

A rapid cycle analytics framework for vaccine safety surveillance within a real-world data network: Experience with enhanced surveillance of the Janssen COVID-19 vaccine

E. Claire Newbern^a, Azza Shoaibi^a, Kevin Haynes^a, Clair Blacketer^a, Corinne Willame^b, Frank DeFalco^a, Gowtham A. Rao^a, Kourtney Davis^a, Luis Anaya Velarde^a, Nicolas Praet^b, Rupa Makadia^a, Yimei Xu^a, Patrick Ryan^a, Martijn Schuemie^{a,*}

^a Janssen Research & Development, LLC, 1125 Trenton Harbourton Rd, Titusville, NJ 08560, United States

^b Janssen Research & Development, LLC, Turnhoutseweg 30, Beerse 2340, Belgium

ARTICLE INFO

Keywords:
Pharmacoepidemiology
Phenotype
Algorithms
Vaccination
Coronavirus (COVID-19)

ABSTRACT

Objective: To complement and support routine pharmacovigilance, Janssen conducted rapid real-world data analyses for near real-time safety monitoring of the Janssen COVID-19 vaccine and to contextualize potential safety signals.

Methods: Analyses were performed in four U.S. healthcare claims databases (February 2022–May 2023) using standardized algorithms for three vaccine exposures, 56 outcomes, and 93 negative controls. Three self-controlled case series and two comparative cohort variants were conducted, each with consideration of multiple at-risk periods following vaccination. Only results that passed pre-determined, standardized diagnostics were unblinded. Two evidence interpretation strategies were employed: 1) Discovery: aimed to support discovering potentially unknown associations for further investigation, correcting for multiple testing and sequential looks over time. 2) Estimation: aimed to quantify the strength of association for specific exposure-outcome pairs and assess statistical uncertainty.

Results: A total of 13 outcomes of interest showed results exceeding the prespecified Discovery threshold. Guillain-Barré Syndrome (GBS) and Bell's palsy had the most consistent signaling over time, analytic methods, and data sources. GBS, an adverse drug reaction that was added to the product information in August 2021, is used as the example to demonstrate the aspects of this rapid analytic framework. Estimation results for GBS were consistent, with effect estimates in the 1–28 day risk window ranging from an incidence rate ratio of 4.0 (95 % confidence interval: 2.1–7.7) in a self-controlled design to a hazard ratio of 6.3 (3.0–13.0) in a cohort design. **Conclusions:** This work demonstrates the value and feasibility of conducting rapid cycle analysis across numerous outcomes in multiple databases employing complementary methodologies over successive time points while maintaining scientific integrity. The scalability of the approach is facilitated by the a priori specification of analytic diagnostics and corresponding thresholds, which excludes analyses likely to yield unreliable results, thereby minimizing subjective interpretation and post-hoc rationalization of failed diagnostic tests.

1. Background

After being identified as a local outbreak in China in late 2019, the World Health Organization rapidly recognized Coronavirus 2019 (COVID-19) as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020 and as a pandemic on 11 March 2020 [1]. Only 11 months after the WHO pandemic declaration, the United States

(US) Food and Drug Administration (FDA) issued an Emergency Use Authorization on 27 February 2021 for the Janssen Ad26.COV2-S vaccine for the prevention of COVID-19. This was followed by conditional marketing authorization in the European Union granted on 11 March 2021 by the European Commission and full or conditional authorization by several other regulatory agencies later. The Janssen COVID-19 vaccine has subsequently been widely administered worldwide. This

* Corresponding author at: 1125 Trenton Harbourton Rd, Titusville, NJ 08560, United States.

E-mail address: mschueme@its.jnj.com (M. Schuemie).

unprecedented and rapid mass vaccination required close and early safety monitoring.

Quantitative pharmacovigilance (PV) approaches to monitoring vaccine safety based on spontaneous reports have limitations, such as under- or overreporting of adverse effects, and the need to estimate exposure rates in external sources capturing vaccine administrations, which might be incomplete or missing [2,3]. In addition, traditional pharmacoepidemiology post-authorization safety studies (PASS) may not generate risk estimates for potential safety signals rapidly enough to address concerns in a timely manner. To overcome these limitations and complement routine PV activities and PASS, rapid safety assessment using real-world data (RWD) from healthcare databases is an alternative to identify and evaluate potential safety signals.

As a part of comprehensive PV activities described in the pharmacovigilance and risk management plans [4] to monitor safety of the Janssen COVID-19 vaccine use under Emergency Use Authorization and (Conditional) Marketing Authorization, Janssen is conducting both routine PV activities and formal PASS. In addition, to support and complement these ongoing safety surveillance activities, the company has conducted rapid and methodologically rigorous real-world data analyses that allowed for near real-time monitoring of the safety of the vaccine and provided additional context for potential safety signals. This framework includes monitoring of Janssen COVID-19 vaccine exposure and sequential analyses in large US real world databases. The aim of this paper is to describe this rapid cycle analytic (RCA) approach and methods used by Janssen.

2. Methods

Following the first Ad26.COV2-S authorization, Janssen launched an initial version of this near real-time real world safety monitoring framework in February 2021 with monitoring in a single US healthcare claims data source and expanded to include four US claims databases in February 2022. Whenever data was refreshed, another 'look' [5] at the data was performed. In total, the data were refreshed four times.

2.1. Data Sources

Janssen first used the HealthVerity COVID-19 Vaccine database because the latency and periodic updates of this data overcame lags in COVID-19 vaccine exposure capture in US closed claims data. From February 2022 forward, Janssen added licensed closed claims data sources including Merative™ (CCAE) MarketScan® Commercial Database, Optum's de-identified Clininformatics® Data Mart Database, and IQVIA™ Adjudicated Health Plan Claims Data. In late 2022, the HealthVerity COVID-19 Vaccine dataset was removed when licensing

ended and was replaced by the IQVIA Pharmetrics Plus Adjudicated Health Plan Claims Data when it was licensed and observed to have sufficient Janssen COVID-19 vaccine exposure. Appendix 1 includes full details of these databases, including descriptions of the data and high-level database characterizations.

When data owners provide updates to databases, they are converted to Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) v5.4 [6] utilizing the Observational Health Data Sciences and Informatics (OHDSI) standardized vocabularies prior to analysis [7]. Databases are characterized and evaluated for quality (Fig. 1). The OHDSI Data Quality Dashboard [8] is used to perform over 3000 data quality checks.

The use of the Merative Marketscan and Optum claims databases was reviewed by the New England Institutional Review Board (IRB) and was determined to be exempt from broad IRB approval, as this research project did not involve human subjects research. Like those databases HealthVerity is deidentified and is HIPAA-compliant, therefore making it non-human subjects research.

2.2. Exposures of Interest

To allow for comparative assessment, three COVID-19 vaccine exposure cohorts were defined based on the first COVID-19 vaccine exposure to either Janssen Ad26.COV2-S (target exposure), Moderna mRNA-1273 (comparator exposure), or Pfizer mRNA-BNT162b2 (comparator exposure). Individuals were included in these COVID-19 vaccine exposure cohorts if they had at least 365 days of observation before and at least one day of observation after this first COVID-19 vaccination starting on or after January 1st, 2021. Cohort entry date was the day of the first vaccine exposure.

2.3. Outcomes of Interest

The outcome list was built using the Safety Platform for Emergency vACCines (SPEAC) COVID-19, [9] and the Biologics Effectiveness and Safety (BEST) Initiative lists of adverse events of special interest (AESI), [10] adding any outcome identified as a potential safety signal in the framework of pharmacovigilance activities. A total of 56 outcomes across multiple body systems (blood / lymphatic, cardiac, immune, and nervous / central) were included and are listed in Table 1.

To develop computable phenotype algorithms for outcomes of interest, a consistent process was followed that relied on clinical descriptions for each outcome and utilized validated definitions identified through systematic literature reviews, when available. When available, existing case definitions and diagnostic codes, such as those from the Brighton Collaboration, US FDA or US Centers for Disease Control and

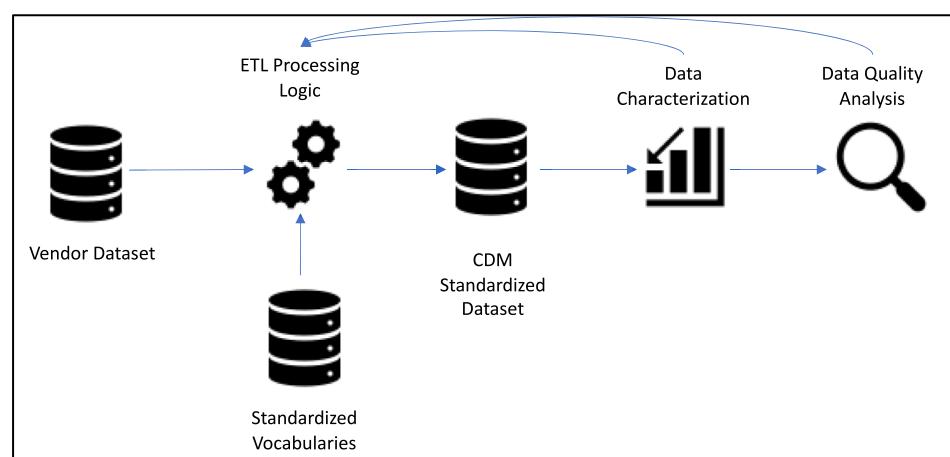


Fig. 1. Data standardization, quality, and characterization pipeline.

Table 1
List of Outcomes of Interest.

Outcome of Interest	Abbreviated Name	Clean Window*
Acute Aseptic Arthritis	Acute Aseptic Arthritis	365
Acute Hepatic Injury	Acute Hepatic Injury	All time prior
Acute Kidney Injury	Acute Kidney Injury	365
Acute Myocardial Infarction	Acute MI	180
Anaphylaxis	Anaphylaxis	30
Arrhythmia (Broad)	Arrhythmia (Broad)	All time prior
Arrhythmia (Narrow)	Arrhythmia (Narrow)	All time prior
Autoimmune Hepatitis	Autoimmune Hepatitis	365
Autoimmune Thyroiditis	Autoimmune Thyroiditis	All time prior
Bell's palsy	Bell's palsy	180
Cerebral Venous Thrombosis	CVT	365
Cerebral Venous Thrombosis with thrombocytopenia	CVT w/ thrombocytopenia	365
Cerebral Venous Thrombosis without thrombocytopenia	CVT w/o thrombocytopenia	365
Composite: Coronary Artery Disease or Non-Hemorrhagic Stroke	CAD or non-H stroke	All time prior for CAD / 365 for non-H stroke
Composite: Deep Vein Thrombosis or Pulmonary Embolism	DVT or PE	365
Composite: DRESS, EM, SJS, TEN (SCAR)	SCAR	365
Composite: Hemorrhagic or Non-Hemorrhagic Stroke	Any stroke	365
Composite: Myocarditis or Pericarditis	Myocarditis or Pericarditis	365
Composite: Thrombosis with Thrombocytopenia (Venous or Arterial or Unspecified vessel type)	Any TWT	365
Coronary Artery Disease, Including Acute Myocardial Infarction	CAD	All time prior
Deep Vein Thrombosis	DVT	365
Deep Vein Thrombosis with thrombocytopenia	DVT w/ thrombocytopenia	365
Deep Vein Thrombosis without thrombocytopenia	DVT w/o thrombocytopenia	365
Disseminated Intravascular Coagulation	DIC	365
Drug reaction with eosinophilia and systemic symptoms (DRESS) syndrome	DRESS	365
Encephalitis, Including ADEM & Meningoencephalitis	Encephalitis	180
Erythema multiforme	Erythema multiforme	365
Generalized Convulsion (Broad Seizure)	Seizure (Broad)	42
Generalized Convulsion (Narrow Seizure)	Seizure (Narrow)	42
Guillain-Barré syndrome	Guillain-Barré syndrome	365
Heart Failure	Heart Failure	All time prior
Hemorrhagic Stroke	Hemorrhagic Stroke	180
Immune Thrombocytopenia or Hemolytic Uremic syndrome	ITP or HUS	365
Microangiopathy	Microangiopathy	365
Minimal Change Disease Nephrotic syndrome	MCD	365
Multiple Sclerosis or Optic Neuritis	MS or OP	All time prior for MS / All ON events
Multisystem inflammatory syndrome (MIS)	MIS	365
Myocarditis	Myocarditis	365
Narcolepsy	Narcolepsy	365
Neuralgic Amyotrophy	Neuralgic Amyotrophy	365
Non-Hemorrhagic Stroke	Non-Hemorrhagic Stroke	180
Pericarditis	Pericarditis	365
Peripheral Arterial Thrombosis	PAT	365
Pulmonary Embolism	Pulmonary Embolism	365

Table 1 (continued)

Outcome of Interest	Abbreviated Name	Clean Window*
Pulmonary Embolism with thrombocytopenia	PE w/ thrombocytopenia	365
Pulmonary Embolism without thrombocytopenia	PE w/o thrombocytopenia	365
Sensorineural Hearing Loss	SHL	365
Small vessel vasculitis	Small vessel vasculitis	365
Stevens-Johnson syndrome (SJS) or Toxic Epidermal Necrolysis (TEN)	SJS or TEN	365
Stress Cardiomyopathy	Stress Cardiomyopathy	365
Thrombocytopenia	Thrombocytopenia	90
Thrombosis with Thrombocytopenia (Arterial)	TWT (Arterial)	365
Thrombosis with Thrombocytopenia (Venous)	TWT (Venous)	365
Thrombosis with Thrombocytopenia (Vessel Type Unspecified)	TWT (Vessel Type Unspecified)	365
Transverse Myelitis	Transverse Myelitis	365
Type 1 Diabetes Mellitus	Type 1 Diabetes Mellitus	All time prior

* Clean window: If a person has multiple occurrences of a specific outcome, occurrences after the first occurrence are not counted until after the pre-determined outcome-specific period (clean window) has been completed. If no clean window is specified (all time prior), the outcome is not allowed to recur. Subjects were excluded if they had the outcome prior to the COVID-19 vaccine exposure (index date) and the index date was within the outcome clean window.

Prevention (CDC) were considered. An internal panel of clinicians, epidemiologists, and informaticians (AS, PR, CW, GR, RM) reviewed code lists used for outcome phenotype algorithms. The OHDSI CohortDiagnostics 3.2.5 R package [11,12] was employed to review code lists, identify overlaps between different outcome phenotypes within data sources, and generate detailed clinical profile summaries for individuals meeting outcome definitions in the data sources, which helped in the evaluation phenotype algorithm validity.

When possible, performance characteristics for algorithms were estimated with the PheEvaluator R package, which estimates the sensitivity, specificity, and positive predicted value of a specific phenotype algorithm against a diagnostic predictive model as the gold standard [13]. Prior research suggests Phevaluator can produce results similar to case review [14,15]. Appendix 2 provides the steps followed to develop and evaluate outcome phenotype algorithms and further details about outcomes, phenotype algorithms, codes, and performance characteristics.

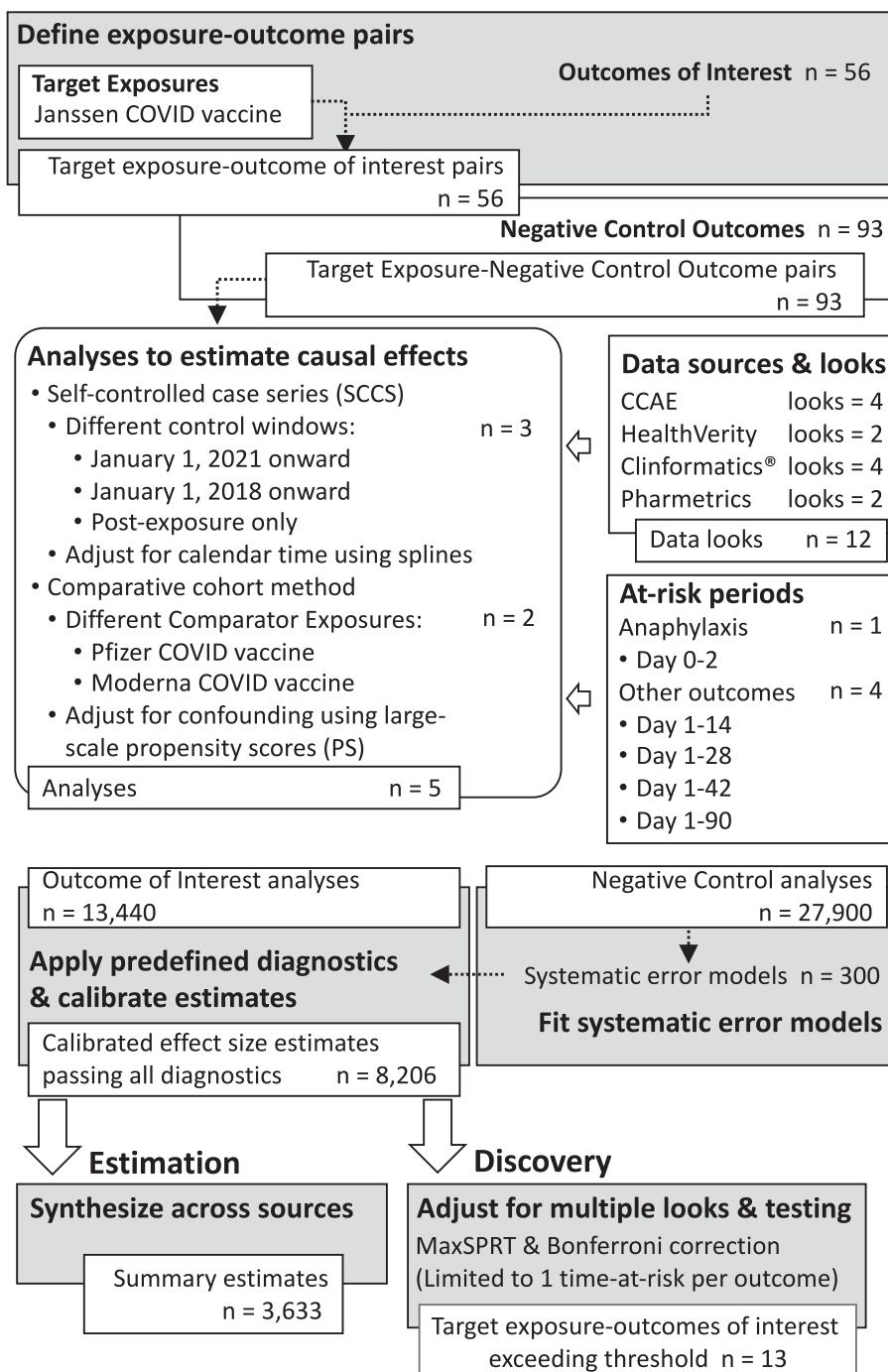
In the remaining part of this manuscript, Guillain-Barré syndrome (GBS), an adverse drug reaction that was added to the Ad26.COV2-S product information in August 2021 [16], is used as the example to illustrate the methods and the results of this rapid analytic framework.

2.4. Negative Control Outcomes

Negative controls were used to assess the validity of the results and to account for residual bias. Negative controls are outcomes believed not to be caused by any of the COVID-19 vaccines and therefore likely would not be flagged as a signal by this safety surveillance system. Effect size estimates for negative controls ideally should be close to the null. Negative control outcomes were defined as the first occurrence of the corresponding negative control SNOMED concept or any of its descendants. The 93 negative controls (Appendix 3) employed have been used previously [17,18].

2.5. Analytic Methods

Over multiple looks at the data, the analysis workflow employed both self-controlled case series (SCCS) and comparative cohort designs (Fig. 2). In both designs, relative risks of outcomes of interest were

**Fig. 2.** Overview of analysis workflow.

The analysis workflow started with definition of exposure-outcome pairs for the Target Exposure (Janssen COVID-19 vaccine) and two sets of outcomes (56 Outcomes of Interest and 93 Negative Control Outcomes). Causal effects for Target Exposure-outcome pairs were estimated using SCCS with three different control windows and a comparative cohort design with two different comparators. These evaluations used outcome-specific times at risk and were applied across four databases with multiple looks over time per database. For Negative Control Outcomes, 27,900 analyses were conducted (93 outcomes \times 5 analytic analyses \times 12 data looks \times 5 at-risk periods). Using the negative control estimates, we fit 300 systematic error models (5 analytic analyses \times 12 data looks \times 5 at-risk periods) that were applied as one of the diagnostics and to calibrate effect size estimates. For Target Exposure-Outcome of Interest pairs, we conducted 13,440 analyses (55 outcomes \times 5 analytic analyses \times 12 data looks \times 4 at-risk periods +1 outcome (anaphylaxis) \times 5 \times 12 \times 4). In total, 8206 analyses for Outcomes of Interest passed all diagnostics thus producing calibrated effect-size estimates. This result set was used to produce 13 target-outcome pairs exceeded the predefined threshold for at least one at-risk period after adjustment for multiple looks and testing (Discovery) and 3633 meta-analytic effect-size estimates (Estimation). CCAE = Merative™ MarketScan® Commercial Database, HealthVerity = HealthVerity COVID-19 Vaccine Database, Clinformatics® = Optum's de-identified Clinformatics® Data Mart Database, Pharmetrics = IQVIA™ Adjudicated Health Plan Claims Data, Data looks from April 2022 through May 2023.

assessed during four overlapping at-risk periods following vaccination (1–14, 1–28, 1–42 and 1–90 days) except for anaphylaxis, which used an at-risk window starting on the vaccination day (0–2 days). Negative control outcome estimates were used to derive systematic error distribution, which were subsequently applied to empirically calibrate estimates and confidence intervals for Outcomes of Interest. The analytical code to characterize outcome phenotype definitions and to conduct analyses can be found here: <https://github.com/ohdsi-studies/CovidVaccineRapidCycleAnalyses>.

2.6. Self-Controlled Case Series (SCCS)

SCCS methodology [19], using conditional Poisson regression with covariates for at-risk status (exposure), pre-exposure (30 days prior to vaccination), and calendar time, was employed to assess the relative incidence of each specified outcome. This assessment contrasts outcomes occurring during an at-risk period immediately following Janssen COVID-19 vaccination (target Exposure) with outcome events happening in a not at-risk period, which can include time before or after vaccine exposure but excludes the defined at-risk period. Multiple SCCS analyses were conducted for each outcome with variations in analytic period start dates, use of only post-exposure data, and the period defined as at-risk (Appendix 3). All SCCS analyses used splines to adjust for calendar time.

2.7. Comparative Cohort

The comparative cohort analysis employed a Cox proportional hazards model to estimate the hazard ratio for a specified outcome. This estimation involves separate comparisons of the first dose of Janssen vaccination (target exposure) to the first doses of the Moderna and the Pfizer Covid-19 vaccines, which serve as two separate comparator exposure cohorts. For each outcome we defined a prior outcome look-back period, and one or more risk window start and end dates (Appendix 3). To adjust for potential confounding, we employ Large-Scale Propensity Scores (LSPS), a data-driven approach that includes all baseline covariates (between 60,000 and 100,000 variables) observed in the data including all demographics, conditions, drug exposures, and procedures observed in the year prior to vaccination in a regularized regression. Prior research suggests LSPS can identify potential confounders missed by experts [20], is superior to other data-driven approaches such as High-Dimensional Propensity Scores (HDPS) [20], and can adjust for confounders that are only indirectly measured [21]. To avoid including strong instrumental variables (IVs) we evaluate univariable correlations. Moreover, our prior research suggests including IVs is unlikely to lead to bias [22]. We apply variable ratio matching on the propensity score [23].

2.8. Evidence Interpretation Strategies

This work relies on two distinct evidence interpretation strategies (Discovery and Estimation) to support two use cases (Table 2).

The Discovery evidence interpretation strategy supports the use case

Table 2
Differences between Discovery and Estimation Evidence Interpretation Strategies

	Discovery	Estimation
Adjustment for multiple testing	Yes	No
Provides point estimate and 95 % confidence interval	No	Yes
Provides <i>p</i> -value	Yes	No
Motivation	Hypothesis testing	Estimate effect magnitude
Number of analyses	Limited at-risk periods	All at-risk periods

of discovering potentially unknown associations for further investigation. The analysis relies on a large, prespecified set of target exposure and outcome of interest combinations. To limit the number of analyses included in Discovery, a single at-risk period for each outcome was considered. Because of uncertainty in possible time-to-onset, for a limited set of outcomes (GBS, myocarditis, myocarditis or pericarditis, and pericarditis) two at-risk periods were considered (Appendix 3). To correct for multiple testing, a Bonferroni correction was applied across outcomes, analyses, and databases. To correct for sequential testing across multiple looks at the data, the Maximized Sequential Probability Ratio Test (MaxSPRT) combined with empirical calibration by negative controls were applied [24]. A threshold was computed aiming at a family-wise type 1 error rate of 50 % to minimize the number of false negatives. Only those analyses that exceeded the predefined Discovery threshold were reviewed.

The Estimation evidence interpretation strategy fits the use case of quantifying the strength of association for a specific target exposure-outcome of interest pair and assesses statistical uncertainty. Estimation results were only reviewed for exposure-outcome pairs for which either Discovery or external evidence such as spontaneous reports suggested a potential effect. Effect-size estimates were combined across databases using Bayesian random-effects meta-analysis with non-normal likelihood approximation to avoid bias due to small counts [25]. Empirical calibration for the Estimation strategy was performed by 1) computing meta-analytic estimates for all negative control outcomes, 2) using these estimates to fit empirical null distributions, and 3) applying these to calibrate the meta-analytic point estimates, confidence intervals, and *p*-values for the outcomes of interest. Only effect-size estimates that passed pre-determined analytic diagnostic thresholds (see Analytic Diagnostics section) were included in the meta-analysis.

2.9. Analytic Diagnostics

For results to be unblinded, analyses had to pass a set of pre-determined, standardized diagnostics. Each diagnostic has a failure threshold, which prevents unblinding if not met. Diagnostics are briefly described below and detailed further in Appendix 4.

All analyses underwent the following analytic diagnostics:

1. Power: The Minimum Detectable Relative Risk (MDRR) for a given observed sample size (after applying analytic exclusions) using an $\alpha = 0.05$, $\beta = 0.20$ was calculated. The diagnostic failure threshold is an MDRR value of 10 or higher, rejecting outcomes with so little data that there's less than 20 % chance of identifying a relative risk of 10 (at a regular significance level of 5 %).
2. Systematic error: A systematic error distribution using negative control estimates [26] was fit and summarized as the Expected Absolute Systematic Error (EASE) (see Appendix 4). An EASE of 0 means all variation in negative control estimates can be explained by random error. The diagnostic failure threshold is an EASE value of 0.25 or higher. When EASE = 0.25 and systematic error is centered on 0, a true relative risk of 1 has a 95 % probability of appearing to be anywhere between 0.54 and 1.85 due to systematic error.

For the SCCS design, these additional diagnostics were employed:

1. Reverse causality: We tested if the risk of the outcome was already increased immediately prior to vaccine exposure (30 days before versus 30 days after). The diagnostic failure threshold for reverse causality test is a *p*-value <0.01 , meaning there is less than 1 % chance of seeing such a high rate of the outcome prior to vaccination (relative to after vaccination) when the rates before and after are in fact the same.
2. Time Trend: Within-subject confounding can occur when the rate of an outcome changes as a function of calendar time, which can occur due to data capture lags (incurred but not documented events are

expected to occur close to the date of data extraction). This diagnostic assesses whether adjustment using a spline function over calendar time in the SCCS design was sufficient to control for these time trends. Monthly outcome rates after spline adjustment were computed and separately compared to the unadjusted mean monthly rate. A family-wise alpha of 0.05 was used, and a Bonferroni correction for the number of months was applied. If at least one monthly rate was statistically different from the mean monthly rate, then this diagnostic failed.

For the comparative cohort design, the following additional diagnostic were used:

1. Attrition: Subjects could be excluded from the Janssen COVID-19 vaccine cohort, for example because they also received the comparator vaccine, did not have a match in the comparator cohort, or did not have the required exposure clean window. The attrition diagnostic would fail if more than 50 % of the Janssen COVID-19 cohort was excluded.
2. Equipoise: A preference score, the linear transformation of the propensity score [27], was calculated. Then an equipoise measure was computed as the percent of the population with a preference score between 0.3 and 0.7. The diagnostic failure threshold is an equipoise value of less than 10 %, meaning less than 10 % of the population was somewhat likely to receive either vaccine based on their baseline characteristics
3. Covariate Balance: We computed the standardized difference of mean (SDM) for every covariate used to balance any two exposure groups. The diagnostic failed when the absolute SDM of any covariate was 0.1 or higher, meaning we would stop if any of the 60,000 to 100,000 covariates was out of balance between the two exposure groups after propensity score matching.

3. Results

The four data looks that assessed the risk of 56 different outcomes of interest for first exposures to the Janssen COVID-19 vaccine in four US claims datasets were conducted between April 2022 and May 2023. The number of persons identified with their first exposure to Janssen vaccine differed across data sources with the largest population identified in the HealthVerity COVID-19 Vaccine database followed by IQVIA, Merative (CCAE), and Clininformatics® (Fig. 3). Clininformatics® had the largest percentage increase in Janssen vaccine exposure over the four

sequential looks (25 %) whereas Merative (CCAE) had the largest absolute increase (59,742 added exposures). Janssen exposures in Merative (CCAE) decreased between the third and fourth sequential look due to changes in insurance plan contract coverage and additional capture of non-Janssen vaccine exposures reducing the number of eligible subjects.

In the most recent data look for each database, over 90 % of Janssen recipients were between 20 and 64 years old with slightly lower numbers of those aged 20 to 44 years than those aged 45 to 64 years (Fig. 4). Merative (CCAE) claims stop at the 65-to-69-year age group because this dataset is based on claims for working-age adults. Just under half of recipients in each database (43.6–47.0 %) were female. These characteristics did not change significantly between looks (data not shown).

Fig. 5 includes the number of analyses that passed the pre-determined analytic diagnostics (Appendix 4) for the 56 Outcomes of Interest assessed across databases in the four data looks. Of the 13,440 analyses conducted for these outcomes, 8206 passed all analytic diagnostics. CVT with thrombocytopenia was the only outcome with analyses that failed analytic diagnostics across all databases and data looks. The analytic diagnostics that accounted for the most failures were the MDRR in the comparative cohort design and Time Trend in SCCS. Appendix 4 includes detailed examples of analytic diagnostics for the GBS outcome.

Only the 13 outcomes of interest highlighted in Fig. 6 (values greater than 0) exceeded the prespecified Discovery threshold. Bell's palsy and GBS were the outcomes with the highest number of analyses reviewed under the Discovery strategy.

Fig. 7 demonstrates the results available for the Estimation evidence interpretation strategy for the GBS outcome. Results are suppressed when prespecified analytic diagnostics were not passed (e.g. at-risk period = 1–14 days, Clininformatics®, all analyses). Effect estimates are elevated across methods and at-risk periods (times at risk) but are highest for the 1–28 and 1–42 days at risk.

4. Discussion

In this effort to support PV activities for the Janssen COVID-19 vaccine, we demonstrated the viability of conducting RCA on real-world data for a number of outcomes using complementary methodologies and databases over successive time points while maintaining scientific integrity. We adhered to best practices in each analytic approach employing methods to mitigate confounding factors including the use of

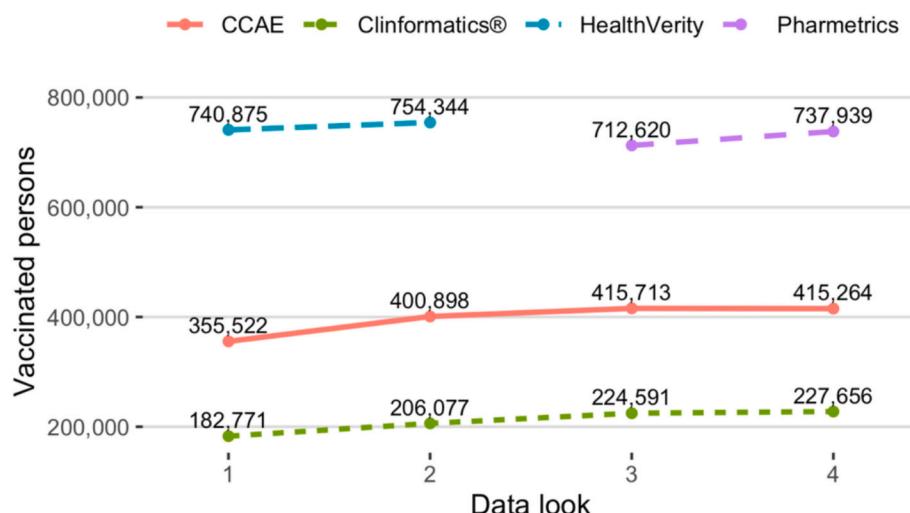


Fig. 3. Count of persons with first Janssen COVID-19 vaccine exposures by database and data look.

CCAE = Merative™ MarketScan® Commercial Database, HealthVerity = HealthVerity COVID-19 Vaccine Database, Clininformatics® = Optum's de-identified Clininformatics® Data Mart Database, Pharmetrics = IQVIA™ Adjudicated Health Plan Claims Data, Data looks from April 2022 through May 2023.

	CCAE	HealthVerity	Clinformatics®	Pharmetrics
Age 15 - 19	2.8%	3.2%	2.5%	2.6%
Age 20 - 24	8.6%	8.8%	7.8%	8.4%
Age 25 - 29	6.3%	6.7%	6.9%	6.5%
Age 30 - 34	7.8%	8.0%	8.9%	7.9%
Age 35 - 39	9.4%	8.9%	9.9%	8.9%
Age 40 - 44	10.6%	9.4%	10.4%	9.8%
Age 45 - 49	11.5%	9.7%	10.6%	10.4%
Age 50 - 54	13.6%	11.7%	12.4%	12.7%
Age 55 - 59	14.3%	13.2%	12.7%	14.1%
Age 60 - 64	13.6%	13.6%	11.3%	13.8%
Age 65 - 69	1.4%	3.9%	3.6%	3.6%
Age 70 - 74		1.5%	1.4%	0.7%
Age 75 - 79		0.8%	0.7%	0.2%
Age 80 - 84		0.4%	0.4%	0.3%
Age 85 - 89			0.3%	0.0%
Age 90 - 94		0.3%	0.0%	
Female	45.3%	47.0%	43.6%	44.8%

Fig. 4. Characteristics of persons with first Janssen COVID-19 vaccine exposure for each database in its final data look.

CCAE = Merative™ MarketScan® Commercial Database, HealthVerity = HealthVerity COVID-19 Vaccine Database, Clininformatics® = Optum's de-identified Clininformatics® Data Mart Database, Pharmetrics = IQVIA™ Adjudicated Health Plan Claims Data.

large-scale propensity scores and self-controlled designs with calendar time adjustments via splines. The scalability of our approach is facilitated by a priori specification of analytic diagnostics and corresponding thresholds, which excludes analyses likely to yield unreliable results, thereby minimizing subjective interpretation and post-hoc rationalization of these diagnostics. Other teams have published rapid RWD analyses on the safety of COVID-19 vaccines including Janssen's, but these approaches used few or underpowered data sources [28,29], historical comparators [30–32], no prespecified diagnostics [29–33], or one analytic approach [28–34].

Contrary to the conventional two-step process of signal detection using a simple, often unadjusted method followed by effect estimation using a more sophisticated design [35], our approach uses best-practice epidemiological methods, including advanced confounding adjustment, to generate the most dependable estimates from the data, which are then utilized for both discovery and estimation purposes. The results of this RCA demonstrate the capacity to independently confirm identified adverse events, such as GBS, indicative of high sensitivity while concurrently minimizing the incidence of false positives, indicative of high specificity. This balance underscores the effectiveness of RCA to allow reliable PV safety monitoring. Here we used a family-wise alpha of 50 %, meaning we accept a 50 % probability of at least 1 false positive. This is already far higher than used by others [36,37], but an open question remains whether one should accept even higher rates of false positives to avoid missing important signals (false negatives).

Evidence generated through this RCA was integrated as a key component of PV activities for the Janssen COVID-19 vaccine following accelerated clinical development and mass distribution at the height of the pandemic. Following the Emergency Use Authorization in both the US and the EU, Janssen committed with Regulatory Authorities to enhanced monitoring of a list of Adverse Events of Special Interest (AESIs) with both spontaneous reporting and RCA during the post-authorization phase. A signal could, therefore, be triggered by either spontaneous reports received and/or from RCA findings. For other medicinal products, this is typically achieved with spontaneous reporting alone. RCA findings were also considered during signal evaluation alongside other sources of evidence, including clinical trial results and summaries of gathered spontaneous PV information, to assess the likelihood of a causal association between the Outcome of Interest under evaluation and the Janssen COVID-19 vaccine. When considering the

potential causal relationship between the Janssen COVID-19 vaccine and an Outcome of Interest, Janssen considered available evidence using the Hill's Criteria for Causality [38] as a framework to support decision-making. RCA evidence was suited to support consideration of specific elements in the Hill's framework including strength, specificity, coherence, and temporality. Using RCA across multiple databases also assesses consistency but should be considered in concert with all other available data.

During the lifetime of the Janssen COVID-19 vaccine, a total of 41 safety signals were opened covering a variety of topics. Of these, 16 were confirmed, leading to labelling changes. Of the 16 confirmed signals, 11 of them were confirmed based only on clinical trial and post-marketing spontaneous data: Facial Paralysis (Bell's Palsy), Capillary Leak Syndrome (CLS), Anxiety-related reactions (including dizziness, paresthesia and loss of consciousness), Diarrhea, Vomiting, Tinnitus, Lymphadenopathy and Thrombosis with Thrombocytopenia. The remaining signals (Myocarditis-Pericarditis, Transverse Myelitis, Venous Thromboembolism, Immune Thrombocytopenia, and Guillain-Barre Syndrome) were confirmed based on disproportionate reporting from spontaneous sources in combination with an increased risk observed in this RWD RCA.

This work was impacted by limitations that commonly affect analysis based on RWD. We relied on both open and closed healthcare claims to try to manage limitations of each source, slow maturity time (up to 6–8 months) for closed claims and missing or duplicate claims in open claims [39]. The rapidity of RCA is highly dependent on the rate of exposure to the analyzed medicinal product. In the case of COVID-19 vaccines, mass vaccination allowed to reach the MDRR threshold for some outcomes relatively rapidly. This may not be the case for products with slower uptake. US claims are unlikely to represent target populations outside of the US, and depending on insurance plans included, they did not represent the US adult population eligible for COVID-19 vaccination. To support real world evidence generation outside of the US, the European Medical Agency is supporting efforts to conduct safety analyses using the Data Analysis and Real World Interrogation Network (Darwin EU) [40]. Overlaps in analyzed databases could not be quantified or resolved due to privacy protections, but database-specific analyses were performed, and summary estimates were based on calculated effect estimates to minimize this impact. The overall analytic design with repeated looks over time did not anticipate the loss of a data source (HealthVerity).

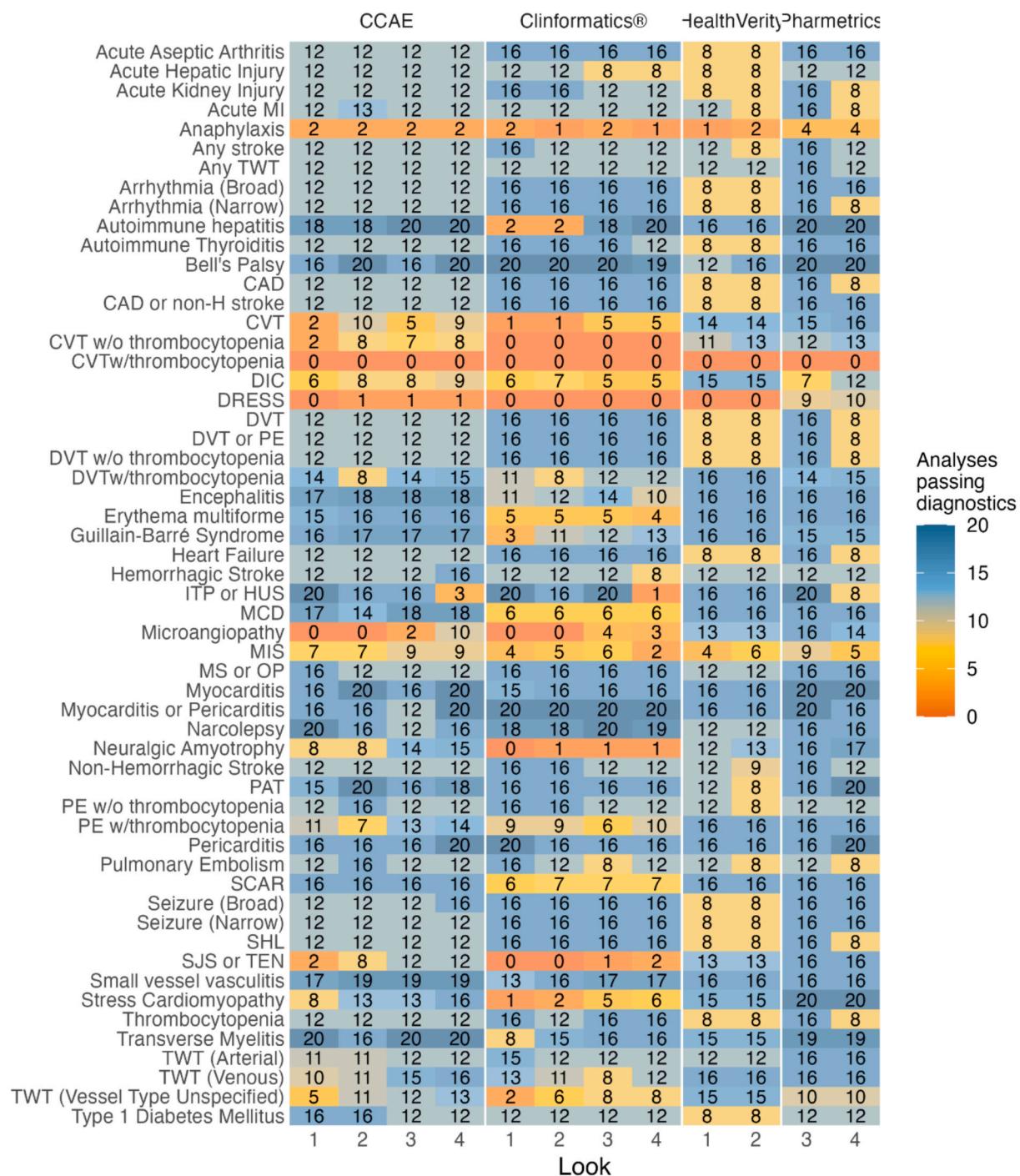


Fig. 5. Number of analyses passing all Analytic Diagnostics per Outcome of Interest, database, and data look.

Full names for Outcomes of Interest are included in Table 1. Details about these outcomes are provided in Appendix 2. Appendix 4 includes detail about applied Analytic Diagnostics. CCAE = Merative™ MarketScan® Commercial Database, HealthVerity = HealthVerity COVID-19 Vaccine Database, Clininformatics® = Optum's de-identified Clininformatics® Data Mart Database, Pharmetrics = IQVIA™ Adjudicated Health Plan Claims Data, Data looks from April 2022 through May 2023.

Future iterations of this framework will consider such impacts. COVID-19 vaccine administration in mass vaccination programs could be missing (exposure misclassification) in claims data sources, however, this impact would likely be non-differential and bias effect estimates toward the null. The size of the Janssen COVID-19 vaccine cohort was limited relative to mRNA cohorts due to lower Janssen vaccine coverage, which affects precision around effect estimates, especially for very rare outcomes (<1/10,000 vaccinations). With 75 % of US adults having received at least one dose of COVID-19 vaccine by late September 2021 and over 50 % receiving booster doses by late May

2022 [41], the opportunity to monitor first dose safety declined rapidly over time. Also with limited recommendation for Janssen as a booster dose [42] and related minimal uptake [41], this work was not expanded to include booster doses, which could have different effect sizes or associations with different outcomes than the primary series. Risks within population subgroups could not be assessed within this framework as designed. Ad hoc analyses were conducted when age- and sex-population-specific risks were identified for COVID-19 vaccines (myopericarditis) [43]. A final limitation involved our example of GBS. Because this outcome was already suspected to be caused by several of

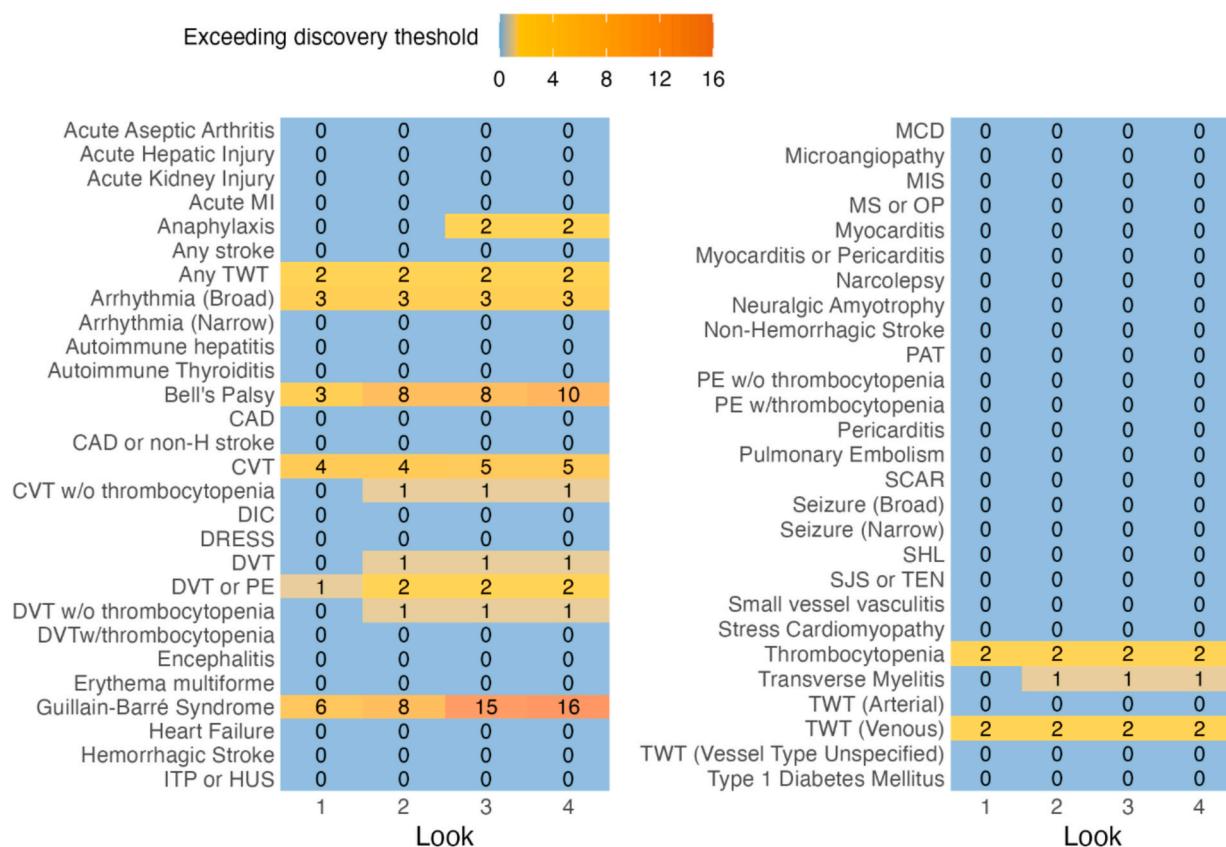


Fig. 6. Analyses exceeding the predefined Discovery Evidence interpretation strategy threshold per Outcome of Interest and data look.

Details about Outcomes of Interest are provided in Appendix 2 and the Discovery Evidence Interpretation strategy in Appendix 3 Section 3. For each outcome, a total of 20 analyses ([3 SCCS +2 cohort method] * 4 databases) were considered for the Discovery Evidence Interpretation strategy. Guillain-Barré syndrome, Myocarditis, Myocarditis or Pericarditis, and Pericarditis had two times-at-risk considered instead of one. The total number of analyses therefore was $2^*20 = 40$.

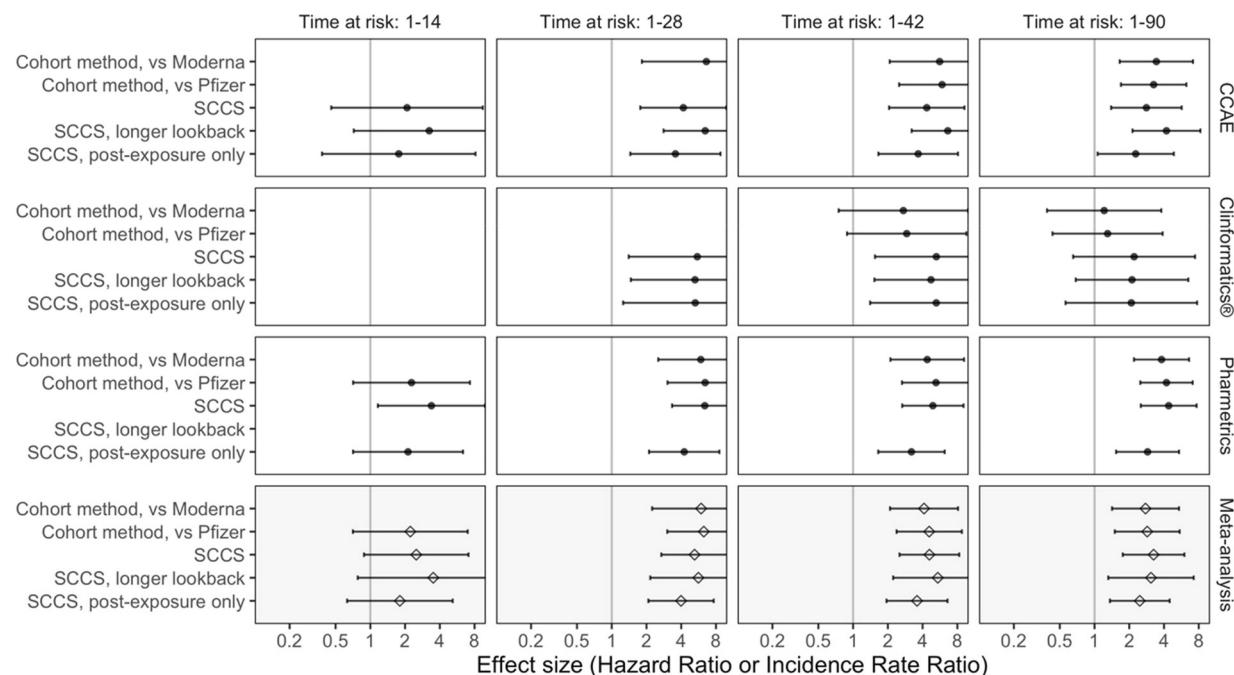


Fig. 7. Effect size estimates across analyses, at-risk periods (times at risk), and databases for Guillain-Barré syndrome at fourth data look.

Details about the Guillain-Barré syndrome outcome are provided in Appendix 2. CCAE = Merative™ MarketScan® Commercial Database, HealthVerity = HealthVerity COVID-19 Vaccine Database, Clininformatics® = Optum's de-identified Clininformatics® Data Mart Database, Pharmetrics = IQVIA™ Adjudicated Health Plan Claims Data, SCCS = Self-controlled case series, Data looks from April 2022 through May 2023

the COVID-19 vaccines early on, detection bias could have occurred; GBS is hard to diagnose, [44] and awareness of the potential side-effect could have increased the sensitivity with which this outcome could be detected in the claims data. Possibly, without this potential bias, this side-effect would have been detected later.

In September 2021, the FDA released draft guidance highlighting the potential impacts of exposure and outcome misclassification in real-world data and recommended medical chart review for validated analysis to control for these effects [45]. In the context of rapid, large-scale vaccination, a tradeoff between accuracy and timeliness for this work had to be made. PheValuator was used to assess performance of most outcome phenotype algorithms across a number of data sources [14,15]. Other approaches to validate analyses, such as other rapid validation methods and sensitivity analyses [46], and possibly pre-specifying minimum thresholds on phenotype sensitivity and specificity need to be further assessed to support their use and acceptance for regulatory-grade real-world evidence generation.

5. Conclusions

Through this work, we demonstrate the value and feasibility of assessing vaccine safety through the conduct of rapid cycle analyses on numerous outcomes in multiple large real-world databases over successive time points. Every aspect of this work, standardization to a common data model, evidence-based development of computable phenotype algorithms, a priori specification of analytic diagnostics, application of complementary advanced analytic methods, and two evidence interpretation strategies, was designed to readily support pharmacovigilance activities with minimal subjective interpretation.

CRediT authorship contribution statement

Atta Shoaibi: Writing – review & editing, Validation, Software, Conceptualization. **Kevin Haynes:** Writing – review & editing, Writing – original draft, Conceptualization. **Clair Blacketer:** Writing – review & editing, Writing – original draft, Data curation. **Corinne Willame:** Writing – review & editing, Methodology, Conceptualization. **Frank DeFalco:** Writing – review & editing, Software, Data curation. **Gowtham A. Rao:** Writing – review & editing, Validation. **Kourtney Davis:** Writing – review & editing, Methodology, Conceptualization. **Luis Anaya Velarde:** Writing – review & editing, Methodology, Conceptualization. **Nicolas Praet:** Writing – review & editing, Methodology, Conceptualization. **Rupa Makadia:** Writing – review & editing, Writing – original draft, Data curation. **Yimei Xu:** Writing – review & editing. **Patrick Ryan:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Martijn Schuemie:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Within the last 3 years, all authors have any personal financial interests or professional relationships with Janssen Pharmaceuticals. All authors were employees of Janssen at this work. The following authors owned stock in Janssen: ECN, NP, KH, KJD, LA, RM, YX, CW. This research received NO additional support from an organization beyond the authors' academic institutions. There are NO patents or copyrights licensed to the author(s) that are relevant to the work submitted for publication to disclose.

There are NO other relationship or activity for authors that may be interpreted as a conflict of interest by the reader.

Acknowledgements

We would like to acknowledge the support of Gayle Murray, who performed literature reviews for Outcomes of Interest case definitions and any existing computable phenotypes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.vaccine.2025.127044>.

Data availability

While we are not allowed to share patient data used in the analyses described here, we did provide all code and other artifacts allowing others with access to the same data to reproduce our results

References

- [1] WHO EURO Region EURO. Coronavirus disease (COVID-19) pandemic - Europe. Dec 2023. <https://www.who.int/europe/emergencies/situations/covid-19>; 2023.
- [2] Hazelz L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29:385–96. <https://doi.org/10.2165/00002018-200629050-00003>.
- [3] Prevots DR, Sutter RW, Strebel PM, Weibel RE, Cochi SL. Completeness of reporting for paralytic poliomyelitis, United States, 1980 through 1991. Implications for estimating the risk of vaccine-associated disease. Arch Pediatr Adolesc Med 1994; (148):479–85. <https://doi.org/10.1001/archpedi.994.02170050037007>.
- [4] European Medicines Agency EMA. European Union Risk Management Plan JCQVDEN (VAC31518 [Ad26.COV2.S]). 2023.
- [5] Bu F, Schuemie MJ, Nishimura A, Smith LH, Kostka K, Falconer T, et al. Bayesian safety surveillance with adaptive bias correction. Stat Med 2024;43:395–418.
- [6] OHDSI. OMOP Common Data Model. Dec 2023, <https://ohdsi.github.io/CommonDataModel/cdm54.html>; 2022.
- [7] Reich C, Ostrompolo A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI standardized vocabularies—a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc., 2024;ocad247: doi: <https://doi.org/10.1093/jamia/ocad247>.
- [8] Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. J Am Med Inform Assoc 2021;28:2251–7. <https://doi.org/10.1093/jamia/ocab132>.
- [9] Brighton Collaboration. Priority list of adverse events of special interest for COVID-19. 2020.
- [10] Food and Drug Administration. Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring. 2021.
- [11] Gilbert J, Rao G, Schuemie M, Ryan P, Weaver J. CohortDiagnostics 3.2.5. Dec 2023, <https://ohdsi.github.io/CohortDiagnostics/>; 2023.
- [12] Rao GA, Shoaibi A, Makadia R, Hardin J, Swerdel J, Weaver J, et al. CohortDiagnostics: phenotype evaluation across a network of observational data sources using population-level characterization. PLoS One 2025;20:e0310634.
- [13] Swerdel JN. PheValuator: Rapid Phenotype Evaluation. Dec 2023, <https://ohdsi.github.io/PheValuator/>; 2023.
- [14] Swerdel JN, Schuemie M, Murray G, Ryan PB. PheValuator 2.0: methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. J Biomed Inform 2022;135:104177. <https://doi.org/10.1016/j.jbi.2022>.
- [15] Swerdel JN, Hripcak G, Ryan PB. PheValuator: development and evaluation of a phenotype algorithm evaluator. J Biomed Inform 2019;97:103258. <https://doi.org/10.1016/j.jbi.2019>.
- [16] Food and Drug Administration. Coronavirus (COVID-19) Update: July 13, 2021. <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-july-13-2021>; 2021.
- [17] Lai LY, Arshad F, Areia C, Alshammari TM, Alghoul H, Casajust P, et al. Current approaches to vaccine safety using observational data: a rationale for the EUAEUS (evaluating use of methods for adverse events under surveillance-for vaccines) study design. Front Pharmacol 2022;13:837632. <https://doi.org/10.3389/fphar.2022>.
- [18] Schuemie MJ, Arshad F, Pratt N, Nyberg F, Alshammari TM, Hripcak G, et al. Vaccine safety surveillance using routinely collected healthcare data—an empirical evaluation of epidemiological designs. Front Pharmacol 2022;13:893484. <https://doi.org/10.3389/fphar.2022>.
- [19] Whitaker HJ, Paddy Farrington C, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med 2006;25:1768–97.
- [20] 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. <https://doi.org/10.1093/ije/dyv120>.
- [21] Zhang L, Wang Y, Schuemie MJ, Blei DM, Hripcak G. Adjusting for indirectly measured confounding using large-scale propensity score. J Biomed Inform 2022; 134:104204.

- [22] Tian Y, Pratt N, Hester LL, Hripcak G, Schuemie MJ, Suchard MA. Evaluating the impact of instrumental variables in propensity score models using synthetic and negative control experiments. arXiv. 2024.
- [23] Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl. 2):69–80. <https://doi.org/10.1002/pds.3263>.
- [24] Schuemie MJ, Bu F, Nishimura A, Suchard MA. Adjusting for both sequential testing and systematic error in safety surveillance using observational data: empirical calibration and MaxSPRT. *Stat Med* 2023;42:619–31. <https://doi.org/10.1002/sim.9631>.
- [25] Schuemie MJ, Chen Y, Madigan D, Suchard MA. Combining cox regressions across a heterogeneous distributed research network facing small and zero counts. *Stat Methods Med Res* 2022;31:438–50. <https://doi.org/10.1177/09622802211060518>.
- [26] Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014;33:209–18. <https://doi.org/10.1002/sim.5925>.
- [27] Walker A, Patrick A, Lauer M, Hornbrook M, Marin M, Platt R, et al. A tool for assessing the feasibility of comparative effectiveness research. *Comparative Effectiveness Research* 2013;2013:11–20. <https://doi.org/10.2147/CER.S40357>.
- [28] Li X, Raventos B, Roel E, Pistillo A, Martinez-Hernandez E, Delmestri A, 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. <https://doi.org/10.1136/bmj-2021->
- [29] Hanson KE, Goddard K, Lewis N, Fireman B, Myers TR, Bakshi N, et al. Incidence of Guillain-Barre syndrome after COVID-19 vaccination in the vaccine safety datalink. *JAMA Netw Open* 2022;5:e228879. <https://doi.org/10.1001/jamanetworkopen.2022.8879>.
- [30] Wong HL, Tworkoski E, Ke Zhou C, Hu M, Thompson D, Lufkin B, et al. Surveillance of COVID-19 vaccine safety among elderly persons aged 65 years and older. *Vaccine* 2023;41:532–9. <https://doi.org/10.1016/j.vaccine.2022.11.069>.
- [31] Lloyd PC, Hu M, Wong HL, Shoaibi A, Ke Zhou C, Lo AC, 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. <https://doi.org/10.1016/j.vaccine.2022.09.060>.
- [32] Sturkenboom M, Messina D, Paoletti O, de Burgos-Gonzalez A, García-Poza P, Huerta C, et al. Cohort monitoring of 29 Adverse Events of Special Interest prior to and after COVID-19 vaccination in four large European electronic healthcare data sources. medRxiv. <https://www.medrxiv.org/content/10.1101/2022.08.17.22278894v1.full-text>; 2022.
- [33] Botton J, Dray-Spira R, Baricault B, Drouin J, Bertrand M, Jabagi MJ, et al. Reduced risk of severe COVID-19 in more than 1.4 million elderly people aged 75 years and older vaccinated with mRNA-based vaccines. *Vaccine* 2022;40:414–7. <https://doi.org/10.1016/j.vaccine.2021.12.009>.
- [34] Li X, Burn E, Duarte-Salles T, Yin C, Reich C, Delmestri A, 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* 2022;379:e071594. <https://doi.org/10.1136/bmj-2022->
- [35] Arshad F, Schuemie MJ, Bu F, Minty EP, Alshammari TM, Lai LYH, et al. Serially combining epidemiological designs does not improve overall signal detection in vaccine safety surveillance. *Drug Saf* 2023;46:797–807. <https://doi.org/10.1007/s40264-023-01324-1>.
- [36] Lloyd PC, Hu M, Wong HL, Shoaibi A, Ke Zhou C, Lo AC, 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.
- [37] Wong HL, Tworkoski E, Ke Zhou C, Hu M, Thompson D, Lufkin B, et al. Surveillance of COVID-19 vaccine safety among elderly persons aged 65 years and older. *Vaccine* 2023;41:532–9.
- [38] Rothman KJ, Greenland S. Encyclopedia of Biostatistics. 2005. <https://doi.org/10.1002/0470011815.b2a03072>.
- [39] Baser O, Samayoa G, Yapar N, Baser E, Mete F. Use of open claims vs closed claims in health outcomes research. *J Health Econ Outcomes Res* 2023;10:44–52. <https://doi.org/10.36469/001c.87538>.
- [40] European Medicines Agency EMA. DARWIN EU® has completed its first studies and is calling for new data partners. Dec 2023, <https://www.ema.europa.eu/en/news/darwin-eur-has-completed-its-first-studies-and-calling-new-data-partners>; 2023.
- [41] US Centers for Disease Control and Prevention CDC. COVID-19 vaccinations in the United States, Jurisdiction. 2023; 2024. <https://data.cdc.gov/Vaccinations/Covid-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc>.
- [42] Mbaeyi S, Oliver SE, Collins JP, Godfrey M, Goswami ND, Hadler SC, et al. The advisory committee on immunization Practices' interim recommendations for additional primary and booster doses of COVID-19 vaccines - United States, 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:1545–52. <https://doi.org/10.15585/mmwr.mm7044e2>.
- [43] US Centers for Disease Control and Prevention CDC. Clinical Considerations: Myocarditis and Pericarditis after Receipt of COVID-19 Vaccines Among Adolescents and Young Adults. Feb 2024, <https://www.cdc.gov/vaccines/covid-19/clinical-considerations/myocarditis.html>; 2023.
- [44] Feit H, Tindall RS, Glasberg M. Sources of error in the diagnosis of Guillain-Barre syndrome. *Muscle Nerve* 1982;5:111–7.
- [45] US Federal Drug Administration FDA. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory DecisionMaking for Drug and Biological Products: Guidance for Industry. 2021; 2024. <https://www.fda.gov/media/152503/download>.
- [46] Layton JB, Ziemiczki R, Danysh HE, Gaber C, Gilsenan A, Johannes CB. Graphical representation of an array of quantitative bias analysis scenarios for differential outcome misclassification. ICPE. Virtual. <https://www.rtihs.org/publications/graphical-representation-array-quantitative-bias-analysis-scenarios-differential>; 2021.