90 days after the index, they leave the database, or the start date of the AESI event. Death was not explicitly used as a censoring criteria, as the capture of death varies within each database, including some databases that do not capture death at all. We do censor patients at the end of observation, whether that be death or other reason for loss-to-follow-up.

We also excluded individuals for a specific outcome if a prior event had occurred during the clean window for that outcome. A clean window is the minimum time between event occurrences to be considered a new event. We used an outcome-specific clean window, within which an outcome is not considered incident. The clean window was 365 days for all outcomes except 30 days for anaphylaxis and 183 days for facial nerve palsy and encephalomyelitis.[21] The clean window was set different for anaphylaxis, facial nerve palsy, and encephalomyelitis to better define incident cases of these AESI.

Incidence rates were calculated as the total number of events divided by the person time at risk. The incidence rates were stratified by age and sex subgroups for each database. The age subgroups were (in years): 0–5, 6–17, 18–34, 35–54, 55–64, 65–74, 75–84, and ≥85. The age-specific rates for the AESIs were pooled across the databases using a random effects meta-analysis, with the DerSimonian-Laird method to estimate between database variation. The 95% prediction intervals were calculated using the R package "meta". Prediction intervals reflect the expected uncertainty if an estimated

rate from another study were included in the meta-analysis.

Indirect standardisation[22] was used to account for differences between the age subgroups and sex distribution in the COVID-19 disease cohort and the 'Pre-Pandemic Background Population'. Within each database, the pre-pandemic cohort was used as the demographic reference, such that an expected incidence rate could be computed using COVID-19-specific incidence rates weighted to reflect the 'Pre-Pandemic Background Population' demographics. COVID-19 cohort observed and expected rates were compared using standardised incidence ratios (SIR) with corresponding exact 95% confidence intervals.

Negative control outcomes were also used to evaluate potential bias in incidence ratio estimates. Twenty negative control outcomes were selected as diseases with no *a priori* evidence of a causal relationship with COVID-19, based on literature review and clinician adjudication (Appendix 3). Incidence rate and SIRs for these negative control outcomes were estimated in the same manner as for the 16 AESIs.

Meta-analytic age- and sex-specific rates were classified using the WHO Council for International Organizations of Medical Sciences thresholds: very common (≥10%), common (>1%–<10%), uncommon (≥0.1%–<1%), rare (≥0.01%–<0.1%), and very rare (<0.01%).[23]

All analyses were performed in R. The study protocol and analysis code are available at https://github.com/ohdsi-studies/Covid19SubjectsAesiIncidenceRate.

Role of funding source

There was no funding source for this study. All authors approved the final version, had access to the data, and accept responsibility to submit the final version for publication.

Results

We included 23,840,986 'Patients with COVID-19' from 26 databases representing a diverse set of care settings from North America, Europe, and Asia (Table 1). The percentage of females found across the databases ranged from 47.4% to 56.1%; all databases had slightly more females than males, except IMASIS and JMDC. The average age of persons in each database ranged from 24 to 72 years, reflecting the differing patient populations covered across the databases (e.g., IBM_MDCD has a large child and childbearing-aged female population, while IBM_MDCR represents retired individuals). The 'Pre-Pandemic Background Population' represented more than 492,730,503 person records. Characterisation results produced by CohortDiagnostics.[24] can be found on an interactive web app (https://data.ohdsi.org/Covid19SubjectsAesiIncidenceRate/).

Fig. 1 plots the incidence rates for the 'Patients with COVID-19' population, stratified by age group (x-axis), sex (half the x-axis represents females and the other half males), and database (color-coded by database type: claims, EMR, and GP centric data), trellised by the 16 AESIs. Similar age trends across databases were observed. Some AESIs had a clear increase in incidence rate with age: AMI, non-haemorrhagic stroke, DVT, PE, haemorrhagic stroke, Bell's Palsy, TWT, ITP, DIC, and GBS. In contrast, some AESIs had a clear decrease in incidence rate with age: appendicitis and anaphylaxis. Finally, some age trends were less clear: myocarditis and pericarditis, narcolepsy, encephalomyelitis, and transverse myelitis. Fig. 1 also shows substantial database heterogeneity within individual AESIs. For example, for DIC in the 90 days after COVID-19 diagnosis, males aged 35–54 in JMDC experienced the event 2417 per 100,000 person-years as compared to CPRD which experienced the event 5 per 100,000 person-years. Additionally, the incidence rates for 'Patients with COVID-19' were compared to the 'Pre-Pandemic Background Population' in Appendix 4. For most outcomes, across all database/age group/sex combinations, the incidence rates are higher for the 'Patients with COVID-19' compared to the 'Pre-Pandemic Background Population'.

To help give an overview of these results, Fig. 2 shows the pooled estimated age- and sex-stratified incidence rates per 100,000 person years (PY) for AESI events in the 90-days after index for 'Patients with COVID-19'. Nevertheless, the incidence of several AESIs differed greatly between age and sex strata. For example, pulmonary embolism was uncommon for patients less than 35 then becoming common for older individuals. The pattern was similar for AMI, non-haemorrhagic stroke, and DVT (with the exception of AMI and non-haemorrhagic stroke becoming common at an older age of 55). The incidence rate of haemorrhagic stroke and TWT was primarily uncommon becoming common later in life (75 years and 65 years respectively). The other AESIs were primarily uncommon with ages between 18 and 74, experiencing the event rarely. For comparison, we additionally produced this figure for the 'Pre-Pandemic Background Population' which can be found Appendix 5.

Fig. 3 reports the meta-analytic estimates of SIRs comparing the 'Patients with COVID-19' to the 'Pre-Pandemic Background Population'. Twelve of 16 AESIs had ratios above 2 and 7 of 16 AESIs had ratios above 5. Pulmonary embolism had the highest SIRs (11.7 [95% confidence interval 10.1–13.7]), suggesting that the observed incidence of pulmonary embolism in the 90 days after COVID-19 diagnosis was over 11 times higher than expected in a 90-day period for the background population in the pre-pandemic period. The database specific results are reported in Appendix 6. The negative control outcome estimates are provided in Appendix 7 and reveal consistent positive bias in estimated SIR values. SIRs ranged from 0.9 to 5.0 with a majority of

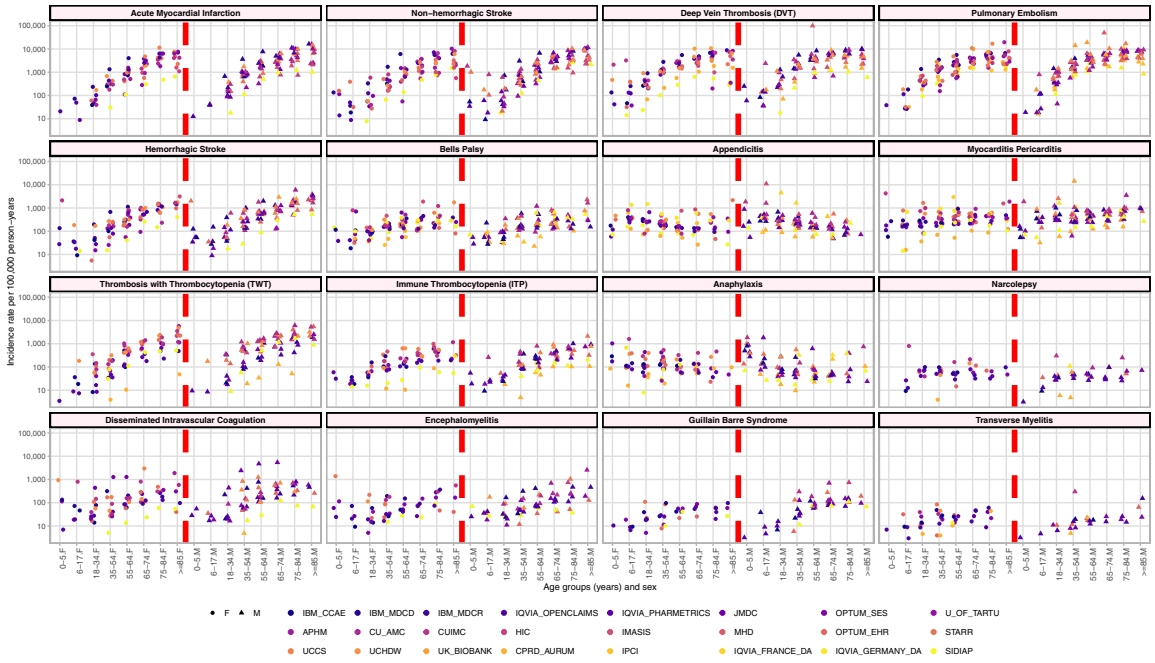

Fig. 1: **For 'patients with COVID-19', age group and sex stratified incidence rates of the 16 adverse events of special interest for 24* databases**.* FIIBAP and IU did not contribute to the 'Pre-Pandemic Background Population' and thus and IR could not be calculated. APHM = Health Data Warehouse of Assistance Publique - Hopitaux de Marseille, CPRD_AURUM = Clinical Practice Research Datalink AURUM, CU_AMC = University of Colorado Anschutz Medical Campus- Health Data Compass, CUIMC = Columbia University Irving Medical Center, F = Female, FIIBAP = Fundación para la Investigación e Innovación Biosanitaria en Atención Primaria COVID19, HIC = Health Informatics Centre, IBM_CCAE = IBM® MarketScan® Commercial Claims and Encounters Database, IBM_MDCD = IBM® MarketScan® Multi-State Medicaid Database, IBM_MDCR = IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database, IMASIS = Parc de Salut Mar Barcelona Information System, IPCI = Integrated Primary Care Information, IQVIA_FRANCE_DA = IQVIA Disease Analyzer France, IQVIA_GERMAN_DA = IQVIA Disease Analyzer Germany, IQVIA_OPENCLAIMS = IQVIA LRxDX Open Claims, IQVIA_PHARMETRICS = IQVIA Pharmetrics, IU = Istanbul Faculty of Medicine, Istanbul University, M = Male, MHD = Medaman Hospital Data, OPTUM_EHR = Optum® de-identified Electronic Health Record Dataset, OPTUM_SES = Optum De-Identified Clinformatics® Data Mart Database–Socio-Economic Status, SIDIAP = The Information System for Research on Primary Care, STARR = STAnford medicine Research data Repository-OMOP, U_OF_TARTU = University of Tartu (U_OF_TARTU), UCCS = University Clinical Center of Serbia, UCHDW = University of California Health Data Warehouse, UK_BIOBANK = UK Biobank.

generated SIRs above 1.7. The meta-analysis was repeated for 'Patients with COVID-19 prior to 2021' to see if inclusion of COVID-19 vaccinated patients may have impacted the results however the results found were similar (Appendix 8).

Discussion

To our knowledge this is the largest study to date on the descriptive epidemiology of AESIs among the COVID-19 population. In this study we estimated age- and gender-specific incidence rates of 16 AESIs among COVID-19 patients using 26 observational data health sources from around the world. To contextualise our findings, we reported on the age- and gender-SIR, comparing the IR among COVID-19 cohort to the 'Pre-Pandemic Background Population' of the databases. Our findings suggest a consistent trend of an increased risk for multiple AESIs among the COVID-19 cohort when compared to the database 'Pre-Pandemic Background Population', though the magnitude of the risk should be interpreted with caution.

We found considerable heterogeneity in the IR among the COVID-19 cohort reflected by the wide prediction intervals of the pooled age- and sex-specific IR, suggesting that caution is needed when these estimates are used. We also observed considerable variability with age and some with sex groupings, emphasising the need for age- and sex-stratification when assessing risks and benefits of COVID-19 vaccines.

The observed magnitude of heterogeneity across sources within age and sex subgroups suggests that residual differences are present. The remaining heterogeneity may be related to differences in healthcare systems, settings, data capture processes, or true differences in subpopulations. A limitation of this work is we have not furthered stratified by comorbidities related to either COVID-19 or outcome (such as type 2 diabetes

Outcome by Sex	0-5 Years	6-17 Years	18-34 Years	35-54 Years	55-64 Years	65-74 Years	75-84 Years	>=85 Years
Acute Myocardial Infarction								
Female	527 (2 to 116345)	151 (5 to 4281)	84 (28 to 253)	335 (123 to 910)	940 (374 to 2365)	2047 (1044 to 4014)	3940 (1969 to 7882)	4960 (2090 to 11772)
Male	469 (2 to 121380)	139 (4 to 5421)	167 (44 to 639)	827 (333 to 2053)	1991 (954 to 4153)	3526 (1945 to 6391)	5721 (2717 to 12046)	8198 (2782 to 24155)
Non-haemorrhagic Stroke								
Female	552 (3 to 95794)	153 (6 to 3663)	96 (21 to 441)	356 (105 to 1209)	839 (223 to 3156)	1790 (829 to 3866)	3723 (1677 to 8265)	4928 (1974 to 12303)
Male	547 (4 to 79312)	162 (8 to 3148)	121 (16 to 924)	575 (155 to 2135)	1422 (476 to 4249)	2965 (1532 to 5739)	5009 (2421 to 10361)	6524 (2225 to 19132)
Deep Vein Thrombosis (DVT)								
Female	478 (5 to 44031)	245 (25 to 2399)	287 (97 to 847)	817 (337 to 1979)	1638 (743 to 3610)	2839 (1549 to 5203)	3908 (1968 to 7763)	3781 (1467 to 9750)
Male	425 (8 to 23054)	217 (24 to 1935)	346 (81 to 1468)	1364 (490 to 3794)	2993 (1479 to 6057)	3994 (2213 to 7211)	4538 (2113 to 9750)	4504 (1789 to 11342)
Pulmonary Embolism								
Female	504 (5 to 53649)	168 (17 to 1697)	409 (124 to 1346)	1119 (493 to 2539)	2025 (1021 to 4014)	3650 (2533 to 5259)	4782 (3044 to 7515)	4718 (2353 to 9460)
Male	444 (3 to 67053)	170 (13 to 2194)	521 (109 to 2491)	1686 (652 to 4358)	3438 (2014 to 5868)	5230 (3953 to 6920)	6504 (4119 to 10269)	6480 (3914 to 10726)
Haemorrhagic Stroke								
Female	574 (3 to 101501)	145 (5 to 4607)	60 (14 to 248)	155 (41 to 590)	328 (123 to 874)	592 (295 to 1186)	935 (401 to 2178)	1175 (556 to 2482)
Male	568 (4 to 72465)	150 (7 to 3149)	109 (25 to 484)	263 (75 to 925)	550 (193 to 1567)	781 (371 to 1644)	1452 (738 to 2855)	1948 (487 to 7798)
Bell's Palsy								
Female	539 (8 to 34829)	189 (17 to 2091)	91 (59 to 143)	153 (77 to 304)	212 (98 to 457)	300 (71 to 1261)	383 (82 to 1787)	403 (65 to 2519)
Male	464 (5 to 46723)	166 (18 to 1513)	91 (35 to 236)	182 (86 to 384)	240 (87 to 666)	310 (98 to 978)	378 (73 to 1958)	648 (68 to 6176)
Appendicitis								
Female	580 (27 to 12353)	355 (175 to 719)	260 (145 to 467)	181 (100 to 326)	159 (71 to 356)	200 (57 to 700)	269 (49 to 1487)	334 (30 to 3778)
Male	561 (29 to 10808)	432 (195 to 961)	345 (153 to 780)	210 (123 to 359)	159 (86 to 292)	201 (69 to 588)	292 (47 to 1834)	450 (31 to 6580)
Myocarditis Pericarditis								
Female	650 (19 to 22601)	268 (74 to 965)	329 (76 to 1418)	290 (122 to 688)	344 (131 to 904)	383 (142 to 1037)	404 (132 to 1242)	425 (88 to 2052)
Male	529 (24 to 11794)	362 (106 to 1237)	469 (164 to 1340)	330 (147 to 742)	374 (167 to 840)	498 (208 to 1191)	709 (207 to 2423)	798 (131 to 4842)
Thrombosis with Thrombocytopenia (TWT)								
Female	457 (3 to 62013)	146 (1 to 17082)	57 (6 to 546)	116 (17 to 794)	312 (47 to 2069)	582 (84 to 4015)	964 (74 to 12496)	1476 (94 to 23167)
Male	430 (2 to 74298)	137 (1 to 21627)	89 (10 to 803)	280 (35 to 2221)	600 (80 to 4488)	1143 (135 to 9644)	1741 (146 to 20754)	2096 (173 to 25442)
Immune Thrombocytopenia (ITP)								
Female	506 (5 to 55485)	161 (4 to 6440)	67 (24 to 188)	110 (36 to 340)	189 (62 to 580)	286 (87 to 947)	452 (99 to 2058)	409 (71 to 2366)
Male	480 (4 to 59121)	150 (3 to 6553)	52 (17 to 160)	127 (45 to 361)	204 (90 to 464)	361 (122 to 1065)	517 (125 to 2145)	754 (136 to 4193)
Anaphylaxis								
Female	546 (34 to 8722)	224 (38 to 1298)	120 (57 to 252)	135 (46 to 393)	134 (39 to 457)	165 (31 to 886)	209 (15 to 2851)	299 (10 to 9177)
Male	670 (105 to 4273)	222 (40 to 1228)	90 (29 to 281)	100 (21 to 487)	101 (17 to 588)	144 (18 to 1139)	272 (7 to 10999)	430 (6 to 31495)
Narcolepsy								
Female	508 (4 to 66259)	182 (2 to 18808)	69 (29 to 163)	73 (27 to 194)	98 (15 to 621)	186 (18 to 1892)	213 (14 to 3336)	337 (7 to 16043)
Male	459 (3 to 81397)	169 (1 to 20837)	59 (16 to 220)	73 (17 to 308)	103 (15 to 710)	167 (12 to 2232)	302 (11 to 8349)	495 (10 to 25617)
Disseminated Intravascular Coagulation								
Female	509 (4 to 65261)	148 (4 to 5549)	48 (10 to 229)	91 (9 to 920)	163 (32 to 820)	302 (41 to 2226)	326 (71 to 1500)	397 (33 to 4751)
Male	452 (4 to 55867)	127 (3 to 5681)	67 (7 to 679)	164 (11 to 2349)	294 (33 to 2653)	487 (68 to 3504)	613 (82 to 4595)	536 (93 to 3076)
Encephalomyelitis								
Female	575 (4 to 93365)	145 (3 to 6015)	49 (7 to 348)	76 (12 to 503)	91 (24 to 347)	148 (31 to 713)	257 (40 to 1673)	344 (24 to 4934)
Male	495 (3 to 84065)	146 (8 to 2571)	56 (13 to 253)	85 (18 to 408)	134 (34 to 520)	180 (56 to 583)	379 (55 to 2614)	563 (43 to 7329)
Guillain Barre Syndrome								
Female	508 (2 to 133569)	131 (1 to 20989)	34 (9 to 130)	52 (9 to 292)	88 (19 to 403)	150 (22 to 1018)	181 (13 to 2611)	295 (7 to 12936)
Male	441 (2 to 96492)	126 (1 to 25674)	51 (3 to 826)	73 (11 to 473)	106 (38 to 301)	162 (44 to 591)	295 (39 to 2229)	478 (14 to 16304)
Transverse Myelitis								
Female	456 (2 to 87635)	117 (1 to 21552)	30 (4 to 206)	35 (6 to 190)	47 (10 to 231)	96 (7 to 1379)	121 (3 to 4442)	168 (2 to 12892)
Male	406 (2 to 66668)	112 (1 to 22746)	40 (1 to 1354)	41 (4 to 414)	52 (7 to 388)	89 (5 to 1743)	172 (3 to 11116)	283 (6 to 12779)
Council of International Organizations of Medical Sciences frequency classification								
Very rare (<1/10000)	Rare (>=10000 to <1/1000)		Uncommon (>= 1/1000 to <1/100)		Common (>= 1/100 to <1/10)		Very Common (>=1 /10)	

Fig. 2: **Pooled estimated age and sex stratified incidence rates per 100,000 person years (with 95% prediction intervals), calculated from meta-analyses for 'Patients with COVID-19'.**

Source	SIR (95% CI)
Pulmonary Embolism	11.77 (10.08 – 13.73)
Disseminated Intravascular Coagulation	8.79 (7.13 – 10.85)
Myocarditis Pericarditis	7.70 (6.61 – 8.98)
Guillain Barre Syndrome	6.79 (5.03 – 9.17)
Encephalomyelitis	5.94 (4.59 – 7.70)
Thrombosis with Thrombocytopenia (TWT)	5.60 (4.74 – 6.62)
Deep Vein Thrombosis (DVT)	5.56 (4.57 – 6.76)
Acute Myocardial Infarction	4.70 (4.07 – 5.43)
Non–hemorrhagic Stroke	4.59 (3.95 – 5.34)
Hemorrhagic Stroke	3.86 (3.26 – 4.57)
Immune Thrombocytopenia (ITP)	3.37 (2.96 – 3.82)
Transverse Myelitis	2.53 (2.01 – 3.17)
Anaphylaxis	1.93 (1.56 – 2.40)
Bells Palsy	1.84 (1.60 – 2.12)
Appendicitis	1.72 (1.47 – 2.02)
Narcolepsy	1.32 (1.05 – 1.66)

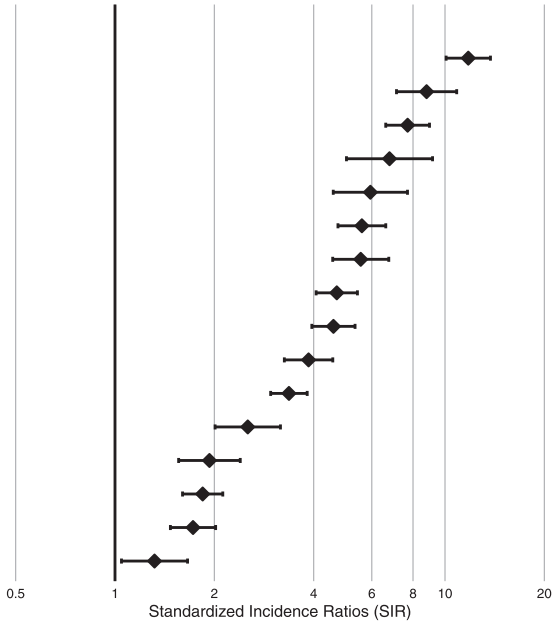

Fig. 3: **Standardised incidence ratios forest plot with meta analysis comparing the 'patients with COVID-19' to the 'pre-pandemic background population'.** CI = confidence interval.

or obesity) or socio-economic status. These differences may also be due to systematic error, selection bias, or differential outcome measurement error between databases.[25] In a similar analysis across 13 databases, Li et al.[5] reported that AESI background rates varied by database and regions. Consistent to our findings, the authors also found considerable differences in the incidence rates by age and sex, suggesting caution is needed when incidence rates are compared across populations.[5]

Among COVID-19 patients, the majority of COVID-19 AESIs were found to be "uncommon" with few "rare" among all age and gender groups investigated. Thrombotic events such as AMI, strokes, DVT, and pulmonary embolism were more frequent compared to other AESIs and were "common" in older COVID-19 patients. These findings are consistent with prior studies suggesting that cardiovascular and thrombotic complications in particular are relatively common post COVID-19,[26,27] especially among older patients. In most databases, the risk of these thrombotic events was higher among COVID-19 patients when compared to the 'Pre-Pandemic Background population' of the databases with a pooled SIR above 3. In particular, the SIR for pulmonary embolism had an elevated trend away from the null in all databases but one. It is well established that COVID-19 is associated with a hypercoagulable state, and older patients with additional risk factors have a worse outcome. Consistent with our results, clinical evidence suggests that COVID-19 is particularly related to pulmonary embolism[28] especially among those with pneumonia. However, until now, the precise incidence of thrombosis in patients with COVID-19 has not been determined, mainly because multiple studies have reported conflicting estimates.[29]

The Centers for Disease Control and Prevention (CDC) lists myocarditis and pericarditis, thrombosis with thrombocytopenia syndrome (TTS), presented in this work as TWT, and GBS as three serious types of adverse events following COVID-19 vaccination, with evidence that suggests, although rare, a link to certain types of COVID-19 vaccinations.[30] These three adverse events are also listed on the label of at least one COVID-19 vaccine authorised for use in the US and Europe. Our findings suggest that, while still rare among COVID-19 patients, the risk of these three events may be higher in the COVID-19 patients when compared to the 'Pre-Pandemic Background Population'. Since the estimated incidence ratios are unadjusted and are not presented for causal inference, future investigation may be warranted to fully contextualise the risk benefit balance of the COVID-19 vaccines as it relates to myocarditis and pericarditis, TWT, and GBS.

Multiple studies have concluded that COVID-19 disease may lead to neurological complications leading to GBS, and some studies also highlighted differences in the presentation of the disease, with greater severity of symptoms in GBS associated with COVID-19.[31–33] Previous studies have also found that myocarditis and pericarditis may be part of the wide spectrum of cardiovascular sequelae of COVID-19 disease.[32,34,35] A CDC network retrospective cohort study[34] using data electronic medical records across multiple databases in the US, reported that among 814,524 COVID-19 patients the incidence of myocarditis or pericarditis varied by age and gender and ranged between 17.6 and 114.0 per 100,000 for males and between 10.8 and 61.7 per 100,000 for females in a 21 day risk window. The same study reported that the incidences of myocarditis or pericarditis after SARS-CoV-2 infection were higher than after the mRNA COVID-19 vaccination for both males and females in all age groups. To our knowledge, our study is the first to report on TWT incidence among 'Patients with COVID-19'. However, prior work has highlighted the potential misclassification error related to identifying TTS (also known as vaccine-induced thrombotic thrombocytopenia [VITT]) in observational data.[36] The incidence rates of TWT reported in this study were higher among men of older ages which is inconsistent with known trends of TTS/VITT. Thus, these rates should be interpreted with caution.

Our findings suggest that COVID-19 disease itself must be considered when assessing the relationship between COVID-19 vaccines and the AESI. To be specific, our findings suggest that COVID-19 disease may be associated with at least some of the AESIs and may consequently exert a confounding or intermediating effect in the observed association between the vaccines and the AESI. When conducting observational studies investigating the association between the vaccines and the AESI, it is important to control for COVID-19.

A particular strength of this study is that it includes a large number of databases from around the world, covering a sizable study population with diversity in geographical location, underlying populations, pandemic status, health systems, and data types. Our analysis enabled a comprehensive and standardised assessment of incidence rates of AESIs among 'Patients with COVID-19' across multiple settings. This was possible due to the use of the OMOP CDM, which enabled use of the same study design and analytical code in all databases and to gather results from participating data partners rapidly and without transferring patient level data. All outcome definitions, clinical codes, and phenotype algorithms have been made open source and are available online for review and to maximise reproducibility and reuse.

A limitation of this analysis was we did not differentiate for the multiple variants of COVID-19, and we did not consider recurrent COVID-19. This means we cannot categorise or compare the differences in AESIs associated with the different variants or in patients who had COVID-19 multiple times. The incidence of the AESIs may change depending on the variant or number

of infections, so this work cannot describe the incidence of an AESI for a given variant, be representative of future variants, or describe what happens for patients who had multiple infections of COVID-19.

An additional limitation of the analysis of the SIRs is that they do not fully account for confounding and were shown to exhibit bias with negative controls and therefore should not be interpreted as causal effect estimates. Also, the use of a historical population as a comparator is known to be associated with an elevated type 1 error, which may bias the results away from the null[37] (e.g., individuals with 'Patients with COVID-19' disease my tend to receive more clinical attention or examination than individuals in the 'Pre-Pandemic Background Population'). Results from negative control outcomes in this work show a consistent positive bias (away from the null). These findings indicate that most associations observed in our study are potentially larger than the magnitude of the bias observed using negative controls. Another limitation of this study is that all outcomes and the COVID-19 definition itself are subject to measurement error. While most of the outcome definitions have been used in prior studies and were reviewed by clinicians and data experts, they were mainly based on the presence of specific diagnostic codes and were not validated further. There was significant heterogeneity in COVID-19 test and disease ascertainment across databases. COVID-19 severity is an important factor that was outside the scope of this analysis but may have affected the risk of AESIs and probably varied by database. When defining the 'Pre-Pandemic Background Population', we 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.[25]

Finally, there are some limitations related to participating databases. Information on hospital admissions was not available in all primary care databases used (Table 1) resulting in inpatient events not being captured. EHR databases are subject to incomplete capture of medical events that may occur but are recorded outside the participating health system. The bias of incomplete information was partially mitigated by including only those patients who had at least one year of continuous observation but defining continuous observation can be challenging across disparate databases. Administrative claims databases offered a potentially complete data capture but lacked some important data elements such as laboratory test results. We mitigated some of these limitations by providing within database incidence rate comparison. Additionally, the repeated influxes of vast numbers of critically ill COVID-19 cases suddenly overwhelmed hospitals and we cannot exclude changes in the organisation and thus the quality of coding in code-based administrative database. Finally, all our databases represented subsets of the population in which they originate, which poses a risk of selection bias.

Our study assessed the descriptive epidemiology of the occurrence of AESIs in the 90-days after COVID-19 disease. These results found large variations in the rates of AESIs in 'Patients with COVID-19' across age groups and sex, showing the need for stratification. Considerable database heterogeneity was found across the AESIs, suggesting individual study estimates should be interpreted with caution. Comparing the 'Patients with COVID-19' to the 'Pre-Pandemic Background Population' showed a fairly consistent elevated rate in experiencing an AESI within 90-days after index. This elevated risk we see both consistently across the database stratified results as well as the meta-analysis. The results of this work are of public health importance as they help put into perspective the risk of the AESIs post vaccination versus post SARS-CoV-2 infection.

Contributors

Using the CRediT, Contributor Roles Taxonomy[38] here are how the co-authors participated in this work. For conceptualization, or formulation of the research goals, the following authors participated: EAV, AS1 (Shoaibi), TA, AGS, GR, DRM, TDS, KN, CR, EPM, NHS, JDP, MTGM, DB, HCJ, ADH, MPJ, PSB, NGB, RV, AJB, AP2 (Patel), LD, CH, FA, AO, FN, GH, MAS, DPA, PRR, PBR. For data curation, or those who worked with research data, the following authors participated: EAV, CB, SvS, TLC, SH, DRM, JMR, JMRA, MAM, MdW, LHJ, TDS, AP1 (Pistillo), RK, FM, SD, VP, MGK, KN, CR, AS2, EPM, NHS, JDP, MTGM, DB, HCJ, ADH, MPJ, NGB, SS, AYU, BE, LB2 (Belmans), GB, DJD, ZC, RV, AJB, AP2, LD, CH, GH, DPA, PRR, PBR. For formal analysis, or those who analysed or synthesised study data, the following authors participated: EAV, AS1, LYHL, CB, AGS, GR, CF, JMRA, MAM, LHJ, AP1, FM, MGK, EPM, SS, AYU, BE, LB2, GB, AO, FN, DPA, MJS, PBR. For investigation or conducting of the research or data/evidence collection, the following authors participated: KH, DRM, JMRA, MAM, RK, SS, AYU, BE. Design of the methodology was performed by the following individuals: EAV, AS1, TA, AGS, GR, DRM, MGK, AS2, EPM, FA, AO, FN, GH, MAS, DPA, PRR, MJS, PBR. Project administration was performed by the following individuals: EAV, CB, MGK, EPM, NHS, GB, RV, AJB, AP2, LD, CH, PBR. Resources, or provision of computing resources or analysis tools, was performed by the following: DRM, JMR, JMRA, MAM, TDS, MGK, AS2, DPA, PRR. Software, or program development, was performed by: EAV, AGS, SH, LHJ, MGK, EPM, JDP, MAS, MJS, PBR. Supervision, or oversight, was performed by: LB1 (Boyer), DRM, TDS, AJB, AP2, LD, CH, MAS, DPA, PRR, MJS, PBR. Validation, or verification of results, was performed by: RM, LB1, SH, MdW, MGK, PBR. Visualization, or creation of visualizations or data presentation, was performed by: EAV, LYHL, CB, FB, PBR. Writing of the original draft was performed by: EAV, AS1, LYHL, TA, TDS, ER, FM, SD, VP, MTGM, DB, HCJ, ADH, MPJ, PSB, NGB, SS, AYU, BE, DJD, ZC, AO, FN, PBR. Editing and reviewing the original writing was performed by: EAV, AS1, LYHL, CB, TA, RM, KH, AGS, GR, SvS, CF, LB1, TLC, DRM, JMR, JMRA, MAM, LHJ, TDS, ER, AP1, FM, SD, MGK, KN, CR, AS2, EPM, NHS, JDP, MTGM, DB, HCJ, ADH, MPJ, PSB, NGB, SS, AYU, BE, LB2, GB, MLNGM, DJD, ZC, RV, AJB, AP2, LD, CH, FB, FA, AO, FN, GH, MAS, DPA, PRR, MJS, PBR. All authors approved the final version, had access to the data, and accept responsibility to submit the final version for publication. No competing interests. All authors approved the final version, had access to the data, and accept responsibility to submit the final version for publication.

Data sharing statement

The results of the analysis can be both found at https://data.ohdsi.org/Covid19SubjectsAesiIncidenceRate/ and an export of the incidence rates can be found in Appendix 4.

Some of the datasets used within this study are available via license. The data that support the findings of this study are available to license from IBM (IBM_CCAE, IBM_MDCD, IBM_MDCR), Optum (OPTUM_EHR, OPTUM_SES), JDMC, IQVIA (IQVIA_FRANCE_DA, IQVIA_GERMANY_DA, IQVIA_OPENCLAIMS, IQVIA_PHARMETRICS). Data are available from IBM at https://www.ibm.com/products/marketscan-research-databases, from Optum at https://www.optum.com/business/solutions/life-sciences/real-world-data.html, from JMDC at https://www.jmdc.co.jp/en/jmdc-claims-database/, and from IQVIA at https://www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights.

Outside the license data previously described, this study was performed as a federated network study, meaning the data remained with the data partner. Individual organisations would need to be contacted in order to gain access to those data assets.

Declaration of interests

EAV, AS, CB, RM, KH, AGS, GR, MJS, and PBR are employees of Janssen Research and Development LLC and shareholders of Johnson & Johnson (J&J) stock. Many data partners (including JMR, JMRA, MAM, MdW, LHJ, GB, PRR) on this project have received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The funding source had no role in writing the manuscript or the decision to submit it for publication. Odysseus Data Services, where SvS is employed, is a beneficiary of the mentioned IMI 2 JU project. DPA has received support from the EMA and Innovative Medicines Initiative. DPA's research group has received grant support from Amgen, Chesi-Taylor, Novartis, GSK, UCB Biopharma and J&J. DPA's department has received advisory/consultancy fees from AstraZeneca, Amgen, Astellas, J&J, and UCB Biopharma. DPA's department has received fees for speaker services from Amgen and UCB Biopharma. DPA's has other financial or non-financial interests in Janssen, on behalf of the IMI-funded EHDEN and European Medical Information Framework (EMIF) consortiums, and Synapse Management Partners have supported training programmes organised by DPA's department that was open for external participants. TDS received financial support from the Institute de Salud Carlos Ill (ISCIII; Miguel Servet 2021: CP21/00023) which is co-funded by the European Union. The funding source had no role in writing the manuscript or the decision to submit it for publication. PRR, MdW, LHJ work for a research group who received/receives unconditional research grants from Chiesi, Amgen, UCB, J&J, European Medicines Agency (EMA), and IMI. GH provided funding for use of the CUIMC database and the OHDSI infrastructure. GH received support through a grant to Columbia University (NIH R01 LM006910). MAS receives grants from the US Department of Veterans Affairs within the scope of this work, and grants and contracts from the US National Institutes of Health, US Food & Drug Administration, and Janssen Research & Development outside the scope of this work. EPM receives salary support from the Department of Medicine at University of Calgary. EPM receives consulting fees from Janssen Research & Development for work that is outside the scope of the current study. FM's main employment is for Heliant d.o.o, an EHR provider in Serbia. JDP received salary support through the Universidad de Norte. DJD and ZC are full-time employees of CPRD, which provides contract research services for government and related healthcare authorities, and the pharmaceutical industry, including Janssen. FN holds some AstraZeneca shares and was employed there until 2019. RV, AJB, AP2, LD, and CH received partial funding for this work by the University of California Health System and its Center for Data-Driven Insights and Innovation. MGK received support for the manuscript via institutional (internal funding) only. AJB's research has been funded by NIH, Peraton (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oréal, and Progenity. AJB receives royalty payments through Stanford University, for several patents and other disclosures licensed to NuMedii and Personalis. AJB is a co-founder and consultant to Personalis and NuMedii; consultant to Mango Tree Corporation, and in the recent past, Samsung, 10x Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, Merck, and Roche. AJB has received honoraria and travel reimbursement for invited talks from J&J, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, many academic institutions, medical or disease specific foundations and associations, and health systems. AJB is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Meta (Facebook), Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, Regeneron, Sanofi, Pfizer, Royalty Pharma, Moderna, Sutro, Doximity, BioNtech, Invitae, Pacific Biosciences, Editas Medicine, Nuna Health, Assay Depot, Vet24seven, and several other non-health related companies and mutual funds. DRM has grants or contracts from Wellcome Trust Clinical Research Fellowship (214588/Z/18/Z). FB has a leadership or fiduciary role as the program chair of j-ISBA.

Acknowledgements

The authors would like to thank Lee Evans for his support of the environment that allowed sharing of results.

The authors thank the Center for Data-driven Insights and Innovation at UC Health (CDI2; https://www.ucop.edu/uc-health/functions/center-for-data-driven-insights-and-innovations-cdi2.html), for its analytical and technical support related to use of the UC Health Data Warehouse and related data assets, including the UC COVID Research Data Set (CORDS).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.101932.

References

1 *WHO coronavirus (COVID-19) dashboard*; 2022. https://covid19.who.int/. Accessed July 5, 2022.
2 Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N Engl J Med*. 2021;384(15):1412–1423.
3 Vasileiou E, Simpson CR, Shi T, et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. *Lancet*. 2021;397(10285):1646–1657.
4 Holder J. Tracking coronavirus vaccinations around the world. *New York Times*. 2022.
5 Li X, Ostropolets A, Makadia R, et al. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: a multinational network cohort study. *BMJ*. 2021;373:n1435.
6 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(10342):2191–2199.
7 Woo EJ, Mba-Jonas A, Dimova RB, Alimchandani M, Zinderman CE, Nair N. Association of receipt of the Ad26.COV2.S COVID-19 vaccine with presumptive guillain-barré syndrome, february-July 2021. *JAMA*. 2021;326(16):1606–1613.
8 Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *N Engl J Med*. 2021;384(22):2092–2101.
9 Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574–578.
10 Kimura S, Sato T, Ikeda S, et al. Development of a database of health insurance claims: standardization of disease classifications

and anonymous record linkage. *J Epidemiol*. 2010;20(5):413–419. https://doi.org/10.2188/jea.je20090066.
11 Datta S, Posada J, Olson G, et al. A new paradigm for accelerating clinical data science at Stanford Medicine. *arXiv preprint arXiv*. 2020:200310534.
12 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. https://doi.org/10.1371/journal.pmed.1001779.
13 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. https://doi.org/10.1093/ije/dyac026.
14 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 Therap*. 2009;47(10):617–626. https://doi.org/10.5414/cpp47617.
15 Recalde M, Rodríguez C, Burn E, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). *Int J Epidemiol*. 2022. https://doi.org/10.1093/ije/dyac068.
16 Voss EA, Shoaibi A, Ostropolets A, et al. *[RESEARCH PROTOCOL] adverse events of special interest within COVID-19 subjects*; 2021. https://ohdsi-studies.github.io/Covid19SubjectsAesiIncidenceRate/Protocol.html.
17 Kostka K, Duarte-Salles T, Prats-Uribe A, et al. Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using charybdis. *Clin Epidemiol*. 2022;14:369–384.
18 Morales DR, Ostropolets A, Lai L, et al. Characteristics and outcomes of COVID-19 patients with and without asthma from the United States, South Korea, and Europe. *J Asthma*. 2023:76–86.
19 Duarte-Salles T, Vizcaya D, Pistillo A, et al. Thirty-day outcomes of children and adolescents with COVID-19: an international experience. *Pediatrics*. 2021;148(3):e2020042929.
20 Prats-Uribe A, Sena AG, Lai LYH, et al. Use of repurposed and adjuvant drugs in hospital patients with Covid-19: multinational network cohort study. *BMJ*. 2021;373:n1038.
21 *Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring Protocol*; 2021, 01.12. 2021. https://www.bestinitiative.org/wp-content/uploads/2021/02/C19-Vaccine-Safety-AESI-Background-Rate-Protocol-FINAL-2020.pdf. Accessed October 14, 2021
22 Chan CK, Feinstein AR, Jekel JF, Wells CK. The value and hazards of standardization in clinical epidemiologic research. *J Clin Epidemiol*. 1988;41(11):1125–1134.
23 *Guidelines for preparing core clinical-safety information on drugs*; 1999. 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.
24 Rao G, Schuemie MJ, Ryan PB, W J, G J. *CohortDiagnostics: diagnostics for OHDSI cohorts*. 2022.
25 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.
26 Chen S, Zhang D, Zheng T, Yu Y, Jiang J. DVT incidence and risk factors in critically ill patients with COVID-19. *J Thromb Thrombolysis*. 2021;51(1):33–39.
27 Chen S, Zheng T, Wang S, et al. Association between risk of venous thromboembolism and mortality in patients with COVID-19. *Int J Infect Dis*. 2021;108:543–549.
28 Chen J, Wang X, Zhang S, et al. Characteristics of acute pulmonary embolism in patients with COVID-19 associated pneumonia from the city of wuhan. *Clin Appl Thromb Hemost*. 2020;26:1076029620936772.
29 Al-Ani F, Chehade S, Lazo-Langner A. Thrombosis risk associated with COVID-19 infection. A scoping review. *Thromb Res*. 2020;192:152–160.
30 Interim clinical considerations for use of COVID-19 vaccines currently approved or authorized in the United States, 3033 https://www.cdc.gov/vaccines/covid-19/clinical-considerations/interim-considerations-us.html. Accessed June 10, 2022.
31 Filosto M, Cotti Piccinelli S, Gazzina S, et al. Guillain-Barré syndrome and COVID-19: an observational multicentre study from two Italian hotspot regions. *J Neurol Neurosurg Psychiatry*. 2021;92(7):751–756.
32 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(12):1078–1090.
33 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(12):2144–2153.
34 Block J, Boehmer T, Forrest C, et al. Cardiac complications after SARS-CoV-2 infection and mRNA COVID-19 vaccination — PCORnet. https://www.cdc.gov/mmwr/volumes/71/wr/mm7114e1.htm; 2022. Accessed June 10, 2022.
35 Patone M, Mei XW, Handunnetthi L, et al. Risks of myocarditis, pericarditis, and cardiac arrhythmias associated with COVID-19 vaccination or SARS-CoV-2 infection. *Nat Med*. 2022;28(2):410–422.
36 Shoaibi A, Rao GA, Voss EA, et al. Phenotype algorithms for the identification and characterization of vaccine-induced thrombotic thrombocytopenia in real world data: a multinational network cohort study. *Drug Saf*. 2022;45:685–698.
37 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.
38 CRediT, Contributor roles Taxonomy. https://groups.niso.org/higherlogic/ws/public/download/26466/ANSI-NISO-Z39.104-2022.pdf; 2022. Accessed February 12, 2023.