

The use of observational research to inform clinical practice

Nigel Paneth, Michael Joyner

J Clin Invest. 2021;131(2):e146392. <https://doi.org/10.1172/JCI146392>.

Viewpoint

A historical perspective Randomized controlled trials (RCTs) are universally accepted in the world of medicine as the preferred design for the analysis of health-related interventions, be they preventive or therapeutic. It may seem odd to realize that this now-axiomatic approach has been standard for only some 70 years of the more than two-millennia history of medicine. While sporadic proto-randomized trials had been published earlier, the RCT became most firmly established as the standard design for treatment evaluation in 1948, when the British Medical Research Council showed in a carefully randomized multi-hospital trial that streptomycin had substantial benefits for patients with pulmonary tuberculosis compared with the then-standard treatment of bed rest (1). The randomized trial became so well established that many often feel that absent one RCT or, ideally several, showing the statistically significant efficacy of an intervention, no recommendation can be given as to the use of that intervention. There is certainly good reason for caution. Observational studies of the efficacy of treatment — comparisons before and after treatment, or of treatment recipients with nonrecipients — have often led to serious errors in medicine. To cite just one egregious example, the use of diethylstilbestrol in pregnancy was justified by deeply flawed observational research (2), whereas the sole RCT (3) — which showed no evidence of effectiveness at all — was [...]

Find the latest version:

<https://jci.me/146392/pdf>



The use of observational research to inform clinical practice

Nigel Paneth¹ and Michael Joyner²

¹Department of Epidemiology and Biostatistics, and Pediatrics and Health Development, Michigan State University, College of Human Medicine, East Lansing, Michigan, USA. ²Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, Minnesota, USA.

A historical perspective

Randomized controlled trials (RCTs) are universally accepted in the world of medicine as the preferred design for the analysis of health-related interventions, be they preventive or therapeutic. It may seem odd to realize that this now-axiomatic approach has been standard for only some 70 years of the more than two-millennia history of medicine. While sporadic proto-randomized trials had been published earlier, the RCT became most firmly established as the standard design for treatment evaluation in 1948, when the British Medical Research Council showed in a carefully randomized multi-hospital trial that streptomycin had substantial benefits for patients with pulmonary tuberculosis compared with the then-standard treatment of bed rest (1).

The randomized trial became so well established that many often feel that absent one RCT or, ideally several, showing the statistically significant efficacy of an intervention, no recommendation can be given as to the use of that intervention. There is certainly good reason for caution. Observational studies of the efficacy of treatment — comparisons before and after treatment, or of treatment recipients with nonrecipients — have often led to serious errors in medicine. To cite just one egregious example, the use of diethylstilbestrol in pregnancy was justified by deeply flawed observational research (2), whereas the sole RCT (3) — which showed no evidence of effectiveness at all — was roundly ignored. The long-term damage to the treated fetus was only uncovered after years of use by millions of pregnant women (4).

There are times when RCTs are difficult to undertake. Some interventions do not lend themselves easily to randomiza-

tion. Large multicomponent systems of care such as coronary care and complex surgical procedures are not easily subject to random assignment. It is difficult to imagine an RCT of the Heimlich maneuver or of door-to-balloon time in angioplasty. The equipoise required to undertake a trial may be undercut by accumulated experience or powerful belief systems that make an untreated control arm seem unethical. Trials at times can be so costly to mount that it may not seem worth the investment of time and resources. For all these reasons, many widely used interventions in medicine have never undergone assessment by RCT.

Lessons from epidemiology

The field of epidemiology, though not bereft of trials, draws its conclusions predominantly from observational data. The first and nearly identical sets of epidemiologic rules of judgment for ascertaining causality were published on both sides of the Atlantic nearly simultaneously (5, 6), apparently not quite independently (7). The US version provides a handy quintet of criteria — consistency, strength, specificity, temporal relationship, and coherence — with which to judge the likelihood of any exposure-disease association being causal. For the purpose of evaluating treatment, temporal association, i.e., that the exposure or treatment preceded the onset of disease is generally a given, and the specific treatment and the specific outcome of interest are usually a clear focus of the research. That leaves three criteria to consider when thinking of observational research in relation to treatments in medicine.

Strength. Strength refers to the size of the observed difference, not the *P* value associated with it, which only assesses the

role of chance in creating the association, whatever the strength of the association. If a judgment about treatment is to be made on the basis of observational data, the effect size had better be substantial. Confounders and biases inevitably arise when study arms are not made comparable through random assignment. Since confounding factors must have effects on the outcome that are larger than the effect being claimed for the treatment, a large effect size puts a cap on the likelihood of a confounder or a bias operating; a 50% reduction in mortality from treatment is much harder to confound than a 20% reduction.

Coherence. Does the intervention make sense in light of what else we know? A prominent component of this criterion is mechanism of action. Few interventions are undertaken without a hypothesized mechanism of action, but the evidence supporting the mechanism can come from a variety of sources, and in vitro does not always translate to in vivo, especially in humans.

Consistency. A treatment repeatedly shown to be effective is more likely to be truly effective than one which seems effective in some studies but not in others.

We suggest one additional criterion that has stood the test of time, and that is the use of total population data to draw conclusions. RCTs are conducted in individuals willing to be enrolled in a study and to accept randomization. Such individuals are usually younger, healthier, more educated, and less likely to be from minority populations. Generalization from trials can therefore be uncertain, but total population data have sample sizes thousands of times larger than any trial and exclude virtually no one.

Some of the best evidence for the effectiveness of cancer screening is the consistent declines in the mortality rates for the four cancers universally screened for in the US — breast, colon, cervix, and prostate — and the correspondence of

Conflict of interest: The authors have declared that no conflict of interest exists.

Copyright: © 2021, American Society for Clinical Investigation.

Reference information: *J Clin Invest*. 2021;131(2):e146392. <https://doi.org/10.1172/JCI146392>.

these declines with the onset of screening and the paucity of alternative explanations for the declines (8, 9). Several studies of whether newborn intensive care reduces neonatal mortality have been based on total population data sources, which, without exception, show lower mortality in high-risk newborns born where intensive care was available (10). These cross-sectional assessments have been amply supported by time-trend findings from vital data in the total US population (11).

A balanced perspective in the midst of a pandemic

In epidemic situations, the problems of conducting phase III RCTs are compounded by the absence of information on the most appropriate patients to treat, dosage to use, and side effects generally uncovered in phase I or II trials. Moreover, the urgency to treat when patients are dying in large numbers can make providers reluctant to wait for the findings of a large trial.

Convalescent plasma in the current coronavirus disease 2019 (COVID-19) pandemic provides a useful example. Trials were slow to be mounted in the early days of the pandemic, and all trials so far published have been small and statistically inconclusive for mortality. Six of the seven trials now in the public domain (none from the US) showed nonsignificantly lower mortality in the treated arms, averaging approximately 50%, but significant clinical improvements in other parameters were noted in some of the trials (12). Results from much larger trials are now on the verge of being reported, and by the time this Viewpoint is published, we may not have to rely on observational data to draw conclusions, but as of now, the most informative data we have are from observational studies.

At least 13 studies have been published comparing patients with COVID-19 treated or not treated with convalescent plasma, and all show reduced mortality in the treated group, often closely mirroring the findings of the RCTs (13). The most convincing evidence, however, is the strong

and significant dose-response relationship of the active agent in convalescent plasma — antibody titer — with mortality in two studies of several thousand patients for whom antibody levels could be assessed in the plasma they received (14, 15). The higher the titer, the lower the mortality. Antibody levels were unknown at the time of transfusion, making the possibility of bias or confounding nearly as remote as in an RCT. These data were a major component of the evidence used by the FDA to issue Emergency Use Authorization for in-patient treatment with convalescent plasma on August 23, 2020 (16).

Turning to the causal criteria, the association of convalescent plasma with mortality found both in trials and observational data — approximately a halving of mortality — is a strong effect. The findings have been both consistent across studies and coherent with what we know of how antibodies work and historical evidence of the effectiveness of convalescent plasma in other infectious diseases (17).

Conclusions

Ultimately, everything we do in medicine is a matter of judgment. Although RCTs should be supported whenever possible, we should not be paralyzed into inaction when they are not available, nor should we willfully ignore important evidence coming from other sources. Federal agencies, professional bodies — indeed, any person or group making therapeutic recommendations — should always consider the totality of the available evidence, including that generated by observational studies.

Address correspondence to: Nigel Paneth, Departments of Epidemiology and Biostatistics and Pediatrics and Human Development, Michigan State University, College of Human Medicine, 909 Fee Road, Room 218, East Lansing, Michigan 48824, USA. Phone: 517.884.3961; Email: paneth@epi.msu.edu.

1. [no authors listed]. Streptomycin treatment of pulmonary tuberculosis. *Br Med J*. 1948;2(4582):769–782.

2. Smith OW, Smith GVS. Use of diethylstilbestrol to prevent fetal loss from complications of late pregnancy. *N Engl J Med*. 1949;241(15):562–568.
3. Dieckmann WJ, et al. Does the administration of diethylstilbestrol during pregnancy have therapeutic value? *Am J Obstet Gynecol*. 1953;66(5):1062–1081.
4. Herbst A, et al. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med*. 1971;284(15):878–881.
5. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295–300.
6. USDHEW. The 1964 Surgeon General's Advisory Committee on Smoking and Health: Smoking and Health. Chapter 3. Criteria for Judgment. <https://biotech.law.lsu.edu/cases/tobacco/nbbmq.pdf>. Accessed December 7, 2020.
7. Blackburn H, Labarthe D. Stories from the evolution of guidelines for causal inference in epidemiologic associations: 1953–1965. *Am J Epidemiol*. 2012;176(12):1071–1077.
8. Siegel RL, et al. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30.
9. Levin TR, et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large, community-based population. *Gastroenterology*. 2018;155(5):1383–1391.e5.
10. Laswell SW, et al. Perinatal regionalization for very low-birth-weight and very preterm infants: a meta-analysis. *JAMA*. 2010;304(9):992–1000.
11. Buehler JW, et al. Birth weight-specific infant mortality, United States, 1960 and 1980. *Public Health Rep*. 1987;102(2):151–161.
12. Libster R, et al. Prevention of severe COVID-19 in the elderly by early high titer plasma [preprint]. <https://doi.org/10.1101/2020.11.20.20234013>. Posted on MedRxiv on November 21, 2020.
13. Klassen SA, et al. Evidence favoring the efficacy of convalescent plasma for COVID-19 therapy [preprint]. <https://doi.org/10.1101/2020.07.29.20162917>. Posted on Medrxiv October 29, 2020.
14. Joyner MJ, et al. Effect of convalescent plasma on mortality among hospitalized patients with COVID-19: initial three-month experience [preprint]. <https://doi.org/10.1101/2020.08.12.20169359>. Posted on Medrxiv on August 12, 2020.
15. U.S. Food and Drug Administration (FDA). Updated evidence to support the emergency use of COVID-19 convalescent plasma — as of 9/23/2020. <https://www.fda.gov/media/142386/download>. Accessed December 7, 2020.
16. U.S. Food and Drug Administration (FDA). Letter from Denise M. Hinton, Chief Scientist, FDA to Robert P. Kadlec, Assistant Secretary for Preparedness, US DHHS. <https://www.fda.gov/media/141477/download>. Accessed December 7, 2020.
17. Casadevall A, Pirofski LA. The convalescent sera option for containing COVID-19. *J Clin Invest*. 2020;130(4):1545–1548.