In this issue of Value in Health, Thompson reviews the merits and potential pitfalls of efforts to replicate randomized controlled trials (RCTs) using real-world data (RWD). We agree with much of the content in the article but disagree with the author’s primary conclusion that there is only downside to the emulation activities because failure to emulate will undermine confidence in real-world-data (RWD) studies. Indeed, the primary objectives of emulation studies are to identify the sources of variability in RWD, understand its limitations and advantages, and thereby appreciate its appropriate utility and value.

There are many reasons why RCTs might generate different treatment effect estimates than observational studies. In fact, it is somewhat surprising that observational studies would ever be able to account for sufficient real-world confounders to emulate the results of a RCT. But this is an empirical question, and it is central to a number of RCT emulations currently underway. Randomization controls for bias from both observed and unobserved confounders. With the exception of instrumental variables techniques, most statistical methods used in observational studies control only for observed confounders. Thus, the efficacy–effectiveness gap largely amounts to whether the important confounders are known and can be measured in available databases; this is likely to vary by data source, disease state, eligibility criteria, outcome measures, and others. Bartlett et al found that in a cross-sectional review of 220 clinical trials published in high-impact journals in 2017, only 15% could potentially be emulated using data available from administrative claims data versus electronic medical records, versus the combination of claims and electronic medical record data; are both or neither enhanced by linkage to other data sources such as pharmacy data? How is it influenced by statistical methods used to estimate the treatment effects with the observational data? How is it influenced by the decision making of researchers about how they identify the inclusion or exclusion criteria of the trial, measure the endpoints, select the statistical methods, and choose the research design for comparing the treatment groups? With these questions in mind, it is important to emphasize that the various emulation approaches—whether statistical or instrumental variables techniques—must be understood.

Despite many studies that have found similar average treatment effect estimates from RCTs and RWD, other studies have documented wide variation in results from RWD studies within the same therapeutic areas, including studies using propensity score-based methods. Before RWE can be reliably used for regulatory decision making as required by the 21st Century Cures Act, the sources of variability in RWD studies need to be better understood. Those sources of variability include not only those included in the data themselves (eg, the potential for non-representativeness and bias within the dataset) but also in the approaches—and assumptions—of the data scientists who choose which data to interrogate and for what questions.

One promising approach to understanding these sources of variability is to compare estimates obtained from RWD studies that attempt to emulate the eligibility criteria, endpoints, and other features of trials as closely as possible. A small number of RWD studies have attempted to emulate the findings of RCTs, and, similarly, the results of several RWD studies have been confirmed by RCTs. In some instances, disagreements between observational studies and RCTs were attributable to avoidable errors in RWD study design. This has led to a focus on the importance of research design in observational RWD studies attempting to draw causal inferences regarding treatment effects. However, there is a need to understand not only the conditions under which it is possible to obtain similar estimates between RCTs and observational studies but also why the results may differ when there is disagreement.

The objectives of RCT emulation efforts such as Randomized Controlled Trials Duplicated using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology (RCT DUPLICATE) and Observational Patient Evidence for Regulatory Approval and Understanding Disease (OPERAND) are to better understand the sources of variation that influence the ability to close the efficacy–effectiveness gap. Are the measured confounders in certain disease states more complete than in others? How is variation influenced by the type of data (eg, claims data versus electronic medical records, versus the combination of claims and electronic medical record data; are both or neither enhanced by linkage to other data sources such as pharmacy data)? How is it influenced by statistical methods used to estimate the treatment effects with the observational data? How is it influenced by the decision making of researchers about how they identify the inclusion or exclusion criteria of the trial, measure the endpoints, select the statistical methods, and choose the research design for comparing the treatment groups? With these questions in mind, it is important to emphasize that the various emulation efforts are not primarily concerned with demonstrating agreement between RCTs and observational studies—or to undermine the usefulness of RWD or real world evidence (RWE). Rather, they seek to understand the sources of variability that may cause an observational study to generate a treatment effect estimate that differs from that of a RCT designed to answer the same question for the same patient population.

As Thompson notes, RCT DUPLICATE is a large-scale, FDA-funded study that is attempting to replicate the average treatment effects of approximately 30 RCTs using administrative medical claims data. For trials where it is possible to mimic the trial design with claims data, this work will provide information on variation in the ability to replicate average treatment effects across different disease categories. OPERAND, on the other hand, is providing complementary insights on sources of variation. The study design differs: OPERAND involves the independent analysis by 2 different research teams of 2 trials using the same initial large dataset. Differences in analytic decisions between the teams are considered, as are prespecification and other methodological approaches. Further,
OPERAND is evaluating treatment effects estimates in the broader RWD population using linked de-identified administrative claims and EHR data. Using EHR data for trial emulation enables the expansion to trials with clinical endpoints (eg, HbA1c, pain scores, lipid levels) typically captured in EHR data. Finally, OPERAND is evaluating how loosening the inclusion/exclusion criteria from those specified in the RCT and replicated by the observational study affects the treatment effect estimates. How different are the treatment effect estimates in the broader real-world population actually treated in clinical practice versus the narrower group studied in the trials? It is through rigorous and empirical analysis that the utility and value of RWE is revealed.

Although RCT emulation efforts such as OPERAND and RCT DUPLICATE are highly structured in their attempts to mimic the designs of trials using RWD, they can be criticized because the trial results are known when the RWD analyses are being planned and conducted. This gives the researchers conducting the emulation analysis a target to which they can “fit the analysis” so as to obtain the published result. But this criticism would not impugn confidence in RWD; rather, the a posteriori concurrence in RWD, may then be more usefully relied upon for subsequent analysis a target to which they can "fit the analysis". This gives the researchers conducting the emulation analyses (eg, of different subpopulations). The concurrent validation of the RWE, revealing whether and how well the RWD, analytic choices, and outcome measures correlate with the RCT, may then be more usefully relied upon for subsequent analyses (eg, of different subpopulations).

It is true that the FDA has a strong preference for evidence from RCTs in making regulatory decisions, but the agency is under a legislative mandate to provide guidance on the use of RWE for regulatory decision making. RWE derived from the analysis of administrative claims or EHR data has been used to obtain safety data on approved medical products under the FDA’s Sentinel Initiative. The 21st Century Cures Act would extend the application of RWE to new indications for previously approved drugs. The FDA’s framework nicely outlines its RWE program. It is clear that the FDA considers RWE to include evidence based on RWD in prospective RCTs conducted in real-world treatment settings, but that does not address the usefulness of analysis of extant RWD routinely collected for other purposes. The results of carefully controlled observational studies using RWD to emulate RCTs will yield empirical data to understand the variability, limitations, and utility of RWE. We trust that subjecting RWE to the scientific rigor of defined analysis will reinforce and validate, not undermine, the trustworthiness of its appropriate use in regulatory decision making.

**Article and Author Information**

**Accepted for Publication:** November 16, 2020

**Published Online:** December 18, 2020

doi: https://doi.org/10.1016/j.jval.2020.11.003

**Author Affiliations:** Florence Heller Graduate School, Brandeis University, Waltham, MA (Crown); Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA (Bierer).

**Correspondence:** Barbara E. Bierer, MD, Division of Global Health Equity, Department of Medicine, Brigham and Women’s Hospital, 75 Francis St, Boston, MA 02115. Email: bbierer@bwh.harvard.edu

**Author Contributions:** Concept and design: Crown, Bierer
Drafting of the manuscript: Crown, Bierer
Critical revision of the paper for important intellectual content: Crown, Bierer
Obtaining funding: Bierer
Administrative, technical, or logistic support: Bierer

**Conflict of Interest Disclosures:** Drs Crown and Bierer reported receiving grants from Amgen, AstraZeneca, Merck, Optum, Pfizer, Sanofi, and UCB BioSciences outside the submitted work.

**Funding/Support:** The authors received no financial support for this research.

**REFERENCES**