VIEWPOINT

Fusing Randomized Trials With Big Data
The Key to Self-learning Health Care Systems?

Derek C. Angus, MD, MPH
Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; and Associate Editor, *JAMA*.

Randomized clinical trials (RCTs) have revolutionized medicine by providing evidence on the efficacy and safety of drugs, devices, and procedures. Today, more than 40 000 RCTs are reported annually, their quality continues to increase, and oversight mechanisms ensure adequate protection of participants. However, RCTs have at least 4 related problems: (1) they are too expensive and difficult; (2) their findings are too broad (average treatment effect not representative of benefit for any given individual) and too narrow (trial population and setting not representative of general practice); (3) randomizing patients can make patients and physicians uncomfortable, especially when comparing different types of existing care; and (4) there are often long delays before RCT results diffuse into practice.

The new alternative is "big data." Because medical care is increasingly digitized in electronic health record (EHR) data sets and linked biological and genetic data banks, proponents suggest that health care systems are at the dawn of an era in which a patient's prognosis and optimal therapy will be generated from rapid analysis of these data sets using sophisticated machine-learning strategies. The information is relatively inexpensive, generated as a by-product of patient care (overcoming the cost problem), and both specific to individuals (ie, adequately narrow) and, en masse, descriptive of the entire delivery system (ie, adequately broad). No individuals are randomized, so the ethical issues appear less complex. The richness and immediacy of these new data could allow tailored treatment decisions in real time, overcoming delays in knowledge translation. As such, although the RCT remains the gold standard for evaluation of experimental therapies, big data is proposed as a better approach for the broad swath of comparative effectiveness questions that arise in clinical practice. Indeed, the Institute of Medicine envisions big data as the engine for so-called learning health care systems.[1]

Yet big data is not a replacement for the RCT. The singular beauty of the RCT is the strength of causal inference that arises from random assignment. All known and unknown factors that can influence outcome, other than the treatment assignment, are distributed randomly—and, if randomization is effective, evenly—among the groups. Thus, any statistical difference in outcome between groups can be attributed to the assigned interventions. In contrast, big data exploits variation that may appear random but is not in fact randomized. For example, big data may compare patients who did or did not receive a particular drug and then attempt to determine the extent to which any difference in patient outcome may be related to receiving the drug. However, for all the allure of inexpensive access to massive amounts of data, the Achilles' heel is lack of causal inference. No matter how detailed the measurement and how sophisticated the adjustment for all known variables, big data cannot eliminate unmeasured factors coincident with a particular treatment assignment that could explain an apparent change in outcome.[2]

Thus, each approach has complementary strengths: RCTs offer causal inference, and big data offers the potential for low-cost, high-volume, nuanced answers with immediate feedback. Rather than debate which is better, the greatest promise may come from fusing them.

Fixing Problem 1—Cost and Difficulty

Conducting RCTs as freestanding enterprises requires considerable infrastructure, much of which is duplicative with clinical care. Many argue better integration with the EHR could reduce RCT costs and overcome some logistic difficulties. For example, EHR screening tools are often used to identify potential patients. For patients enrolled in RCTs, the burden of data collection is frequently eased by automated download from the EHR. The interventions being tested also could be nested in the EHR, including both the order entry for the intervention and the recording of intervention delivery. There are also examples of RCTs of diagnostic tests whereby patients are randomly assigned to have test results presented in the EHR to their physicians (NCT02130986). One simple integration is to randomize by cluster (eg, a hospital ward) and capture all data from the EHR.[3]

However, full integration of patient-level randomization has remained relatively elusive. Fiore et al[4] reported one of the first examples of so-called point-of-care clinical trials. Although modest in scale, this study demonstrated physicians could be prompted by the EHR to consider enrolling their patients in an RCT. If the physician and patient agreed, the EHR contained the structure both to record enrollment and to generate random treatment assignment. There is desire to accelerate this integration, especially in federally funded efforts, such as the National Institutes of Health Clinical and Translational Science Awards program. Nonetheless, the barriers are not trivial and include data system integration and security, protections for research participants, and adequate oversight. Perhaps the most important challenge is that the individuals providing care at the bedside, and those responsible for maintaining an EHR that supports clinical care, are neither compensated for nor comfortable with participation in clinical research. Thus, true success may only come when health care delivery systems are adequately motivated to answer research questions.

Corresponding Author: Derek C. Angus, MD, MPH, Department of Critical Care Medicine, University of Pittsburgh, 3550 Terrace Street, 614 Scaife Hall, Pittsburgh, PA 15261 (angusdc@upmc.edu).

Fixing Problem 2—Too Broad and Too Narrow Answers

Nesting RCTs in the EHR to lower cost is not enough; they must also be smarter. In the era of precision medicine, there is a desire for estimates of treatment effects for individual patients, rather than the "average" effect of treatment A vs B. Answers also are needed not only for treatment A vs B but for A vs B vs C, conditional on exposure to D, E, or F. Adaptive trials, which incorporate rules to adjust aspects of the trial during enrollment, are well suited for such questions. For example, enrichment strategies allow entry criteria to change if the trial "learns" that certain patients, such as those identified by genetic or molecular markers, are better suited to certain therapies. However, these designs have been used mainly in phase 2 trials.

An extension of adaptive trials, known as "platform" trials, expands the scope to focus not on a particular intervention but on an entire disease or syndrome, comparing multiple interventions, even within different domains of treatment, adding or dropping interventions over time.[5] For example, a large European study will target all patients admitted to intensive care with severe pneumonia.[6] The study will simultaneously test multiple antibiotic regimens, host immunomodulation, and mechanical ventilation strategies and is designed to generate separate estimates of benefit conditional on both exposure to the different combinations of treatments and different phenotypes, such as those with and without shock. However, extending this model further may be necessary. The best mechanism to generate generalizable estimates of treatment effect (ie, broader answers) and understand heterogeneity of treatment effect (ie, narrower answers) is to have extremely broad enrollment, ideally representative of the entire breadth of a disease or condition.[7] With this approach, embedding the RCT in the EHR is not only efficient but essential to meet the scientific goals.

Fixing Problem 3—Discomfort With Randomization

When fully actualized, big data could continuously analyze prior EHR data to provide bedside probabilities of benefit associated with different therapeutic options. For example, a big-data solution could be an EHR prompt or decision aid that says, "based on analysis of similar patients, there is a 70% probability that this patient will fare better with treatment A vs treatment B." Understandably, this approach may be perceived as more reassuring and safer than asking the patient to be randomized 50:50 to treatment A vs B in an RCT, even though the uncertainty (70%, from observational data) is considerable.

However, the RCT does not need to randomize 50:50. Instead, it can use the adaptive trial technique known as response-adaptive randomization, which allows the random allocation ratios to change over time based on accruing information. Thus, although individuals may still be assigned to poorly faring treatment groups, their odds are reduced over time. Therefore, faced with the same odds as above, the RCT would metaphorically toss a weighted coin with a 70% probability of landing on treatment A. Response-adaptive randomization can be tailored to specific subgroups of patients, again similar to the promise of big data, such that the probability of receiving a given therapy is based on the odds of success seen in similar patients. In other words, this approach would mimic some of the purported benefits of tailored big-data decision support, yet do so on randomized evidence and therefore with stronger causal inference. When testing therapies for which the conventional wisdom is equipoise and exposure outside the trial is 50:50, the patient, on average, is safer participating in the trial than not.

Fixing Problem 4—Sluggish Knowledge Translation

The grandest leap would be to fuse all these elements in a new kind of RCT that can be called a randomized, embedded, multifactorial, adaptive platform (REMAP) trial. A REMAP trial emulates many of the strengths of big data while retaining random assignment and, used in a health care system to address comparative effectiveness questions, would function not only as a research study but also as a continuous quality-improvement program. For example, if a health care system enrolled all willing patients with severe pneumonia into a pneumonia REMAP trial, then, de facto, the patients would be preferentially assigned to the best-performing treatment regimen, consistent with the principles of continuous quality improvement. Thus, the health care system could be self-learning, learning would be based on strong causal inference, and patients could be cared for with increasing odds of benefit and reduced odds of harm as knowledge is accrued. In other words, knowledge could be generated and translated simultaneously, eliminating the knowledge-translation gap for any participating health care system.

Although the REMAP trial model is aspirational, most of its challenges involve the existing components: conducting RCTs, incorporating adaptive designs, and implementing the promise of big data. The most significant challenge is the EHR. A major effort is still required to ensure data are accurate, patients are adequately phenotyped, and information is accessible in a timely fashion. In addition, REMAP trials will require novel partnerships of trialists, big-data experts, health care systems, clinicians, and patients. Adequate protection of human research participants is essential. But beyond that, it is possible that patients, informed of the inherent uncertainty that underpins many clinical decisions, may drive this change. Similarly, government and commercial insurers might appreciate that this fusion could serve their beneficiaries and consider incentive schemes for health care system participation. Perhaps the greatest argument for such a program is consideration of the alternative: failing to fully leverage the power of the EHR by restricting the use of the strongest instrument for inferring causality—randomization—to a tiny proportion of the vast majority of clinical treatment decisions that continue to be made under considerable uncertainty.

ARTICLE INFORMATION

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Grumbach K, Lucey CR, Johnston SC. Transforming from centers of learning to learning health systems: the challenge for academic health centers. *JAMA*. 2014;311(11):1109-1110.

2. Titiunik R. Can big data solve the fundamental problem of causal inference? *Pol Sci Politics*. 2015; 48(1):75-79.

3. Noto MJ, Domenico HJ, Byrne DW, et al. Chlorhexidine bathing and health care–associated infections. *JAMA*. 2015;313(4):369-378.

4. Fiore LD, Brophy M, Ferguson RE, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials*. 2011;8(2):183-195.

5. Berry SM, Connor JT, Lewis RJ. The platform trial. *JAMA*. 2015;313(16):1619-1620.

6. Platform for European Preparedness Against (Re-)emerging Epidemics (PREPARE). PRACTICE C: Adaptive Randomised Trial SARI—Workpackage 5. PREPARE website. http://www.prepare-europe.eu/About-us/Workpackages/Workpackage-5. Accessed June 15, 2015.

7. Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat Med*. 1999;18 (12):1467-1474.