SOUNDING BOARD

The Magic of Randomization versus the Myth of Real-World Evidence

Rory Collins, F.R.S., Louise Bowman, M.D., F.R.C.P., Martin Landray, Ph.D., F.R.C.P., and Richard Peto, F.R.S.

Nonrandomized observational analyses of large electronic patient databases are being promoted as an alternative to randomized clinical trials as a source of "real-world evidence" about the efficacy and safety of new and existing treatments.[1-3] For drugs or procedures that are already being used widely, such observational studies may involve exposure of large numbers of patients. Consequently, they have the potential to detect rare adverse effects that cannot plausibly be attributed to bias, generally because the relative risk is large (e.g., Reye's syndrome associated with the use of aspirin, or rhabdomyolysis associated with the use of statin therapy).[4] Nonrandomized clinical observation may also suffice to detect large beneficial effects when good outcomes would not otherwise be expected (e.g., control of diabetic ketoacidosis with insulin treatment, or the rapid shrinking of tumors with chemotherapy).

However, because of the potential biases inherent in observational studies, such studies cannot generally be trusted when — as is often the case — the effects of the treatment of interest are actually null or only moderate (i.e., less than a twofold difference in the incidence of the health outcome between using and not using the treatment).[4-6] In those circumstances, large observational studies may yield misleading associations of a treatment with health outcomes that are statistically significant but noncausal, or that are mistakenly null when the treatment really does have clinically important effects. Instead, randomized, controlled trials of adequate size are generally required to ensure that any moderate benefits or moderate harms of a treatment are assessed reliably enough to guide patient care appropriately (Box 1).[5-7]

Reliance on nonrandomized observational studies risks inadequate assessments of both safety and efficacy because the potential biases with respect to both can be appreciable. For example, the treatment that is being assessed may well have been provided more or less often to patients who had an increased or decreased risk of various health outcomes. Indeed, that is what would be expected in medical practice, since both the severity of the disease being treated and the presence of other conditions may well affect the choice of treatment (often in ways that cannot be reliably quantified). Even when associations of various health outcomes with a particular treatment remain statistically significant after adjustment for all the known differences between patients who received it and those who did not receive it, these adjusted associations may still reflect residual confounding because of differences in factors that were assessed only incompletely or not at all (and therefore could not be taken fully into account in adjusted analyses). Modeling studies indicate that potential biases in observational studies may well be large enough to lead to the false conclusion that a treatment produces benefit or harm, with none of a range of statistical strategies capable of adjusting with certainty for bias. Those findings are consistent with findings from reviews that compared estimates of treatment effects from observational studies with estimates from randomized trials, with examples in which results for the same intervention were similar but also many in which the results were importantly different.[8-12]

Such discrepancies are illustrated by a database analysis involving the entire Danish population that found that the relative risk of death from cancer was 15% lower (95% confidence interval, 13 to 18) among patients who had taken statin therapy for only a few years than among those who had not taken statin therapy, even after statistical adjustment for what was

Box 1. Facilitation of Randomization to Enhance Patient Care and Protect Public Health.

Randomization Provides Evidence about Treatment Effects That Can Be Trusted

Randomization results in groups of patients that are balanced (give or take the play of chance) with respect to their risks of all types of health outcomes. Consequently, in sufficiently large randomized trials, the effects of a treatment can be reliably assessed.

Nonrandomized observational studies may be able to detect large treatment effects. However, the potential biases can be appreciable, so such studies cannot be trusted when the benefits or harms of a treatment are actually null or only moderate.

Obstacles to Randomized Trials Should be Removed to Protect Patients

Increased focus on adherence to rules rather than on the scientific principles that underlie randomized trials has substantially increased the complexity and cost of trials.

Promotion of nonrandomized analyses of databases as a rapid source of "real-world evidence" about the effects of treatments is a false solution to the problems caused by the bureaucratic burdens imposed on randomized trials.

Instead, obstacles to randomized trials should be removed to allow more new treatments to become available and to facilitate the reliable assessment of existing treatments.

known about potential confounding factors.[13] Likewise, in some other nonrandomized studies, statin therapy has been associated with a reduced incidence of cancer (e.g., in one such study, the incidence of colon cancer was about half as high as the incidence among patients not taking a statin).[5] In contrast, in a meta-analysis of individual patient data from randomized trials involving more than 10,000 cases of incident cancer, there were no apparent effects of statins on the incidence of cancer or death from cancer — either overall or at any particular trial site — during an average of 5 years of statin therapy (longer exposure than in the observational studies) or during prolonged follow-up thereafter.[5] Conversely, in contrast to the compelling evidence for the beneficial effects of statins on cardiovascular mortality observed in randomized trials, the incidence of death from cardiovascular causes in the Danish study was approximately one quarter higher among the patients who had taken a statin than among those who had not (presumably because increased risk had led to statin therapy being prescribed). Although this increased incidence of death was reduced after various statistical adjustments were made, the study was still not able to detect the reduction in cardiovascular risk that is known to be produced by statin use.[5]

The "magic" of randomization is that it is guaranteed to result in groups of patients that are balanced (give or take the play of chance) with respect to both known and unknown risk factors (regardless of whether those risk factors have been assessed) and, hence, with respect to their risks of any type of health outcome.[5] Unbiased assessment of the effects of the trial treatment can then be obtained by ensuring that health outcomes are ascertained similarly among the patients randomly assigned to the treatment under investigation and among those who are not. For subjective health outcomes (such as symptoms or mood), this process often needs to be enhanced by masking the treatment assignment (which is not possible in observational studies that make use of clinical databases). Continued follow-up of all the patients included in a randomized trial (even if some of them stop their assigned treatment) maintains the like-with-like comparison produced by randomization (even if the characteristics of the patients who do not adhere to their assigned treatment differ between the randomized groups). Consequently, differences in the incidence of health outcomes between the treatment groups in a randomized trial based on intention-to-treat comparisons can be attributed as causal to the treatment being evaluated (subject to statistical tests that indicate the differences are not likely to be due to chance and the avoidance of unduly data-dependent emphasis on results in selected trials or subgroups within trials[14]).

In generalizing the results of a randomized trial, the assumption is not that the patient population studied is representative of all patients but rather that the proportional effects of the treatment studied on each specific health outcome should be similar in different circumstances, unless there is good reason to expect otherwise.[15] Consequently, valid estimates of the absolute benefits and harms of a treatment can be obtained by applying reliable randomized evidence for its separate proportional effects on each outcome of interest to the absolute incidence of these outcomes in observational studies conducted within a particular population. For example, information from randomized trials of secondary prevention strategies involving patients

Box 2. Opportunities to Improve the Quality and Efficiency of Randomized Trials of New and Existing Interventions.

Appropriate trial guidelines
Based on scientific principles: Focus on issues that can materially affect the reliability of the results (including randomization with concealed assignment, adherence to trial intervention, completeness of follow-up, and intention-to-treat analyses).
Developed in partnership: Create new guidelines that can be adapted for many different types of trials through a collaboration of regulators, investigators, patients, and funders.

Enhanced recruitment
Faster and more predictable: Access electronic health care record systems and specialized registries to identify large numbers of potentially eligible patients.
Broader and more generalizable: Avoid unduly restrictive inclusion and exclusion criteria so that the results are relevant to a wide range of patients.

Improved quality
Better adherence: Implement interactive electronic case-report forms to help ensure complete and consistent data collection and to enhance adherence to the protocol and safety procedures.
Centralized monitoring: Improve patient safety and trial performance through real-time monitoring and analysis of electronic data from local trial sites.

Effective follow-up
Complete and comprehensive: Minimize loss to follow-up and facilitate prolonged follow-up of health outcomes by linkage to electronic health record systems.
Extended range of outcomes: Enhance the assessment of the safety and efficacy of treatment by incorporating technological advances (e.g., smartphones and digital sensors).

at high risk for occlusive vascular events can usefully inform estimates of the effects of primary prevention in a lower-risk general population.[5]

Part of the drive toward using nonrandomized observational studies to assess the effects of treatment comes from the current costs and complexity of conducting randomized trials.[16,17] During the past 25 years, there has been an enormous increase in the rules and related bureaucracy governing clinical trials, with the intention of improving the safety of the participants in trials and the reliability of the results. However, undue focus on adherence to rules (exacerbated by overinterpretation of those rules) rather than on the scientific principles that underlie randomized trials does not necessarily improve either a trial's quality or the patients' safety, but it does increase complexity.[18] As a consequence, pharmaceutical companies have become far more dependent for the conduct of their clinical trials on the contract research organization industry, which has grown exponentially from an annual revenue of approximately $2 billion in the early 1990s to $40 billion in 2019.[19] In parallel, the scientific contribution of academic researchers to industry-funded trials has been reduced, with the previous model of creative partnerships largely replaced by service contracts involving a burgeoning academic research organization industry.

Moreover, the direction of drug development has changed in ways that may adversely affect public health. For example, in the past decade, the revenue from the 10 top-selling drugs in the United States increased by a factor of 2.5, but the patient population that those medications target decreased by a factor of 7.5 (Meanwell C: personal communication). This trend may reflect the current high costs of conducting large randomized trials to detect important incremental effects in common conditions,[7,16,17] leading to a shift toward seeking treatments with larger effects in less common conditions that could be detected in smaller trials. There is also evidence that eligibility criteria are being made more restrictive and the durations of trials are being abbreviated in order to contain costs; both these factors reduce the generalizability and reliability of the evidence about efficacy and safety.[20] However, the solution to the problems caused by the bureaucratic burdens that have been increasingly imposed on randomized trials during the past 25 years is not to replace randomization with unreliable nonrandomized database analyses. Instead, unnecessary obstacles to the reliable assessment of the efficacy and safety of treatments in randomized trials of appropriate size should be removed (Box 2).

One consequence of this bureaucratic burden has been increasing difficulty in recruiting patients into trials, which has resulted in a trend toward small numbers of patients being enrolled at each of hundreds of sites in many countries.[20,21] As a better alternative, rapid recruitment can be

achieved at far fewer sites (with consequent cost reductions and enhanced trial quality) by using electronic patient records, which are increasingly widely available, to identify large numbers of eligible patients. For example, nationwide searches of hospital records in the United Kingdom have been used effectively to identify and recruit eligible patients into a series of randomized trials of cholesterol-modifying treatments,[22,23] and several studies have been conducted within a Swedish registry of patients with heart disease that is now being extended across Europe.[24] In addition, recruitment can be facilitated by avoidance of unduly restrictive or specific eligibility criteria (which often require costly and time-consuming collection and verification of qualifying information); this approach also helps to ensure that results from randomized trials are more widely generalizable to relevant patient populations.[25]

Expensive but relatively ineffective trial monitoring strategies (such as checking source documents and making frequent site visits) are espoused by regulatory guidelines and enforced by many research funders, auditors, and regulatory inspectors. However, rather than focusing on the detection of problems after they have occurred (when it is often too late to rectify them), systems that prevent material deviations should be built into the design of trials.[26] For example, the use of interactive electronic case-report forms can not only help to ensure the collection of all the required data (because items cannot be missed and additional data can be sought when required) and to improve the consistency of the data collected, but also enhance adherence to the trial protocol (e.g., through built-in eligibility checks and prompts when particular actions, such as laboratory safety assays, are required).[23] In addition, real-time electronic transfer of data allows efficient centralized monitoring of patient safety and site performance, with rapid feedback helping to improve performance.[27-30]

Linkage to centralized record systems can be used to enhance the detection of various types of health outcomes, not only during the treatment period of the trial but also in the longer term, yielding a more complete evaluation of a treatment. Assessment of the effects of the trial treatment on these outcomes can often be based directly on the record systems, since the reliability of comparisons between randomized groups is generally not materially improved by adjudication of the recorded health outcomes (which typically involves time-consuming data collection and assessment based on unduly specific definitions).[31-33] For example, in one trial of statin therapy, the randomized comparisons based on outcomes identified retrospectively through health record systems were not materially different from those based on adjudicated outcomes recorded during the trial treatment period; moreover, these comparisons were able to show additional benefits during a prolonged follow-up period.[34] Technological advances are also allowing assessment of an extended range of health outcomes (e.g., smartphone-supported evaluations of quality of life, mood, and cognition, and digital sensors to monitor functional measures).[35]

In summary, the replacement of randomized trials with nonrandomized observational analyses is a false solution to the serious problem of ensuring that patients receive treatments that are both safe and effective. The Clinical Trials Transformation Initiative, which is supported by the Food and Drug Administration, has shown that it is possible to develop guidance that can help improve specific aspects of the design and conduct of randomized trials.[26,30] There is now an urgent need to develop comprehensive guidelines based on the scientific principles underlying randomized, controlled trials that focus on those aspects that really matter for both generating reliable findings and ensuring patient safety, and that take advantage of technological advances to increase the scope of randomized evidence. Such guidelines would be relevant not only for the various phases of clinical development that lead to regulatory approval of new interventions (since reduction of wasteful practices could allow more new treatments to become available) but also for noncommercial randomized trials of existing treatments (since making more such trials affordable could lead to better patient care and improved public health).[36]

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom.

1. Galson S, Simon G. Real-world evidence to guide the approval and use of new treatments. Washington, DC: National Academy of Medicine, 2016 (https://nam.edu/wp-content/uploads/2016/10/Real-World-Evidence-to-Guide-the-Approval-and-Use-of-New-Treatments.pdf).

2. Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials. Clin Pharmacol Ther 2017;102:924-33.
3. National Academies of Sciences, Engineering, and Medicine. Examining the impact of real-world evidence on medical product development: proceedings of a workshop series. Washington, DC: National Academy of Sciences, 2019 (http://nationalacademies.org/HMD/Reports/2019/examining-impact-real-world-evidence-on-medical-product-development-proceedings.aspx).
4. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007;334:349-51.
5. Collins R, Reith C, Emberson J, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet 2016;388:2532-61.
6. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 2018;320:867-8.
7. Eapen ZJ, Lauer MS, Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. JAMA 2014;311:1397-8.
8. Ioannidis JPA, Haidich A-B, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821-30.
9. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:1-173.
10. Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol 2007;166:646-55.
11. Giordano SH, Kuo YF, Duan Z, Hortobagyi GN, Freeman J, Goodwin JS. Limits of observational data in determining outcomes from cancer therapy. Cancer 2008;112:2456-66.
12. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol 2010;63:64-74.
13. Nielsen SF, Nordestgaard BG, Bojesen SE. Statin use and reduced cancer-related mortality. N Engl J Med 2012;367:1792-802.
14. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ 2012;344:e1553.
15. Roberts I, Prieto-Merino D. Applying results from clinical trials: tranexamic acid in trauma patients. J Intensive Care 2014;2:56.
16. Reith C, Landray M, Devereaux PJ, et al. Randomized clinical trials — removing unnecessary obstacles. N Engl J Med 2013;369:1061-5.
17. Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. J Clin Oncol 2010;28:2925-35.
18. Assessment of the functioning of the "Clinical Trials Directive" 2001/20/EC. Public consultation paper. Brussels: European Commission, 2009. (Publication no. ENTR/F/2/SF D(2009) 32674) (https://ec.europa.eu/health/sites/health/files/files/clinicaltrials/docs/2009_10_09_public-consultation-paper.pdf).
19. Analytical Research Cognizance. Global contract research organization (CRO) market 2019 by company, regions, type and application, forecast to 2024. Report ID 296079. March 11, 2019 (http://www.arcognizance.com/report/global-contract-research-organization-cro-market-2019-by-company-regions-type-and-application-forecast-to-2024).
20. Waters DD, Hsue PY. PCSK9 inhibition to reduce cardiovascular risk: tempering expectations. Circ Res 2017;120:1537-9.
21. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. N Engl J Med 2019;380:2295-306.
22. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 2002;360:7-22.
23. The HPS3/TIMI55-REVEAL Collaborative Group. Effects of anacetrapib in patients with atherosclerotic vascular disease. N Engl J Med 2017;377:1217-27.
24. James S, Rao SV, Granger CB. Registry-based randomized clinical trials — a new clinical trial paradigm. Nat Rev Cardiol 2015;12:312-6.
25. Kim ES, Bruinooge SS, Roberts S, et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research joint research statement. J Clin Oncol 2017;35:3737-44.
26. Meeker-O'Connell A, Glessner C, Behm M, et al. Enhancing clinical evidence by proactively building quality into clinical trials. Clin Trials 2016;13:439-44.
27. Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials 2008;5:49-55.
28. Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials 2012;9:257-64.
29. Food and Drug Administration. Oversight of clinical investigations — a risk-based approach to monitoring. Guidance for industry. August 2013 (http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM269919.pdf).
30. Lindblad AS, Manukyan Z, Purohit-Sheth T, et al. Central site monitoring: results from a test of accuracy in identifying trials and sites failing Food and Drug Administration inspection. Clin Trials 2014;11:205-17.
31. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. Clin Trials 2009;6:239-51.
32. Ndounga Diakou LA, Trinquart L, Hróbjartsson A, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. Cochrane Database Syst Rev 2016;3:MR000043.
33. Hlatky MA, Ray RM, Burwen DR, et al. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. Circ Cardiovasc Qual Outcomes 2014;7:157-62.
34. Ford I, Murray H, McCowan C, Packard CJ. Long-term safety and efficacy of lowering low-density lipoprotein cholesterol with statin therapy: 20-year follow-up of West of Scotland Coronary Prevention Study. Circulation 2016;133:1073-80.
35. Herrington WG, Goldsack JC, Landray MJ. Increasing the use of mobile technology-derived endpoints in clinical trials. Clin Trials 2018;15:313-5.
36. Landray MJ, Bax JJ, Alliot L, et al. Improving public health by improving clinical trial guidelines and their application. Eur Heart J 2017;38:1632-7.

DOI: 10.1056/NEJMsb1901642