

Use of Real World Data to Assess Cardiovascular Safety Synopsis of the 19 October 2016 CSRC/FDA Think Tank Meeting

A CSRC/FDA sponsored Think Tank was convened at the FDA White Oak Headquarters on 19 October 2016 to evaluate the potential use of real world data (RWD) to identify cardiovascular (CV) safety risks associated with drugs and to consider how such data might complement findings from randomized controlled trials (RCTs) and when it would provide sufficient and credible evidence to inform decision making across various stakeholders. To accomplish this task, the day was divided into the following sections that included brief expert presentations followed by discussion: 1) a review of existing data types; 2) assessing causality; 3) evaluating study designs and analytical methods including CV event ascertainment and drug exposure; and 4) methodologies to reduce bias and estimate effect size.

The FDA kicked off the day with a perspective on the use of RWD to assess CV safety, emphasizing the need to obtain valid, well defined and reliable data and the need to distinguish data collection from the specific use of the data (e.g., observational studies or controlled trials). There was a discussion of effectiveness studies using randomized trials conducted within the healthcare setting and acknowledged there are many possibilities. However, whether RWD could be used in a non-randomized setting (especially safety trials) is a subject where their use may well depend on the scenario. There are situations in which epidemiological “safety” studies may be indicated, for example, when there is insufficient time to conduct a new RCT and/or when the effect size might be relatively large (e.g., HR>2.0). Observational evidence regarding cardiac safety specific to testosterone agents, peripherally acting mu-opioid receptor antagonists, and dabigatran were presented as case examples to highlight some of the challenges with real world evidence. In particular, results from observational studies may be inconsistent thus limiting the confidence of the findings and /or observational study findings may be discordant with evidence from RCTs making interpretation challenging. These factors can influence the decisions rendered or actions stipulated by regulatory authorities or others.

There are many real world data types, including but not limited to prospective cohort studies, observational registries, electronic health records, and claims/payor databases. Each has strengths and limitations, especially with regards to CV event ascertainment. Given the diversity of healthcare systems in the US, data linkage across databases (e.g., electronic health records, insurance claims and pharmacy databases) might prove more yielding in the identification of CV events and other relevant data, but data linkage is not a trivial undertaking. Showing similar results from different data sources would be reassuring. The Medical Device Epidemiology Network (MDEpiNet) is one example of a national (US) system designed to integrate real world data and enable the conduct of such studies. In countries where a single payer system exists, data collection/ascertainment may be easier; however, the size of the database will depend on the population, and that may be a limiting factor. Ultimately, the data type(s) used in a real-world study should be “fit for purpose,” meaning it is appropriate for the intended use.

Generating reliable observational evidence using electronic healthcare data is critical to estimating a population-level effect. A variety of analytic methods exist, along with performance measures to objectively measure the reliability of observational evidence (e.g., consistency, discrimination and calibration). Recommended best practices for population-level effect estimation include: evidence generation (protocol, use of validated software, replicate across databases); evidence evaluation (produce standard diagnostics, include negative/positive controls, calibrate confidence interval); and evidence dissemination (provide more than the effect estimate; share protocol/evaluation). The “optimal method” is one that is transparent, fully reproducible, and empirically demonstrated across multiple databases to discriminate between true effects with unbiased and well-calibrated estimates.

Pragmatic trials are of increasing interest, especially as healthcare decision makers search for clinically effective and cost effective treatments for budgetary reasons. Pragmatic trials are designed to show real world effectiveness of an intervention in a broad patient group thus the study population is more heterogeneous than in a RCT. Consent and randomization (but not blinding) are feasible and outcomes are measured according to standard of practice. Designing and implementing these trials requires thoughtful, prospective planning as evidenced in the case example presented of the ADAPTABLE trial (a prospective randomized trial comparing the effectiveness of two daily doses of aspirin). This study is being conducted through PCORnet (developed by PCORI), and large amounts of health data are being leveraged (as a result of data linkage and extensive data management). The SALFORD Lung Study was also presented to exemplify the conduct of a pragmatic trial conducted with an investigational drug using pharmacies and electronic health records in a community based setting. This was the first ever phase 3 pragmatic trial.

Endpoint selection in electronic healthcare data is critical. The most reliable endpoints are those that are “serious and result in an immediate healthcare episode” such as MI, stroke, CHF, major bleed or revascularization. Mortality per se is highly reliable in claims but the specific cause is generally not captured. It is collected in the CMS database and the National Death Index (NDI); but, in the NDI, there is an approximate 18-month lag time for the information to be available. Procedure based endpoints are highly reliable due to the need for reimbursement. Hospital based endpoints such as MI and stroke are dependent upon the coding of the condition. Case identification algorithms may be helpful. These events can generally be validated with a high degree of certainty but “adjudication” (i.e., meeting prespecified criteria) may not be feasible because laboratory, ECG and imaging data are not usually captured (as compared with RCT). CV events, such as hospitalization for unstable angina, may be less reliable endpoints due to variations in coding. Validation of a sample of the endpoints is encouraged.

Defining exposure to a specific drug of interest is important. Pharmacy claims databases are useful but only capture medications paid for by insurance carriers and do not include over-the-counter (OTC) therapies. Start and stop dates for medications are seldom available. In addition, several situations may confound the estimation of exposure. For example, if a patient pays cash for a drug, a claim may not be submitted to the insurance provider, or a patient may

be provided “free samples” for a period of time before filling a prescription. Consequently, there may be a moderate amount of “missing data” in pharmacy claims databases.

Bias can originate from many sources. For example, bias may result from prognostic indicators that are not measured (e.g., smoking status) or be due to changes in access to care (e.g., formulary changes). Healthy individuals may be more adherent to a drug than less healthy persons thus introducing selection bias. It is critical to understand potential sources of bias that might influence study results at the design stage, institute measures to minimize these sources of bias, measure them, and address them in the analysis plan. For example, prospectively plan to conduct a wide range of sensitivity analyses varying the definitions for exposure, covariates and outcomes of interest. Use of quantitative bias analysis (an estimate of uncertainty arising from systematic errors) also was proposed.

Effect size estimates should be reliable and reproducible. Consider if the estimates generalize to different populations (databases) and if the confidence intervals reflect uncertainty about the effect size. Calibrate the study evidence under the null and alternative hypothesis; negative outcome controls help expose and control bias.

Causality cannot effectively be determined. A drug might be associated with a given outcome for any number of reasons (e.g., causation, chance, confounding/bias). Interpreting study findings can be challenging due to factors such as overconfidence (in which the degree of certainty is overestimated), personal preconceptions, cognitive dissonance, and conflicts of interest. To mitigate these issues, overfitting (i.e., overestimating our certainty) should be avoided, conflicts of interest should be stated, and knowledge about how the data systems were built and the data collected is essential (since these platforms were not built for systematic data collection as in a RCT).

In conclusion, it was agreed that there is a path forward to building an evidence base to show that real world studies can provide credible and reliable evidence regarding CV safety risks for drugs that could be used to supplement RCT findings and inform decision making, but we are not there yet. This topic and potential next steps will be further discussed in a forthcoming White Paper.