DIA

Integrative Analysis of Randomized Clinical Trial and Observational Study Data to Inform Post-marketing Safety Decision-Making

Li-An Lin[1] · Yafei Zhang[1] · Walter Straus[2] · William Wang[1]

Abstract

Safety evaluation is a continual and iterative process throughout the drug development life cycle and requires long time horizons and large amounts of data to fully understand the safety profile of a medical product. Although randomized clinical trials (RCT) provide high-quality data for an initial assessment of safety signals, the safety signals may not all have been known at the time of approval because safety data collected from RCT only involve a relatively small number of subjects during a relatively short follow-up period. The increased accumulation of post-marketing real-world data (RWD) presents an opportunity to utilize them for safety decision-making; these include identifying new safety signals, further characterization of safety concerns that are raised in pre-marketing RCT, and further generalization of RCT findings to the broader patient populations not previously studied in RCT. In this paper, we use cardiovascular safety outcome trial for antidiabetic therapies as an illustrative example and discuss how integrative analysis of RCT and observational study data can answer regulatory concerns about cardiovascular risk in a post-marketing setting. A novel statistical analysis strategy is proposed to combine both sources of safety data in a data fusion approach. The proposed approach includes three stages: (1) feasibility analysis that uses an RCT to validate an observational study, applying estimand framework and emulating RCT with RWD; (2) integrative analysis that combines evidence from the RCT and observational study data cooperatively; and (3) sensitivity analysis that examines the consistency of the previous analyses. Two potential utilities of the proposed integrative analysis for the cardiovascular safety outcome trial are discussed.

Keywords Real-world data · Observational study · Randomized clinical trial · Cardiovascular outcome trial · Type 2 diabetes mellitus · Causal inference

Background

Randomized clinical trials (RCTs) are often considered the gold standard for establishing treatment efficacy and safety of a new intervention and are critical for the purposes of establishing the evidence supporting product licensure. The risks associated with the new intervention may not be all known at the time of approval, because safety data are collected from studies that involve a relatively small number of subjects during a relatively short follow-up period of time. Furthermore, RCTs typically involve stringent study

designs, employing carefully selected populations, and have been criticized for not providing results that can be generalized to routine clinical practice [1]. In contrast, the use of real-world effectiveness and safety data collected outside of highly controlled traditional randomized trials has received increasing attention in the medical literature [2–4]. In response to requirements under the 21st Century Cures Act and the sixth Prescription Drug User Fee Act (PDUFA VI) [5, 6], the FDA is developing the process and guidance for using real-world data (RWD) and its resultant real-world evidence (RWE) to support the assessment of safety and effectiveness in regulatory submissions.

Real-world data, as defined by U.S. FDA RWE framework [7], are data related to patient health status and/or the delivery of health care collected in the setting of routine clinical care. Examples of RWD include data derived from electronic health records (EHRs), medical claims, and billing data (used for administrative purposes) and data from

✉ Li-An Lin
leeang.ling@gmail.com

1 Clinical Safety Statistics, Merck & Co., Inc, Kenilworth, NJ, USA

2 Clinical Safety, Moderna, Cambridge, MA, USA

product and disease registries [7]. These data sources often include a large number of patients and can be considered as representative of real-world patient populations. Based on research objectives and questions, the specific target patient population can be defined for whom to collect information and then construct the study sample from RWD [8]. The rich data sources in a real-world setting also promote external validity which is a commonly raised concern in RCTs due to the enrolled participants not fully representing those patients seen in routine clinical practice. Meanwhile, the increasing development of such large databases, as well as the maturation of information technology invites their use in health care and clinical decision-making. An acknowledged limitation of real-world databases is that treatment use is simply observation of physician/patient choice, not randomized. As a result, baseline covariates of treated subjects often differ systematically from those of untreated subjects. Randomization is routinely used in experimental/clinical research to reduce or eliminate the effect of baseline covariates that might otherwise occur due to treatment assignment and which can influence the treatment use and/or outcomes between treatment groups (also known as confounding factors).

Considering the respective strengths and limitations of RCT and RWD, there is a growing interest in utilizing data from both sources of evidence to inform regulatory decision-making. Safety evaluation is a continual and iterative process throughout the drug development life cycle and requires long time horizons and large amounts of data to fully understand the safety profile of a medical product. While clinical trials provide high-quality data for an initial assessment of the safety signals of a new drug, they alone cannot fully characterize the safety profile. Post-marketing RWD plays a critical role in further understanding the safety profile of a drug once it has been licensed and is being used in clinical practice. A well-designed observational study (either retrospective or prospective design), together with RCT, can be used to answer broader regulatory safety questions.

There is little methodological work assessing the statistical analysis strategy in conjunction with RCT and observational study design. And, significant challenges remain regarding potential study design and analytic methodologies to determine whether observational studies can reliably generate evidence on the safety and effectiveness of new interventions to inform regulatory decision-making. In this paper, we focus on the integrative analysis of RCTs and observational studies to inform post-marketing safety decision-making. We propose a three-stage statistical analysis strategy that utilizes data fusion techniques (originally developed for computer science) to integrate these two data sources in a cooperative manner, designed to enhance clinical trial results while mitigating several of the limitations associated with observational data. The remainder of this paper is organized

as follows. In Section 2, we introduce statistical issues and methods for synthesizing RCT and RWD evidence. Section 3 presents the Cardiovascular Outcome Trial (CVOT) as a motivating example. Section 4 describes the proposed three-stage statistical analysis strategy. Practical considerations and related issues are discussed in Section 5.

Introduction

Meta-analysis is widely used in medical research to derive a pooled effect estimate based upon a synthesis of findings of multiple studies that frequently employ diverse designs [9]. Assuming the included studies are random samples drawn from a hypothetical population of pooled studies, the meta-analysis result can be interpreted as an objective estimation of the mean treatment effect across this hypothetical population [10]. Due to the heterogeneity across studies, this assumption may not hold and random-effect models may be able to provide "an overall summary of what has been learned" and "a quantitative measure of how results differ, above and beyond sampling error" [11]. In meta-analysis of RCTs and observational studies, more complicated models have been proposed to accommodate additional heterogeneity between these two types of studies, including naïve pooling that does not differentiate between-trial designs, inclusion of observational study data as prior information, a power prior approach where information from the observational study is down-weighted to reflect confidence in the study findings, and a three-level hierarchical model where one level of the model accounts for differences in RCT and observational study designs in addition to study level and participant level [12–16]. A more detailed discussion of statistical modeling can be found from the review paper by Schmitz et al. [12] and Efthimiou et al. [15].

As described, meta-analysis provides a quantitative measure of treatment effect across RCTs and observational studies. However, its applicability in combining these two study types is limited since subjects in observational studies are typically from a different population as from RCTs and the treatment effect may not be the same across those populations [15, 17]. There are restrictive procedures in the selection of RCT participants which makes the RCT population not a random sampling of the general patient population [18]. In particular, the distribution of confounding factors that modify the treatment effect and treatment choice in RCTs often differs from subjects seen in routine clinical practice. Moreover, the clinical question and trial objective may differ from the observational study. The design and conduct of RCT make RCT participants more likely to be adherent and less likely to deviate from the prespecified protocol schedules. The observational study is constrained by the data observed in actual practice

patterns in real life. Directly combining these two types of studies together in a meta-analysis cannot produce estimates with a clear causal interpretation for any reasonable hypothetical population as meta-analyses combine study-specific effects based on the precision of the estimates rather than their relevance to the target population [17]. Moreover, the design of an RCT is different from that of an observational study, usually having shorter follow-up time, a smaller sample size, a bigger departure from routine clinical practice, and ethical restrictions. When an observational study data is included in a meta-analysis, it amplifies concerns about whether the efficacy and safety of the corresponding interventions can be translated into real-world effectiveness and safety [15]. In considering the challenge for defining the common clinical question across studies and large heterogeneity (including unexplained inconsistency) of treatment effect, meta-analysis may be inappropriate.

An alternative approach is presenting RCT and observational study results side by side where the observational study is considered to complement RCTs or address some of their limitations [17]. There is a stronger scientific justification for deriving evidence of a drug effect from RCTs as compared to observational studies [7]. The data produced from observational study studies can be useful complementary to those generated by RCTs and help to establish real-world effectiveness, harm, use, and value of treatments in a broad population of patients from routine care [19, 20]. In the literature, there are examples where RCT and observational study studies have reached similar conclusions about a treatment effect, and there are also examples when an effect identified in an observational study is discordant with the previously characterized effect (effect size, direction, or magnitude) [7]. Observational study results that are concordant with the RCT can provide regulators with greater confidence in new drug application review and approval process, whereas discordant results could warrant deeper reexamination of the RCT or observational study to identify reasons for this discordance and additional research may be necessary to further explore [3, 21]. This approach has been used in regulatory approvals. For example, an FDA advisory meeting was held in January 2019 to discuss the new drug application for Sotagliflozin as an adjunct to insulin therapy to improve glycemic control in adults with type 1 diabetes mellitus (T1DM). The risk for diabetic ketoacidosis (DKA), a life-threatening complication caused by a lack of insulin in the body, with the use of Sotagliflozin for treatment of T1DM has been raised as RCTs which showed an increased risk of DKA with Sotagliflozin compared with placebo [22]. The U.S. FDA presented both RCTs and FDA Sentinel data to support the review where analyses with Sentinel data showed that DKA rates observed in off-label use were higher than expected based on RCTs.

The complementary approach provides a qualitative way to review these two data sources together and draws the joint findings, but it does not provide an integrative quantitative assessment to accommodate both types of data sources. Furthermore, it is still questionable whether or not the observational study is sufficiently credible, interpretable, and ultimately acceptable for the regulatory purpose [16]. Due to the lack of randomization and the presence of confounding bias in an observational study, statistical methods (such as propensity score adjustment and instrumental variable analysis) have been developed to adjust for confounding bias and conduct causal inference. The commonly used methods include multivariable risk models, propensity score adjustment, and instrumental variable analysis [23, 24]. The propensity score for a subject, defined as the conditional probability of being treated given the subject's covariates, can be used to balance the covariates between treatment groups and thus reduce estimation bias. In the absence of random assignment (which is typical of RWD study designs), propensity scoring is a tool designed to mitigate inter-group differences in important known baseline covariates. It is important to note that propensity score methods work best in large datasets in which one can obtain a reasonable overlap of confounding factors between treatment groups. In addition, propensity score will not adjust for the impact of unknown or unmeasured confounders that are not included in the propensity score model [23, 25]. Instrumental variable is a third variable that is influenced by explanatory and outcome variables in the causal model. Instrumental variable methods are not based on the hard-to-verify assumption of "no unmeasured confounding," but require identification of a valid instrumental variable that needs to be correlated with treatment status and does not independently affect the outcomes of interest (except through a treatment effect). Often, proper instrumental variables are notoriously difficult to identify so that propensity score adjustment may be the only available practical choice. The causal estimation may also be compromised by measurement bias, i.e., when data items are measured with error. Assessment of data sources (including completeness, consistency, and trends over time) is needed to make sure the medical codes or combinations of codes, such as the World Health Organization's International Classification of Diseases Coding and Coding and ICD-9-CM Codes Used in Vaccine Safety Research, adequately capture the underlying medical concepts they are intended to represent [7, 26].

Addressing residual confounding and other biases in casual estimation is critical to improving the validity of observational studies and subsequently drawing causal conclusions with both RCTs and observational studies. Current gaps or perceived obstacles to increasing use of RWD include insufficient confidence in data sources and also integrating these with clinical data [27]. Although there is considerable growth in the development of advanced

causal inference methods in the past decades, it is still not guaranteed that causal inference methods for observational studies can address all of the potential bias and achieve the same level of evidence as RCTs. Increasing the credibility of observational studies can be at the design and/or analysis stage of the study. It is best to adjust for the potential for bias early, as statistical methods cannot replace a good research design. Since a well-designed RCT can provide high-quality evidence to establish causality, it can provide the foundation for framing observational studies. Several observational studies have successfully emulated the target trial using RWD and demonstrated that this approach can help avoid common methodological pitfalls and explain between-study differences [28–32]. By systematic comparisons of matched observational studies that emulate RCTs as closely as possible in terms of populations, exposure, and outcomes, it can provide confidence that observational study design and analysis plan are sufficiently valid to draw casual effect with RWD [33].

Motivating Example

Diabetes mellitus is a metabolic disorder characterized by the presence of hyperglycemia (elevated blood glucose) due to defective insulin secretion, insulin action, or both. Over time, poorly controlled serum glucose damages multiple organ systems, including elevating the risk for cardiovascular (CV) disease. Numerous antidiabetic drugs have been approved in Europe and the USA to control serum glucose. Although the correlation between poor glycemic control and cardiovascular risk is clear, it has proven difficult to demonstrate a causal relationship between improved diabetes control and reduced cardiovascular risk in type 2 diabetes mellitus (T2DM) [34]. Concerns about this gap in the assessment of CV safety grew after two highly controversial meta-analyses of CV risk were published [35, 36]. Given these concerns and the prevalence of CV disease in diabetic patients, both the U.S. FDA (in 2008) and the European Medicines Agency (EMA, in 2012) issued guidance for industry to address CV risk in for new antidiabetic therapies to treat T2DM [37, 38]. In particular, the U.S. FDA advised that pre-marketing phase 2 and phase 3 trials should rule out an 80% excess risk (upper bound of the rate ratio of the 2-sided 95% confidence interval less than 1.8) for major cardiovascular events (MACE). If the upper bound of the 95% confidence interval is between 1.3 and 1.8, a dedicated post-marketing cardiovascular outcome trial (CVOT) will be required to rule out a 30% excess risk after approval. The EMA guideline is similar to the FDA guidance but does not prospectively define any pre- or post-approval CV event risk margins.

After the publishing of the 2008 U.S. FDA guidance, every novel antidiabetic agent approved has undergone a dedicated CVOT, typically involving 5,000–15,000 people with type 2 diabetes and high-CV risk and planned to lasting 3–5 years [34]. And, none of the CVOTs completed and published to date have identified an increased risk of CV events; some of the CVOTs have instead demonstrated a reduced risk [39]. On October 24, 2018, the FDA held a 2-day Advisory Committee meeting to review the utility and impact of the 2008 guidance. The issues addressed by the FDA advisory board were as follows: (1) the impact of the recommendations in the 2008 guidance on the assessment of CV risk for drugs indicated to improve glycemic control in patients with T2DM; (2) the transferability of CV safety findings from members of a drug class that were studied to all drugs in the class (class effect), and (3) whether an unacceptable increase in CV risk needs to be excluded for all new drugs to improve glycemic control in patients with T2DM, regardless of the presence or absence of a signal for CV risk in the development program [40]. The panels made several recommendations for future regulatory guidance and CV outcome trials regarding antidiabetic therapies, include requiring only the 1.3 non-inferiority margin for regulatory approval, conducting trials for longer durations, considering head-to-head active comparator trials, increasing the diversity of patient populations, collecting safety data beyond cardiovascular events, and identifying ways to improve translation of trial results to general practice. Following the discussions at the Advisory Committee meeting, the U.S. FDA has revisited the recommendations of 2008 guidance and published new draft guidance in March 2020 [41]. The new draft guidance recommends sponsors to enroll a broader range of patients with comorbidities and diabetes-associated conditions, including patients with chronic kidney disease and older patients. It also removed the stringent criterion that requires sponsors to uniformly rule out a specific degree of risk margin for CV adverse outcomes (as recommended in previous 2008 guidance), but focuses more on the meaningful and reliable estimation of risk. However, the question still remains—how to rule out the excessive CV risk if a clinical trial study shows some degree of CV risk for the new antidiabetic treatment. In the new draft guidance, the importance of substantial safety information to support indication for glycemic control is still well recognized, and U.S. FDA continues to support CVOTs for patients with diabetes and will intend to work with sponsors to find efficient designs. Given the increasing availability of RWD supported by sophisticated statistical analytic tools along with regulatory interest in the use of such data to inform regulatory decision-making, observational study data could be an appealing alternative for a post-marketing CVOT.

Statistical Analysis Strategy

To address the acknowledged limitations of RCTs and RWD, we will explore the utility of data fusion techniques that were originally developed for computer science to integrate these two data sources in a cooperative manner. By definition, data fusion is the process of integrating multiple datasets collected under heterogeneous conditions (i.e., different populations, regimes, and sampling methods) to produce more consistent, accurate, and useful information than that provided by any individual dataset in isolation [42, 43]. It considers data integration as transporting heterogeneous datasets from one experimental condition to another via a causal diagram. Bareinboim and Pearl have derived a theoretical framework for causal analysis in combined experimental and observational datasets and extended it to clinical research settings [43]. In this manuscript, we present a novel statistical analysis strategy in applying data fusion for RCT and observational study evidence integration. Based on the relationship between the data sources, data fusion classifies the analysis methods into three categories: complementary fusion, where the data sources are not directly dependent on each other, represents non-overlapping information and can be combined to give a more complete image of the object under observation; competitive fusion (also called redundant fusion or meta-analysis), where data sources provide independent and overlaying information about the same target and can thus be fused to increment the confidence and robustness in a system; and cooperative fusion, where the data sources are used to create dependency with each other (e.g., one source of information is used to guide the search of new observations for another source) and eventually create new information that is more complex and not available from the original information. These three categories of data fusion methods are not mutually exclusive. Many

applications implement aspects of more than one of the three types. The first two methods have been discussed in Section 2. When RCT and observational study are conducted on a different population and under a different set of conditions, the cooperative fusion aims to leverage exploit design distinctions among the available studies and synthesize an aggregate measure of "targeted" effect size instead of "averaging out" differences. Using the cooperative fusion idea to integrate RCT and RWD evidence, we propose a three-stage statistical analysis strategy: (1) feasibility analysis that uses an existing RCT to validate and plan an observational study, applying estimand framework and emulating RCT with RWD; (2) integrative analysis that combines evidence from the RCT and observational study data cooperatively; and (3) sensitivity analysis that examines the consistency of the previous analyses. Figure 1 summarizes the working process of the proposed statistical analysis strategy.

Feasibility Analysis

Feasibility is necessary before fully considering an observational study for regulatory purposes. Through feasibility analysis, whether the data source is fit-for-use and whether the study design is sufficient to answer the intended research questions will be examined. To ensure the reliability and relevance of RWD submitted for use in regulatory decisions, FDA requires an assessment of whether the data capture relevant data on exposure, outcomes, and covariates [7]. Moreover, whether the observational study can provide credible casual estimation will need to be verified. Recently published EMA draft guideline on registry-based studies specified that analysis of potential information bias, selection bias, and potential confounding bias should be submitted either separately or as part of the proposed protocol [44]. An RCT can provide an excellent venue for the evaluation of
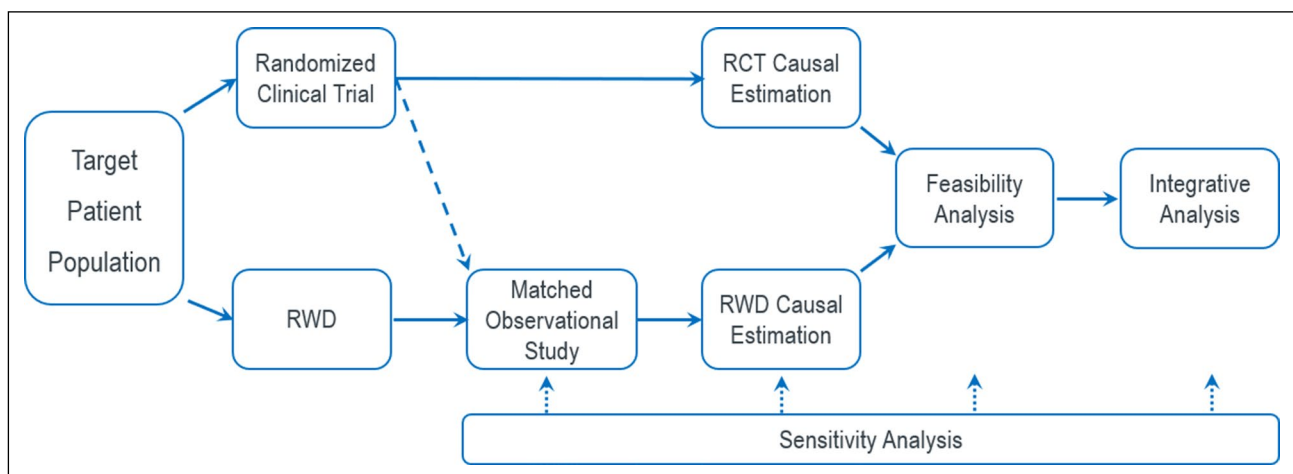

Fig. 1 Flow chart of proposed statistical analysis strategy

an observational study. Although there is significant existing research on the validity of observational studies, the most credible approach to evaluating their utility to inform regulatory decision-making is to compare the findings of such studies with the findings of RCTs addressing similar clinical questions, under the assumption that the RCTs represent a reference standard for the underlying true treatment effect [32]. As an example, a study assessed the correspondence of cardiovascular events that were ascertained by algorithms applied to Medicare data with the original Women's Health Initiative (WHI) trial that was conducted with the protocol-driven data collection and physician adjudication using linked records of trial participants to Medicare claims data [45]. Agreement between the two data sources was quite high, and the resultant hazard ratios and 95% confidence intervals were quite comparable. These encouraging results provided the needed support to launch a new embedded pragmatic trial that will rely heavily on Medicare claims to ascertain outcomes data. We propose the following three steps of feasibility analysis.

The first step of feasibility analysis is to emulate the target trial with RWD. Recent advances in comparative effectiveness research methods can help us design an observational study that mimics an RCT [46]. Hern*a´*n and Robins developed a framework for comparative effectiveness research using large observational databases to emulate the target trial [30]. Several observational studies have successfully emulated the target trial with RWD and demonstrated that this approach can help avoid common methodological pitfalls and explain between-study differences [28, 29]. In the process of designing an observational study to emulate an existing RCT, the clinical question of the study can be framed and the study design can be adjusted to match with the RCT as much as possible. The estimand framework, described in ICH E9(R1) Guideline, can be applied to frame the estimand of the observational study [47, 48]. For summarizing data with RCT, the estimand of the observational study should be aligned with the RCT estimand which will form the basis to establish feasibility analysis. The five estimand attributes

should be carefully examined by a multi-disciplinary team, including (1). treatment condition of interest; (2). population of patients targeted by the clinical question; (3). variable (or endpoint) to be obtained for each patient that is required to address the clinical question; (4). intercurrent events; and (5). population-level summary [47]. Similar to the five estimand attributes, data fusion considered four dimensions in transporting findings across studies with heterogeneous conditions (population, observational/experimental, sampling, measure) [43].

Completely replicate RCT is not possible. One important reason is that the RCT population is not a random sampling from the target population. The different distributions of patient baseline characteristics may vary the casual estimation between studies. Table 1 provides an extreme hypothetical example of such differences. A direct comparison of treatment effect from RCT and observational study that matched with RCT design, in the upper portion of Table 1, suggests that a lower treatment effect estimation from the observational study than the RCT. However, when stratifying the comparison by baseline disease severity, as in the lower portion of Table 1, the response rate from both treatments (A and B) is revealed to be the same in the RCT and observational study. The observed difference drawn from the naïve comparison between the RCT and observational study is an example that the heterogeneity in patient population between studies can be difficult to account for. Thus, the second step is to match the distribution of patient population between the observational study and RCT data. If the balance of the patient population cannot be achieved after adjusting the observational study design, the assessment of agreement between RCT data and RWD is not feasible. Similarly, if balance is achieved only in a very small subset of the RWD population, the analysis is again infeasible in the given RWD source. The third step of feasibility analysis is comparing the casual estimation of the RCT and matched observational study. If assignment to the treatment and outcome by treatment in the observational study were unconfounded, the treatment effects in the two studies should be

Table 1 A hypothetical example of the observed difference in treatment effect estimation due to differences in patient baseline characteristics

Outcome	RCT		Observational study		Between-study difference[a]
	Treatment A	Treatment B	Treatment A	Treatment B	
Aggregated overall response rates[b] (%)					
All patients	60	25	40	15	10
Response rates (%) by baseline severity					
Severe	30	10	30	10	0
Non-severe	70	30	70	30	0

[a]Observed difference in treatment effect estimation between RCT and observational study

[b]Assuming 75% severe (25% non-severe) in RCT and 25% severe (75% non-severe) in observational study

expected to be similar. The differences in the estimated causal effects between the two studies can be considered as evidence of unobserved confounders.

By comparing the findings from RCT and matched observational study, whether the data, estimand, study design, and analysis plan are sufficiently valid to detect known causal associations in the RCT can be identified. If no substantial inconsistencies are found in the matched observational study, one can proceed with the integrative analysis of RCT and matched observational study. Disagreements might occur due to residual confounding, important differences in routine clinical practice and treatment in RCT, or systematic differences in matching observational study with RCT. If inconsistent causal effects are seen, further examination of their magnitude and reasons for such differences are warranted to gain a broader understanding of RWD and its corresponding on whether the observational study can be used to support regulatory safety decision-making.

Integrative Analysis

In this section, we will describe two applications in combining RCT data and RWD information together once the observational study can be validated via feasibility analysis. If the matched observational study can be established to match with RCT estimand, the meta-analysis can be applied to combine RCT data and matched observational study data. Moreover, it is more reliable to further expand the design of the validated observational study design to a broader patient population that was not previously studied in RCT. Other applications are also possible but will not be discussed in this paper.

The first application is meta-analysis of RCT and the matched observational study. As we discussed in Section 2, in the absence of a common target population, assessing causality when directly combining RCT and observational study findings in a meta-analysis approach may not be feasible. By matching the observational study with the RCT, the clinical question of the study can be framed to a similar question as in the RCT. Meanwhile, the study design and patient population are similar across the matched observational study and RCT. This meta-analysis approach can provide a quantitative measure of the pooled causal effect across the RCT and observational study and focus on the RCT population. As discussed in 3 with respect to the CVOT example, sponsors need to compare the incidence of major CV events occurring with the investigational agent to those in the control group to show that the upper bound of the 2-sided 95% for the estimated risk ratio is less than 1.8 for the pre-marketing submission. However, the number of CV events accrued during a typical phase 2 or 3 development program is usually insufficient to provide high statistical confidence. Thus, sponsors typically proposed to conduct

a CV meta-analysis that includes completed Phase 2 and Phase 3 studies [49]. Although the 2020 U.S. FDA draft guidance removed the uniform degree of risk for CV events, it is still not clear how to exclude an unacceptable increase in CV or other identifiable risks associated with the new agent if the data are lacks of precision. To answer this question, the meta-analysis of an RCT and matched observational study could be performed to increase the precision of estimation and exclude a much lower degree of CV risk (such as the 1.3 criteria for the post-marketing setting that is defined in 2008 U.S. FDA guidance).

As described, the meta-analysis approach is only applicable when the analysis population can be properly defined. When performing integrative analysis of an RCT and the matched observational study under the meta-analysis framework, we can only provide answers to the RCT population which may not be generalizable for RWD population. The second application is to generalize RCT findings. As illustrated in Fig. 1, the RCT population can substantially differ from the RWD population. In particular, RCT findings generally only include patients who are adherent to therapy, monitoring, and follow-up and that is not part of real-world clinical practice. For example, current CVOTs typically have studied in patients with advanced CV risk or already established CV disease to ensure accrual of sufficient events in a timely manner and have sufficient statistical power in detecting CV risk differences [50, 51]. They cannot be entirely representative of the general population and therefore there is less certainty that the new treatment will improve CV outcomes for patients with a shorter duration of diabetes or without established CV complications. Lower-risk populations could determine whether diabetes medications offer CV protection for those who do not yet have CV disease [50]. Leveraging the finding established by feasibility analysis, the design of the matched observational study can be modified to reflect patients that are not well represented in the clinical trial. Since the study design and its causal relationship has been examined by the corresponding RCT, the extended observational study (with, for example, broader inclusion/exclusion criteria, longer follow-up times) could yield more credible results. In this sense, the RCT is used to validate RWD data source and observational study design, which may not necessary to combine its data with RWD.

Sensitivity Analysis

Despite attempts to emulate RCT design as closely as possible, differences between the RCT and corresponding RWE study populations remained. When emulating an RCT with RWD, a number of compromises will have to be made, such as eligibility criteria, endpoints definition, treatment strategies to be compared, and follow-up time [30]. Given the constraints of the RWD, adjusting the design and analysis

strategy in matching the target RCT is needed. Then, we will choose the one that is closest to the target trial that can best answer the clinical question of interest. The difference between the matched observational study and the corresponding RCT should be documented and explored by evaluating alternative approaches to controlling the remaining confounding and its resultant bias. The sensitivity analysis can assess whether possible variations would lead to different results by altering some aspects of the study design and analysis. This is aligned with ICH E9 (R1) addendum for clinical trials and EMA draft guideline on registry-based studies [44, 47]. For example, death is recorded in Medicare claims but the cause of death is not [32]. The primary endpoint of CVOT typically is a composite of CV death, non-fatal MI, and non-fatal stroke. In addition, some CVOTs added hospitalization for unstable angina and heart failure as primary endpoints. It is expected that a significant proportion of all-cause deaths may be CV deaths in diabetic patients with high-CV risk [32]. To match the observational study with the existing CVOT, it is possible to include all-cause death as the primary endpoint. However, the assumptions about the cause of death should be further explored via sensitivity analysis with all-cause death excluded. The cause of death may able to be asserted via linked data from HER or National Death Index (NDI). In this case, the robustness of linked data can also be examined by sensitivity analysis. By extensive sensitivity analyses, the robustness of findings will be explored to identify the range of plausible effect estimates under varying assumptions about the untested assumption.

Discussion

In this paper, we discussed the utility of RWD for post-marketing safety decision-making, in conjunction with RCT data. The observational study can provide complementary evidence for the RCT in several dimensions, such as routine clinical practice, a broader patient population, and longer-term follow-up. If the questions that both studies are trying to answer are clinically identical or similar, meta-analysis methods can be used to synthesis study findings. However, there are still considerable obstacles in increasing the use of RWD, such as insufficient confidence for validating RWD, causal interpretation of observational studies, and also integrating these with RCTs. We proposed a three-stage statistical analysis strategy in this paper: feasibility analysis that uses an RCT to validate an observational study, integrative analysis that combines evidence from an RCT and an observational study cooperatively, and sensitivity analysis that examines the consistency of previous analyses. The proposed approach allows us to validate the causal effects in an observational study by matching the study with an RCT in all dimensions, such as distribution of patient population

and treatment strategy, outcome. The forthcoming integrative analysis can only be feasible and interpretable if the observational study can be consistent with the RCT in causal estimation for a target population. Finally, the sensitivity analyses provide an additional examination of the robustness of the findings by exploring assumptions in the study design and analysis. A potential application of the proposed approach for the post-marketing CVOTs to alleviate the CV risk for antidiabetic therapies has been discussed.

Although there are some recent attempts examples where RWD was successfully used to emulate RCT, not every RCT can be closely emulated by RWD. For example, the treatment regimens in RCTs are often highly structured and may be difficult to be replicated in normal practice settings. It also has been argued that RCTs do not accurately reflect real-world circumstances under which patients are treated which limit the applicability of RCT results to real life [52, 53]. In order to yield meaningful data, the RCT must adapt and evolve so that the clinical knowledge generated from an RCT can increase the clinical relevance and overall usefulness of trial results. Such differences in study settings contribute to different treatment effect estimates from an RCT and the matched observational study. Furthermore, the results of a matched observational study are not always consistent with the RCT even when RWD can successfully emulate RCT and provides valid casual estimation. There have been several instances in which the efficacy and/or safety of marketed drugs differs substantially from that assessed in RCTs. Such phenomena can be considered as a problem of variability in drug response (efficacy or toxicity), either due to biological (e.g., previously unidentified patients' genetic biomarkers and other intrinsic/extrinsic factors) or behavioral (e.g., inappropriate prescribing and drug handling, poor patient adherence) sources [54]. Overall, it is expected that the variability in a real-world setting is higher than that in clinical trial conditions which may reduce the likelihood to detect a true effect for the matched observational study. For these reasons, assessing concordance between RCTs and observational studies requires the application of quantitative methods which are interpreted and informed by medical judgment. Several statistical methods may be applied to assess agreement between RCT and observational study results, such as estimation agreement, standardized difference, and difference-in-difference [32, 55]. As these methods are applied to assess concordance between RCT and observational study results, there is the opportunity to further refine them within the setting of medical judgment. Nevertheless, by comparing an RCT with a matched observational study, it provides insight into whether and to what extent differences in treatment effect estimates are due to bias related to the study setting, patient population, treatment strategy, clinical practice, randomization, and/or other differences between RCT and observational study. The

additional sensitivity analysis provides even more exploratory insights and generates hypotheses for future studies.

Previous studies, comparing the results of RCTs and observational studies, suggested that the lack of concordance can be attributed to differences in the populations being investigated or bias in the observational studies as a result of lack of randomization [56]. The proposed approach may rule out the possible differences in study populations and have a closer assessment of bias due to imbalanced unmeasured confounding between treatment arms in the observational study. If the feasibility analysis indicated that the matched observational study cannot provide consistent estimates to the RCT and no clinical rationale can explain such discordance, a randomized study will need to be conducted instead. This type of randomized study, such as a pragmatic randomized controlled trial (pRCT) or registry-based randomized controlled trial (R-RCT), that combined the advantages of a prospective RCT design with real-world natural in real-world features, could be seen as expanded use of RWD that meets regulatory criteria [27]. Nevertheless, the proposed approach could provide feasibility assessment for conducting such real-world trial and also inform trial design for the study. Such a staggered approach in developing an appropriate study type possesses flexibility for drug developers and regulators in evaluating which study is most appropriate to answer post-marketing safety questions, while still maintaining the highest standard of evidence to support regulatory decision-making.

References

1. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet. 2005;365(9453):82–93.
2. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, et al. Real-world evidence—what is it and what can it tell us. N Engl J Med. 2016;375(23):2293–7.
3. Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? Clin Pharmacol Ther. 2017;102(6):924–33.
4. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clin Pharmacol Ther. 2019;105(4):867–77.
5. Bonamici S, H.R.34 - 114th congress (2015–2016): 21st Century Cures Act. https://www.congress.gov/bill/114th-congress/house-bill/34. 2016.
6. U.S. Food and Drug Administration, Prescription Drug User Fee Act(PDUFA) – PDUFA VI: Fiscal Years 2018–2022. https://www.fda.gov/industry/prescription-drug-user-fee-amendments/completed-pdufa-vi-deliverables.
7. U.S. Food and Drug Administration, Framework for FDA's Real-World Evidence Program. 2018.
8. Girman CJ, Ritchey ME, Zhou W, Dreyer NA. Considerations in characterizing real-world data relevance and quality for regulatory purposes: a commentary. Pharmacoepidemiol Drug Saf. 2019;28(4):439.
9. Glass GV. Integrating findings: the meta-analysis of research. Rev Res Educ. 1977;5(1):351–79.
10. Manski CF. Toward credible patient-centered meta-analysis. Epidemiology. 2020;31(3):345–52.
11. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials. 2015;45:139–45.
12. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Stat Med. 2013;32(17):2935–49.
13. Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. Res Synthesis Methods. 2015;6(1):45–62.
14. Cameron C, Fireman B, Hutton B, Clifford T, Coyle D, Wells G, Dormuth CR, Platt R, Toh S. Network meta-analysis incorporating randomized controlled trials and non-randomized comparative cohort studies for assessing the safety and effectiveness of medical treatments: challenges and opportu- nities. Syst Rev. 2015;4(1):1–8.
15. Efthimiou O, Mavridis D, Debray TP, Samara M, Belger M, Siontis GC, Leucht S, Salanti G, G. W. P. 4, Combining randomized and non-randomized evidence in network meta-analysis. Stat Med. 2017;36(8):1210–26.
16. Jenkins D, Bujkiewicz S, Martina R, Dequen P, Abrams KR. Methods for the inclusion of realworld evidence in network meta-analysis. arXiv:1805.06839.
17. Dahabreh IJ, Petito LC, Robertson SE, Hern´an MA, Steingrimsson JA. Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. Epidemiology. 2020;31(3):334–44.
18. Dahabreh IJ, Hern´an MA. Extending inferences from a randomized trial to a target population. Eur J Epidemiol. 2019;34(8):719–22.
19. Schneeweiss S, Eichler H-G, Garcia-Altes A, Chinn C, Eggimann A-V, Garner S, Goettsch W, Lim R, Löbker W, Martin D, et al. Real world data in adaptive biomedical innovation: a framework for generating evidence fit for decision-making. Clin Pharmacol Therap. 2016;100(6):633–46.
20. Kosiborod M, Cavender MA, Fu AZ, Wilding JP, Khunti K, Holl RW, Norhammar A, Birkeland KI, Jørgensen ME, Thuresson M, et al. Lower risk of heart failure and death in patients initiated on sodium-glucose cotransporter-2 inhibitors versus other glucose-lowering drugs: the CVD- REAL study (comparative effectiveness of cardiovascular outcomes in new users of sodium-glucose cotransporter-2 inhibitors). Circulation. 2017;136(3):249–59.
21. Fralick M, Kesselheim AS, Avorn J, Schneeweiss S. Use of health care databases to support supple- mental indications of approved medications. JAMA Intern Med. 2018;178(1):55–63.
22. Hampp C, Swain RS, Horgan C, Dee E, Qiang Y, Dutcher SK, Petrone A, Tilney RC, Maro JC, Panozzo CA. Use of sodium–glucose cotransporter 2 inhibitors in patients with type 1 diabetes and rates of diabetic ketoacidosis. Diabetes Care. 2020;43(1):90–7.
23. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265–81.
24. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91(434):444–55.
25. Zhang X, Stamey JD, Mathur MB. Assessing the impact of unmeasured confounders for credible and reliable real-world evidence. Pharmacoepidemiol Drug Saf. 2020;29(10):1219–27.
26. Mullooly J, Donahue J, DeStefano F, Baggs J, Eriksen E, V. D. Q. W. Group, et al. Predictive value of icd-9-cm codes used in vaccine safety research. Methods Inform Med 47(04), 328–335 (2008).

27. Lamberti MJ, Kubick W, Awatin J, McCormick J, Carroll J, Getz K. The use of real-world evidence and data in clinical research and postapproval safety studies. Therap Innov Regul Sci. 2018;52(6):778–83.

28. Caniglia EC, Robins JM, Cain LE, Sabin C, Logan R, Abgrall S, Mugavero MJ, Hern´andez- D´ıaz S, Meyer L, Seng R, et al. Emulating a trial of joint dynamic strategies: an application to moni- toring and treatment of HIV-positive individuals. Stat Med. 2019;38(13):2428–46.

29. Dickerman BA, Garc´ıa-Alb´eniz X, Logan RW, Denaxas S, Hern´an MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25(10):1601–06.

30. Hern´an MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758–64.

31. Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, Babiker A, Law M, Chu H, Byrne D, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. Am J Epidemiol. 2019;188(8):1569–77.

32. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, Schneeweiss S. Non- randomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. Clin Pharmacol Ther. 2020;107(4):817–26.

33. Kim SC, Schneeweiss S. When randomized clinical trials and real-world evidence say the same: tocilizumab and its cardiovascular safety. Arthrit Rheumatol. 2020;72(1):4–6.

34. Regier EE, Venkat MV, Close KL. More than 7 years of hindsight: revisiting the FDA's 2008 guidance on cardiovascular outcomes trials for type 2 diabetes medications. Clin Diabetes. 2016;34(4):173–80.

35. Nissen SE, Wolski K, Topol EJ. Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. JAMA. 2005;294(20):2581–6.

36. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med. 2007;356(24):2457–71.

37. FDA Guidance for Industry, Diabetes mellitus—evaluating cardiovascular risk in new antidiabetic ther- apies to treat type 2 diabetes, withdrawed. 2008.

38. European Medicines Agency, Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129256.pdf . 2012

39. Singh M, Sharma R, Kumar A. Safety of SGLT2 inhibitors in patients with diabetes mellitus. Curr Drug Saf. 2019;14(2):87–93.

40. FDA Background Document, Endocrinologic and metabolic drugs advisory committee meeting. https://www.fda.gov/media/121272/download. 2018.

41. FDA Guidance for Industry, Type 2 diabetes mellitus: evaluating the safety of new drugs for improving glycemic control. https://www.fda.gov/media/135936/download. 2020.

42. Durrant-Whyte HF. Sensor models and multisensor integration. In: Autonomous robot vehicles. Springer, New York. 1990;73–89.

43. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. Proc Natl Acad Sci. 2016;113(27):7345–52.

44. European Medicines Agency, Guideline on registry-based studies (draft). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf. 2020.

45. Anderson GL, Burns CJ, Larsen J, Shaw PA. Use of administrative data to increase the practical- ity of clinical trials: Insights from the women's health initiative. Clin Trials. 2016;13(5):519–26.

46. Pencina MJ, Rockhold FW, D'Agostino RB. Deriving real-world insights from real-world data: biostatistics to the rescue. Ann Intern Med. 2018;169(6):401–2.

47. ICH Harmonised Guideline E9 (R1). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. 2019.

48. Fang Y, Wang H, He W. A statistical roadmap for journey from real-world data to real-world evidence. Therap Innov Regul Sci. 2019;1–9.

49. Marchenko O, Jiang Q, Chakravarty A, Ke C, Ma H, Maca J, Russek-Cohen E, Sanchez-Kam M, Zink RC, Chuang-Stein C. Evaluation and review of strategies to assess cardiovascular risk in clinical trials in patients with type 2 diabetes mellitus. Stat Biopharm Res. 2015;7(4):253–66.

50. Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, Green JB, Buse JB, Inzucchi SE, Leiter LA, et al. Cardiovascular outcomes trials in type 2 diabetes: where do we go from here? reflections from a diabetes care editors' expert forum. Diabetes Care. 2018;41(1):14–31.

51. Birkeland KI, Bodegard J, Norhammar A, Kuiper JG, Georgiado E, Beekman-Hendriks WL, Thuresson M, Pignot M, Herings RM, Kooy A. How representative of a general type 2 diabetes pop- ulation are patients included in cardiovascular outcome trials with SGLT2 inhibitors? a large European observational study. Diabetes Obes Metab. 2019;21(4):968–74.

52. Cohen AT, Goto S, Schreiber K, Torp-Pedersen C. Why do we need observational studies of everyday patients in the real-life setting? Eur Heart J Suppl. 2015;17(suppl D):D2–8.

53. Derksen JW, May AM, Koopman M. The era of alternative designs to connect randomized clinical trials and real-world data. Nat Rev Clin Oncol. 2019;16(9):589–589.

54. Eichler H-G, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, Rowland M, Schneider CK, Bloechl-Daum B. Bridging the efficacy–effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discovery. 2011;10(7):495–506.

55. Abadie A. Semiparametric difference-in-differences estimators. Rev Econ Stud. 2005;72(1):1–19.

56. Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? BMC Med. 2020;18(1):1–2.