JCI The Journal of Clinical Investigation

# Generating evidence for therapeutic effects: the need for well-conducted randomized trials

Robert M. Califf, … , Adrian F. Hernandez, Eric D. Peterson

*J Clin Invest.* 2021;131(2):e146391. https://doi.org/10.1172/JCI146391.

**Viewpoint**

The impact of the severe acute respiratory syndrome coronavirus 2 pandemic is placing tremendous pressure on government and the biomedical research enterprise to quickly develop safe, effective therapies and move them into practice. Amid these unprecedented circumstances, it is worth reflecting on the best pathways for answering questions about which treatments and strategies work, both for individuals and populations. Medical practice often hinges on a key question: should a patient with a specific diagnosis and set of clinical characteristics receive a particular therapy or not? The challenge, then, is to provide patients and clinicians with the best information about an intervention's benefits and risks, thereby helping them make informed decisions. These same considerations help policymakers develop clinical practice guidelines and quality measures and make formulary and payment decisions. Lessons from acute myocardial infarction Scientists and clinical investigators have learned much about the underlying causes of acute myocardial infarction (AMI) (e.g., thrombosis in the coronary artery) as well as complications that lead to adverse outcomes (e.g., ventricular arrhythmia, heart failure). This underlying pathobiology suggests a myriad of potential avenues for prevention and treatment (1). However, many putative therapies failed when examined in prospective randomized controlled trials (RCTs), due to unexpected toxicity or lack of efficacy. Based on these RCTs, we have established what works in the care of patients with AMI […]



# Generating evidence for therapeutic effects: the need for well-conducted randomized trials

Robert M. Califf,[1,2,3] Lesley H. Curtis,[3,4] Robert A. Harrington,[2] Adrian F. Hernandez,[3,4] and Eric D. Peterson[5]

[1]Verily Life Sciences (Alphabet), South San Francisco, California, USA. [2]Stanford University School of Medicine, Stanford, California, USA. [3]Duke University School of Medicine, Durham, North Carolina, USA. [4]Duke Clinical Research Institute, Durham, North Carolina, USA. [5]University of Texas Southwestern Medical School, Dallas, Texas, USA.

The impact of the severe acute respiratory syndrome coronavirus 2 pandemic is placing tremendous pressure on government and the biomedical research enterprise to quickly develop safe, effective therapies and move them into practice. Amid these unprecedented circumstances, it is worth reflecting on the best pathways for answering questions about which treatments and strategies work, both for individuals and populations. Medical practice often hinges on a key question: should a patient with a specific diagnosis and set of clinical characteristics receive a particular therapy or not? The challenge, then, is to provide patients and clinicians with the best information about an intervention's benefits and risks, thereby helping them make informed decisions. These same considerations help policymakers develop clinical practice guidelines and quality measures and make formulary and payment decisions.

## Lessons from acute myocardial infarction

Scientists and clinical investigators have learned much about the underlying causes of acute myocardial infarction (AMI) (e.g., thrombosis in the coronary artery) as well as complications that lead to adverse outcomes (e.g., ventricular arrhythmia, heart failure). This underlying pathobiology suggests a myriad of potential avenues for prevention and treatment (1). However, many putative therapies failed when examined in prospective randomized controlled trials (RCTs), due to unexpected toxicity or lack of efficacy. Based on these RCTs, we have established what works in the care of patients with AMI and have substantially reduced the morbidity and mortality associated with AMI over several decades (2). If not for these trials, clinicians would probably still be using ineffective or even dangerous therapies and strategies in practice.

## Challenges for implementing new therapies in the pandemic

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic presents similar issues but with even greater urgency. A plethora of therapies have been proposed to treat coronavirus disease 2019 (COVID-19), based on our current understanding of disease pathophysiology and the mechanisms of action of putative therapeutics. As with AMI, RCTs have shown that many of these candidates either confer no benefit or, worse, cause net harm (3).

At the same time, the pandemic reveals shortcomings in traditional research models. Many studies of COVID-19 therapies have been too small to definitively answer the questions asked—a perennial problem for our clinical research system (4), but one that has become especially acute. Typically conducted at larger research centers, traditional RCTs can be slow (5), tend to enroll younger, healthier volunteers (6), and in the US, often cost tens to hundreds of millions of dollars.

Given this background, it's worth considering alternatives to traditional RCTs that have been promoted as more efficient, less expensive options for generating reliable evidence to inform practice (Table 1). One option is to return to the use of pathophysiological reasoning to guide treatment. On one hand, the ability to measure biological processes and create better models of disease continues to improve, and this progress has yielded new therapeutic targets, better drug dosing, and even improved selection of patients likely to respond to an intervention. Nevertheless, the vast majority (80%–95%) of molecules that enter human trials fail to win authorization for marketing for clinical use (7), despite being accompanied by a compelling scientific rationale and enough preliminary evidence to convince investors, regulators, and institutional review boards that experiments in humans were warranted. One has only to look at the drugs that showed activity against SARS-CoV-2 or the accompanying inflammatory response in the laboratory, but failed to improve outcomes in the clinic. Thus, while biological plausibility remains a cornerstone of rational drug development, it does not suffice to assure that the benefits of a putative therapy outweigh the risks (8).

Another option is to allow rapid access to therapies based on preliminary data and then simply record what happens in practice. This approach—using real-world observational data to evaluate safety and effectiveness—has been enabled by the widespread availability of electronic health record (EHR) and claims data. Additionally, modern analytical tools such as machine learning can identify relationships between therapies and subsequent outcomes, further expanding the potential to use existing data to furnish clinical evidence. Although statistical methods of adjusting for factors that may influence treatment selection and thus bias outcome comparisons have improved, neither they nor the data in which they are applied are robust enough to routinely serve as the primary means for evaluating therapies.

**Conflict of interest:** RMC is an employee of Verily Life Sciences and Google (both Alphabet companies) and serves on the board of Cytokinetics. RAH has received consulting fees from Bridgebio and holds equity in Element Science. He also receives personal compensation for serving on the board of directors for SignalPath. AFH receives personal fees from Amgen, AstraZeneca, Bayer, Cytokinetics, and Merck, and grants from AstraZeneca and Novartis. EDP has received research support from Amgen, Janssen, Esperion, Bristol Myers Squibb, Sanofi, AstraZeneca, and Merck. He also receives consulting fees from Amgen, Janssen, Novartis, Boehringer Ingelheim, AstraZeneca, Amerin, Sanofi, Abiomed, Merck, Livongo, Cerner, and Eko, and has ownership in Swift Clinical Sciences LLC.



**Reference information:** *J Clin Invest*. 2021;131(2):e146391. https://doi.org/10.1172/JCI146391.

**Table 1. Necessary attributes of high-quality practical randomized controlled trials**

| Attribute | Comment |
|---|---|
| Entry criteria reflecting population likely to be treated | • Balance internal validity with generalizability<br>• Meet ethical obligation to offer participation to people with the condition of interest<br>• Ensure that participants recruited to a trial are reflective of community practice |
| Proper randomization | • Randomization with appropriate randomization scheme<br>• Ensure that trial participants have an accurate working understanding of the trial and the available treatment options |
| Adequate number of events to provide reliable measure of treatment effect | • Statistical power is driven by number of outcomes, not by number of participants |
| Measure adherence to assigned treatment | • Treatment effects are most likely seen during period of treatment<br>• Although intention to treat is the key principle for primary analysis, understanding results requires knowledge of adherence |
| Define and accurately measure key outcomes of interest | • Less precision is needed on baseline characteristics or ancillary data<br>• Focusing on fidelity of outcome measurements optimizes likelihood of answering the question posed by the trial<br>• Focus on trial outcomes most meaningful to patients |
| Complete follow-up for sufficient time to measure benefits and risks | • Benefits and risks often occur in different time frames<br>• The balance of benefit and risk over time is essential |
| Collect unexpected serious adverse events | • After collection of all adverse events in an adequate number of participants, little is added by continuing this practice and it distracts research staff from more important issues<br>• All unexpected serious adverse events should be collected to ensure identification of serious new issues |
| Transparency | • Statistical analysis plan should be in place at trial inception |
| Data should be made available for external validation | |

Challenges such as imbalances in unmeasured confounders and difficulty identifying inception time (time zero) for counting outcomes and for measuring when the therapy would have been started in the untreated group remain daunting (9). Comparative evaluations of observational data will play a growing role in confirming the results of RCTs in broader populations, but these challenges limit their standalone use for weighing safety and efficacy.

A third option would be to use registries to systematically apply therapies with a clear inception time and then create a synthetic control group to compare registry outcomes with a historical or virtual array of expected outcomes. This approach has proven useful in rare genetic diseases with clear, reproducible outcomes and no effective treatments, particularly when death is common early in the disease course. However, because most diseases have heterogeneous outcomes and one or more partially effective treatments, the risk of a faulty control group is high and the same general methodological caveats for observational comparisons apply.

Finally, one could consider randomization, which has the fundamental advantage of balancing prognostic factors (whether measured or not) at inception. Although some argue that RCTs are too slow, there is nothing about randomization as such that causes delay, and well-designed RCTs may actually decrease the time needed to deliver reliable knowledge. The protocol clearly establishes an inception time, both for measuring outcomes without time bias and for measuring baseline characteristics so that inferences can be made about heterogeneity of treatment effect. An increasing array of options for randomization are available depending on the trial's goal. For most individual therapeutic decisions in the context of clinical care, individual randomization is optimal. But for systematic application of policies in healthcare delivery, cluster randomization or quasi-experimental designs such as a stepped-wedge approach may be preferred.

## A modern approach to trial design

To establish which treatments are effective for particular patients, trials should meet general standards for quality and efficiency while limiting extraneous burdens on participants or investigative teams (10), an approach that has been called a practical or simple clinical trial (11). Enough outcomes should be included to confer adequate statistical power, randomized treatment allocation should be monitored, a study population representative of patients who will actually be treated in practice should be enrolled, and follow-up should be complete. Trials should focus on a randomized comparison of alternatives in practice, carefully measuring relevant outcomes but eschewing an onerous and expensive collection of extraneous data. In the context of a pandemic, this presents an attractive option for rapidly testing repurposed drugs otherwise likely to be used off-label without generating useful evidence.

To meet these criteria, the clinical trials enterprise must continue to evolve. Networks of research sites organized to answer important questions with appropriate incentives for clinicians and health systems are essential. Readily available claims and EHR data, combined with access to data collected directly from trial participants via apps and wearable technologies, could reduce strains on budgets and human resources and expand data collection at low cost while enabling research personnel to focus on precise ascertainment of a limited set of critical data. Approaches for prioritizing trials so that clinicians and patients are steered toward studies with the greatest potential impact are also needed.

## What is in the patient's best interest?

One concern is that a clinician's duty to the patient's welfare may conflict with the

mandate of researchers to create generalizable knowledge (12). More specifically, if a treatment is likely to have benefits, how could a clinician not offer the patient that treatment, versus allowing the patient to possibly receive a placebo or alternative standard of care? The counter-argument is that, as noted, most experimental treatments fail because risks outweigh benefits. In fact, a person is more likely to have a good outcome with the control than with the experimental treatment. Further, even when the benefits of the new therapy versus control outweigh the risks, most have modest effects—on the order of a 15% to 25% reduction in the outcome of interest—so the degree of risk from missing out is often overestimated. Finally, when participation in trials is considered, there is an understandable desire to protect people from experimentation, but evidence indicates that study participants do at least as well and possibly better than those who do not participate—a likely benefit of the rigor of a prospective protocol and standard of care, whether a person is in the intervention or control group. Nevertheless, people are often uncomfortable with randomization, even after receiving explanations for why it is essential and usually does not entail excessive risk. Shifting to the default mode of learning about which treatments are best will require a reframing of ethical issues in clinical studies, as has been suggested by some ethicists (13).

## Perspectives

The urgency of the COVID-19 pandemic has been used to justify abandoning randomization. However, a review of evidence from previous pandemics by the National Academy of Medicine concludes that RCTs, preferably on a national scale with an interoperable system for rapid-cycle evidence generation, should be initiated as quickly as possible (14, 15). These lessons are just as pertinent in more ordinary circumstances. People with common chronic conditions or those with rare diseases suffer from inadequate evidence to support critical decisions that affect health and well-being.

The search for methods that can bypass the need for prospective RCTs will continue and progress will be made in specific circumstances, such as the ones described herein. However, in most situations, if we seek reliable knowledge about the benefits and risks of a therapy versus alternatives, well-designed RCTs will continue to be essential for providing reliable evidence. The critical issue for public health—one that the current crisis has starkly illuminated—is how to create a culture and set of incentives that lead to the logical conclusion: if the answer to a clinical question is uncertain, randomize!


1. Fanaroff AC, et al. Randomized trials versus common sense and clinical observation: JACC review topic of the week. *J Am Coll Cardiol.* 2020;76(5):580–589.
2. Krumholz HM, et al. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995-2006. *JAMA*. 2009;302(7):767–773.
3. RECOVERY Collaborative Group, et al. Effect of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med*. 2020;383(21):2030–2040.
4. Califf RM, et al. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007–2010. *JAMA*. 2012;307(17):1838–1847.
5. Eisenstein EL, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials*. 2008;5(1):75–84.
6. Kennedy-Martin T, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16:495.
7. Lowe D. The latest on drug failure and approval rates. In The Pipeline. https://blogs.sciencemag.org/pipeline/archives/2019/05/09/the-latest-on-drug-failure-and-approval-rates. Updated May 9, 2019. Accessed November 2, 2020.
8. Krumholz HM, et al. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995–2006. *JAMA*. 2009;302(7):767–773.
9. Hernán MA, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–75.
10. Richesson R, et al. Developing and refining the research questions. Rethinking Clinical Trials. https://rethinkingclinicaltrials.org/chapters/design/using-electronic-health-record-data-pragmatic-clinical-trials-top/developing-refining-research-questions/. Updated October 20, 2020. Accessed December 1, 2020.
11. Sugarman J, Califf RM. Ethics regulatory complexities for pragmatic clinical trials. *JAMA*. 2014;311(23):2381–2382.
12. Guarino B. New York's bioethics experts prepare for a wave of difficult decisions. *Washington Post*. March 28, 2020. Accessed November 11, 2020. https://www.washingtonpost.com/world/national-security/new-yorks-bioethics-experts-prepare-for-a-wave-of-difficult-decisions/2020/03/28/4501f522-7045-11ea-a3ec-70d7479d83f0_story.html.
13. Faden RR, et al. An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. *Hastings Cent Rep*. 2013;Spec No:S16–27.
14. National Academies of Sciences, Engineering, Medicine. Conducting clinical research during an epidemic. In: Busta ER, et al., eds. *Integrating Clinical Research into Epidemic Response: The Ebola Experience.* National Academies Press; 2017:37–82.
15. Dean NE, et al. Creating a framework for conducting randomized clinical trials during disease outbreaks. *N Engl J Med*. 2020;382(14):1366–1369.