Comments on "Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring"

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, "Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring," and for sharing it for public comment. The protocol is well thought out and detailed.

Our main comment relates to the large magnitude of the expected systematic error. Differences in age groups and differences in the anchoring of the index date can create differences in incidence rates exceeding 10 times. Not only do databases vary in age distribution, but the age of those vaccinated for COVID-19 will be biased by selection procedures that change over time. Therefore, we do not believe that a single incidence rate can be reasonably defined and age-adjustment should be a primary objective of the background rate protocol. Other factors like sex produce a smaller effect but may also be adjusted. The anchoring of the index date of the time at risk also produces large differences in incidence rate estimates. The protocol uses an arbitrary date, such as January 1 or whatever date the patient first qualifies after the clean window, but the actual COIVD-19 vaccinations may be anchored to medical encounters with a huge difference in background rates. We recommend that background rates be estimated both ways to better understand the variability likely due to this choice.

We comment on other details, such as the use of incidence proportion in the EHR databases, whereas it may be appropriate to use both approaches in both sets of databases. We believe that a large set of negative outcome controls should be added, and we believe it may be possible to expand the number of outcomes studied. Additional comments are included.

INTRODUCTION

We applaud the FDA for its transparency in drafting the document, "Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring" and sharing it for public comment. The plan includes a detailed description of the databases, study populations, timing, and limitations in calculating background rates of adverse events in preparation for monitoring the safety of COVID-19 vaccines. We provide the following comments on the plan, with recommendations where appropriate.

The estimation of background rates from observational data is a complex process with many possible biases. The use of large databases implies that systematic error is likely to far exceed

error due to sampling variability. The proposal navigates the many design choices well based on knowledge and experience, but we believe that empirical testing is the only way to prioritize which design choices are most important and to arrive at the reliable method to discover adverse event signals. We have carried out some preliminary work to better understand these design choices, and that work is reflected in our comments below. Based on such work, it can then be decided whether to reduce bias by matching populations (e.g., calculate database-specific incidence rates), by performing corrections (e.g., age adjustment recommended below), or by using the known variance to better interpret the results.

The main concern relates to the generalizability of the population and healthcare dynamics that gave rise to the background rate, and the extent to which the estimates can reliably serve as a proxy for the counterfactual of the observed rate. Most important, we find that two issues are most critical. Many choices result in variations of the rate estimate of up to a factor of two, but for two of them--age distribution and anchoring events--we see variations in estimates over a factor of 10, a rate that renders estimation unreliable. We address these two important factors first in our comments.

AGE DISTRIBUTION AND OTHER EFFECTS

Even within adults, we can see variations in incidence rates by age group over a factor of 10, depending on the outcome. For outcomes like stroke, wide variation by age is in fact expected. Therefore, any estimate is very highly dependent on the age distribution of the sample. The incidence rate is so dependent, in fact, that we do not believe it is reasonable to talk about an overall incidence rate, other than as an academic exercise. The incidence rate must always be stated along with the age distribution of the sampled population. Databases vary in their age distribution, with large effects on incidence rates. More important and more challenging, the age distribution of patients vaccinated for COVID-19 is currently unknown, it is likely to differ strongly from the underlying database due to selection bias related to vaccination with groups like health care workers and the elderly being prioritized, and it will shift over time as different groups are emphasized. There may also be a bias in the recording of vaccination for different age groups.

Comparing rates of outcomes in COVID-19 vaccinated patients to incidence rates estimated from the same database will help but not solve the problem. Instead, the incidence rates must be adjusted for age on the fly as data are gathered. This can be done by calculating age group-specific incidence rates in each database and then age adjusting the incidence rates to the age distribution of the COVID-19 vaccine group. Therefore, we believe that age-based estimates of incidence rates should become the primary objective of the protocol rather than a secondary objective.

For many of the AESI, there is substantial patient-level heterogeneity beyond the age effect noted above. For example, acute myocardial infarction is associated with age, sex and race; non-hemorrhagic and hemorrhagic stroke is associated with age and race; pulmonary embolism and deep vein thrombosis rates have imbalance between Black and White races. Given this empirical observation, even with age adjustment, a single composite overall background rate may not serve as a suitable ‘expected rate’ for the observed COVID vaccinated population and be free from bias due to confounding. Given that the pmaxSPRT framework using historical rates is proposed be used by the FDA (as opposed to contemporaneous comparative cohort design using large-scale propensity score adjustment as previously suggested by the OHDSI community in its prior comments), it seems some adjustment will be required. Therefore, real-time standardization may need to be extended beyond age: at each time interval that the observed

rates are tabulated, also summarize the COVID vaccine exposed population demographics (proportion of persons by age group, sex, and race), and use these observed baseline characteristics to weight the baseline stratified rates to compute an 'age/sex/race-standardized expected rate'. Given the dynamic nature of the vaccination campaign, this will necessarily mean that the expected rate will change at each data collection period.

Recommendation: Calculate age-based incidence rates as a primary objective with the intent to age-adjust the incidence rates when the age distribution of the COVID-19 cohort becomes available, with the possibility to extend this to other characteristics.

ANCHORING EVENTS

The difference between incidence rates calculated from a time at risk that is anchored to a medical event can differ by over a factor of 10 from rates calculated from a time at risk set at an arbitrary time (like January 1 or the expiration of a one-year clean window), with the largest difference occurring for short times at risk, such as that for anaphylaxis. The current plan for the baseline rates to be calculated for long times at risk anchored at January 1 or the expiration of a clean window. These will be compared to incidence rates calculated from time of vaccination, which is a form of health care encounter. The question is--and the answer remains unclear--whether vaccination events are more like medical events or like arbitrary times. The truth may be a combination, depending on the context in which a person is vaccinated. For example, if an employee is called to be vaccinated, then there may be no correlation with medical illness, and a baseline rate based on arbitrary date may be best. If a person got a vaccination as part of another encounter or spurred by concern about a recent comorbidity, then a rate based on an anchoring medical event may make more sense. Because it is difficult to know which is best, we suggest estimating rates based on both to assess the sensitivity to the choice. Incidence rates in the influenza vaccination cohort can compared to these rates, but the mechanism for vaccination likely differs between influenza and COVID-19.

A related concern is the length of the time at risk. The rates of the COVID-19 vaccination cohort will be calculated by identifying the population exposed to a COVID vaccine, with the cohort entry index event of the vaccine dosing date, and applying an outcome-specific time-at-risk that varies from 0-2 days for anaphylaxis to 1-90 days for transverse myelitis. Yet the background rates rely on a constant time horizon that is up to 365 days.

Recommendation: Calculate incidence rates both based on an arbitrary time (as currently designed) and on a medical encounter to assess the sensitivity to anchoring. The arbitrary time may end up being the best, but the comparison will temper conclusions drawn from outcomes with a large difference between the two approaches.

ICD10-CM

Section 5.1 discusses that ICD10-CM based definitions have been developed for each AESI listed. We encourage the FDA to share these definitions publicly so they can be used by other researchers interested in replicating this work or performing additional supplemental research. As OHDSI provided in its prior response, we have developed phenotype algorithms for all of the FDA AESI except Multi-system inflammatory syndrome, and we are eager to collaborate with FDA to compare and further evaluate these algorithms. As previously noted, it may be possible that more refined phenotype algorithms will be required to improve sensitivity or specificity of the definitions. OHDSI has developed a set of tools that could be useful to assist with this evaluation process.

Recommendation: Publish the outcome definitions.

USE OF INCIDENCE RATES AND PROPORTIONS

The protocol details how incidence rates (from claims) and incidence proportions (from EHRs) will be computed, overall and stratified by baseline demographics. It is unclear how this characterization analysis will be directly applied in the context of the prior Master Protocol, which details how pmaxSPRT will be applied to compare observed rates in the COVID vaccine exposed population to expected background rates over time.

PREGNANCY

Pregnancy outcomes are listed in this protocol, but were not previously reported in the Master protocol, so it is unclear how they will be used in the maxSPRT analyses. We do not have preliminary data yet on pregnancy in this context, but we suggest it is worthy of further study.

We note that the pregnancy episode algorithm proposed was adapted from prior work by members within the OHDSI community by Matcho et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5794136/) and was applied to multiple databases in the OMOP CDM, so may provide a useful starting point to facilitate other data partners.

COVID-19 DISEASE

It is unclear how the incidence of COVID-19 diagnosis and hospitalizations will be used. Given that codes for diagnosis and lab tests are not provided, we offer the phenotype definition that OHDSI has previously developed to identify COVID-related hospitalizations https://atlas.ohdsi.org/#/cohortdefinition/198.

MAGNITUDE OF SYSTEMATIC ERROR

It is worth noting that confidence intervals based on the methods proposed only represent sampling variability. With large datasets such as those used here, random error converges to 0, yielding confidence intervals that are extremely narrow, which can be potentially misleading given the large potential for uncertainty due to systematic error due to design considerations (such as target cohort definition and time-at-risk). Systematic analysis of the variance due to these design choices can inform a more reasonable credible interval to consider. OHDSI's preliminary methods research into incidence characterization to support FDA suggests design choices can have an order-of-magnitude impact on incidence estimates. We therefore believe that reporting the variance among databases and design choices will be more informative than the classic statistical confidence intervals.

USE OF INCIDENCE PROPORTION

Incidence will be summarized as a rate in claims data, and as a proportion in EHR data. Incidence rate is defined as number of outcomes, divided by person-time at risk, and it allows for the potential that one person can contribute multiple events based on the cohort re-entry logic around a clean window. Section 10.1.1 suggests that incidence proportion will be defined as the 'number of incident cases among individuals', divided by the 'count of eligible individuals'. To clarify, incidence proportion should be defined as the 'number of persons with an outcome,

divided by the number of persons at risk', meaning that one person cannot contribute multiple events to the numerator.

EHR target population is defined as persons with at least one visit in the calendar year to be used as risk. This may induce a bias, as healthy patients not seeking healthcare will not be counted (analogous to those in claims with enrollment but without activity), but also may inflate estimates because requiring a future visit ensures that some activity takes place after the index date. It may also tend to increase the age of the population. It seems reasonable to consider calculating incidence rates and incidence proportions for both the claims and EHR data partners in a more consistent manner, using both target cohort definitions: 1) all persons with observation period, and 2) all persons with a visit in the calendar year. This would provide a more direct comparison between the sources. As it stands, using different statistics and different target populations assures that claims and EHR incidence estimates cannot be directly compared.

To carry out the observation period version in the EHR databases, most groups consider the observation period to extend from the first instance of data in the database to the last incidence of data in the database for that patient. While this can lead to an underestimate of the incidence rate if patients leave the health care facility and come back later, but are actually absent during the time at risk, it avoids the bias described above (loss of healthy persons).

We recognize the goal to include patients who visit urgent-care centers in one year without other visits, but we wonder if this inclusion adds more bias than it removes. Patients who go to an urgent-care center for an acute injury may go elsewhere for other outcomes like myocardial infarction.

The linked claims-EHR databases may be used to assess the accuracy of different approaches to setting observation periods for EHR databases.

Defining the EHR cohort by requiring a visit during the year can cause a look-ahead bias. The qualifying visit may happen to be the outcome of interest. Therefore, the incidence proportion numerator will include all patients at risk including those who had no other health care issues during the year. The denominator, however, will exclude patients at risk who did not have the outcome and who had no other health care issues, because they had to have at least one visit to get into the denominator. This might be fixed by ensuring that there is at least one visit other than the outcome being studied.

NEGATIVE CONTROL EVENTS

It is unclear how negative control events are intended to be used. Section 5.3 suggests there is a hypothesized effect of calendar time for some negative controls but not others. It is unclear how observing temporal instability in these outcomes will inform the dynamics of the AESI incidence.

We propose that a large set of negative control outcomes should also be added, which are outcomes that are explicitly not expected to be caused by COVID-19 vaccination. If a population-level effect estimation study were designed to produce a relative risk, then these negative controls could be useful for identifying systematic error in the data and methods, and also enable empirical calibration.

OTHER AESI

With proper correction for multiple hypotheses, it may be desirable to study a large number of other AESI. Many can be pulled from the OHDSI phenotype library. A larger number could be monitored by basing this secondary analysis on diagnoses codes rather than on fully defined phenotypes.

At the very least, a protocol should be formally defined for adding individual AESIs as adverse event reports come in. The content would include the criteria for adding an AESI, a description of how the phenotype will be defined, how the phenotype will be validated, how clean windows and times at risk will be determined, and what comparisons will be done. (We recognize that this protocol refers to the baseline rates for a surveillance activity and that the confirmation procedures may be in a separate protocol. Therefore, this may already be covered.)

PRE- AND POST-COVID COHORTS

We agree with the plan to separate the pre- and post-COVID cohorts, but we note that our initial results show only a small difference between the groups for the selected AESI.

ENTRY DATE

By picking January 1 as the base from which claims data are anchored with some patients starting after January 1 if the observation period starts after January 1 of the previous year, a seasonal bias will occur. Patients who start late will preferentially miss winter, and patients who are censored will preferentially miss fall, leading to an emphasis of the spring and summer. An alternative is to compare the analysis to starting July 1 instead of January 1. Furthermore, patients who have entry events earlier in the year (particularly January 1) contribute more time than those who enter later in the year. An alternative is to provide same 365-day time at risk for all entries, letting the time at risk extend beyond the end of the year, at least for the earlier years that would not overlap 2020.

CUMMULATIVE INCIDENCE RATES

We are not sure how the monthly cumulative incidence rates will be used. It may be to adjust for seasonal effects that will occur as new vaccine data are accumulated from January 2021 forward. We note, however, that vaccine data will not be collected uniformly but instead likely rise over time whereas the baseline rates will have data collected more or less uniformly. Therefore, there may still be mild seasonal effects.

GENERALIZABILITY

We believe that this work will be useful to the international community and that work currently going on in the international community will be useful here. For example, within OHDSI, there is current work in the UK and Spain on baseline incidence rates. Collaboration can facilitate generalizability of the results in terms of vaccine type, ethnicity, health care systems, etc.