Development of Interoperable Computable Phenotype Algorithms for Adverse Events of Special Interest to Be Used for Biologics Safety Surveillance: Validation Study

Ashley Holdefer, Jeno Pizarro, Patrick Saunders-Hastings, Jeffrey Beers, Arianna Sang, Aaron Zachary Hettinger, H. Joseph Blumenthal, Erik Martinez, Lance Daniel Jones, Matthew Deady, Hussein Ezzeldin, Steven Anderson

Submitted to: JMIR Public Health and Surveillance
on: June 09, 2023

Table of Contents

Development of Interoperable Computable Phenotype Algorithms for Adverse Events of Special Interest to Be Used for Biologics Safety Surveillance: Validation Study

Ashley Holdefer[1] MS; Jeno Pizarro[1] BBA; Patrick Saunders-Hastings[2] PhD; Jeffrey Beers[1] MD; Arianna Sang[1] BA; Aaron Zachary Hettinger[3, 4] MD, MS; H. Joseph Blumenthal[3] BA; Erik Martinez[1] MBA; Lance Daniel Jones[1] MA; Matthew Deady[1] BA; Hussein Ezzeldin[5] PhD; Steven Anderson[5] PhD

[1]IBM Consulting IBM Bethesda US
[2]Accenture Inc. Ottawa CA
[3]Center for Biostatistics, Informatics and Data Science MedStar Health Research Institute MedStar Health Columbia US
[4]Department of Emergency Medicine Georgetown University School of Medicine Washington US
[5]Center for Biologics Evaluation and Research United States Food and Drug Administration Silver Spring US

Corresponding Author:
Hussein Ezzeldin PhD
Center for Biologics Evaluation and Research
United States Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring
US

Abstract

Background: Adverse events (AEs) associated with vaccination have been evaluated by epidemiological studies, and more recently gained additional attention with the Emergency Use Authorization (EUA) of several COVID-19 vaccines. As part of its responsibility to conduct post-market surveillance, the U.S. Food and Drug Administration (FDA) continues to monitor several adverse events of special interest (AESIs) to ensure vaccine safety, including for COVID-19.

Objective: This study is part of the Biologics Effectiveness and Safety (BEST) Initiative, which aims to improve FDA's post-market surveillance capabilities while minimizing public burden. This study looks to enhance active surveillance efforts through a rules-based, computable phenotype algorithm to identify five AESIs: anaphylaxis, Guillain-Barré syndrome (GBS), myocarditis/pericarditis, thrombosis with thrombocytopenia syndrome (TTS), and febrile seizure. AESI phenotype algorithms can be developed to apply to electronic health record (EHR) data at health provider organizations across the country by querying for standard and interoperable codes. The codes queried in the rules represent symptoms, diagnoses, or treatments of the AESI sourced from published case definitions and input from clinicians.

Methods: To validate the performance of the algorithms, we applied them to EHR data from a United States academic health system and clinicians evaluated a sample of cases. Performance was assessed using positive predictive value (PPV).

Results: Our anaphylaxis algorithm was the best performing, having a PPV of 93.3%. The PPVs for our febrile seizure, myocarditis/pericarditis, TTS, and GBS algorithms were 89.0%, 83.5%, 70.2%, and 47.2%, respectively.

Conclusions: Given our algorithm design and performance, our results support continued research into using interoperable algorithms for widespread AESI post-market detection.

(JMIR Preprints 09/06/2023:49811)
DOI: https://doi.org/10.2196/preprints.49811

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users. Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Development of Interoperable Computable Phenotype Algorithms for Adverse Events of Special Interest to Be Used for Biologics Safety Surveillance: Validation Study

Abstract

Background:
Adverse events (AEs) associated with vaccination have been evaluated by epidemiological studies, and more recently, have gained additional attention with the Emergency Use Authorization (EUA) of several COVID-19 vaccines. As part of its responsibility to conduct post-market surveillance, the U.S. Food and Drug Administration (FDA) continues to monitor several adverse events of special interest (AESIs) to ensure vaccine safety, including for COVID-19.

Objectives:
This study is part of the Biologics Effectiveness and Safety (BEST) Initiative, which aims to improve FDA's post-market surveillance capabilities while minimizing public burden. This study aims to enhance active surveillance efforts through a rules-based, computable phenotype algorithm to identify five AESIs being monitored by the Center for Disease Control and Prevention (CDC) for Covid-19 or other vaccines: anaphylaxis, Guillain-Barré syndrome (GBS), myocarditis/pericarditis, thrombosis with thrombocytopenia syndrome (TTS), and febrile seizure. This study examines whether these phenotypes have sufficiently high positive predictive value (PPV) to ensure that the cases selected for surveillance are reasonably likely to be a post-biologic adverse event. This allows patient privacy, and security concerns for data sharing of non-adverse event patients can be properly accounted for when evaluating the cost / benefit of our approach.

Methods:
AESI phenotype algorithms were developed to apply to electronic health record (EHR) data at health provider organizations across the country by querying for standard and interoperable codes. The codes queried in the rules represent symptoms, diagnoses, or treatments of the AESI sourced from published case definitions and input from clinicians. To validate the performance of the algorithms, we applied them to EHR data from a United States academic health system and provided a sample of cases for clinicians to evaluate. Performance was assessed using PPV.

Results:
With a PPV of 93.3%, our anaphylaxis algorithm performed the best. The PPVs for our febrile seizure, myocarditis/pericarditis, TTS, and GBS algorithms were 89.0%, 83.5%, 70.2%, and 47.2%, respectively.

Conclusions:
Given our algorithm design and performance, our results support continued research into using interoperable algorithms for widespread AESI post-market detection.

Keywords: adverse event; vaccine safety; computable phenotype; post-market surveillance system; real world data; validation study; Food and Drug Administration; electronic health records; COVID-

19 vaccine

Introduction

The U.S. Food and Drug Administration (FDA) Center for Biologics Evaluation and Research (CBER) is responsible for ensuring the safety, purity, potency, and effectiveness of biological products. This includes vaccines, allergenics, blood and blood products, and cells, tissues, and gene therapies for the prevention, diagnosis, and treatment of human diseases, conditions, or injuries [1]. FDA's history of safety surveillance for vaccines includes creation and monitoring of the Vaccine Adverse Event Reporting System (VAERS). VAERS, jointly administered by FDA and the Centers for Disease Control and Prevention (CDC), accepts spontaneous reports of suspected vaccine Adverse Events (AEs) after administration of any vaccine licensed in the United States.

VAERS has been successfully used as an early warning system to identify rare AEs; however, it has limitations. VAERS is a passive surveillance system that relies on individuals, patients, and clinical staff to send in reports, as opposed to automatically collecting them based on clinical data. This can lead to undercounting AEs. Additionally, a causal relationship cannot be established using information from VAERS reports alone [2]. Because of VAERS's limitations, more robust data systems are needed to enhance AE detection. This system would be especially important for detecting the most severe AEs that require medical attention so FDA and CDC can offer guidance for these potentially life-threatening events and ensure product labeling reflects known risks.

To address this gap, CBER established the BEST Initiative in 2017 to build data assets, analytics, and infrastructure for an active, large-scale, efficient post-market surveillance system that can evaluate the safety and effectiveness of biologic products and develop innovative methods [3]. The BEST system is a collection of real-world data (RWD) sources—data related to patient health status and the delivery of health care that are routinely collected from several sources, such as EHRs or claims data [4]. EHR databases, specifically, are a rich source of information. They include data such as clinical notes, which can help address the limitations of VAERS. They also include entire populations of patients to identify whether cases are underreported. Additionally, they may include patients' entire clinical history, which can help establish a causal relationship for an AE. BEST has reached agreements with a limited number of foundational data partners. Access to these data partnerships does not fully address the possible undercounting of AESIs. However, these partnerships allow accelerated development and testing of AESI detection algorithms.

BEST is currently researching a system of distributed computable phenotype algorithms that could be applied at scale to many or all EHR systems across the United States to semi-automatically detect and report potential AESIs from Real-World Data (RWD). Such a system could increase the speed and scope of AE surveillance beyond what is currently available to public health agencies through data partner agreements. To be candidate phenotypes for distributed surveillance use, the phenotypes need to identify probable AEs and avoid false detections. This reflects the need to balance the correct detection of AESIs with the protection of privacy and minimizing burden on health provider systems. For the wider population of health providers to consider deploying such detection algorithms, these phenotype algorithms need to have reasonably high performance (measured by positive predictive value [PPV]) to ensure the cases identified as AEs are likely to be verifiable cases with the outcome of interest. Toward this goal, the computable phenotypes in this study focus on existing EHR data reflecting a detected AE, which are reportable events for public health purposes. The algorithmic identification of undetected AEs or AEs that were not coded properly is not in scope for this study.

Such research must include data from non-AE patients to fully evaluate the performance of a computable phenotype algorithm. Although scientifically desirable in the long term, the inclusion of non-AE cases falls outside of initial goals for a distributed surveillance system – assessing performance (measured by PPV) of the phenotypes for widescale surveillance purposes. The goal of distributing the phenotypes also poses limitations on designing the algorithms. Specifically, the components and complexity of the underlying algorithms need to take into account current EHR standards and technology because they must be deployable and executable across EHR databases without imposing large overhead on health provider systems. If the phenotypes have sufficient PPV and are sufficiently easy to implement at health provider sites, FDA could share the phenotypes to detect AESIs following vaccination in EHRs across the country, which could then be reported to FDA for further review. The ability to detect AESIs using RWD could create an active surveillance system that enhances overall vaccine safety and helps make recommendations to minimize risks for post-vaccination AESIs. The implementation of algorithmic detection and automated reporting of AESIs found in RWD has been shown to increase the odds of submitting a VAERS report by more than 30 times the pre-implementation rate [5].

Although there is a history of studies around post-vaccination AESIs, including those for influenza [6-8] and COVID-19 vaccines [9-13], there has been increased interest in the analysis of vaccine safety and surveillance since the EUA of three COVID-19 vaccines in the United States (Pfizer-BioNTech, Moderna, and Novavax) and their subsequent boosters (e.g., bivalent boosters). FDA hopes to contribute to this research through the development and performance validation of phenotypes for five post vaccination AESIs to identify potential vaccine safety events within EHR databases for this study. The five AESIs chosen include: myocarditis/pericarditis, anaphylaxis, GBS, intracranial or intra-abdominal thrombosis with TTS, and febrile seizure. These AESIs were chosen because they are documented priorities of the CDC's vaccine surveillance [14] for COVID-19 vaccine safety. Additionally, several of these AESIs (anaphylaxis, GBS, febrile seizure) are found following exposure to other vaccines, such as influenza, shingles, pneumococcal conjugate, and measles, mumps, and rubella (MMR). This study describes the methods to develop and validate these five computable phenotype algorithms on an EHR database and the validation results. It is part of FDA's efforts to improve post-market surveillance and is valuable for public awareness, safety, and transparency.

Methods

Computable Phenotype Development

Five AESIs were selected to develop computable phenotypes for our validation study. The study's main focus was detecting COVID-19 vaccine AESIs, so we selected AESIs that CDC specifically identified for monitoring after COVID-19 [14] or for an AESI that has been reported for some sub-populations [15]. Given the uncertainty about future use of COVID-19 seasonal boosters, FDA also wanted to ensure that the AESIs selected had broad applicability to the safety surveillance of other widely used vaccines like influenza, shingles, pneumococcal conjugate, Diphtheria-Tetanus-Pertussis (DTaP) and MMR. Three of our five selections met those criteria given the CDC's documented monitoring of anaphylaxis [16], GBS [17] and febrile seizures [18] for at least one of the vaccines listed.

The phenotype algorithms were designed to be relatively simple and interoperable so any new

healthcare organization's IT department could translate and run them on their EHR database. They were built to query only structured data for interoperable, standard codes – such as Logical Observation Identifiers Names and Codes (LOINC), Systematized Nomenclature of Medicine Clinical Terms (SNOMED), RxNorm – so the algorithm can be generalized or translated across different EHR systems. Historically, this has been a challenge for developing algorithms, since EHR databases often contain their own local code systems specific to the EHR vendor. For example, for this effort we worked with the study partner to map Cerner Multum medication and observation codes to standard RxNorm and LOINC, respectively.

Recent regulation now requires each EHR database to have an application programming interface (API) endpoint that translates any EHR data and many of the EHR's proprietary codes to the U.S. Core Data for Interoperability (USCDI) implementation of the Fast Healthcare Interoperable Resources (FHIR) specification [19]. This specification requires use of interoperable, published code lists [20] (Table 1). These code systems cover almost all clinical events for detection of AEs, such as medical diagnoses, medication prescriptions, lab tests or vital signs taken, and procedures performed. These APIs currently focus on supporting use cases where a single patient's data is queried as opposed to aggregate searches across patients, so we were unable to use them to identify the cohort that our phenotype would select. We were, however, able to use the FHIR API endpoints to pull data for each patient in our validation samples so the participating clinicians could have data with the standard, interoperable code sets for their review.

To facilitate health provider organizations' ability to implement these queries on their EHR, phenotypes were rules-based, used only certain types of structured data, and used common logic across AESIs. The general phenotype logic has been used previously for several post-vaccination AESI studies at the FDA to identify potential AESI cases [21-22] and reuses concepts and methods from past literature from U.S.-based collaborative health research groups, like Observational Health Data Sciences and Informatics (OHDSI) [23], or from similar efforts in the United Kingdom [24], to develop computable phenotype libraries. A health organization only needs to write the general query logic once and then this logic would be able to detect different types of AESIs by referencing different lists of medical codes that represent the different medical events providing evidence that the various AESIs occurred. The logic common to all phenotypes is shown in Figure 1. The code lists that we developed for necessary types of medical evidence are described in more detail below and listed in Table MA2 (Multimedia Appendix 1). The circled items in Figure 1 represent a search for a FHIR resource element containing a code in one of the developed code lists. These were applied within the windows of time denoted by the brackets identifying windows of time before and after a condition diagnosis. The concepts in Figure 1 are described in additional detail below and include:

- **AESI diagnoses and problem list items:** The algorithm first looks for evidence of the AESI represented by a coded final or discharge diagnosis. Only final or discharge diagnoses are used since they best represent the ultimate determination of what was diagnosed during the patient's care. The variability of admitting, working, and other diagnosis types lack the specificity required for the algorithm in this study.
- **Care setting filters:** Additionally, the care setting for every diagnosis was collected based on the medical encounter type for the diagnosis. All diagnosis care settings values were grouped into inpatient, outpatient, or emergency care setting types. Care setting was used to filter out diagnosis codes made during encounters with care settings unlikely to have the specific AESI

diagnosis in the phenotype. The included care settings are defined by case definition and clinician input.

- **Clean window:** Next, a clean window (i.e., a period before the coded diagnosis identified in step 1) is checked to ensure the target diagnosis is the first known diagnosis of its type. This prevents inclusion of historical or ongoing conditions. For all algorithms in this paper, the clean window is defined by all historical patient data in our dataset. To make sure all patient cases had at least a one-year clean window, we pulled an additional historical year of data from our data partner before the study period. Cases where there were multiple occurrences of an AESI diagnosis suggested possible evidence of a chronic condition unrelated to vaccine exposure, and thus were excluded.
- **Condition window:** Lastly, the algorithm searches for sufficient supportive evidence within a condition window. The condition window is defined around the AESI diagnosis date and includes the entire medical encounter period when the condition was diagnosed, as well as two days before and ten days after a condition is diagnosed. Clinical subject matter experts defined condition windows as the timeframe around a diagnosis that supportive evidence would likely present itself in the medical record.
- **Supporting evidence:** Within this the condition window period, the algorithm may filter cases based on supporting evidence of an AESI. This filter looks for either lab test results found in observations, AESI treatment procedures, AESI treatment medications, or procedure, or a combination of the three supporting evidence with a code that meets matches a code on to the phenotypes' concepts code lists. These code lists in the lists aim to include all medical codes that could represent a particular concept, such as administration of epinephrine for an anaphylactic reaction. This AESI supporting evidence filter was applied to all phenotypes except for our Febrile Seizure AESI phenotype because Review of existing research showed febrile seizure algorithms, in general, had the highest PPV among the selected AESIs. The concepts to build code lists for the supporting evidence were identified using case definitions Following this, we prioritized improving specificity in the other AESI phenotypes by including filters requiring additional supporting evidence [25-28] and our clinician's input.
- **Vaccine Exposure:** In real-world operation, the algorithm would also include a vaccine exposure and risk window, or period surrounding vaccination in which diagnoses are searched. For the study's purposes of having sufficient volume and statistical power to estimate operating characteristics of the algorithm, these exposure rules were not included.

Ideally, to assess whether these algorithms generalize to other sites, we would have a multisite validation study. Because of the high cost of data agreements, however, we only had data available for a single EHR site. To avoid overfitting and ungeneralizable results, we designed our algorithm development methods to only use our EHR data as a validation set and not use any of it to train, develop, or fine tune the algorithm. While this does not remove the need for additional external validation, it reduces the likelihood of finding ungeneralizable results. To identify what medical concepts the algorithm should use as evidence, clinicians identified observations, medications, conditions, and procedure concepts from the AESI's case definition, their relevant clinical experience, or other research from their literature review. A brief description of the AESI and the reference of the case definition used is captured in Table 1 below and additional information on the case definition is saved in Table MA1 in the Multimedia Appendix (Multimedia Appendix 1) document.

Table 1. AESI case definitions and descriptions.

AESI	Description	Case definition reference
Myocarditis / Pericarditis	Myocarditis and pericarditis are inflammatory processes involving the myocardium, pericardium, or both (myopericarditis).	Morgan, 2008 [29]
Anaphylaxis	Anaphylaxis is an acute hypersensitivity reaction with multi-organ-system involvement that can present as, or rapidly progress to, a life-threatening reaction. It may occur following exposure to allergens from a variety of sources, including food, aeroallergens, insect venom, drugs, and immunizations.	Rüggeberg, 2007 [30]
Guillain-Barré syndrome (GBS)	GBS constitutes an important proportion of acute flaccid paralysis cases worldwide. It is a condition characterized by various degrees of weakness, sensory abnormalities, and autonomic dysfunction due to damage to peripheral nerves and nerve roots.	Sejvar, 2011 [31]
Intracranial or intra-abdominal TTS	Several cases of unusual thrombotic events and thrombocytopenia have developed after vaccination with the recombinant adenoviral vector encoding the spike protein antigen of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (ChAdOx1 nCov-19, AstraZeneca). More data were needed on the pathogenesis of this unusual clotting disorder [32].	Brighton Collaboration, 2021 [33]
Febrile seizure	There is no Brighton Collaboration definition of febrile seizure, so we used both the fever and seizure case definitions. Fever is defined as an elevation of body temperature above normal. It is usually caused by infection, but can also be associated with several immunologic, neoplastic, hereditary, metabolic, and toxic conditions. Seizures are episodes of neuronal hyperactivity, most commonly resulting in sudden, involuntary muscular contractions.	Marcy, 2004 [34]; Bonohoeffer, 2004 [35]

An analyst completed a text search for a list of terms for these identified concepts, a list of which is captured in Table MA2 (Multimedia Appendix 1), to build the code lists of relevant codes from selected interoperable coding libraries (Table 2). This was accomplished by searching the open-sourced OHDSI Observational Medical Outcomes Partnership (OMOP) concepts table and ATLAS tool [36] which is a collection of thousands of interoperable codes and their definitions/descriptions. The table was searched for any definition or description that matched the identified concept for the interoperable code systems that we listed in Table 2 below and then was reviewed by a clinician for their suitability for the algorithm.

Table 2. Codes used for each type of clinical data.

Clinical data	Interoperable code lists used
Diagnosis	International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), SNOMED
Medication / Immunization	National Drug Code (NDC), RxNorm
Procedures	Current Procedural Terminology (CPT), ICD-10 Procedure Coding System (ICD-10-PCS)
Observations	LOINC

The immunization and the diagnosis ICD-10-CM and SNOMED code lists have been published on the Value Set Authority Center (VSAC) [37] and the additional observation, medication, and evidence code lists may be added in the future after this study is published.

For a surveillance use case, the algorithms need be run regularly (e.g., daily, weekly) to collect batches of historical cases once all the data are available (as opposed to a real-time implementation to collect cases as they are happening). Because the algorithms were created to prioritize simplicity and interoperability rather than maximize total performance (e.g., metrics beyond PPV such as sensitivity, negative predictive power, etc.), this study aimed for improved performance (measured by PPV) to existing AESI claims-based algorithms. Given our knowledge of how some crucial distinguishing information is part of unstructured clinical notes, which are not considered by the algorithms in this study, we expect further analysis is needed to improve accuracy [38-39]. Natural language processing (NLP) techniques can improve algorithm performance, but greatly increase the deployment complexity across healthcare organizations. Therefore, no NLP techniques were used for any phenotypes designed for this study.

Study Period

The study period was January 1, 2018 through May 1, 2022, to ensure the study's data sampled patients both before and after FDA issued the EUA and full licensure for COVID-19 vaccines. We also pulled at least one year of historical data for all patients, so our dataset includes historical information from January 1, 2017, to January 1, 2018, for all patients with medical encounters in the study period. Patients were included even if there were no clinical events in their historical period.

Data

The study population came from a single academic health system in the United States, with EHR medical encounter data from more than 2.6 million patients and greater than 20.7 million medical encounters for the study period. Table 3 shows the demographic breakdown for age, gender, race, and ethnicity of this population.

Table 3. Demographics of academic health system for study population.

Category	Demographic group	Patients	
		N	%

Total	**Total**	**2,666,974**	**100.0%**
Age	Under 5 years	96,146	3.6%
	5 to 17 years	224,941	8.4%
	18 to 24 years	224,631	8.4%
	25 to 44 years	840,395	31.5%
	45 to 64 years	689,075	25.8%
	65 years and above	591,497	22.2%
	Missing	289	0.0%
Gender	Male	1,167,374	43.8%
	Female	1,494,096	56.0%
	Missing	5,504	0.2%
Race	White	1,030,834	38.7%
	Black or African American	748,746	28.1%
	Asian / Pacific Islander	53,666	2.0%
	American Indian or Alaska Native	5,834	0.2%
	Other	198,265	7.4%
	Unknown	629,608	23.6%
	Declined to Answer	21	0.0%
Ethnicity	Hispanic	94,207	3.5%
	Non-Hispanic	1,866,561	70.0%
	Unknown	706,206	26.5%

The entire EHR population during the study period was eligible to be selected by one of our developed phenotype algorithms. There were no age, medical condition, or other exclusions on the population for the algorithm to select cases. Clinical data necessary to select and validate cases selected by the algorithm were provided to the study team through a series of EHR data extracts for all patients in the study period. The algorithm required the following clinical data categories:

- Demographic
- Encounter
- Condition
- Procedure
- Medication
- Observation

EHR data extracts were mapped and loaded into an OHDSI OMOP database [40]. Medication, observation, and procedure data extracts were requested and loaded into the database only for patients who would not be disqualified by other algorithm criteria. For patients selected to be in the validation sample, these data along with clinical data for allergies, immunizations, and clinical notes, were pulled from the EHR's FHIR API endpoints, patient by patient, using a custom Python script to loop through the patients in the sample. The data were loaded into a Health Level 7 (HL7) application programming interface (HAPI) FHIR server. We only pulled FHIR data for cases not initially disqualified by the vaccination and diagnosis filters to avoid unnecessary large data transfers and storage. The algorithm flagged potential AESIs that met the specified criteria. Samples of these cases were sent to physicians for validation.

Validation Sample

Once the algorithm identified cases, a random sample was drawn for each AESI for clinician adjudication. We used stratified sampling to ensure cases during pre- and post-COVID-19 EUA periods were represented (Figure 2). This was due to concerns regarding potential confounding introduced by the COVID-19 vaccines, when attention to possible AESI or medical charting of AESI may have shifted. Where possible for each AESI, 100 cases were sampled from the pre-COVID-19 EUA period and 35 from the post-COVID-19 EUA period. If there were less than 100 or 35 cases in either period, the sample would contain all cases the algorithm selected. Febrile seizure was the exception, as we believe the COVID-19 vaccine EUA should not affect the algorithm's performance because febrile seizure AEs are usually associated with pediatric populations, and the COVID-19 vaccine was not approved for these populations during the study period [35].

Case counts sampled in each period were based on the incidence of diagnosis code occurrence within each period, as well as the period covered. In addition, we added negative controls selected randomly from every encounter in the period to establish a baseline comparison for the case validation process. We included negative controls as a quality control step, to reduce the chance of quality issues with the data and to review the methods our clinicians were following, and not for the purpose of making inferences about the phenotypes' performance for non-AE cases (e.g., through metrics such as sensitivity, negative predictive power, or an overall metric for performance). This study did not focus on the algorithmic identification of undetected AEs or AEs that were not coded properly. The focus of this study was to determine the phenotypes' PPV. Given the expense of clinicians' time for validations and the rarity of the AESIs, there would be minimum benefit to this study to have a negative control sample large enough to draw strong inferences. Further, negative case controls would not further validate the utility of the phenotypes as tools for identifying probable AESIs through distributed surveillance. We added 20 negative controls from the pre-COVID-19 EUA period and seven from the post-COVID-19 EUA period. Physicians were blinded to which cases were controls and which were not.

Chart Review Process

The sample of cases used to validate the algorithm were loaded into a chart review tool for clinician review. This allowed the clinicians to sort through the clinical information for a case and record the determination. Each case was assigned to two clinicians for review. The clinical validation used a patient's full clinical history, which included EHR data, including all clinical notes for each case. The full EHR data used for clinician review included data unused by the detection algorithm described above, including different types of data (e.g., allergies, clinical notes) and data filtered out (e.g., admitting diagnosis, encounters with different care settings).

For each case, the clinician evaluated whether the clinical data evidence met the specified case definition criteria. Relevant patient data for the case window were available and presented to the clinicians in an easy-to-use, browser-based tool with a custom user interface. In the tool, clinicians were able to group items by type, search across all items and text, and request additional chart data to expand the window and access any available historical patient data if desired.

All suspected AEs were validated using published case definitions [29-31,33-35] according to the levels of diagnostic certainty: level 1 (definite), level 2 (probable), and level 3 (possible). If a case did not meet one of the levels in the case definition, it was assigned as level 4 (doubtful) or level 5

(ruled out). “Ruled out” is distinct from “doubtful” in that “ruled out” cases have definitive evidence disqualifying them from being a correct diagnosis. If a case was determined to be “definite” or “probable”, it was considered a positive case of the AESI.

In the event of a disagreement between a positive and negative clinical review, a third clinician made a final determination by reviewing the case EHR data. If the clinicians found the structured or unstructured EHR data was insufficient, clinicians marked this in their review by creating a level 3 (possible, insufficient evidence) designation, where an AESI could have occurred, but where there was not enough documentation to fulfill the requirements of the case definition.

Statistical Analysis

Positive Predictive Value (PPV)

Each algorithm’s PPV was the proportion of positive AEs the algorithm identified that were confirmed by clinical adjudication. PPVs were calculated for each AESI overall, as well as stratified by pre- and post-COVID-19 EUA periods and care setting (inpatient, emergency department, outpatient). Sensitivity analyses were performed to evaluate the impact of medication use, different case definitions, and levels of evidence. PPVs were calculated in two different ways for each AESI algorithm. The first PPV calculated removed all possible cases with insufficient evidence from the denominator (Cases labeled “definite” and “probable” / Total cases minus any labeled “possible, insufficient evidence” by clinicians). PPV was then calculated with the cases with insufficient evidence added back into the denominator (Cases labeled “definite” and “probable” / Total cases). Reporting both PPV calculations can help with understanding performance for different algorithm uses. Algorithm performance should ideally be compared to past literature of detection algorithms for the same AESI.

Confidence Intervals

Because PPV is a binominal proportion, we calculated confidence intervals (CIs) for the PPV using the Agresti-Coull interval [41], which is the recommended method for estimating accurate CIs for binomial proportions like PPV [42].

Inter-Rater Reliability

Inter-rater reliability is used to measure the extent to which two physicians agreed in their AESI assessment. It is calculated using Cohen’s kappa between the first two reviewers. Cohen’s kappa measures the agreement between two raters classifying instances into mutually exclusive groups [43].

Stratification Analysis and Sensitivity Analysis

After validation was completed, we conducted a stratification and sensitivity analysis. We selected two stratification variables that could reasonably impact the generalizability of the results. First, we stratified the data by pre- and post-EUA date to confirm that the algorithm behavior did not change for AESIs after the COVID-19 vaccine was approved and administered to a large portion of the population. Ideally, the algorithms would perform consistently across these eras, but there are multiple factors that could impact performance over these time periods. We also stratified the data by

the care setting of the AE diagnosis, given care setting may be associated with varying EHR data elements in settings such as emergency departments compared to inpatient settings. Algorithm performance was computed using PPV within each stratum.

We also completed a post-hoc sensitivity analysis where we investigated whether the algorithm could be improved, as measured by PPV, through small changes to it or by updating the process for evaluation. These changes were based on insight from clinicians or data analysts reviewing validation results, so results may not generalize to other datasets. However, we did attempt to limit our analysis to decisions that could have been feasibly made without post-validation insights. The changes to the algorithms were either removing medications, observations, procedures, or diagnosis codes that are not specific enough to the AESI in question or adding logic to further filter out cases by requiring more supporting evidence.

We also completed a sensitivity analysis on the GBS algorithm to calculate the PPV if we relaxed some of the specific case definition evaluation criteria if more general evidence was available. We found that two pieces of evidence the case definition required were often missing in the chart review tool: lack of cerebrospinal fluid (CSF) white blood cell (WBC) count in cases of elevated CSF protein and limited or inconsistent documentation of diminished or absent reflexes. In some of these cases, we saw evidence that a neurologist was consulted and felt there was strong suspicion of GBS despite the missing documentation for these tests. This could be explained by two mechanisms.

First, and most likely, this could be due to data loss during the delivery or translation of EHR data to our chart review tool. Because we did not have direct access to the data, our process for obtaining, translating to different common data models/standards, and presenting the data to clinicians using the chart review tool could cause the data for these tests to be incorrectly mapped.

Second, case definition requirements for GBS are extremely strict, and physicians in this study believed some of these might have represented valid GBS cases while not meeting every requirement. For example, several of the cases with missing CSF WBC count did mention cytoalbuminologic dissociation (or similar); in the presence of such a clinical statement, we might infer that CSF WBC count was performed and acceptable to meet the case definition criteria despite a missing test result.

Also, in cases where a neurologist felt strongly that GBS was a likely diagnosis, along with other supporting evidence, it may be acceptable to rely on documented progressive and significant muscle weakness, especially with conflicting reflex findings. In these instances, we placed more weight on the clinician review (which may account for any unforeseen difficulties in data processing and the strictness of the case definition), rather than relying solely on the available (non-missing) data types of the algorithm, for assigning case diagnostic certainty.

Table 4. Total list of all sensitivity analyses for each AESI.

AESI	Data type	Sensitivity analysis	Reasoning
Myocarditis / Pericarditis	Medication	Removal of non-steroidal anti-inflammatory drugs (NSAIDs) from our list of	NSAIDs are medications that can be used to treat many different conditions besides myocarditis

		qualifying medication supporting evidence	and pericarditis.
Myocarditis / Pericarditis	Diagnostic code	Stratification by diagnostic code (myocarditis vs. pericarditis)	Diagnostic criteria differ for these related conditions and may lead to different performance.
GBS	Medication	Removal of gabapentin from our list of qualifying medication supporting evidence	Gabapentin was originally used as supporting evidence of a GBS episode due to its use for nerve pain associated with GBS events [44]. However, it is also used for a variety of other conditions with neuropathic pain and is not specific to GBS.
GBS	Case definition	Update case definition criteria to allow for a case to be validated as positive if missing documentation for absent or diminished reflexes in the weak limbs, CSF WBC count with neurology consult, or clinical note indicating evidence of the test result of GBS more generally	Documentation required for definite or probable GBS as defined by the case definition diagnosis was often missing from our dataset due to failure to capture in EHR or failure to translate to our dataset and can be supplemented by an expert's judgement (e.g., a neurologist).
Febrile seizure	Medication	Addition of medications used to treat fever	The original febrile seizure algorithm did not filter out cases without suggested evidence, but we believed adding suggested evidence could improve PPV.
Febrile seizure	Observation	Addition of observation of clinician describing symptoms of seizure activity	The original febrile seizure algorithm did not filter out cases without suggested evidence, but we believed adding suggested evidence could improve PPV.
TTS	Diagnostic code	Stratification by most prevalent diagnostic code I81 vs. all other codes	Diagnostic criteria differ for these related conditions and may lead to different performance.

The stratification or sensitivity analyses are meant as exploratory analyses to prompt additional research, but subgroups often have too small a sample size that have narrow enough CIs for meaningful results.

Results

Population Sample

Figure 2 illustrates the identification of the study populations and validation sample. From the study population of 20.7 million medical encounters for 2.6 million patients over the study period, the

algorithm selected 1,195 cases of myocarditis/pericarditis, 550 of anaphylaxis, 123 of GBS, 626 of febrile seizure, and 395 of TTS. Of these patient cases, a stratified, random sample of 135 cases each was selected from myocarditis/pericarditis, anaphylaxis, and TTS populations. All 75 pre-EUA cases of GBS and a random sample of 35 post-EUA cases were selected to be validated. A random selection of 100 cases from the pre-EUA period were sampled to validate febrile seizure. An additional 27 negative control cases were sampled for each algorithm from the roughly 20.7 million medical encounters not selected by the algorithm in our study period. Twenty of these cases were sampled from the period prior to the COVID-19 vaccine EUA, and the remaining seven came from the period following.

Overall PPV and Inter-Rater Reliability Results

Table 5 presents algorithm performance measured by PPV for each of the five AESIs using cases that had sufficient evidence and all cases (i.e., including cases unable to be confirmed as positive by clinicians due to insufficient evidence). Counts for the number of cases included in each PPV calculation can be found in Table MA3 in the multimedia appendix.

Table 5. Total validation PPV results.

AESI	Metric	Detected Cases	
		PPV	95% CI
Myocarditis / Pericarditis			
	Cases w/ sufficient evidence only	83.5%	74.9%, 89.6%
	All cases	63.7%	55.2%, 71.4%
Anaphylaxis			
	Cases w/ sufficient evidence only	93.3%	86.4%, 97.0%
	All cases	72.6%	64.4%, 79.5%
GBS			
	Cases w/ sufficient evidence only	47.2%	35.8%, 58.9%
	All cases	30.9%	22.9%, 40.3%
TTS			
	Cases w/ sufficient evidence only	70.2%	61.4%, 77.6%
	All cases	64.4%	55.9%, 72.1%
Febrile seizure			
	Cases w/ sufficient evidence only	89.0%	80.0%, 94.4%
	All cases	89.0%	80.0%, 94.4%

Overall PPVs, when removing all cases with insufficient evidence, were highest for anaphylaxis (93.3%; 95% CI: 86.4%, 97.0%) and febrile seizure (89.0%; 95% CI: 80.0%, 94.4%), followed by myocarditis/pericarditis (83.5%; 95% CI: 74.9%, 89.6%) and TTS at unusual sites (70.2%; 95% CI: 61.4%, 77.6%). Lowest was for GBS (47.2%; 95% CI: 35.8%, 58.9%). All negative control cases

across the five phenotypes were correctly classified by the algorithms.

The PPV results from the chart reviews of the validation sample for each AESI are reported for all cases as well as for only cases with sufficient evidence to make a clear by chart reviewers. The frequencies and percentages for insufficient evidence are presented with the stratification results, below, in Table7. The inter-rater reliability scores for clinician chart reviews all showed substantial agreement between the clinicians (Table 6). Inter-rater reliability, measured by Cohen's kappa, suggests substantial reliability when the value is greater than 0.61, with many similar texts recommending a higher threshold of 0.80 [43].

Table 6. Inter-rater reliability.

AESI	Total cases validated	Inter-rater reliability
Myocarditis / Pericarditis	162	0.814
Anaphylaxis	162	0.770
GBS	137	0.832
TTS at unusual sites	162	0.851
Febrile seizure	120	0.965

Stratification

To evaluate consistency across pre- and post-EUA periods and care settings, we reported true positive (TP) and PPV results for each stratum (Table 7).

Table 7. Stratification analysis: validation sample results.

AESI	AESI and metric	Detected cases	Pre-EUA	Post-EUA	Inpatient	Outpatient	Emergency department
Myocarditis / Pericarditis							
	Total TP cases	86	68	18	72	10	4
	Total cases w/ sufficient evidence	103	79	24	79	16	8
	PPV w/ sufficient evidence, 95% CI	83.5% (74.9%, 89.6%)	86.1% (76.3%, 92.4%)	75.0% (53.4%, 89.5%)	91.1% (82.3%, 96.0%)	62.5% (36.0%, 84.0%)	50.0% (15.4%, 84.7%)
	Total cases	135	100	35	91	26	18
	PPV total cases, 95% CI	63.7% (55.2%, 71.4%)	68.0% (58.1%, 76.5%)	51.4% (34.7%, 67.8%)	79.1% (69.4%, 86.4%)	38.5% (21.2%, 58.8%)	22.2% (6.7%, 47.9%)
	% Insufficient evidence	23.7%	21.0%	31.4%	13.2%	38.5%	55.6%
Anaphylaxis							
	Total TP cases	98	70	28	17	N/A	81

	Total cases w/ sufficient evidence	105	74	31	19	N/A	86
	PPV w/ sufficient evidence, 95% CI	93.3% (86.4%, 97.0%)	94.6% (86.2%, 98.4%)	90.3% (73.4%, 98.0%)	89.5% (65.6%, 99.7%)	N/A	94.2% (86.6%, 97.9%)
	Total cases	135	100	35	27	N/A	108
	PPV, 95% CI	72.6% (64.4%, 79.5%)	70.0% (60.2%, 78.3%)	80.0% (63.0%, 90.9%)	63.0% (42.9%, 79.7%)	N/A	75.0% (65.8%, 82.4%)
	% Insufficient evidence	22.2%	26.0%	11.4%	29.6%	N/A	20.4%
GBS							
	Total TP cases	34	24	20	34	N/A	N/A
	Total cases w/ sufficient evidence	72	52	20	72	N/A	N/A
	PPV w/ sufficient evidence, 95% CI	47.2% (35.8%, 58.9%)	46.2% (32.9%, 60.0%)	50.0% (28.1%, 71.9%)	47.2% (35.8%, 58.9%)	N/A	N/A
	Total cases	110	65	45	110	N/A	N/A
	PPV, 95% CI	30.9% (22.9%, 40.3%)	36.9% (25.9%, 49.5%)	44.4% (30.4%, 59.4%)	30.9% (22.8%, 40.3%)	N/A	N/A
	% Insufficient evidence	34.5%	30.7%	42.9%	34.5%	N/A	N/A
TTS							
	Total TP cases	87	64	23	86	1	0
	Total cases w/ sufficient evidence	124	91	33	122	1	1
	PPV w/ sufficient evidence, 95% CI	70.2% (61.4%, 77.6%)	70.3% (60.0%, 78.9%)	69.7% (51.6%, 83.5%)	70.5% (61.7%, 78.0%)	100.0% (0.0%, 100.0%)	0.0% (0.0%, 100.0%)
	Total cases	135	100	35	133	1	1
	PPV, 95% CI	64.4% (55.9%, 72.1%)	64.0% (54.0%, 72.9%)	65.7% (48.2%, 80.0%)	64.7% (56.1%, 72.4%)	100.0% (0.0%, 100.0%)	0.0% (0.0%, 100.0%)
	% Insufficient evidence	8.1%	9.0%	5.7%	8.3%	0.0%	0.0%
Febrile seizure							

	Total TP cases	73	73	N/A	0	N/A	73
	Total cases w/ sufficient evidence	83	83	N/A	1	N/A	82
	PPV w/ sufficient evidence, 95% CI	88.0% (78.8%, 93.6%)	88.0% (78.8%, 93.6%)	N/A	0.0% (0.0%, 100.0%)	N/A	89.0% (80.0%, 94.4%)
	Total cases	100	100	N/A	1	N/A	99
	PPV, 95% CI	73.0% (63.3%, 80.9%)	73.0% (63.3%, 80.9%)	N/A	0.0% (0.0%, 100.0%)	N/A	73.7% (64.1%, 81.6%)
	% Insufficient evidence	18.0%	18.0%	N/A	100.0%	N/A	17.2%

None of the algorithms had notable differences between the pre-and post-EUA periods, since all 95% CIs had some overlap. However, there were some differences between the PPVs for the two periods that could be significant with a larger validation sample. The difference in PPV for myocarditis/pericarditis varied from 68.0% in the pre-EUA period to 51.4% in the post-EUA period, while anaphylaxis showed the opposite pattern with a 70.0% PPV in the pre-EUA period that jumped to 80.0% PPV in the post-EUA period.

We also reported stratified results by care setting (Table 7). For myocarditis/pericarditis, the PPV of cases with an inpatient care setting (79.1%; 95% CI: 69.4%, 86.4%) was notably higher than those from outpatient (38.5%; 95% CI: 21.2%, 58.8%) or emergency department (22.2%; 95% CI: 6.7%, 47.9%) care settings.

Anaphylaxis did not have a large difference across care settings, as the 95% CIs overlap between the two care settings. However, they did show better performance with cases in an emergency department (75.0%; 95% CI: 65.8%, 82.4%) care setting over cases with an inpatient care setting (63.0%; 95% CI: 42.9%, 79.7%). The other AESI algorithms filtered for only one care setting or had a vast majority of cases in one care setting.

Sensitivity Analysis

Medication/Observation Algorithm Changes

We analyzed whether changes to medication code lists for the myocarditis/pericarditis and GBS algorithms could improve performance. For the myocarditis/pericarditis algorithm, removal of NSAIDs from the medication code lists showed no change in PPV at 83.5% (Table 8), but PPV values were higher for cases selected with the pericarditis instead of myocarditis ICD-10 codes.

Table 8. Sensitivity analysis: myocarditis/pericarditis validation sample results.

AESI and sensitivity analysis	Total TP cases	Selected cases (change)[a]	PPV, 95% CI (change)[a]	Selected cases w/ sufficient evidence (change)[a]	PPV, 95% CI (change)

Removal of NSAIDs	86	135 (0)	63.7% (55.2%, 71.4%) (0.0%)	103 (0)	83.5% (74.9%, 89.6%) (0.0%)
Pericarditis diagnosis[b]	59	82 (-53)	72.0% (61.1%, 80.8%) (+8.3%)	67 (-36)	88.1% (77.6%, 94.3%) (+4.6%)
Myocarditis diagnosis[b]	27	53 (-82)	50.9% (37.4%, 64.3%) (-12.8%)	36 (-67)	75.0% (57.9%, 87.1%) (-8.5%)

[a] Values in parentheses reflect the change due to the modified algorithm features.
[b] All ICD-10-CD codes the algorithm used were broken into two groups: Myocarditis (I40.0 Infective myocarditis, I40.1 Isolated myocarditis, I40.8 Other acute myocarditis, I40.9 Acute myocarditis, unspecified, I51.4 Viral myocarditis) and pericarditis (B33.22 Viral pericarditis, B33.23 Acute rheumatic pericarditis, I30.0 Acute nonspecific idiopathic pericarditis, I30.1 Infective pericarditis, I30.8 Other forms of acute pericarditis, I30.9 Acute pericarditis, unspecified, I32 Pericarditis in diseases classified elsewhere, I41 Meningococcal pericarditis).

For the GBS algorithm, when cases were removed where gabapentin (used for post-GBS pain management) was the only supporting evidence, PPV increased to 38.1% (95% CI: 28.2%, 49.1%) from 30.9% (95% CI: 22.9%, 40.3%) (Table 9).

Table 9. Sensitivity analysis: GBS validation sample results.

AESI and sensitivity analysis	Total TP cases	Selected cases (change)[a]	PPV, 95% CI (change)[a]	Selected cases w/ sufficient evidence (change)[a]	PPV, 95% CI (change)[a]
Removal of gabapentin	33	86 (-24)	38.4% (28.6%, 49.2%) (7.5%)	53 (-19)	62.3% (48.3%, 74.5%) (+15.0%)
Adjusted case definition	49	110 (0)	44.5% (35.4%, 54.0%) (+13.6%)	72 (0)	68.1% (56.3%, 78.0%) (+20.8%)
Adjusted case definition + Removal of gabapentin	49	86 (-26)	57.1% (46.2%, 67.4%) (+26.2%)	68 (-4)	72.1% (60.0%, 81.6%) (+24.8%)

[a] Values in parentheses reflect the change due to the modified algorithm features.

Our initial febrile seizure algorithm did not use any supporting evidence to filter out possible false positives since we believed we could get adequate PPV without it.

For our sensitivity analysis, we tested requiring supporting evidence in the condition period, such as the presence of medications for reducing fever like acetaminophen, observation evidence when the patient's chief complaint was related to fever or seizure, or the presence of both. When filtered to only cases with either medication or observation evidence, febrile seizure PPV increased significantly to 93.3% (95% CI: 84.7%, 97.6%) from the original algorithm PPV of 73% (95% CI: 63.3%, 80.9%), with no overlap in 95% CIs and a p-value less than 0.01 (Table 10). When the algorithm required both medication and observation evidence, it performed even better (PPV 96.9%; 95% CI: 88.5%, 99.9%).

Table 10. Sensitivity analysis: febrile seizure.

AESI and sensitivity analysis	Total TP cases	Selected cases (change)[a]	PPV, 95% CI (change)[a]	Selected cases w/ sufficient evidence (change)[a]	PPV, 95% CI (change)[a]
Cases w/ either medication or observation	70	75 (-25)	93.3% (84.7%, 97.6%) (+20.3%)	73 (-10)	95.9% (87.9%, 99.2%) (+7.9%)
Cases w/ both medication and observation evidence	63	65 (-35)	96.9% (88.5%, 99.9%) (+23.9%)	63 (-20)	100.0% (92.8%, 100.0%) (+12.0%)

[a] Values in parentheses reflect the change due to the modified algorithm features.

Diagnostic Code List Changes

We also analyzed if changing which diagnostic codes were used to identify the AESI might lead to higher performance for the myocarditis/pericarditis and TTS algorithms.

For myocarditis/pericarditis, we found that an algorithm only looking for the myocarditis code (PPV 50.9%; 95% CI: 37.4%, 64.3%) underperformed an algorithm with just pericarditis codes (PPV 72.0%; 95% CI: 61.1%, 80.8%) (Table 8). For TTS, we found that the main ICD-10-CM code I81 for “portal vein thrombosis” (73.5%; 95% CI: 64.0%, 81.3%) outperformed all other codes in our code list, including G08 (intracranial and intraspinal phlebitis and thrombophlebitis), I82.0 (Budd-Chiari syndrome), I82.3 (embolism and thrombosis of renal vein), and I82.890 (acute embolism and thrombosis of other specified veins), with a PPV of 36.4% (95% CI: 21.3%, 54.4%) (Table 11).

Table 11. Sensitivity analysis: TTS.

AESI and sensitivity analysis	Total TP cases	Selected cases (change)[a]	PPV, 95% CI (change)[a]	Selected cases w/ sufficient evidence (change)[a]	PPV, 95% CI (change)[a]
I81	75	102 (-33)	73.5% (64.0%, 81.3%) (+9.1%)	96 (-28)	78.1% (68.6%, 85.4%) (+8.0%)
All other TTS ICD codes[b]	12	33 (-102)	36.4% (21.3%, 54.4%) (-28.0%)	28 (-96)	42.9% (25.4%, 62.1%) (-27.3%)

[a] Values in parentheses reflect the change due to the modified algorithm features.
[b] All other TTS ICD codes include G08, I82.0, I82.3, I82.890.

Case Definition Validation Criteria

Lastly, we analyzed whether a small update to our case definition criteria for the GBS algorithms described above in the methods sensitivity analysis section would improve reported performance in Table 9. When we re-calculated PPV based on this new case definition criteria, the PPV was 44.5% (95% CI: 35.4%, 54.0%). When we applied this validation criteria change to the algorithm when gabapentin was removed, as discussed above, the algorithm achieved a PPV of 57.1% (95% CI: 46.2% 67.4%).

Discussion

Principal Results

The results of this study show that, for four out of five AESIs, we can build an interoperable computable phenotype with comparable or increased performance to algorithms in the existing literature. These algorithms are developed using a rules-based approach to facilitate application and increase generalizability of performance across EHR databases. For the phenotypes with poorer performance, the issues were often that the case definition required documentation of a test that was either lost in our data pipeline, was not completed, or was not recorded by the treating physician or nurse. While these cases are marked as false positives based on our methodology, they may be true AEs that are lacking the documentation to meet the case definition. Some small updates to the algorithms or the case definition evaluation method could be made to potentially improve the algorithms' performances, but a more important next step would be to validate our algorithms on other data partners to ensure generalizability of the original algorithms and any updates. Given the need for active AE surveillance, this study is still an important first step toward building an algorithm that can be distributed and implemented on health provider EHR databases and accurately detect AEs.

The PPV results of the phenotypes, negative control groups, and stratification and sensitivity analysis are discussed in more detail below. Note our negative control groups and many of the stratification and sensitivity analyses have sample sizes too small to draw strong conclusions as illustrated by the width of the 95% CIs for those results. These were exploratory analyses completed as a supplement to the main findings of the paper around the PPV of the algorithms.

Myocarditis/Pericarditis

The myocarditis/pericarditis algorithm showed strong PPV performance using cases with sufficient evidence. The literature appears to lack good comparison studies against which to evaluate this algorithm's performance. A meta-analysis from 2013 reviewed myocarditis/pericarditis algorithm studies and found none of them evaluated their algorithm by calculating PPV [45].

When myocarditis/pericarditis was segmented via care settings, algorithm performance was highest for inpatient settings, with a PPV of 79.1%. This can be attributed to the availability of supporting clinical data needed for accurate case detection in such settings. Given that inpatient testing is necessary to meet the criteria of the case definition, the algorithm performance matches clinical expectations and adds to its public health importance.

In emergency care settings, myocarditis/pericarditis is often diagnosed for patients with a history of inpatient visits to other health system(s). This increases the probability of these patients having additional documentation necessary to meet the case definition. This highlights the role of health information exchanges in supporting public health use cases, improving AE reporting, and enhancing postmarket surveillance.

Myocarditis/pericarditis had a notable difference in PPVs for pre- and post-EUA date. The post-EUA date strata of the sample had a higher percentage of cases coming from the emergency department, which had few cases pre-EUA. This could be explained by patients being diagnosed during previous inpatient stays in other health systems, and a lower threshold to give a preliminary diagnosis with limited information. This category had a lower PPV on average for myocarditis/pericarditis, likely

due to less documentation than an inpatient care setting. This highlights the need for further validation of the algorithm in these settings for an effective public health benefit, and to gain confidence that our algorithm is fit-for-purpose.

Because the aim of the algorithms is post-vaccination AESI detection in support of public health safety surveillance, any potential degradation in performance in the post-EUA period is a concern. If performance decrease in the post-EUA period is driven by post-vaccination myocarditis/pericarditis being more likely to have confounding physical findings that could affect how quickly and in which care setting it gets diagnosed, the PPV from this study may not be applicable to a post-vaccination version of the phenotype. There is a small overlap in the two periods' PPV 95% CI, and a two-sample proportion test returns a p-value of 0.08. This suggests that the difference could also be due to statistical noise. However, given the importance of the post-EUA period to the algorithm's future task and the size of the difference, we suggest validating additional cases in the post-EUA period to confirm whether the algorithm is actually less effective.

Anaphylaxis

In cases with sufficient evidence, our anaphylaxis algorithm performed strongly with a PPV score of 93.3% (95% CI: 86.4%, 97.0%). This shows possible slight improvement over previous anaphylaxis research, although both results are within the 95% CI [25-26]. When stratified by care setting, the algorithm performed better in emergency department care settings. This can be explained due to the anaphylaxis symptoms and treatment being more likely to be well-documented in this setting. Availability of additional evidence increases the PPV of the algorithm. Since anaphylaxis cases related to vaccination are more likely to end up in the emergency department, the better performance of the algorithm would provide a better public health benefit.

Overall, the performance of the algorithm was moderate compared to that seen in literature. With no obvious avenues for improvement available, s no additional sensitivity analyses were applied.

GBS

Our initial GBS algorithm showed weak performance for GBS with a PPV of 47.2% (95% CI: 35.8%, 58.9%). Given existing research on GBS validations, this result is not surprising, since our result is comparable with a study result showing GBS algorithm validation PPV (PPV 29.0%; 95% CI: 24.0%, 34.0%) [27]. We hoped our algorithm would improve on this study's results, allowing us to meet the "moderate" performance threshold defined in the methods section, given that we added additional logic to require suggested evidence and filter out historical diagnoses. However, we believe the algorithm's performance could be improved based on the sensitivity analysis results.

An increase in performance was observed when adjusting the case definition interpretation of GBS to allow for more general written clinical notes or neurology consult evidence to replace specific documented test results. The lack of standardization in laboratory results is fraught with challenges such as inconsistent data. The observed improvement in the GBS phenotype highlighted the need for further standardization to have a better impact on public health benefit.

Further, the performance of the GBS algorithm was improved by exclusion of non-specific medications such as gabapentin, increasing its public health benefit. Gabapentin is often used to treat generalized neuropathic pain for a variety of conditions other than GBS, including diabetes, and can confound the results.

With both case definition and medication adjustments to the algorithm, the PPV rose to be closer to the moderate performance threshold and an increase over the cited historical study. Because these changes were informed by the cases in the validation study post-hoc, they might be overfitted to this validation sample and may not be generalizable. They should be tested in other EHR systems.

The GBS algorithm performed slightly better in the post-EUA period, but the performance of both periods was well within the 95% CI of the other period. The GBS algorithm only applies to the inpatient care setting, so no care setting stratification analysis was performed.

Febrile Seizure

Our febrile seizure algorithm performed strongly, with a PPV score of 89.0% using cases with sufficient evidence. This performance is in line with existing febrile seizure algorithm validation research [28], where a febrile seizure validation study on the FDA Sentinel database showed a PPV of 70.0% (95% CI: 64.0%, 76.0%). Our sensitivity analysis suggests that even better performance could possibly be achieved by adding additional filters to select cases with supporting medication and observation evidence, which are well-documented in EHRs. The better performance of the algorithm provides better public health benefit, and further supports the use of EHRs in detection of AEs. For cases that met either or both of those criteria, the PPV increased. Since these changes to the algorithm happened after the validation was completed, they overstate the general performance increases when applied to a new EHR setting but offer avenues for a future validation study. Future research can test whether stronger performance is possible with these filters and focus on reviewing the algorithm's application to AEs following pediatric vaccinations.

TTS

The TTS algorithm showed moderate performance for PPV at 70.2% which is similar to a separate FDA TTS validation study which estimated performance at 76.1% (95% CI 67.2-83.2%) [22]. TTS had consistent performance across both pre- and post-EUA periods and did not have enough cases in the outpatient and emergency department care settings for any defensible findings around diagnosis care setting stratification. Our sensitivity analysis revealed that when the AESI was diagnosed with the ICD-10 code I81 (portal vein thrombosis), the algorithm showed a significant increase when compared to the performance of all other ICD codes, 73.5% (95% CI: 64.0%, 81.3%) compared to 36.4% (95% CI: 21.3%, 54.4%). Although if an increase to specificity is desired at the cost of some sensitivity, the TTS algorithm could be limited to only select the higher performing I81 diagnosis code.

Limitations

There are several limitations to this study. First, it only evaluates general AESIs, and not post-vaccination AESIs specifically since the algorithms do not require evidence of vaccine administration as criteria. While this was necessary due to the rareness of the post-vaccination AESIs in our data, it is possible the algorithms perform worse detecting post-vaccination AESIs specifically since they will often present slightly different and in different populations when occurring after a vaccine administration. For example, the major presenting symptoms appeared to resolve faster in cases of myocarditis after COVID-19 vaccination than in typical viral cases of myocarditis [9]. To guard against this, we included both pre- and post-COVID-19 EUA data with the hope that post-EUA cases would include some post-vaccination AESIs. However, we did not have enough post-EUA cases available to build a large enough sample size for a comparison with enough statistical power to provide definitive evidence on this topic. Another limitation in this vein is the general small sample

size for all stratification, sensitivity, and negative control analyses. We make sure to state that these analyses are exploratory in nature and the reader should not form strong conclusions from them given their small samples size and large confidence interval range. Future research could address these concerns by identifying a data source with enough post-vaccination AESI cases to complete a comparably large validation study.

An additional limitation of this study is that it only measures algorithms' PPVs instead of investigating other metrics which could give a better picture of the algorithm's holistic performance like sensitivity/specificity. Specifically, these other metrics would estimate how many of the total positive cases are being identified and how well the algorithm is able to identify cases without the AESI. However, we believe this limitation is necessary because (a) the main purpose of this study is to assess the PPV of phenotypes because it answers the most relevant public health question, if the algorithms will generate a quality detected set of adverse event cases for the public health surveillance, and (b) a much higher cost and more extensive data sharing is needed to properly estimate sensitivity and specificity because of the required validation sample size necessary for a negative control group. To calculate PPV, one only needs a sample of the cases selected by the algorithm. To estimate the sensitivity and specificity, however, it would be necessary to also validate an extremely large negative control group sample since the AESI conditions the algorithms are trying to detect are often rare events. We would expect it to be even more rare for these conditions of interest for AESI to happen and not be recorded with types of structured data elements that are being used in the phenotypes. In fact, the lack of structured data elements in some negative control cases lead to a clinician asking the research team if something was wrong because their case had no relevant charted events to be reviewed. A much larger validation study would also expose clinicians to a larger set of patient data for cases that have a low likelihood of having an adverse event. This approach limits interaction with protected health information (PHI) data until the algorithms' PPVs support continued research with broader samples and methodologies.

Another limitation is that, although they were designed to be simple to deploy, the algorithms are still time-consuming to apply to different EHR systems. Although a hallmark of this algorithm is its interoperability, the algorithm logic still must be applied to the EHR common data model or extracted and translated into another common data model as was done for this study. Interoperable codes should be available for all patients, given the requirement to provide patient data in an interoperable FHIR standard. However, given the recency of the requirement, they might not be available in all systems and require some code translation on the health organization side, especially when analyzing at the population level. Additionally, since the interoperable codes will only be available through a FHIR API, this adds another data pull and integration with the EHR system to obtain these codes for the algorithm.

In the future, the evolving landscape of health IT may facilitate the public health use cases of detecting and reporting post-vaccination AESIs in a safe and secure manner that protects patient privacy. This could be achieved by EHRs supporting secure querying of patient cohorts with probable post-vaccination AESIs using clinical query language (CQL) [46] or other interoperable query language. Reducing the burden of automatic detection of post-vaccination AESIs would help public health organizations improve AE surveillance with minimal additional burden to healthcare organizations and providers.

A final limitation of this study is that the algorithms were only applied to one site. Going forward, algorithm performance should be validated at other sites to ensure they generalize. Although the algorithms were generated without prior input from the data, the study is still limited to one healthcare organization and this method could have different operating characteristics (PPV,

sensitivity, etc.) at a second location.

Future research can be performed to improve algorithm accuracy, and as stated previously, would require additional partner EHR data systems. To create a better performing algorithm, machine learning techniques could be used to train the model to identify specific patterns of data instead of relying on rules-based methods that incorporate published case definition criteria and clinical subject matter expert experience. When given enough data, machine learning approaches generally outperform rules-based approaches across domains and some prior research suggests this is true in the medical domain as well [47].

However, machine learning methods will not generalize across EHR systems because the data patterns that machine learning identifies could be specific to an individual healthcare organization. Trying to build a large dataset that combines multi-site data is extremely difficult and costly due to concerns over infrastructure, regulations, privacy, and data standardization. A method such as federated learning could be explored to alleviate this problem. Federated learning allows multiple sites to collaboratively train a global model without directly sharing data and has been used to train machine learning algorithms at EHR sites previously [48].

Conclusions

In summary, this study presents strong initial evidence that creating simple, interoperable, rules-based phenotypes can detect AESIs on a new data source and that the phenotypes outperform the PPV outcomes for historical validations studies for these conditions. The study validates five different AESIs to prove that this approach can work for a broad range of AESIs, while also highlighting where the approach might be less successful. For example, the GBS algorithm was built using ICD-10 codes that previous validation studies have demonstrated are not accurate predictors of a GBS case that meets case definition criteria; subsequently, our GBS algorithm performed poorly. The validation study sample sizes for all AESIs allowed for adequate precision to evaluate algorithm PPV against historical studies.

An active surveillance system can enhance vaccine safety and aid in the development and use of safer vaccines and recommendations to minimize the AE risks after vaccination [49]. The algorithms were developed using a method that should be able to be applied to and generalize performance for new EHR databases, but more research is needed to confirm this. If the methodology can be successfully used to detect post-vaccination AESI cases across EHR databases, these algorithms could be deployed widely to inform FDA decision-making, promote public safety, and improve public confidence. Going forward, further research and investigation are needed to enhance algorithm performance and integrate the algorithms across healthcare organizations for active surveillance in the interest of public health.

Acknowledgments

Development of the manuscript benefitted from significant engagement with FDA CBER team members and their partners. We thank them for their contributions. Additional feedback on the manuscript was provided by IBM Consulting (Stella Muthuri, Brian Goodness); Accenture Consulting (Shayan Hobbi); and Korrin Bishop - Writing & Editing (Korrin Bishop). Generative AI was not used for any portion of the manuscript writing.

Data Availability Statement

The data analyzed in this study is subject to the following licenses/restrictions: Data were available to the FDA for the purpose of evaluating algorithms for AESI outcomes and are not available as a de-identified public data set. For inquiries or questions regarding the data, we encourage individuals to direct their queries to Hussein Ezzeldin, Hussein.ezzeldin@fda.hhs.gov.

Funding

This research was funded through the FDA CBER BEST Initiative. Several co-authors hold commercial affiliations with Accenture, IBM Consulting, and MedStar Health Research Institute. Accenture (P.S.-H.), IBM Consulting (A.A., J.P., J.B., A.S., E.M., L.J., and M.D.), and Medstar Health Research Institute (A.H. and J.B.) provided support in the form of salaries for authors but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

Authors Ashley Arnold, Jeno Pizarro, Jeff Beers, Arianna Sang, Erik Martinez, Lance Jones, and Matt Deady are/were employed by IBM while participating in the study. Author Patrick Saunders-Hastings is employed by Gevity Consulting Inc., part of Accenture. Authors Aaron Zachary Hettinger and H. Joseph Blumenthal are employed by MedStar Health Research Institute, and Aaron Zachary Hettinger holds an appointment with Georgetown University School of Medicine. These authors have delivered clinical and epidemiology consulting engagement for public and private sector partners. These affiliations did not impact the study design, data collection and analysis, decision to publish, or preparation of the manuscript, and do not alter our adherence to *JMIR* policies on sharing data and materials. The opinions expressed are those of the authors, and do not necessarily represent the opinions of their respective organizations. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

AE: Adverse event
AESI: Adverse event of special interest
AI: Artificial intelligence
API: Application programming interface
BEST: Biologics Effectiveness and Safety Initiative
CBER: Center for Biologics Evaluation and Research
CDC: Centers for Disease Control and Prevention
CI: Confidence interval
COVID-19: Coronavirus disease 2019
CPT: Current Procedural Terminology
CQL: Clinical query language
CSF: Cerebrospinal fluid
DTaP: Diphtheria-Tetanus-Pertussis
EHR: Electronic health record
EUA: Emergency Use Authorization

FDA: U.S. Food and Drug Administration
FHIR: Fast Healthcare Interoperable Resources
GBS: Guillain-Barré syndrome
HAPI: Health Level 7 application programming interface
HL7: Health Level 7
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification ICD-10-PCS: ICD-10 Procedure Coding System
IT: Information technology
IVIG: Intravenous immunoglobulin
LOINC: Logical Observation Identifiers Names and Codes
MMR: Measles, mumps, and rubella
nCoV-19: novel coronavirus 2019
NDC: National Drug Code
NLP: Natural language processing
NSAIDs: Non-steroidal anti-inflammatory drugs
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
PHI: Protected health information
PPV: Positive predictive value
RWD: Real-world data
SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2
SNOMED: Systematized Nomenclature of Medicine Clinical Terms
TP: True positive
TTS: Thrombosis with thrombocytopenia syndrome
USCDI: U.S. Core Data for Interoperability
VAERS: Vaccine Adverse Event Reporting System
VSAC: Value Set Authority Center
WBC: White blood cell

References

1. CBER vision & mission. Washington, DC: U.S. Food & Drug Administration; 2019. https://www.fda.gov/about-fda/center-biologics-evaluation-and-research-cber/cber-vision-mission [accessed Feb 22, 2024]
2. Vaccine Adverse Event Reporting System (VAERS). Atlanta: Centers for Disease Control and Prevention; 2022. https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vaers/index.html#anchor_1616772696807 [accessed Feb 22, 2024]
3. CBER Biologics Effectiveness and Safety (BEST) System. Washington, DC: U.S. Food & Drug Administration; 2022. https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/cber-biologics-effectiveness-and-safety-best-system [accessed Feb 22, 2024]
4. Framework for FDA's Real-World Evidence Program. Washington, DC: U.S. Food & Drug Administration; 2018. https://www.fda.gov/media/120060/download [accessed Feb 22, 2024]
5. Baker MA, Kaelber DC, Bar-Shain DS, Moro PL, Zambarano B, Mazza M, et al. Advanced clinical decision support for vaccine adverse event detection and reporting. Clin Infect Dis. 2015 Sep 15;61(6):864-870. doi:10.1093/cid/civ430
6. Polakowski LL, Sandhu SK, Martin DB, Ball R, Macurdy TE, Franks RL, et al. Chart-confirmed Guillain-Barré syndrome after 2009 H1N1 influenza vaccination among the Medicare population, 2009-2010. Am J Epidemiol. 2013 Sep 15;178(6):962-973. doi:10.1093/aje/kwt051

7. Martin D, Menschik D, Bryant-Genevier M, Ball R. Data mining for prospective early detection of safety signals in the Vaccine Adverse Event Reporting System (VAERS): A case study of febrile seizures after a 2010-2011 seasonal influenza virus vaccine. Drug Saf. 2013 Jul;36:547-556. doi:10.1007/s40264-013-0051-9
8. Erlewyn-Lajeunesse M, Bonhoeffer J, Ruggeberg J, Heath P. Anaphylaxis as an adverse event following immunisation. J Clin Pathol. 2007 Jul;60(7):737-739. doi:10.1136/jcp.2006.037457
9. Oster ME, Shay DK, Su JR, Gee J, Creech CB, Broder KR, et al. Myocarditis cases reported after mRNA-based COVID-19 vaccination in the US from December 2020 to August 2021. JAMA. 2022 Jan 25;327(4):331-340. doi:10.1001/jama.2021.24110
10. Vu S, Bertrand M, Jabagi MJ, Botton J, Drouin J, Baricault B, et al. Age and sex-specific risks of myocarditis and pericarditis following Covid-19 messenger RNA vaccines. Nat Commun. 2022 Jun 25;13:3633. doi:10.1038/s41467-022-31401-5
11. Hanson KE, Goddard K, Lewis N, Fireman B, Myers TR, Bakshi N, et al. Incidence of Guillain-Barré syndrome after COVID-19 vaccination in the Vaccine Safety Datalink. JAMA Netw Open. 2022 Apr 1;5(4):e228879. doi:10.1001/jamanetworkopen.2022.8879
12. Kim JE, Min Y, Shin JY, Kwon Y, Bae J, Sung JJ, et al. Guillain-Barré syndrome and variants following COVID-19 vaccination: Report of 13 cases. Front Neurol. 2022 Jan 27;12:820723. doi:10.3389/fneur.2021.820723
13. MacIntyre C, Veness B, Berger D, Hamad N, Bari N. Thrombosis with thrombocytopenia syndrome (TTS) following AstraZeneca ChAdOx1 nCoV-19 (AZD1222) COVID-19 vaccination – A risk-benefit analysis for people < 60 years in Australia. Vaccine. 2021 Aug 9;39(43):4784-4787. doi:10.1016/j.vaccine.2021.07.013
14. Selected Adverse Events Reported after COVID-19 Vaccination. Atlanta: Centers for Disease Control and Prevention; 2023. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html [accessed Feb 22, 2024]
15. COVID-19 Vaccine Safety in Children and Teens. Atlanta: Centers for Disease Control and Prevention; 2023. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/vaccine-safety-children-teens.html [accessed Feb 22, 2024]
16. Preventing and Managing Adverse Reactions: General Best Practice Guidelines for Immunization. Atlanta: Centers for Disease Control and Prevention; 2022. https://www.cdc.gov/vaccines/hcp/acip-recs/general-recs/adverse-reactions.html [accessed Feb 22, 2024]
17. Guillain-Barré Syndrome and Vaccines. Atlanta: Centers for Disease Control and Prevention; 2023. https://www.cdc.gov/vaccinesafety/concerns/guillain-barre-syndrome.html [accessed Feb 22, 2024]
18. Febrile Seizures and Childhood Vaccines. Atlanta: Centers for Disease Control and Prevention; 2020. https://www.cdc.gov/vaccinesafety/concerns/febrile-seizures.html [accessed Feb 22, 2024]
19. Standards-based application programming interface (API) certification criterion [Internet]. Washington, DC: Office of the National Coordinator for Health Information Technology; 2020. https://www.healthit.gov/sites/default/files/page2/2020-03/APICertificationCriterion.pdf [accessed Feb 22, 2024]
20. US Core Data for Interoperability USCD V3. Ann Arbor: HL7 International; 2023. https://build.fhir.org/ig/HL7/US-Core/uscdi.html [accessed Feb 22, 2024]
21. Shoaibi A, Lloyd PC, Wong HL, Clarke TC, Chillarige Y, Do R, et al. Evaluation of potential adverse events following COVID-19 mRNA vaccination among adults aged 65 years and older: Two self-controlled studies in the U.S. Vaccine. 2023 Jul 19;41(32):4666-4678. doi:

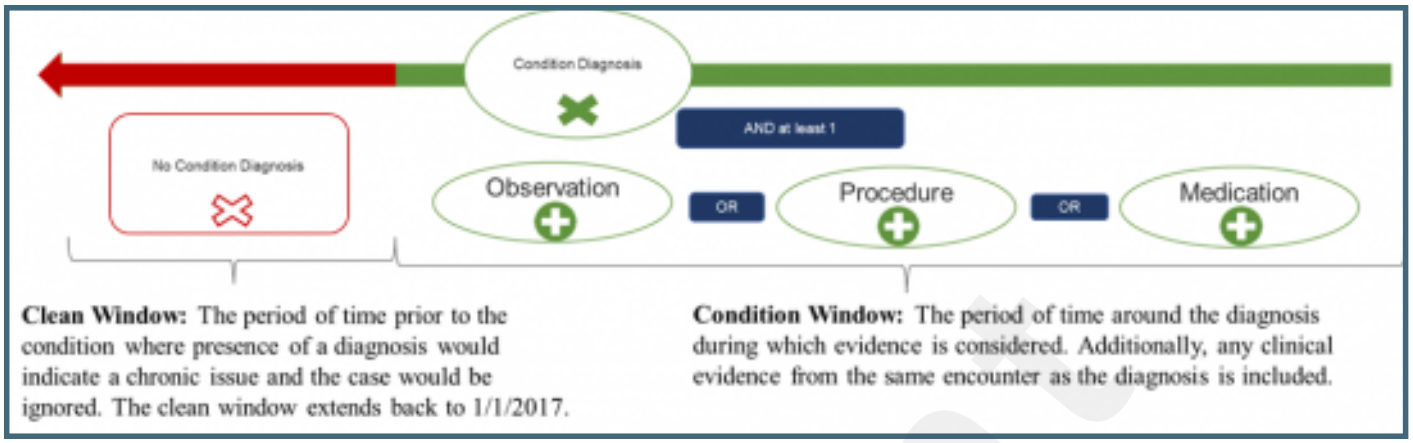
10.1016/j.vaccine.2023.06.014. Epub 2023 Jun 14. PMID: 37344261
22. Hobbi S, Saunders-Hastings P, Zhou CK, Beers J, Srikrishnan A, Hettinger A, et al. Development and Validation of an Algorithm for Thrombosis with Thrombocytopenia Syndrome (TTS) at Unusual Sites. Int J Gen Med. 2023 Jun 15;16:2461-2467.PMID: 37342408
23. Kostka K. Chapter 10 Defining cohort. In: Abedtash H, Ascha M, Beno M, Blacketer C, Blatt, D, Christian B, et al., editors. The book of OHDSI. New York: Observational Health Data Sciences and Informatics; 2021. https://ohdsi.github.io/TheBookOfOhdsi/Cohorts.html [accessed Feb 22, 2024].
24. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J Am Med Inform Assoc. 2019 Dec 1;26(12):1545-1559. doi:10.1093/jamia/ocz105
25. Walsh KE, Cutrona SL, Foy S, Baker MA, Forrow S, Shoaibi A, et al. Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2013 Nov;22(11):1205-1213. doi:10.1002/pds.3505
26. Bann MA, Carrell DS, Gruber S, Shinde M, Ball R, Nelson JC, et al. Identification and validation of anaphylaxis using electronic health data in a population-based setting. Epidemiology. 2021 May 1;32(3):439-443. doi:10.1097/EDE.0000000000001330
27. Funch D, Holick C, Velentgas P, Clifford R, Wahl PM, McMahill-Walraven C, et al. Algorithms for identification of Guillain-Barré syndrome among adolescents in claims databases. Vaccine. 2013 Apr 12;31(16):2075-2079. doi:10.1016/j.vaccine.2013.02.009
28. Kawai AT, Martin D, Henrickson SE, Goff A, Reidy M, Santiago D, et al. Validation of febrile seizures identified in the Sentinel Post-Licensure Rapid Immunization Safety Monitoring Program. Vaccine. 2019 Jul 9;37(30):4172-4176. doi:10.1016/j.vaccine.2019.05.042
29. Morgan J, Roper MH, Sperling L, Schieber RA, Heffelfinger JD, Casey CG, et al. Myocarditis, pericarditis, and dilated cardiomyopathy after smallpox vaccination among civilians in the United States, January–October 2003. Clin Infect Dis. 2008 Mar 15;46 Suppl 3:S242-S250. doi:10.1086/524747
30. Rüggeberg JU, Gold MS, Bayas JM, Blum MD, Bonhoeffer J, Friedlander S, et al. Anaphylaxis: Case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine. 2007;25(31):5675-5684. doi:10.1016/j.vaccine.2007.02.064
31. Sejvar JJ, Kohl KS, Gidudu J, Amato A, Bakshi N, Baxter R, et al. Guillain-Barré syndrome and Fisher syndrome: Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2011;29(3):599-612. doi:10.1016/j.vaccine.2010.06.003
32. 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. doi:10.1056/NEJMoa2104840
33. Chen, R., & Buttery Monashm, J. (2021). DRAFT - TTS: Case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Zenodo. https://doi.org/10.5281/zenodo.6697333
34. Marcy SM, Kohl KS, Dagan R, Nalin D, Blum M, Jones MC et al. Fever as an adverse event following immunization: Case definition and guidelines of data collection, analysis, and presentation. Vaccine. 2004 Jan 26;22(5-6):551-556. doi:10.1016/j.vaccine.2003.09.007
35. Bonhoeffer J, Menkes J, Gold MS, de Souza-Brito G, Fisher MC, Halsey N, et al. Generalized convulsive seizure as an adverse event following immunization: Case definition and guidelines for data collection, analysis, and presentation. Vaccine. 2004 Jan 26;22(5-6):557-562. doi:10.1016/j.vaccine.2003.09.008

36. Schuemie M, DeFalco F. Chapter 8.3 ATLAS. In: Abedtash H, Ascha M, Beno M, Blacketer C, Blatt, D, Christian B, et al., editors. The book of OHDSI. New York: Observational Health Data Sciences and Informatics; 2021. https://ohdsi.github.io/TheBookOfOhdsi/OhdsiAnalyticsTools.html#atlas [accessed Feb 22, 2024]
37. Value Set Authority Center. Bethesda, MD: National Library of Medicine, 2022. https://vsac.nlm.nih.gov/ [accessed Feb 22. 2024]
38. Whitaker B, Pizarro J, Deady M, Williams A, Ezzeldin H, Belov A, et al. Detection of allergic transfusion-related adverse events from electronic medical records. Transfusion. 2022;62(10):2029-2038. doi:10.1111/trf.17069
39. Deady M, Ezzeldin H, Cook K, Billings D, Pizarro J, Plotogea AA, et al. The Food and Drug Administration Biologics Effectiveness and Safety Initiative facilitates detection of vaccine administrations from unstructured data in medical records through natural language processing. Front Digit Health. 2021 Dec 22;3:777905. doi:10.3389/fdgth.2021.777905
40. Standardized data: The OMOP Common Data Model. New York: Observational Health Data Sciences and Informatics; 2023. https://www.ohdsi.org/data-standardization/ [accessed Feb 22, 2024].
41. Agresti A, Coull B. Approximate is better than “exact” for interval estimation of binomial proportions. Am Stat. 1998;52(2):119-126. doi:10.1080/00031305.1998.10480550
42. Brown L, Cai T, DasGupta A. Interval estimation for a binomial proportion. Statist Sci. 2001 May;16(2):101-133. doi:10.1214/ss/1009213286
43. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-282. PMID:23092060
44. Pandey C, Bose N, Garg G, Singh N, Baronia A, Agarwal A, et al. Gabapentin for the treatment of pain in Guillain-Barre Syndrome: A double-blinded, placebo-controlled, crossover study. Anesth Analg. 2002 Dec;95(6):1719-1723. doi:10.1097/00000539-200212000-00046
45. Idowu RT, Carnahan R, Sathe NA, McPheeters ML. A systematic review of validated methods to capture myopericarditis using administrative or claims data. Vaccine. 2013 Dec 30;31 Suppl 10:K34-K40. doi:10.1016/j.vaccine.2013.08.074
46. CQL – Clinical Quality Language. Washington, DC: eCQI Resource Center; 2022. https://ecqi.healthit.gov/cql?qt-tabs_cql=0 [accessed Feb 22, 2024]
47. Ong MS, Klann JG, Lin KJ, Maron BA, Murphy SN, Natter MD, et al. Claims-based algorithms for identifying patients with pulmonary hypertension: A comparison of decision rules and machine-learning approaches. J Am Heart Assoc. 2020 Oct 20;9(19):e016648. doi: 10.1161/JAHA.120.016648
48. Dang T, Lan X, Weng J, Feng M. Federated learning for electronic health records. ACM Trans Intell Syst Technol. 2022 Jun;13(5):72, 1-17. doi:10.1145/3514500
49. Manual for the surveillance of vaccine-preventable diseases. Atlanta: Centers for Disease Control and Prevention; 2021. https://www.cdc.gov/vaccines/pubs/surv-manual/chpt21-surv-adverse-events.html [accessed Feb 22, 2024]

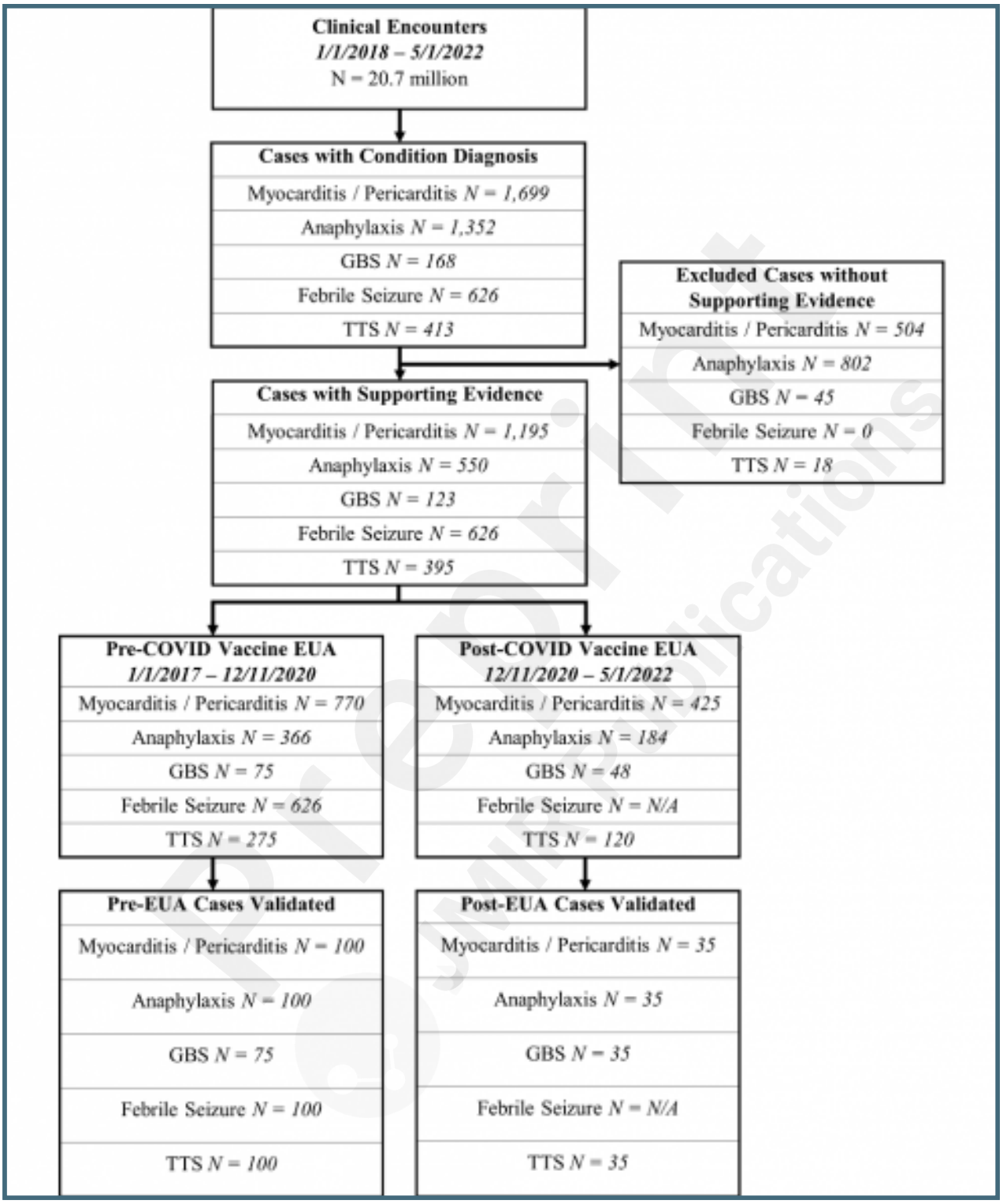
Supplementary Files

Figures

Composition algorithm modified for each AE.



Study population CONSORT diagram.



Multimedia Appendixes

Case definition criteria used for algorithm development and validation.
URL: http://asset.jmir.pub/assets/7e6895420bd9e8d71456a4694849fa85.docx

Untitled.
URL: http://asset.jmir.pub/assets/11b5bfc859d5b192ee16ac359b562b15.docx