CellPress

Trends in Molecular Medicine

Opinion

Finding evidence for treatment decisions in a pandemic

Nigel Paneth,[1,*] Michael J. Joyner,[2] and Arturo Casadevall[3]

The randomized controlled trial (RCT) is the mainstay of treatment evidence in medicine because it is the most rigorously conducted and analyzed form of human research. Yet, the intense and detailed methodology that has evolved to implement RCTs has reduced their value in providing timely and useful evidence of the effectiveness of medical or public health interventions in pandemic conditions. The crisis conditions of a pandemic illustrate the need for medicine to take a broader view of the evidentiary landscape to include evidence from sources other than conventional RCTs. Such sources include analyses of vital data, observational research, quasi-experiments, and flexible RCTs carefully designed to address underlying biological and clinical realities and to avoid unnecessary rigidity.

Highlights

During pandemics, health care and public health professionals rarely have all the information required for decision-making. Randomized trials are the key source of regulatory decisions and clinical recommendations, but they can be misleading, especially when conducted under pandemic conditions. Problems include the high likelihood of false-negative trials, lack of generalizability, and trial interruptions because of pandemic waning of cases.

Some authors have recommended that more attention be paid in decision-making to 'real-world evidence', including vital data, propensity-matched cohort studies, some forms of administrative data, and quasi-experiments. In addition, trials can be improved by taking a fuller account of the pathophysiology of the condition under investigation and its relation to the agent studied. More use can be made of adaptive trials that can flexibly be modified to changing circumstances. Pandemic conditions reveal the need to broaden the evidentiary base beyond the conventional randomized trial.

Sources of evidence for treatment decisions

The RCT has become firmly established as the most important source of evidence in medicine for the evaluation of treatments. Guidance panels, specialty society recommendations, and, most important, federal regulators nearly always require one or more RCTs to support a treatment endorsement. In recent years, voices have been raised suggesting that reliance on RCTs alone is not ideal and that a broader evidentiary base should be considered in evaluating therapies [1–4]. Concerns have been raised about the generalizability of RCTs, about the length of time required to complete them, the regulatory burden, and much else. Even the FDA has noted the need for real-world evidence and data, though most FDA treatment approvals still require RCT evidence (https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence). However, RCTs cannot guarantee the safety of approved drugs [5], nor can they always assure effectiveness in the real world [6]. It has been argued that nonrandomized observational studies frequently [7], though not always [8], have results nearly identical to trials. In this opinion article, we argue that pandemic circumstances point out the limitations of over-reliance on the RCT and the need for medicine to broaden its evidentiary base for choosing therapies.

The conventional RCT in ordinary times

The preparation required for developing an RCT is substantial. The specification of a clear and measurable hypothesis is critical because it determines the population in which to undertake the trial, the outcome of interest, and the exclusion and inclusion criteria. Power must be estimated, and sources of qualified participants must be sought. A randomization scheme that cannot be apparent to participant, investigator, or data analyst [so-called triple **masking** (see Glossary)] is usually developed, and a placebo may need to be synthesized. The protocol must include an analysis plan, including stopping rules. The overview functions of an **institutional review board**, a data safety committee, and sometimes a **clinical monitor** must be employed. Partnerships must be sought if the trial is to be multicenter. Funding must be obtained to staff a

[1]Epidemiology & Biostatistics and Pediatrics & Human Development, Michigan State University, Lansing, MI, USA
[2]Anesthesiology, Mayo Clinic, Rochester, MN, USA
[3]Department of Molecular Microbiology and Immunology, Johns Hopkins University School of Public Health, Baltimore, MD, USA

*Correspondence:
paneth@msu.edu (N. Paneth).

team of clinicians, research scientists, trained nursing staff, research coordinators, statisticians, laboratory personnel, and more.

The model RCT, pioneered in the study of streptomycin in tuberculosis [9], is of a hospital-administered drug to treat a well-defined disease where signs of improvement are likely to be clear, if not unmistakable. Another model, pioneered for polio [10], was the vaccine trial, conducted in **outpatients** and in very large numbers, but again with the advantage of a single administration of the agent under clinical supervision.

The further one moves away from those models, the more complex the design, execution, and analysis of the trial. Trials of oral drug regimens to outpatients cannot easily assure compliance with the intervention. Trials that require testing of participants after randomization (e.g., when administering blood products) create a time lag referred to as 'immortal time bias' [11]. Trials aiming to prevent rather than to treat disease require much larger sample sizes and longer time horizons. Interventions other than drugs – nutritional/dietary interventions, surgical procedures, education programs – raise greater possibilities of noncompliance, crossover, and difficulties in masking.

The time from trial conception to first enrollment is usually at least 1 year, and in one study of 109 trials with a substantial follow-up period, the lag from first enrollment to publication of findings was more than 5 years [12]. Further years elapse before a positive trial finding is adopted by the medical community, in part because of conflicting trial findings [13] and uncertainty about generalizability [14].

The conventional RCT during a pandemic

Epidemic circumstances create several tensions that can further limit the value of RCTs. Because proven treatments cannot be available at the onset of a pandemic, repurposed drugs with little pathophysiologic rationale, such as **hydroxychloroquine** and **ivermectin,** may become the subjects of trials. It is difficult to specify trial outcomes when the clinical course of a new disease is not yet fully understood, and this uncertainty complicates estimation of the number of participants needed and calculation of statistical power. Epidemics often come in waves, and the waxing and waning of eligibles can wreak havoc in even the best planned trial, as seen in some coronavirus disease 2019 (COVID-19) trials stopped for lack of participants [15].

In the first major published COVID-19 RCT of **remdesivir**, the authors commented:

> *The trial was implemented during a time of restricted travel, and hospitals restricted the entrance of nonessential personnel. Training, site initiation visits, and monitoring visits often were performed remotely. Research staff were often assigned other clinical duties, and staff illnesses strained research resources. Many sites did not have adequate supplies of personal protective equipment and trial-related supplies, such as swabs* [16].

During an epidemic, the highest priority is given, understandably, to the sickest patients. Moreover, hospitalized patients are the most accessible to the specialty clinicians who conduct trials and have the most adverse outcomes to study. Not surprisingly, then, the sickest patients often become the focus of trials, though when antivirals are needed – as in the COVID-19 pandemic – early treatment is critical. The traditional preference of the medical community for cure over prevention is amplified during an epidemic.

This **inpatient** bias can be seen in nearly all RCTs of therapies for COVID-19. In trials of steroids [17], the severely ill inpatient was the appropriate target. In fact, in the largest trial, mechanically

Glossary

Clinical monitor: clinician who ascertains adverse events during a trial.
Hydroxychloroquine: antimalarial drug used to treat COVID-19.
Inpatient: hospitalized patient.
Institutional review board: the body responsible for ethical research conduct.
Ivermectin: antiparasitic drug used to treat COVID-19.
Masking: concealing the trial assignment.
Medicaid: US federal medical insurance for patients with low incomes.
Outpatients: patients treated outside the hospital.
Remdesivir: antiviral drug used to treat COVID-19.

ventilated patients benefited more from steroids than did patients not receiving oxygen [17]. But it was inappropriate to study remdesivir [16] and convalescent plasma [18] – both antivirals – largely in hospitalized patients. The few trials of these antivirals conducted in outpatients showed far greater effectiveness [19–21].

By contrast, pharmaceutical companies tested their two major products – vaccines and monoclonal antibodies – nearly entirely outside of the hospital. Vaccines were, of course, studied in unaffected individuals, whereas most monoclonal antibody trials were conducted in outpatients, where a striking difference was seen: success in outpatients [22] but ineffectiveness with inpatients [23], as predicted for an antiviral therapy.

In summary, several RCTs conducted during the COVID-19 epidemic had difficulty achieving their goals and were often flawed by selection of the wrong populations to study. Although negative RCTs were useful in stemming the use of ineffective drugs such as hydroxychloroquine and ivermectin and demonstrated the value of steroids in very sick patients, antivirals such as remdesivir and convalescent plasma found themselves being misapplied in both trials and in practice by being given too late in the disease course.

Alternative sources of evidence

Numerous forms of data exist from which observations may be gleaned that should be incorporated into thinking about clinical policy and treatment decisions, not to replace but to supplement RCT data. These include both observational and experimental findings.

Observational data

Vital data and registries

We must recognize that many important components of the population – prisoners, the institutionally disabled, those in military service, the uninsured – are weakly represented in many forms of 'big data'. Data systems that are independent of individual medical care providers, however, and that exist by virtue of legal mandates (vital data, state disease registries) have the powerful advantage of including disease and mortality data from virtually the entire population. Yet, the value of vital data in showing trends in mortality supporting or opposing the value of widely used interventions is often overlooked.

The best evidence for the value of cancer screening comes from the remarkable declines in mortality for screened cancers. Where screening is for early disease (breast and prostate cancer), declines in age-specific mortality independent of patterns of incidence shown in state and national registry data parallel the onset of screening [24,25]. Where screening is for precancerous lesions (cervical and colon cancer), declines in both incidence and mortality have been documented [26,27].

Population-wide declines in mortality from leukemia in childhood likewise parallel the revolutionary developments in chemotherapy for this disorder beginning in the 1950s and 1960s [28]. Newborn intensive care as a whole has never been studied via RCT, but unprecedented declines in mortality for low birthweight infants became apparent in vital data as early as the 1970s [29]. The failure to pay attention to vital data trends is a weakness of many high-level clinical policy recommendations and may be linked to the recent increase in prostate cancer mortality and in metastasis at presentation [30].

During the recent epidemic, vital data were the best source for understanding the impact of the epidemic on overall mortality in the USA and also showed that areas with COVID-19-related mask mandates had lower age-specific COVID-19 mortality rates than areas that did not [31]. It was also possible to show, by linking hospitalization data, mortality data from vital records,

and blood bank data, a powerful inverse correlation between convalescent plasma use and mortality between September 2020 and March 2021 [32].

Cohort studies with propensity matching

Comparisons of recipients and nonrecipients of treatments are fraught with difficulties, not the least of which is confounding by indication [33], meaning that choice of treatment is dependent on disease severity and comorbidities. Nonetheless, such comparisons are frequently made in medicine. During the epidemic, several hospitals designed cohort studies within their institutions to examine whether convalescent plasma was helpful or not in COVID-19 patients, doing their best to match recipients and nonrecipients on a variety of factors affecting the risk of adverse outcomes. Studies from Houston Methodist Hospital [34] and New York's Mount Sinai Hospital [35] were good examples of the careful propensity-matched cohort study, and they revealed that, as expected, benefit was found only in patients treated early.

Administrative data and electronic medical records

The use of large administrative databases that cross hospital boundaries or are based on other sources such as **Medicaid** files are single-hospital cohort studies writ large but generally with less ability to control for confounding. They also frequently encounter difficulties with the lack of comparability of recorded items across the entire data file. Historically, the most productive forms of administrative data analysis of treatments have come from large medical care systems, such as Kaiser Permanente or the Henry Ford Health System, which use a single format for all components of the system and have research offices to curate the data and understand their limitations. Data from heterogeneous hospital sources using different medical record systems and amalgamated by convenience have been less successful in providing convincing evidence of treatment effectiveness. During the current epidemic, the Hospital Corporation of America, which uses a uniform data system and has a sophisticated research office, was able to provide very convincing evidence of the value of convalescent plasma, again showing the value of early treatment and its uselessness in advanced disease [36].

Experimental evidence

Quasi-experiments

It may not be widely appreciated that historically important health policy decisions were made on the basis of experiments that were not randomized, but that instead introduced an intervention to one population while not introducing it to another, examining health states of interest in both populations before and after the time of the intervention, so-called quasi-experiments. A good example is water fluoridation, adopted nationally after quasi-experiments in Michigan [37] and New York [38] showed dramatic improvements in childhood caries prevalence in towns receiving water fluoridation while control towns showed no such difference.

Natural experiments

A natural experiment refers to environmental circumstances that assign individuals in a process that seems largely random to receive or not to receive an intervention. Perhaps the best-known example is from an epidemic: John Snow's comparison of epidemic cholera deaths in houses in London that received their water from a company that had moved its supply upstream in the Thames well before the epidemic to one that was supplied from water in London below the sewer outflow. Snow imagined the experiment to be the equivalent of an RCT:

> *No experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera.… No fewer than 300,000 people, of both sexes,*

> *of every age and occupation, and of every rank and station … were divided into two groups, … one group being supplied with water containing the sewage of London, and amongst it, whatever might have come from a cholera patient, the other having water quite free from such impurity* [39].

The 14-fold difference in death rates found was unlikely due to confounding and provided a key rationale for investing in clean water supply systems in the industrialized world.

More sensible trials

A critical element of RCT design is to match the treatment with understanding of disease pathophysiology. During the COVID-19 pandemic, steroids were tested as anti-inflammatories in patients who were likely suffering from cytokine storm, and they worked. Antivirals used in patients in the same pathophysiologic state were ineffective, which would have been no surprise to clinicians active in the heyday of antiserum therapy [40].

Epidemic RCTs need to be able to adapt to changing information. While the strength of an RCT is consistency in application, if a trial cannot learn from developments revealed by the trial or from changing circumstances in the world in which the participants live, it cannot be helpful during a pandemic. None of the RCTs launched in 2020 adapted to information either from their own trials or from other information.

Adaptive or Bayesian RCTs allow changes in study design resulting from findings – either from the trial itself or from outside the trial – that suggest that the trial should be modified. Accrual processes can be adapted to increase or decrease enrollment in some study arms to favor therapies that seem better or even to omit some treatment arms [41]. During this recent pandemic, RCTs of antibody therapies that inadvisably enrolled very sick patients with advanced disease in studies [42] could have focused their attention earlier on disease after a study of more than 3000 plasma recipients showed a dose–response relationship to antibody content of infused convalescent plasma, but only in patients treated early who were not mechanically ventilated [43].

Concluding remarks

The evidence-based movement, which has done so much to sensitize clinicians to the need for firmer foundations for treatment decisions, has inadvertently created a world in which the broader tool kit of evidence is too often ignored (see Outstanding questions). The RCT is an especially effective tool in studying well-understood pharmaceutical treatments in conditions where the agent is likely to be effective, but the negative RCT is far less informative than the positive RCT because of the many ways in which a trial can fail to be a fair test of the underlying hypothesis.

The pandemic has illustrated the pressing need to integrate data from different sources to make the wisest decisions. Judicious consideration of all available evidence is likely to lead to better policies than reliance in simple yes-or-no fashion on methodologically correct but pathophysiologically flawed or underpowered RCTs [4]. As has been emphasized in observational research in epidemiology, focusing on consistency of findings, the coherence of findings with basic pathophysiology, and the value of observing dose–response relationships can help guide clinical policies [44].

Outstanding questions

How can medicine increase its use of adaptive randomized trials?

How can randomized trials be made more generalizable to the whole population?

Do the human subject protection rules around randomized trials reduce participation by some population groups?

Should meta-analyses formally consider the match of therapy to disease pathophysiology as a criterion for judging randomized trials?

Why do task force and professional society recommendations rely entirely on randomized trial data?

How can trends in vital data and registry data, which are population-wide sources of information, be made better use of in policy recommendations?

How can our education systems inculcate in trainees and students greater use of judgment and avoid the rigidity produced by insistence on one form of evidence?

Declaration of interests

The authors have no interests to declare.

References

1. Pearce, W. *et al.* (2015) Randomised trials in context: practical problems and social aspects of evidence-based medicine and policy. *Trials* 16, 394
2. Naidoo, P. *et al.* (2021) Real-world evidence and product development: opportunities, challenges and risk mitigation. *Wien. Klin. Wochenschr.* 133, 840–846
3. Liu, M. *et al.* (2022) Toward a better understanding about real-world evidence. *Eur. J. Hosp. Pharm. Sci. Pract.* 29, 8–11
4. Paneth, N. (2022) The fossilization of randomized clinical trials. *J. Clin. Invest.* 132, e158499
5. Downing, N.S. *et al.* (2017) Postmarket safety events among novel therapeutics approved by the US Food and Drug Administration between 2001 and 2010. *JAMA* 317, 1854–1863
6. Gyawali, B. *et al.* (2021) Regulatory and clinical consequences of negative confirmatory trials of accelerated approval cancer drugs: retrospective observational study. *BMJ* 374, n1959
7. Benson, K. and Hartz, A.J. (2000) A comparison of observational studies and randomized controlled trials. *N. Engl. J. Med.* 342, 1878–1886
8. Odgaard-Jensen, J. *et al.* (2011) Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst. Rev.* 4, MR000012
9. Streptomycin in Tuberculosis Trials Committee (1948) Streptomycin treatment of pulmonary tuberculosis. *Br. Med. J.* 2, 769–782
10. Francis, T. *et al.* (1957) Evaluation of the 1954 Field Trial of the Poliomyelitis Vaccine: Final Report. Edwards Brothers
11. Renoux, C. *et al.* (2021) Biases in evaluating the safety and effectiveness of drugs for the treatment of COVID-19: designing real-world evidence studies. *Am. J. Epidemiol.* 190, 1452–1456
12. Ioannidis, J.P. (1998) Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 279, 281–286
13. Kent, D.M. and Thaler, D.E. (2008) Stroke prevention – insights from incoherence. *N. Engl. J. Med.* 359, 1287–1289
14. Cartwright, N. (2011) A philosopher's view of the long road from RCTs to effectiveness. *Lancet* 377, 1400–1401
15. Avendano-Sola, C. *et al.* (2021) Convalescent plasma for COVID-19: a multicenter, randomized clinical trial. *J. Clin. Invest.* 131, e152740
16. Beigel, J.H. *et al.* (2020) Remdesivir for the treatment of COVID-19 – final report. *N. Engl. J. Med.* 383, 1813–1826
17. The Recovery Collaborative Group (2021) Dexamethasone in hospitalized patients with COVID-19. *N. Engl. J. Med.* 384, 693–704
18. Klassen, S.A. *et al.* (2021) The effect of convalescent plasma therapy on mortality among patients with COVID-19: systematic review and meta-analysis. *Mayo Clin. Proc.* 96, 1262–1275
19. Gottlieb, R.L. *et al.* (2022) Early remdesivir to prevent progression to severe COVID-19 in outpatients. *N. Engl. J. Med.* 386, 305–315
20. Libster, R. *et al.* (2021) Early high-titer plasma therapy to prevent severe COVID-19 in older adults. *N. Engl. J. Med.* 384, 610–618
21. Sullivan, D.J. *et al.* (2021) Randomized controlled trial of early outpatient COVID-19 treatment with high-titer convalescent plasma. *medRxiv* Published online December 21, 2021. https://doi.org/10.1101/2021.12.10.21267485
22. Dougan, M. *et al.* (2021) Bamlanivimab plus etesevimab in mild or moderate COVID-19. *N. Engl. J. Med.* 385, 1382–1392
23. ACTIV-3/TICO LY-CoV555 Study Group (2021) A neutralizing monoclonal antibody for hospitalized patients with COVID-19. *N. Engl. J. Med.* 384, 905–914
24. Collin, S.M. *et al.* (2008) Prostate-cancer mortality in the USA and UK in 1975-2004: an ecological study. *Lancet Oncol.* 9, 445–452
25. Taylor, R. *et al.* (2019) Breast cancer mortality and screening mammography in New Zealand: incidence-based and aggregate analyses. *J. Med. Screen.* 26, 35–43
26. Hakama, M. *et al.* (1985) Evaluation of screening programmes for gynaecological cancer. *Br. J. Cancer* 52, 669–673
27. Brown, J.J. *et al.* (2021) Decreased colorectal cancer incidence and mortality in a diverse urban population with increased colonoscopy screening. *BMC Public Health* 21, 1280
28. La Vecchia, C. *et al.* (1998) Trends in childhood cancer mortality as indicators of the quality of medical care in the developed world. *Cancer* 83, 2223–2227
29. Buehler, J.W. *et al.* (1987) Birth weight-specific infant mortality, United States, 1960 and 1980. *Public Health Rep.* 102, 151–161
30. Jemal, A. *et al.* (2021) Prostate cancer incidence 5 years after US Preventive Services Task Force recommendations against screening. *J. Natl. Cancer Inst.* 113, 64–71
31. Lafzi, A. *et al.* (2021) Analysis of the effectiveness of face-coverings on the death ratio of COVID-19 using machine learning. *Sci. Rep.* 11, 21675
32. Casadevall, A. *et al.* (2021) Convalescent plasma use in the USA was inversely correlated with COVID-19 mortality. *eLife* 10, e69866
33. Kyriacou, D.N. and Lewis, R.J. (2016) Confounding by indication in clinical research. *JAMA* 316, 1818–1819
34. Salazar, E. *et al.* (2021) Significantly decreased mortality in a large cohort of coronavirus disease 2019 (COVID-19) patients transfused early with convalescent plasma containing high-titer anti-severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein IgG. *Am. J. Pathol.* 191, 90–107
35. Liu, S.T.H. *et al.* (2020) Convalescent plasma treatment of severe COVID-19: a propensity score–matched control study. *Nat. Med.* 26, 1708–1713
36. Arnold Egloff, S.A. *et al.* (2021) Convalescent plasma associates with reduced mortality and improved clinical trajectory in patients hospitalized with COVID-19. *J. Clin. Invest.* 131, e151788
37. Arnold, F.A. *et al.* (1956) Effect of fluoridated public water supplies on dental caries prevalence. *Public Health Rep.* 71, 652–658
38. Ast, D.B. *et al.* (1950) The Newburgh-Kingston Caries Fluorine Study I. Dental findings after three years of water fluoridation. *Am. J. Pub. Health* 40, 716–724
39. Snow, J. (1855) On The Mode of Communication of Cholera, Second Edition, Much Enlarged. Churchill, London, p. 75
40. Cecil, R.L. (1936) Effects of early serum treatment on pneumococcus type I pneumonia. *Trans. Am. Clin. Climatol. Assoc.* 52, 52–63
41. Berry, D.A. (2006) Bayesian clinical trials. *Nat. Rev. Drug Discov.* 5, 27–36
42. RECOVERY Collaborative Group (2021) Convalescent plasma in patients admitted to hospital with COVID-19 (RECOVERY): a randomised controlled, open-label, platform trial. *Lancet* 397, 2049–2059
43. Joyner, M.J. *et al.* (2021) Convalescent plasma antibody levels and the risk of death from COVID-19. *N. Engl. J. Med.* 384, 1015–1027
44. Hill, A.B. (1965) The environment and disease: association or causation? *Proc. R. Soc. Med.* 58, 295–300