

# The Promise and Perils of Big Data in Healthcare

Austin B. Frakt, PhD, and Steven D. Pizer, PhD

**S**purred by the accuracy with which companies like Google and Netflix use large amounts of data to anticipate our interests, there is growing investment in Big Data applications to healthcare. For example, the FDA's postmarket surveillance program—the Sentinel Initiative—analyzes billions of drug prescriptions for adverse events.<sup>1</sup> However, if we're not careful, big data healthcare could cause harm.

Data are “big” if they either: 1) represent many more subjects than a typical randomized clinical trial (RCT)—tens of thousands or more—and/or 2) they include a broad range—hundreds or more—of clinically relevant patient and provider characteristics. Such data can extend the reach of clinical research to include study of rare events, heterogeneous treatment effects, long-term outcomes, and other topics difficult or impossible to study with RCTs.

Some have suggested that big data will rapidly improve healthcare delivery. For instance, finding insufficient guidance in the medical literature, physicians at Stanford's Packard Children's Hospital used electronic medical record data to help make an urgent decision about using anticoagulation medication in a lupus patient.<sup>2</sup> The strongest proponents of such big data applications believe that with enough information, causal relationships reveal themselves without an RCT.

Are they right? For clinical applications, this is a vital question. For instance, for every 5 million packages of x-ray contrast media distributed to healthcare facilities, about 6 individuals die from adverse effects.<sup>3</sup> With big data, we learn that such deaths are highly correlated with electrical engineering doctorates awarded, precipitation in Nebraska, and per capita mozzarella cheese consumption (correlations 0.75, 0.85, and 0.74, respectively).

However, because we cannot conceive of a causal mechanism, it is obvious that these variables play no causal role in x-ray contrast media deaths. That such high correlations can be easily mined from big data is concern-

ing nonetheless, because it is not always trivial to assess whether they are telling us something useful. For example, observational data reveal that proton pump inhibitor (PPI) use is associated with pneumonia incidence.<sup>4</sup> This could be causal because a mechanism is plausible—gastric acid reduction could increase bacterial colonization—but perhaps the association arises because other factors drive both PPI use and pneumonia incidence.

Faced with this kind of uncertainty, there is temptation to insist that only an RCT can convince us of causation; however, the very promise of big data is its potential to see what RCTs won't, thereby improving care in ways that RCTs cannot. Although caution is warranted, we should not dismiss big data too quickly. The way forward requires careful selection of observational research designs coupled with rigorous testing for violations of key assumptions on which causal inference relies.

A useful technique for testing a big data finding—like whether PPIs increase the risk of pneumonia—is to probe for other factors that might be driving it, called a “falsification test.” A falsification test begins with the question, “If something else is driving this observational finding, where else would it show up?” It explores the association of the suspected causal variable (eg, PPI use) with other outcomes (eg, not pneumonia) or in other populations that we have good reason to believe should not be affected by it. The key is to select other outcomes or populations that are likely to also be affected by factors that could be driving the suspected causal relationship.<sup>5</sup> For example, Prasad and Jena<sup>4</sup> raise doubts that the PPI-pneumonia relationship is causal by showing that its use is also associated with outcomes for which no clear causal mechanism exists: chest pain, urinary tract infections, osteoarthritis, rheumatoid arthritis flares, and deep venous thrombosis.

Prentice et al<sup>6</sup> applied big data and falsification tests to analyze outcomes from use of common second-line medi-

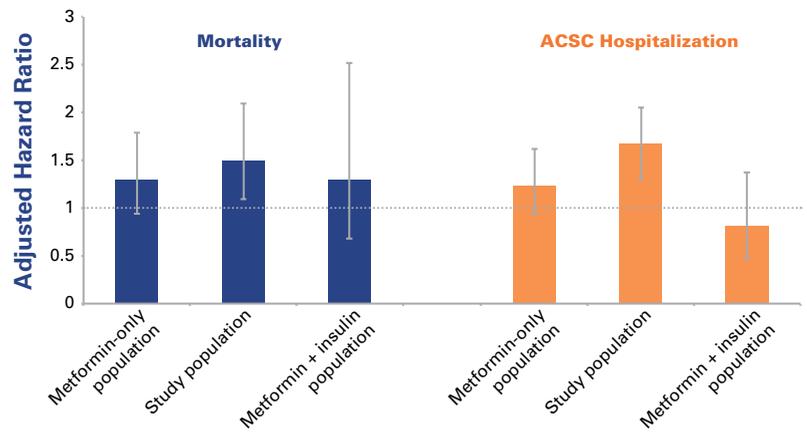
cations for type 2 diabetes: sulfonylureas (SUs) and thiazolidinediones (TZDs). Using 10 years of Veterans Health Administration data merged with Medicare claims, they analyzed over 80,000 patients for mortality and hospitalization outcomes—end points for which no prior RCT was adequately powered. The study found that SUs are associated with a 50% increase in risk of mortality and a 68% increase in avoidable hospitalizations compared with TZDs. Study findings were based on variation in the relative rates with which physicians prescribe the 2 medications. Can we interpret them as causal? Put another way, are prescribing patterns random in a way akin to an RCT’s group assignment?

To probe the assumption that they are, the authors applied several falsification tests. First, if prescribing patterns were random, like an RCT’s group assignment, observable patient and provider characteristics should be balanced across groups stratified by high and low rates of physician prescribing. If factors aren’t balanced—whether in an RCT or in an observational study like this—one should suspect flawed randomization and reject a causal interpretation. Prentice et al observed such balance in demographics, diagnoses, and provider-quality variables, increasing confidence that prescribing pattern served a valid randomizing role.

Next, they also conducted falsification tests using 2 populations that were similar to the study population in every way except disease severity: a healthier population taking only metformin and a sicker population that had been prescribed metformin and then insulin. For neither of these populations was the SU/TZD prescribing pattern statistically significantly related to outcomes (see the **Figure**<sup>6</sup>). This could not have been the case had an omitted factor been the missing link in the association of SUs with bad outcomes. These falsification test results further increase confidence that the associations of SUs with greater mortality and avoidable hospitalization risks are causal.

Big data can be useful in healthcare; they can expand the reach of evidence-based medicine into domains not accessible with RCTs, including postmarket surveillance of drugs. But to fulfill that ambition, big data must be coupled with rigorous observational methods. Falsification tests can illuminate when a relationship is less likely to be causal, potentially sparing practitioners from making grave mistakes. Without them, we run the risk of letting our enthusiasm about big data get ahead of the science and what is best for patients.

**Figure.** Effects of SU/TZD Prescribing Patterns on Mortality and Hospitalization for Patients in 3 Populations (adjusted hazard ratios)<sup>6</sup>



ACSC indicates ambulatory care-sensitive condition; SU, sulfonylurea; TZD, thiazolidinedione. Lines indicate 95% CI.

**Author Affiliations:** VA Boston Healthcare System, Boston University, and Harvard University (ABF), Boston, MA; Northeastern University (SDP), Boston, MA.

**Source of Funding:** This work supported by Health Services Research and Development Service of the US Department of Veterans Affairs (grant no. IIR-10-136), which did not in any way participate in the writing of or exert any influence over the content of this manuscript.

**Author Disclosures:** The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article. The contents do not represent the views of the US Department of Veterans Affairs, Boston University, Harvard University, or Northeastern University.

**Authorship Information:** Concept and design (ABF, SDP); acquisition of data (ABF, SDP); analysis and interpretation of data (ABF, SDP); drafting of the manuscript (ABF, SDP); critical revision of the manuscript for important intellectual content (ABF, SDP); obtaining funding (ABF, SDP).

**Address correspondence to:** Austin B. Frakt, PhD, VA Boston Healthcare System, 150 S Huntington Ave (152H), Boston, MA 20130. E-mail: frakt@bu.edu.

## REFERENCES

- Findlay S. Health policy brief: the FDA’s Sentinel Initiative. *Health Affairs* website. [http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief\\_id=139](http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=139). Published June 4, 2015. Accessed June 8, 2015.
- Frankovich J, Longhurst CA, Sutherland SM. Evidence-based medicine in the EMR era. *N Engl J Med*. 2011;365(19):1758-1759. doi:10.1056/NEJMp1108726.
- Wysowski DK, Nourjah P. Deaths attributed to x-ray contrast media on US death certificates. *AJR Am J Roentgenol*. 2006;186(3):613-615.
- Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA*. 2013;309(3):241-242. doi:10.1001/jama.2012.96867.
- Pizer SD. Falsification testing of instrumental variables methods for comparative effectiveness research [published online August 21, 2015]. *Health Serv Res*. doi:10.1111/1475-6773.12355.
- Prentice JC, Conlin PR, Gellad WF, Edelman D, Lee TA, Pizer SD. Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Value Health*. 2014;17(8):854-862. doi:10.1016/j.jval.2014.08.2674. ■