C19 Vaccine Safety Protocol OHDSI Comments 2020

Observational Health Data Sciences and Informatics (OHDSI)
Contact:
George Hripcsak, MD, MS
Columbia University Irving Medical Center
OHDSI Coordinating Center
622 W 168th St PH20
New York, NY 10032
gh13@cumc.columbia.edu

SUMMARY

The Observational Health Data Sciences and Informatics (OHDSI) vaccine researchers applaud the U.S. Food and Drug Administration (FDA) and its Biologics Effectiveness and Safety team for their transparency in drafting the document, "COVID-19 Vaccine Safety Surveillance: Active Monitoring Master Protocol," and for sharing it for public comment. This will both build public trust and engage outside researchers to contribute to the protocol; collaborators within the OHDSI community stand ready and are eager to support the FDA. The protocol represents an excellent and thoughtful approach based on the concrete experience of the FDA team.

An early task is generating phenotype definitions for the adverse event outcomes of special interest (AESI) proposed in the protocol. Based on the referenced definitions, where provided, or previous literature and OHDSI experience where not, we have created formal definitions with full diagnostics for 17 of the outcomes (skipping one that is not yet formally defined). The definitions and diagnostics have been placed in the OHDSI public repository for review. We describe the phenotypes and our process below, and we illustrate our interpretation of the diagnostics for three of the outcomes. The definitions are based on the OHDSI Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

We also provide comments on the analytic plan. Our main comments relate to the variance of the baseline incidence estimates. We believe that wide variations in estimates for baseline incidence rates will be seen across databases, populations, and times, and this variance may swamp other corrections applied in the analysis. A study of the magnitudes of different sources of this variance will help to optimize the analytic plan. We believe a more extensive causal framework is useful even for signal detection, and without that, matching between the vaccination group and the baseline group will be important. Adjustments to the times at risk are noted, as is the need for diagnostics via controls.

1. INTRODUCTION

We applaud the FDA for their transparency in drafting the document, "COVID-19 Vaccine Safety Surveillance: Active Monitoring Master Protocol" and sharing it for public comment. The FDA's mandate for ensuring public safety in the responsible use of medical products is profound. Never before has this responsibility been more important than in addressing the global COVID-19 pandemic through vaccination of large populations. We commend the FDA for taking the proactive step of communicating their scientific approach to monitoring the safety of COVID vaccines, both because of the real-world evidence it will generate but also because of the need to

continue to build public trust by demonstrating how the FDA and its partners are actively working to preserve public health. To that end, collaborators within the OHDSI community stand ready and eager to support the FDA, the European Medicines Agency (EMA) and other public health organizations around the world in generating reliable real-world evidence that can promote better health decisions and better care.

We note with interest that the FDA has identified 16 adverse events of special interest (AESI) in the general population, and 3 AESI for examination in pediatric populations. As the FDA is well aware, a critical challenge in analyzing healthcare data captured during routine clinical care, such as administrative claims and electronic health records, is the development and evaluation of phenotype algorithms that can be used to identify outcomes within a population. Phenotype algorithms may be subject to measurement error, both due to misclassification of the events of interest – failing to identify all true cases (imperfect sensitivity) and falsely identifying non-cases (imperfect specificity) - as well as imprecisely assigning event dates and times. The concern about bias due to measurement error is one of the primary motivations behind the legacy use of clinical adjudication of source chart records in observation research, though this process itself is subject to substantial error and may be infeasible or prohibitively time and resource intensive. Recognizing this problem, the OHDSI community has sought to develop standardized processes to support the development and evaluation of phenotype algorithms that provides transparency for implementation and empirical evidence for verification. The components of this standardized processes are described below.

The OHDSI community has adopted the OMOP CDM, which is a deep information model for standardizing both the structure and content of patient-level longitudinal observations. The OMOP CDM supports data elements commonly captured in administrative claims data, such as procedure and diagnosis codes from billing records of inpatient and outpatient medical services, as well as prescription drug dispensing from outpatient pharmacies. It also allows standardization of data captured in electronic health record systems, such as physician orders, vital signs and laboratory measurements, problem lists, medication history, and free text clinical notes. Documentation about the OMOP CDM is available here: https://ohdsi.github.io/CommonDataModel/index.html. We note with pride that most of the data sources intended to be used within the FDA study have previously been standardized to the OMOP CDM and used in prior OHDSI network studies, including IBM MarketScan, Optum Clinformatics, Optum EHR, and PEDSNet, and observed that OneFlorida has adopted OMOP CDM within their consortium. The OHDSI community is prepared to support other FDA/CBER/BEST data partners in standardizing their data to ensure FDA can execute their analyses on a consistent platform across its network.

The OMOP CDM has enabled the OHDSI community to develop open-source standardized analytics to design and execute various types of studies, including clinical characterization for describing disease natural history and treatment utilization, population-level effect estimation for safety surveillance and comparative effectiveness, and patient-level prediction for precision medicine and disease interception. All of these tools have been successfully applied to network research in 2020 around COVID-19 (https://www.ohdsi.org/covid-19-updates/). Common to all of these open-source analytic packages is the use of cohort definitions as a standardized input to represent the target and outcome populations of interest. Cohort definitions are OHDSI's instantiation of the idea of 'phenotype algorithms' and implement rule-based heuristics which can be applied to OMOP CDM to identify the set of persons who satisfy one or more inclusion criteria for a duration of time.

The OHDSI community developed an open-source tool, ATLAS, which provides a web-based graphical user interface to design cohort definitions by constructing concept-sets (collections of concepts and source codes to represent a particular clinical entity) and imposing inclusion criteria and temporal logic around the presence or absence of concept-sets across the various data domains within the CDM to identify when patients enter and exit the cohort. ATLAS translates the user design into human-readable text and standardized JSON expression that can be automatically translated into computer-executable SQL to run against any OMOP CDM instance across multiple relational database dialects.

2. OUTCOME DEFINITIONS

To support FDA's current efforts, OHDSI has designed cohort definitions for all but one of the FDA AESI outcomes and have made these definitions publicly available at http://atlas.ohdsi.org.

Table 1 below provides the URLs to the 17 ATLAS cohort definition, where the concept-sets and code can be reviewed and executed against any OMOP CDM instance. Note, the first time you open one of the ATLAS URLs, you will be redirected to authenticate using a Google account, then you'll be able to explore these cohort definitions. If you have any technical issues or would like to review the definitions with the OHDSI team, feel free to reach out to Patrick Ryan directly at ryan@ohdsi.org.

Table 1: FDA AESI Phenotype list with OHDSI cohort definition links

Phenotype	**Cohort definition URL**
Guillain-Barre syndrome	https://atlas.ohdsi.org/#/cohortdefinition/348
Bell's Palsy	https://atlas.ohdsi.org/#/cohortdefinition/347
Anaphylaxis	https://atlas.ohdsi.org/#/cohortdefinition/349
Encephalomyelitis	https://atlas.ohdsi.org/#/cohortdefinition/346
Narcolepsy	https://atlas.ohdsi.org/#/cohortdefinition/345
Appendicitis	https://atlas.ohdsi.org/#/cohortdefinition/344
Non-hemorrhagic stroke	https://atlas.ohdsi.org/#/cohortdefinition/342
Hemorrhagic stroke	https://atlas.ohdsi.org/#/cohortdefinition/341
Acute myocardial infarction	https://atlas.ohdsi.org/#/cohortdefinition/340
Myocarditis/pericarditis	https://atlas.ohdsi.org/#/cohortdefinition/339
Deep vein thrombosis (DVT)	https://atlas.ohdsi.org/#/cohortdefinition/338
Pulmonary embolism (PE)	https://atlas.ohdsi.org/#/cohortdefinition/337

Disseminated intravascular coagulation (DIC)	https://atlas.ohdsi.org/#/cohortdefinition/336
Immune thrombocytopenia (ITP)	https://atlas.ohdsi.org/#/cohortdefinition/335
Transverse myelitis	https://atlas.ohdsi.org/#/cohortdefinition/334
Febrile seizure	https://atlas.ohdsi.org/#/cohortdefinition/332
Kawasaki disease	https://atlas.ohdsi.org/#/cohortdefinition/331

We highlight the content of each cohort definition, by providing the illustrative example of Anaphylaxis. The cohort definition is shown in the ATLAS design editor in Figure 1:

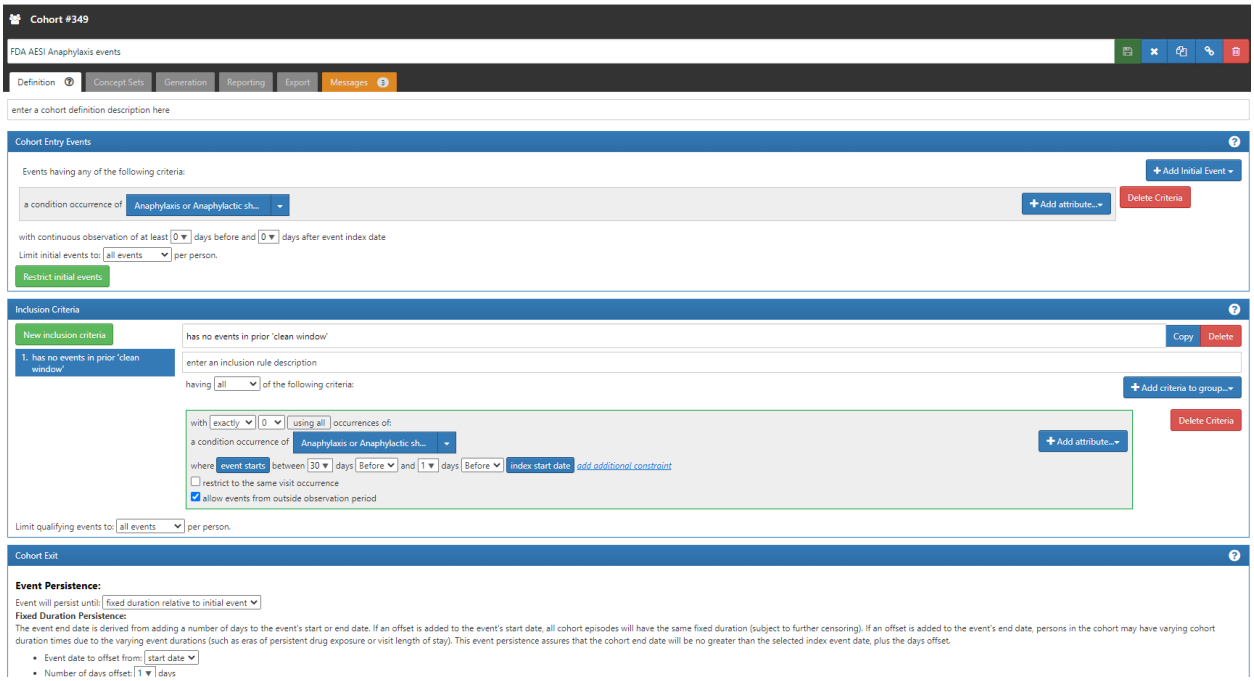


Figure 1: ATLAS cohort definition editor: Anaphylaxis events

In this definition, persons have a cohort entry event on the date of a condition occurrence record containing a concept in the 'Anaphylaxis or Anaphylactic shock' concept-set. All such occurrences are considered candidate entry events. The definition imposes one inclusion criteria: 'has no events in prior 'clean window', which is implemented by requiring that exactly 0 occurrences of 'Anaphylaxis or Anaphylactic shock' are observed in the 30 days preceding the entry event. Persons exit the cohort one day after the entry event. Under this definition, one person may have multiple anaphylaxis events, so long as the duration between successive diagnosis codes is more than 30 days.

The concept-set for 'Anaphylaxis or Anaphylactic shock' can be explored on the 'Concept Sets' tab at the top of the ATLAS design editor, as shown in Figure 2:

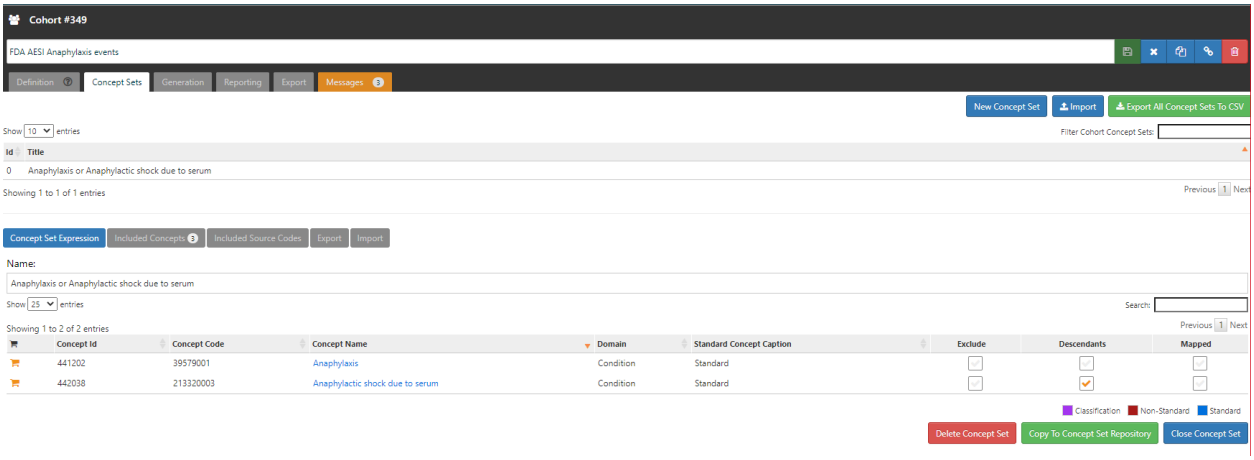


Figure 2: ATLAS concept sets: Anaphylaxis standard concepts

The concept-set expression involves 2 standardized concepts from the SNOMED vocabulary: 'Anaphylaxis' (CONCEPT_ID 441202) and 'Anaphylactic shock due to serum' (CONCEPT_ID 442038). The second concept also includes its descendants in the SNOMED hierarchy. The expression results in 3 included standard concepts in total ('Shock co-occurrent and due to anaphylaxis caused by serum' is also identified). All databases transformed into the OMOP common data model utilize standardized vocabularies so that it becomes possible to identify clinical entities in a consistent fashion.

Databases may be natively coded in the vocabulary standard (here, SNOMED-CT), but may also use source codes that require mapping into the standard. For example, US administrative claims can contain ICD-9-CM and ICD-10-CM codes in diagnosis fields. The OMOP standardized vocabulary allows for review of all source codes that map into the standardized concept-set expression. The 'Included source codes' button within 'Concept Sets' tab of the ATLAS design editor allows users to search and filter to specific source vocabularies and review the corresponding codes, as shown in Figure 3:

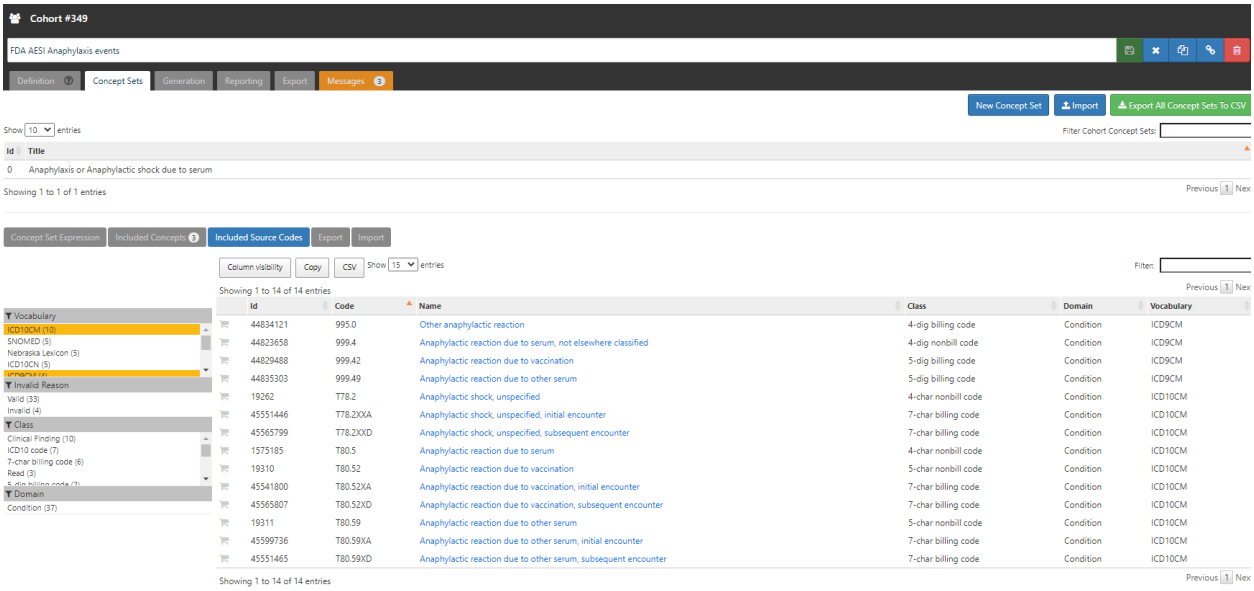


Figure 3: ATLAS included sources codes for Anaphylaxis conceptset

The concept-set expression for 'Anaphylaxis or Anaphylactic shock', which was produced using only 2 standard concepts, covers 37 distinct source codes across 9 source vocabularies. Filtering to ICD-9-CM and ICD-10-CM reveals that this concept-set resolves to 4 ICD-9-CM codes (995.0 'Other anaphylactic reaction', 999.4 'Anaphylatic reaction due to serum, not elsewhere classified', 999.42 'Anaphylactic reaction due to vaccination', and 999.49 'Anaphylactic reaction due to other serum') and 10 ICD-10-CM codes (T82.2, T82.2XXA, T78.2XXD, T80.5, T80.52, T80.52XA, T80.52XD, T80.59, T80.59XA, T80.59XD). ATLAS cohort definitions and concept-sets can be exported from the links and imported into other ATLAS instances for direct use against any OMOP CDM instance. Additionally, code-lists can be exported from ATLAS as comma-separated, variable-length text table (*.csv) for review and use by data partners who have not yet adopted OMOP CDM.

We further highlight one remaining AESI for which we did not yet provide a definition is 'Multi-system inflammatory syndrome', both in children (MIS-C) and adults (MIS-A). Given that MIS-C and MIS-A are rare but unique constellations of symptoms and our understanding the disease has been evolving over the course of the COVID pandemic, it is unclear how MIS will be represented in observational data or whether it is possible to reliably estimate incidence in a pre-COVID period when the syndrome was not recognized. As such, MIS appears to be qualitatively differently in scope from other outcomes, and will require further phenotype development and validation before it could be included. Members of the OHDSI community are interested in collaborating on this important line of methodological research.

The OHDSI community has developed an open-source tool, CohortDiagnostics (https://ohdsi.github.io/CohortDiagnostics/), which allows data partners to evaluate cohort definitions through a suite of design and empirical diagnostics as part of the phenotype development process. Design diagnostics include recommendations for concepts to consider for inclusion in concept-sets based on lexical search and ontology relationships. Data diagnostics include: examination of incidence rates, stratified by age, sex, and year; distribution of time before, during, and after cohort entry; breakdown of the specific concepts that qualify entry into

the cohort; context around the types of healthcare encounters were observed before/during/after cohort entry; and, temporal characterization of the prevalence of other observations such as conditions, drugs, procedures, and measurements- within time windows relative to cohort entry. The final results of CohortDiagnostics, which are aggregate summary statistics with no patient-level identifying data, can then be exported and shared with other collaborators to compare the performance of phenotype algorithms across data sources. In 2020, OHDSI released the PhenotypeLibrary as a central open-source repository of phenotypes, cohort definitions, and CohortDiagnostics results contributed from across the community, containing information for >200 phenotypes.

To support FDA's efforts as part of this COVID vaccine surveillance protocol, OHDSI has added the 17 FDA AESI cohort definitions listed above to the PhentoypeLibrary and executed CohortDiagnostics across 10 collaborating data sources. The current set of results for FDA AESI are available for public exploration at: https://data.ohdsi.org/PhenotypeLibrary/.

To demonstrate the value of the CohortDiagnostics results, we highlight key findings from 3 phenotypes in Appendix 1: Anaphylaxis, Transverse myelitis, and Bell's palsy.

To further the aims of transparency and verification, we encourage other data partners to similarly execute CohortDiagnostics against the FDA AESI definitions within the OHDSI PhenotypeLibrary and to share the results for inclusion into this community resource. The fully executable study as a R-package is here https://github.com/ohdsi-studies/PhenotypeLibraryDiagnostics . Opportunities for further collaboration to support FDA include: evaluate alternative cohort definitions for the identified FDA AESI phenotypes with an aim to reduce measurement error, develop definitions for additional phenotypes that warrant further investigation as potential outcomes of interest, run diagnostics across more databases to verify consistency and explore population heterogeneity, and to collaboratively interpret the existing diagnostics results to ensure an necessary context is established when interpreted observed events within the COVID vaccine exposed populations as data become available.

Development of phenotypes is a pre-requisite step to FDA's objective of characterizing incidence rates of events within the COVID vaccine exposed population, and to compare incidence rates against referent populations. Incidence characterization also requires definition of the population at-risk, and the time horizon that will be surveilled to identify new cases of an outcome. Further clarification may be required to define the referent populations of interest to estimate the 'background rate' used in the pMaxSPRT computation. It is likely that the OHDSI cohort definition framework, as implemented in ATLAS, can be similarly helpful for constructing the at-risk populations of interest. By defining entry events, inclusion criteria, and an exit strategy, a cohort can be produced that provides a clear index date to use to anchor a time-at-risk in the referent population, analogous to the date of COVID vaccine administration used in the target population.

Multiple comparator cohorts could be considered, each carry with them their own limitations and potential for bias:

- Persons observed in a database on a specific date (ex: persons with an observation period that spans 1 JAN 2017 (index date)).
- Persons with a healthcare visit in a given time interval (ex: persons indexed on their first visit that falls within 2017)

- Persons with a specific exposure in a given time interval (ex. persons indexed on influenza vaccine administration date during 2017-2018 flu season)

As the FDA protocol acknowledges as a limitation, the key challenge in comparing the incidence rate derived from a referent population to the observed rate of the COVID vaccinated population is the generalizability of the referent population, both in terms of patient composition and their temporal proximity to healthcare utilization. The comparability of patient composition can be evaluated as a diagnostic by characterizing baseline features at the index date, such as demographics (age, sex, race, calendar month), conditions, drugs, procedures and measurements. With baseline characteristics consistently evaluated for the target cohort and comparator cohort(s), it would be possible to compute standardized difference of mean for each baseline feature and determine which characteristics represent a potential threat to validity. Even in absence of performing statistical adjustment, such as matching or stratification on one or more observed variables or a propensity score, the covariate balance diagnostic could be informative to determine how much confidence to place in any comparison. The threat of validity due to different temporal proximity to healthcare utilization can be evaluated by using comparator groups with defined index dates that serve as suitable proxy for the counterfactual of COVID vaccine exposure, and using a consistent time-at-risk window between the target and comparator groups.

To support FDA's efforts, we illustrate how ATLAS can be used to characterize incidence rates with overall populations and defined subgroups of interest, by implementing a design specification that replicates the FDA protocol, as shown in Figure 4. The user defines a list of target cohorts (which in this case can include multiple alternative comparator populations to consider for the COVID vaccine exposed cohort), the list of outcome cohorts, and a time-at-risk interval. Additionally, if one wishes to examine incidence within subpopulations, she can do so by creating stratification criteria. The FDA protocol specifies subgroups defined by sex, age group, race and ethnicity, so this has been similarly implemented here.

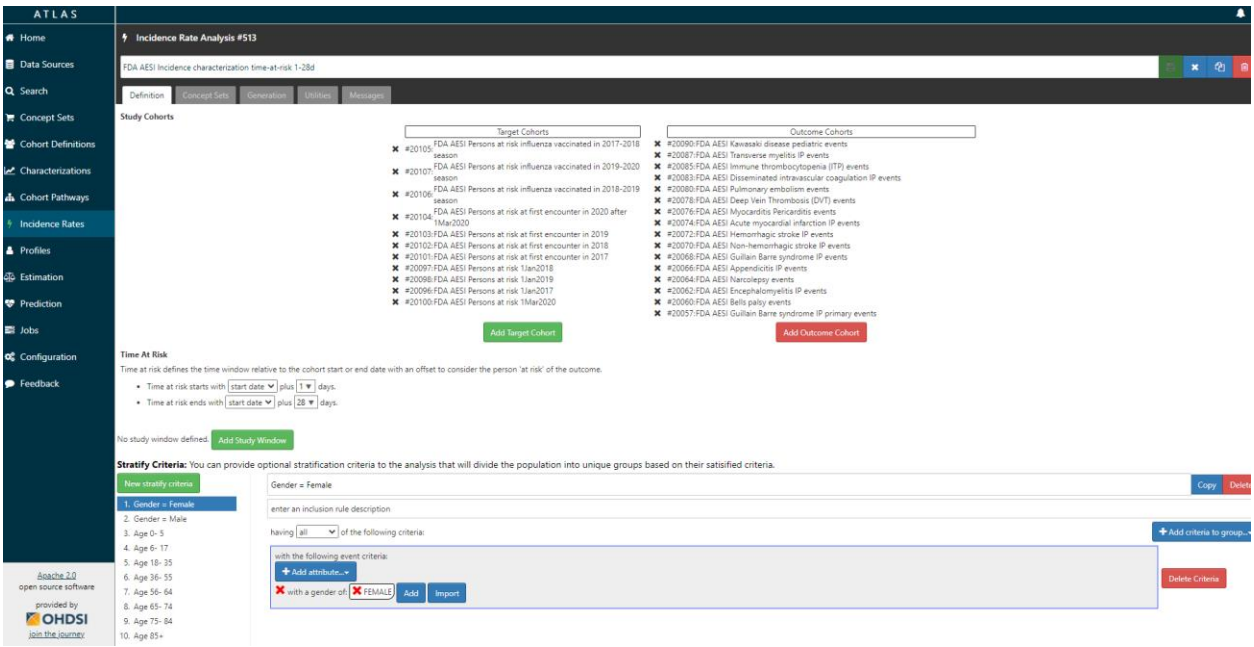


Figure 4: ATLAS incidence rate designer

Incidence analyses have been created by the OHDSI community and made publicly available to support investigations into multiple time-at-risk windows.

It is unclear if annual incidence rates in comparator populations will be consistent with rates estimated from defined time windows. The assumption in incidence rate estimation is a constant risk over the time horizon, and this analysis may be useful as a diagnostic to test that assumption.
Time-at-risk : 0d-365d: https://atlas.ohdsi.org/#/iranalysis/7

Anaphylaxis and febrile seizures are rare adverse events that, if they occur, would be expected in short proximity relative to exposure. Background rates of these events must appropriately account for time and likely needs a sufficiently comparable proxy of a similar healthcare encounter on the index date. Note, given the short time horizon, incidence proportion and incidence rates will be substantially different.
Time-at-risk: 0-2d: https://atlas.ohdsi.org/#/iranalysis/8

Other FDA AESI listed risk windows of 28, 42, or 90 days post-index. In order to evaluate the sensitivity of the time-at-risk analysis design choice, it is possible to characterize incidence for each of the time windows.
Time-at-risk: 1-28d: https://atlas.ohdsi.org/#/iranalysis/10
Time-at-risk: 1-42d: https://atlas.ohdsi.org/#/iranalysis/11
Time-at-risk: 1-90d: https://atlas.ohdsi.org/#/iranalysis/12

Figure 5 shows how the ATLAS output from an incidence rate design produces the evidence consistent with the FDA's desired intent, using the IBM MarketScan Multi-state Medicaid (MDCD) database for the target cohort of 'Persons at risk influenza vaccinated in 2017-2018 season' and the outcome cohort of 'Deep vein thrombosis.' In addition to providing the number of persons, total person-years at risk and the number of cases during the time-at-risk window, ATLAS computes the incidence proportion and incidence rate. The analysis is performed across the entire target cohort, shown in the summary at the top, and then additional for each subpopulation, shown in the table to the left. The treemap visualization on the right allows the user to explore the n-way interactions between all stratification criteria to identify subpopulations that have differential risk. In this illustration, we can see females who are 85+ yo and White have the high incidence rate, >7x higher than the overall rate. The substantial disparities in the incidence rate by gender and by age in this example underscore the general need to consider some adjustment when comparing baseline rates to observe rates, because any differences could otherwise be attributable to imbalance on baseline characteristics.

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	1,142,237	1,895	1.66	988,667	1.92

	Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1.	Gender = Female	629,306	1,232	1.96	543,888	2.27
2.	Gender = Male	512,931	663	1.29	444,775	1.49
3.	Age 0- 5	284,868	15	0.05	246,063	0.06
4.	Age 6- 17	462,879	44	0.10	401,627	0.11
5.	Age 18- 35	117,999	192	1.63	98,050	1.96
6.	Age 36- 55	132,155	627	4.74	114,498	5.48
7.	Age 56- 64	86,205	489	5.67	75,736	6.46
8.	Age 65- 74	35,340	291	8.23	32,181	9.04
9.	Age 75- 84	17,426	169	9.70	15,817	10.68
10.	Age 85+	4,324	54	12.49	3,760	14.36
11.	Race = Black	316,931	662	2.09	279,095	2.37
12.	Race = White	539,174	997	1.85	477,047	2.09
13.	Race = Asian	0	0	0.00	0	0.00
14.	Ethnicity = Hispanic	121,575	45	0.37	107,581	0.42

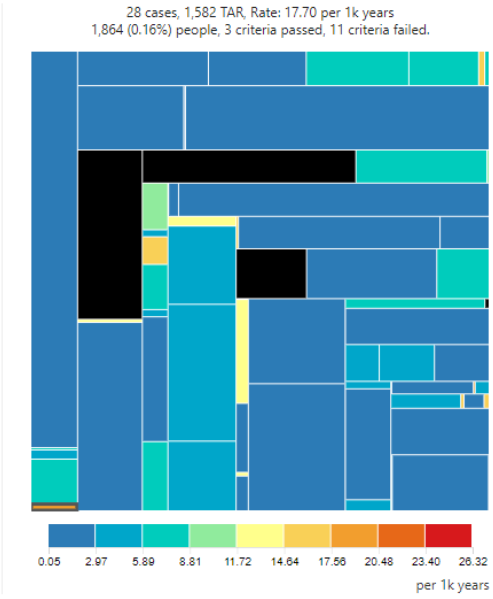


Figure 5: ATLAS incidence summary output

To support FDA's effort to characterize the background rates of COVID AESIs, one collaborative opportunity that could be undertaken is to design and execute an OHDSI network study that computes incidence rates across a range of comparator populations, calendar times and subgroups of interest for the entire parent of AESI phenotypes. The analysis procedure could be executed across multiple data partners within the OHDSI community to compare results. Such information could be used to quantify the uncertainty in the background rate estimation and to properly contextualize that will accrue from the COVID vaccine exposed population in 2021. Given the work that the OHDSI community has already prepared in response to the FDA's protocol and request for comment, such a study should be feasible in a reasonable time period to inform future steps.

3. COMMENTS ON THE ANALYTIC PLAN

Once again, we applaud the detailed analytic plan, which provides a concrete and detailed description of the planned analysis. The plan includes a detailed description of the time windows to be employed and analysis to be applied. The plan accounts for the strengths and limitations of the data sources, corrects for bias where appropriate, and acknowledges the inherent limitations in the analysis. We provide the following comments on the plan.

Variance in incidence rates

Baseline incidence rates are a key component of the analysis. Based on our experience, including a previous assessment of incidence rate of all diseases after exposure to each marketed drug over many databases, estimated incidence rates will vary widely by database, population, and time, in addition to the natural chance variation seen especially with rare outcomes. It is not clear in the

formulation in the protocol's section 4.6.1 how such variation will be included in the analysis (i.e., how variation in mu_t is accounted for). We believe that the largest component of this variance will be systematic error, resulting potentially in several-fold differences in incidence estimates. The protocol describes procedures for addressing some of this systematic error, such as looking at cohorts who underwent influenza vaccination to better match to the COVID-19 vaccination population and comparing rates before COVID-19 with those during COVID-19, which are subject to changes in health-seeking behavior due to lockdowns or hesitancy. In addition to these factors, there are also natural differences in data collection, coding by health care providers, geographic focus, age group, and comorbidities, as well as a variation across states in who receives which COVID-19 vaccine and when. Furthermore, there may be severe detection bias, in which patients and ordering physicians become hypervigilant to adverse events suspected for the vaccine, increasing reporting beyond that observed in the baseline period. These variations will produce differences in rate estimates that may significantly exceed some of the other corrections applied in the analysis (e.g., selection of the test margin, alpha spending plan).

We suggest three approaches to addressing this concern: (1) better characterize the sources of variance in the baseline incidence rates and include that variance in the analysis, (2) consider a full causal model, and (3) better match the baseline population to the vaccination population.

Characterize the variance

We strongly believe that a series of experiments can characterize the sources of variance of the baseline incidence rates and steer the analytic process. These sources, as mentioned above, include demographics, eligibility criteria (e.g., whether they must have had a visit), comorbidities, the precise definition of index date and time at risk, database, time of observation (e.g., with relation to COVID-19 introduction and impact across states), and other health care processes. Our very preliminary analysis reveals that patient-level heterogeneity may be substantial for many of the AESIs, with baseline characteristics such as demographics (age, sex, race) and comorbidities demonstrating large swings in estimated incidence rates. This heterogeneity may even be larger than variance across time with respect to COVID-19 impact or referent influenza season.

Given large patient-level heterogeneity, attempts can be made to better match the populations to improve discrimination (either through a full causal model or through other matching, as noted below), the variance can be included directly in the analytic model, or the results of the analysis can be calibrated post-hoc, widening the confidence intervals to account for the extra variance.

Once vaccination data become available, they, too, can be characterized both through a similar analysis as above and through chart review, where available.

Causal inference with a full propensity model

We acknowledge the decision to avoid full causal inference and to consider this analysis a hypothesis generating study. Nevertheless, even in generating hypotheses, causal inferences are being made. We believe that it is worth exploring the possibility of using a full propensity model even in real-time surveillance. With a systematic approach like large-scale propensity score adjustment that does not require manual selection of confounders, and with modern software implementations and computing power, this is a feasible solution on an ongoing basis. An alpha spending plan could be developed for it. (We acknowledge that even a full causal model will still be subject to issues like detection bias.)

Matching populations and matching analyses to reduce bias

Barring a full analysis, the protocol recognizes the need to match populations between the baseline incidence rates and the observed COVID-19 vaccine recipients, looking at patients vaccinated for influenza and looking at pre- and post-COVID-19 observation periods. It is unclear from the protocol if the influenza vaccine baseline incidence group is being used just to select similar patients or if the baseline incidences are being calculated with identical time windows to the COVID-19 vaccination group. We believe that in measuring the baseline incidence rates, the times at risk may need to parallel those used in the vaccinated group. For example, if the vaccinated group has times at risk as little as 30 days, then calculating a baseline risk using a three-year window (2017 to 2019) corrected to 30 days may not give a comparable rate. We suggest instead measuring the sensitivity of choice of time at risk, comparing the longer interval to using a triggering event, such as influenza vaccination or drug refill, and applying the same protocol as illustrated in Figure 1.

It will be important to match populations (even if short of propensity matching) as closely as possible to get appropriate incidence rates. For example, if early vaccine recipients are nursing home residents and baseline rates are calculated from entire database populations, the expected incidence rates could differ by several fold or an order of magnitude. Therefore, there ought to be an attempt to match the population used to calculate the baseline incidence rates to the population used in the vaccine surveillance. If full adjustment were not possible, one could generate baseline incidence rates stratified by a small number of factors such as age, sex, comorbidity level, and season and do a population-level standardization adjustment on the baseline incidence rate once the vaccine population is observed and characterized.

Diagnostics via controls

Diagnostics will be critically important in assessing the validity of the analysis, and one approach to generate diagnostics is through control exposures and outcomes. Negative exposure controls can test the false positive rate of background incidence rates. Examples here might include drug refills and visits for medical reasons not related to the outcomes. Influenza vaccination events can serve as a comparison exposure rather than as a negative control. Negative outcome controls can also test the false positive rate of the analytic approach.

We note here that “control” have several meanings: the differences among a comparator cohort of patients likely to have similar characteristics to those vaccinated, a formal comparison of event rates for influenza vaccine versus COVID-19 vaccine, a negative control cohort of a disease like hypertension to assess the effect of COVID-19 on health seeking behavior, negative control exposures to test the discrimination of the analytic approach, and negative control outcomes also to test the discrimination of the analytic approach.

Time at risk

The use of a moving clean window as shown in the protocol’s Figure 1 reduces some biases that could affect approaches that are based on a fixed pre-vaccination clean window. For example, it eliminates the bias of late events requiring a longer effective clean window than earlier ones because the proposed moving clean window is of constant length.

The moving clean window may produce immortal time bias in the results, however, unless it is accounted for in the time-at-risk calculation. For example, for anaphylaxis, if the patient has an episode of anaphylaxis 29 days before vaccination, then there is a day of immortal time after the vaccination where anaphylaxis is not possible (or it would have counted as part of the previous event). This can be solved by adjusting the time at risk to avoid the immortal time. In the example, that patient's time at risk begins on day 1 instead of day 0, going from 1 to 2 days instead of 0 to 2 days. This implies that information about the previous events must be carried forward into the analysis, either as the events themselves or as eras produced from the events.

On requiring an inpatient setting

We clarify that when an outcome is limited to the inpatient setting, this implies that all other references to the outcome in other settings are being assumed to be non-events and are instead follow up to a disease that is formally ended or simply errors. The outcomes chosen for this treatment, such as myocardial infarction, appear to be appropriate.

Population at risk

Some patients are not subject to risk of a disease, such as appendicitis not at risk of occurring in patients who have previously had an appendectomy. This will reduce the event rate to a small degree. Such patients can be eliminated in the target cohort definition, or, more simply, they can be accounted for as long as the probability is low by including them in both the baseline incidence rates and the COVID-19 vaccination group.

Other patients are at such high risk for an outcome that a vaccine would never be blamed for the event, such as transverse myelitis in a patient with previous multiple sclerosis. There is a choice of doing manual review after the analysis to assign likely causality, but eliminating these known cases before the analysis to avoid false positive signals.

Use of test margins to reduce false positives

We wonder if test margins will vary by outcome not just by rate and risk window but also by the seriousness or cost of the outcome, as opposed to equal weight for adding the same number of events per period of time.

Multiple events

It is not currently possible to have multiple events because the clean windows are longer than the times at risk, but if further outcomes are added with relatively longer times a risk, multiple events for the same patient will need to be considered.

Further work on outcome definitions

In addition to the outcome phenotyping work described above, OHDSI can further assist in the definition and validation of outcome phenotypes using its concept prevalence data from across its network. These data reveal the usage of concepts across the network and can assist in understanding the potential variance in coding that may be seen within the FDA databases.

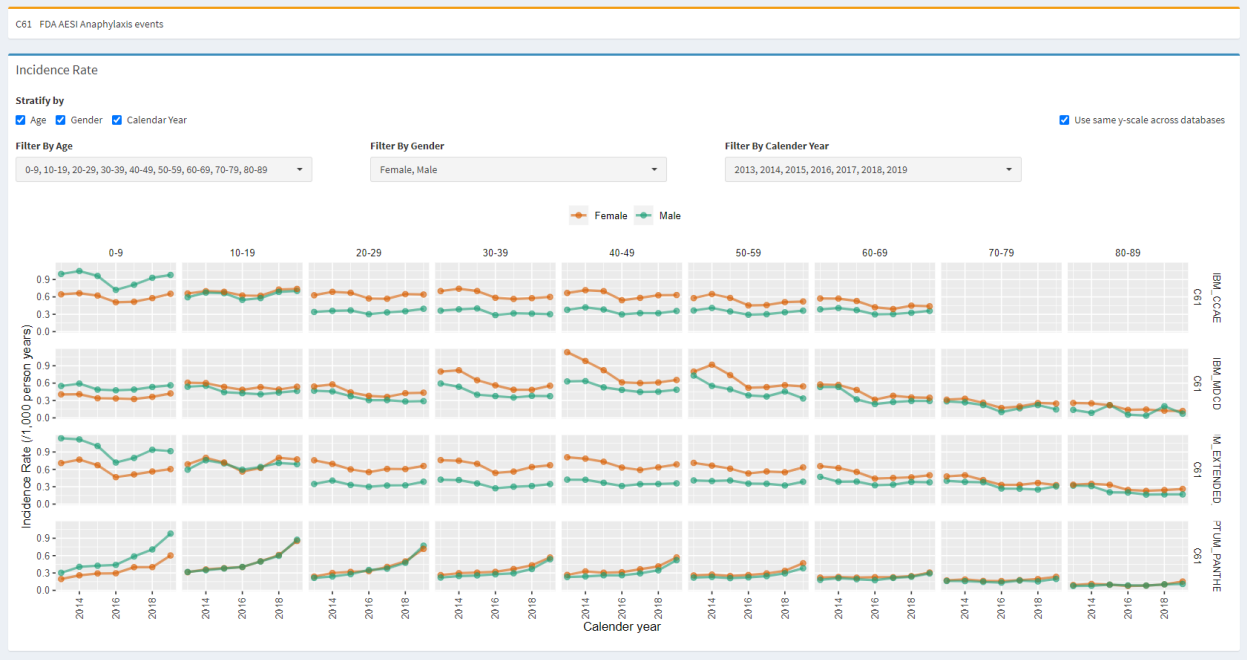
EHR databases

EHR databases are currently proposed to only be used to generate descriptive statistics. OHDSI can assist in the comparison of the performance of signal detection and signal evaluation methods between claims databases and EHR databases, illustrating whether—due to issues like incomplete follow up—EHR databases show reduced discrimination.

Appendix 1: Key findings from preliminary review of OHDSI PhenotypeLibrary results for selected FDA AESI phenotypes

Anaphylaxis:

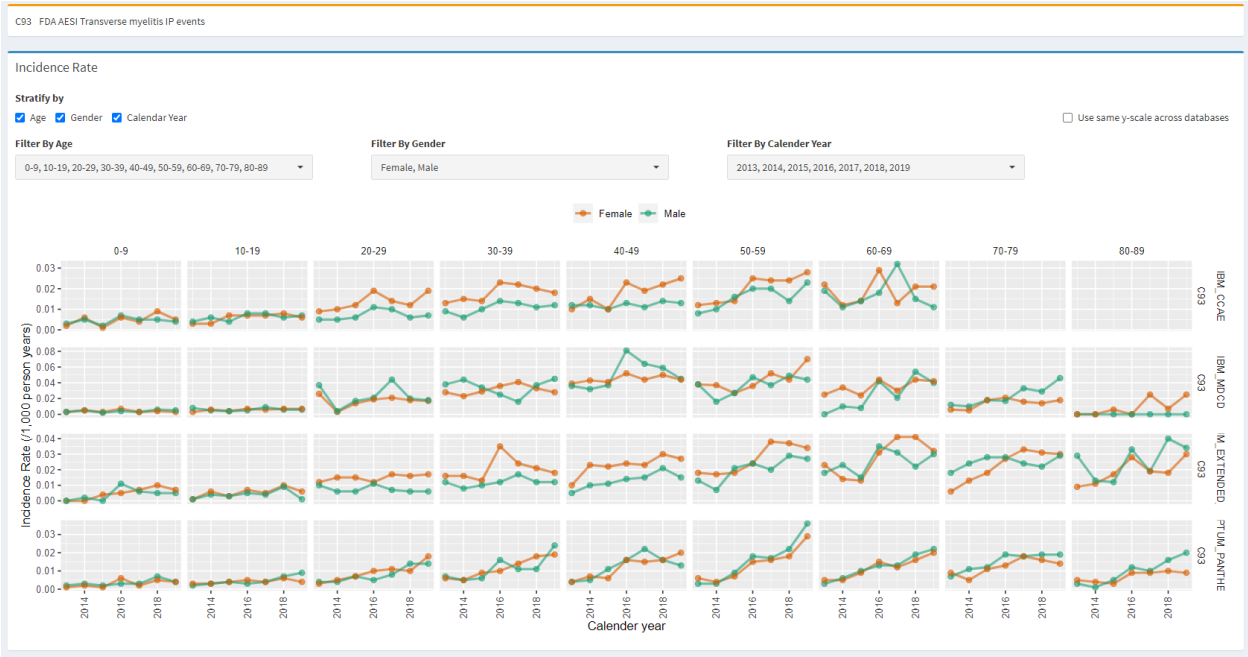
- Overall annual Incidence rate: 0.25-0.5 per 1,000 person-years, across age/sex/years
- Incidence higher in age < 20yo
- Incidence is higher in males < 10yo, but higher in females > 20yo
- Consistent magnitude across US claims databases in age strata, and relatively stable after 2015. (Rates appeared higher pre-2015….coding practice change?)
- Optum EHR shows increasing pattern



- Visit context: majority of event occurring in outpatient visit (<5% are inpatient visits). May be useful to characterize incidence of ER/IP, to provide greater context around severity.
- Temporal characterization: we see allergy tests occur ~12% on index, may want to exclude as part of definitions.
- Epinephrine, prednisone, albuterol are common treatments observed, but <50% total
- Observe ~10% prevalence of epinephrine in 30d prior, potential concern on index date error or just susceptibility (prescription PRN)?
- Time distribution - >75% time all care on index date, unlikely to have subsequent anaphylaxis code within 30d

Transverse myelitis:

- Overall annual incidence rate: 0.006 – 0.015 events per 1,000 person-years, across all age/sex/years
- No consistent changes over time since 2016, but pre-2015 was lower (so may not be able to rely on literature estimates based on ICD-9-CM coding)
- CCAE, Optum Clinformatics show higher incidence in female vs male, but not observed in MDCD or Optum HER
-

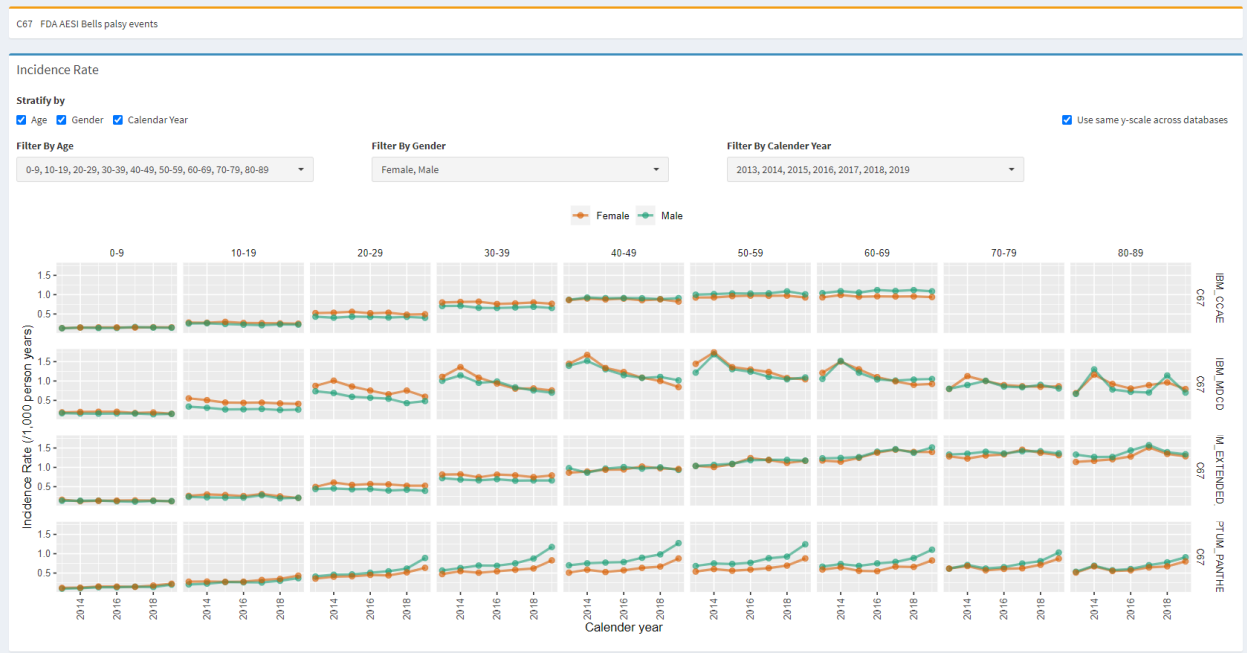


Temporal characterization:

- MRIs commonly observed before/during/after diagnosis (plausible to be part of work up, consider when defining event index date to reduce misclassification error)
- 15% had TM diagnosis in 30d prior to hospitalization code
- 6% in prior year, 10% have multiple sclerosis (MS) in 30d prior, 9% on index, 17% within 30d…..do we exclude persons with MS prior or within 30d?
- ~5-10% with non-specific codes around encephalitis, myelitis, encephalomyelitis

Bell's Palsy:

- Overall annual incidence rate 0.5 – 1.1 per 1000 person years, across all age/sex/years
- Stable over time in CCAE, MDCR, but decreasing in MDCD, increasing in Optum
- No consistent gender differences



- Visit: Context:
  - CCAE: most present outpatient visit, <10% in ER/IP
  - MDCR, Optum: ~30% have initial presentation at ER
- Temporal characterization:
  - Prednisone in ~30% on index, 10% valacyclovir, 8% acyclovir
  - CT scans for ~19% on index, >5% in 30d prior so may consider revising index date
  - Diagnosis codes for ischemic stroke (cerebral infarction, TIA) occur ~5%
  - >25% have some follow-up care within 30d