What Have We Learned from Clinical Trial Emulations?

William Crown, PhD

Distinguished Research Scientist

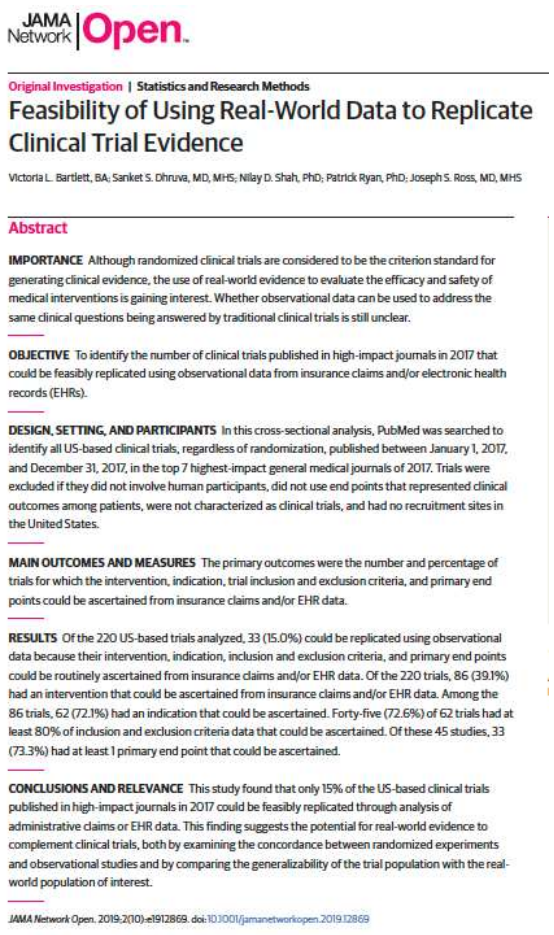
Brandeis University

MAY 18, 2022

Why Try to Emulate Trials at All?

- Goal is not to show that it is possible to get the same answer
- Emulation efforts help us understand when it is possible to replicate findings from the target trial and when it is not. We learn more from the failures than the successes!
- However, there are spillover effects when emulation is successful—loosening the inclusion/exclusion criteria of the index trial to examine treatment effect estimates in the broader population treated in actual clinical practice.

Most Trials Can’t Be Emulated with RWD



JAMA Network | Open.

Original Investigation | Statistics and Research Methods

Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence

Victoria L. Bartlett, BA; Sanket S. Dhruva, MD, MHS; Nilay D. Shah, PhD; Patrick Ryan, PhD; Joseph S. Ross, MD, MHS

Abstract

IMPORTANCE Although randomized clinical trials are considered to be the criterion standard for generating clinical evidence, the use of real-world evidence to evaluate the efficacy and safety of medical interventions is gaining interest. Whether observational data can be used to address the same clinical questions being answered by traditional clinical trials is still unclear.

OBJECTIVE To identify the number of clinical trials published in high-impact journals in 2017 that could be feasibly replicated using observational data from insurance claims and/or electronic health records (EHRs).

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional analysis, PubMed was searched to identify all US-based clinical trials, regardless of randomization, published between January 1, 2017, and December 31, 2017, in the top 7 highest-impact general medical journals of 2017. Trials were excluded if they did not involve human participants, did not use end points that represented clinical outcomes among patients, were not characterized as clinical trials, and had no recruitment sites in the United States.

MAIN OUTCOMES AND MEASURES The primary outcomes were the number and percentage of trials for which the intervention, indication, trial inclusion and exclusion criteria, and primary end points could be ascertained from insurance claims and/or EHR data.

RESULTS Of the 220 US-based trials analyzed, 33 (15.0%) could be replicated using observational data because their intervention, indication, inclusion and exclusion criteria, and primary end points could be routinely ascertained from insurance claims and/or EHR data. Of the 220 trials, 86 (39.1%) had an intervention that could be ascertained from insurance claims and/or EHR data. Among the 86 trials, 62 (72.1%) had an indication that could be ascertained. Forty-five (72.6%) of 62 trials had at least 80% of inclusion and exclusion criteria data that could be ascertained. Of these 45 studies, 33 (73.3%) had at least 1 primary end point that could be ascertained.

CONCLUSIONS AND RELEVANCE This study found that only 15% of the US-based clinical trials published in high-impact journals in 2017 could be feasibly replicated through analysis of administrative claims or EHR data. This finding suggests the potential for real-world evidence to complement clinical trials, both by examining the concordance between randomized experiments and observational studies and by comparing the generalizability of the trial population with the real-world population of interest.

JAMA Network Open. 2019;2(10):e1912869. doi:10.1001/jamanetworkopen.2019.12869

Emulation Starts with Defining the Target Trial

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

OCTOBER 7, 2021

FUNDAMENTALS OF PUBLIC HEALTH

Methods of Public Health Research — Strengthening Causal Inference from Observational Data

Miguel A. Hernán, M.D., Dr.P.H.

Choosing wisely among possible courses of action requires knowledge about the effects of those actions. Public health and medical decision makers therefore need sound causal inferences

to know what works and what harms people. Decision makers

ulation groups, and trials may take years to complete.

protocol, which incorporates eligibility criteria, treatment strategies, treatment assignment, start and end of follow-up, outcomes, causal contrasts (or estimands), and a data-analysis plan. These elements define the causal question and how it will be answered;

We’ve learned a lot about how to do comparisons correctly

1. Active comparator, same treatment modality
2. New users
3. High-dimensional proxy adjustment
4. Control for medication exposure
5. Avoiding design flaws:
    a. reverse causation
    b. adjustment for causal intermediaries
    c. immortal time bias
    d. depletion of susceptibles

When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?

Jessica M. Franklin[1] and Sebastian Schneeweiss[1]

Regulators consider randomized controlled trials (RCTs) as the gold standard for evaluating the safety and effectiveness of medications, but their costs, duration, and limited generalizability have caused some to look for alternatives. Real world evidence based on data collected outside of RCTs, such as registries and longitudinal healthcare databases, can sometimes substitute for RCTs, but concerns about validity have limited their impact. Greater reliance on such real world data (RWD) in regulatory decision making requires understanding why some studies fail while others succeed in producing results similar to RCTs. Key questions when considering whether RWD analyses can substitute for RCTs for regulatory decision making are WHEN one can study drug effects without randomization and HOW to implement a valid RWD analysis if one has decided to pursue that option. The WHEN is primarily driven by vulnerabilities not controlled by investigators, whereas the HOW is focused on avoiding known mistakes in RWD analyses.

Design Trumps Statistics!

The Annals of Applied Statistics 2008, Vol. 2, No. 3, 808–840 DOI: 10.1214/08-AOAS187

FOR OBJECTIVE CAUSAL INFERENCE, DESIGN TRUMPS ANALYSIS[1]

BY DONALD B. RUBIN
Harvard University

For obtaining causal inferences that are objective, and therefore have the best chance of revealing scientific truths, carefully designed and executed randomized experiments are generally considered to be the gold standard. Observational studies, in contrast, are generally fraught with problems that compromise any claim for objectivity of the resulting causal inferences. The thesis here is that observational studies have to be carefully designed to ap- proximate randomized experiments, in particular, without examining any final outcome data. Often a candidate data set will have to be rejected as inadequate because of lack of data on key covariates, or because of lack of overlap in the distributions of key covariates between treatment and control groups, often revealed by careful propensity score analyses. Sometimes the template for the approximating randomized experiment will have to be altered, and the use of principal stratification can be helpful in doing this. These issues are discussed and illustrated using the framework of potential outcomes to define causal effects, which greatly clarifies critical issues.

OPERAND

- Study Objective:
- (1) to better understand sources of variability in treatment effect estimates from observational health care data through comparisons with RCTs
- (2) to examine heterogeneity in treatment effect estimates as the inclusion/exclusion criteria of the RCTs are relaxed to reflect the real world patient population

Co-Leads

OPTUM Labs

MULTI-REGIONAL CLINICAL TRIALS

Sponsors

Amgen
AstraZeneca
Merck
Optum
Pfizer
Sanofi
UCB BioSciences, Inc.

Research Partners Selected

Brown University

Harvard Pilgrim Health Care Institute

Technical Expert Panel

Sponsor representatives

+

9 representatives from academia, pharmaceutical companies, professional societies, etc.

FDA participant as observer

OPERAND Overview

Teams and Approach	Two research teams independently attempt to emulate the same two trials: 1. ROCKET AF 2. LEAD-2 Diabetes
Data	OptumLabs Data Warehouse. (1) claims data alone and (2) claims plus EMR. Initial analyses restricted to inclusion/exclusion criteria of the trials. Followed by relaxation of inclusion/exclusion criteria but within approved indication
Approach	1. Each team used study design documentation provided in the original pivotal publications of the trial results. 2. Given a prescribed set of methods. 3. Allowed to use methods of their own choosing
Decision-making of researchers	Each team documented analytic decisions in research design

Emulation Measures

- 1. Regulatory Agreement—statistically significant result with directional equivalence between RCT and observational estimate
- 2. Statistical Agreement—defined as the point estimate from the observational study falling within the 95% confidence interval of the ATE of the RCT using the reported standard errors of the RCT to define the confidence interval

Findings from OPERAND

- Despite having access to the same RCT documentation and the same OptumLabs data, the two research teams identified somewhat different study cohorts, used different statistical methods but, ultimately, came to similar conclusions of concordance with the target trials.
- This reflected more than just differences in operationalizing RCT inclusion/exclusion criteria and was related to differences in their interpretation of the emulation exercise itself. For example, the endpoint for the LEAD-2 diabetes trial was change in HbA1c. As a result, one team identified a study cohort requiring baseline and follow-up HbA1c levels. The other team, using the target trial emulation framework assumed that the target trial would have no missing HbA1c data. As a result they required only baseline HbA1c levels and dealt with missing values using inverse probability weights.
- This illustrates how researcher decision making can influence study design and statistical methods. In this case, two experienced research teams used appropriate methods. But in the broader world of observational studies, researcher decision-making can lead to selection of poor researcher design and statistical methods.

What Have We Learned From Clinical Trials Emulation Efforts?

- In emulation efforts, the target trial is defined by the RCT being emulated.
- Several emulation efforts have shown that it is possible to estimate similar treatment effects with observational data—at least in certain disease areas. Although rarely done in emulation efforts, relaxing the inclusion and exclusion criteria of the target trial enables treatment effect estimates to be obtained for patients treated in routine care.
- Most RCTs cannot be emulated because of complex treatment regimens and/or clinical inclusion/exclusion criteria that cannot be emulated with available observational data. However, when the data are available, emulations using strong research design and statistical methods have shown that it is possible to estimate similar treatment effects with RCTs and observational studies.
- It is reasonable to expect that the data necessary to emulate economic target trials is generally available in observational data such as claims. As a result, such studies should usually produce reliable results even when there is no actual trial to emulate, as long as they use appropriate research design and statistical methods.

What Have We Learned (continued)

- Variability in researcher decision-making is understudied. In OPERAND we found evidence of variability between two experienced research teams even though the emulation exercise was highly structured and used a common dataset.
- It is reasonable to assume some variability in decision-making on design and statistical methods by experienced researchers. This is unlikely to undermine the reliability of study results as long as appropriate designs and methods are used.
- However, there is wide variance in the quality of observational studies in general. It is important to recognize that, ultimately, this is driven by researcher decision making about data, research design, and statistical methods.
- Additional efforts on fit-for-purpose data, provision of harmonized research protocols, and other efforts are needed to reduce inappropriate decision-making by inexperienced researchers.