The Ongoing Debate about the Merits of RCTs versus Observational Studies

William H. Crown, PhD, 2013-2014 ISPOR President and Chief Scientific Officer, Optum Labs, Cambridge, MA, USA

Few topics engender stronger responses from researchers than their opinions about whether randomization is necessary in order to draw causal inferences from observational data. In one camp are those who argue that, due to the many potential sources of bias in observational studies, the best one can hope for is to draw inferences about correlation, rather than causation. Few would disagree with their conclusion that because randomization distributes both observed and unobserved characteristics of patients across treatment groups (assuming the sample size is sufficient), it is the strongest research design for the unbiased estimation of treatment effects.

On the other hand, even though researchers in the observational data camp would readily agree that randomization is the strongest research design, many would assert that it does not necessarily follow that inferences about causality are impossible with observational studies. Such studies offer the opportunity to generate timely evidence on many research questions provided the data are sufficiently complete and well-measured, and appropriate research design and statistical methods are used.

Leaving the issue of causality aside for the moment, there is general agreement in the research community that RCTs generate estimates of treatment efficacy while observational studies generate estimates of effectiveness. Beyond this one point of common ground, however, there is a chasm between those who advocate strongly for RCTs versus those who feel that observational studies can often provide reliable evidence less expensively and in a more timely fashion.

Estimates of treatment effects from randomized trials often differ substantially from those generated from observational studies; this is no surprise for a variety of reasons. Indeed, why should we expect estimates of treatment efficacy in carefully controlled settings to be the same as estimates of treatment effectiveness? RCTs, especially registration trials, typically have stringent inclusion and exclusion criteria that place limits on comorbid conditions, concomitant drugs, etc. And trial protocols typically strive to make sure that the patients are actually taking the medication being studied. In contrast, once approved for marketing, treatments are generally prescribed to a broader population than that studied in the registration trials and there is no one looking over the shoulder of the patient to make sure that they even take the medication as prescribed. In light of these factors it seems unlikely that the results from randomized and observational studies should be expected to be similar. Consequently, lack of agreement with RCTs hardly seems to be an indictment of the value of observational studies. On the other hand, when different results are found between observational studies and RCTs, we often have little evidence about why this is the case.

Observational studies are subject to the risk of bias from a host of sources. In particular, bias can be introduced by a variety of different types of measurement error—the most prevalent probably being missing or poorly-measured variables. A common example of this are studies using medical claims databases to estimate treatment effects for conditions where controls for clinical severity influence both treatment selection and outcomes. For instance, lack of data on HER-2 status, metastasis status, and cancer stage for studies of breast cancer treatment and outcomes would likely result in biased estimates of treatment effects. It is important to make the distinction, however, between data availability and observational versus randomized designs. As data availability continues to improve, the ability to include relevant control variables will also improve. From a statistical standpoint, an observational model containing a complete set of well-measured control variables should result in similar treatment effect estimates to those of an RCT designed to evaluate the same treatment in a similar patient population.

Of course, even randomized trials of particular interventions conducted on similar patient populations are not always consistent. We would not expect them to be; this is why we have the concept of sampling distributions in statistics! One trial may be drawn from a sample that generates a treatment effect in the right hand side of the sampling distribution, another in the left hand side, and a third right in the middle. The same is, of course, true with the estimates obtained from observational studies. Because we cannot expect consistency even within a given method, we certainly should not expect any particular observational study to yield results similar to any particular RCT. The best we could hope for would be similarity in the means of the respective sampling distributions.

This suggests that we should be thinking about comparing results from meta analyses of similar observational studies and RCTs if we hope to draw conclusions about the similarity or difference in results associated with the two major types of designs. Several studies using this approach have, indeed, found that the treatment effects obtained from RCTs and observational studies on similar samples tend to be very similar. Ioannidis et al. [1] compared the results from 240 randomized trials and 168 observational studies spanning 45 different medical treatments and found a strong correlation among the study results (R=0.75, p<0.001). Similarly, in an analysis of 99 study reports across 5 disease areas over 1991-1995, Concato, Shah, and Horwitz [2] concluded that the results from the observational studies and clinical trials compared were “remarkably similar.” In fact, they found higher variation in the treatment estimates associated with the randomized studies than the observational studies. Similarly, Benson and Hart [3] found “little evidence that estimates of treatment effects in observational studies reported after 1984 are either consistently larger or qualitatively different than those obtained in randomized, controlled trials.”

Despite evidence that RCTs and well-conducted observational studies often lead to similar conclusions, Hannan [4] points out that there are numerous reasons why conclusions from observational studies may diverge from those of randomized trials. Perhaps the most important of these reasons is selection bias—the bias introduced by unobserved patient characteristics that are correlated with both treatment and outcomes. Most of the multivariate methods for estimating treatment effects focus upon the observed variables but
economists, in particular, have developed methods that attempt to test for missing variables or other sources of measurement error. The method of instrumental variables is one of the best known of these methods but it is often difficult to implement because it requires the identification of variables that are correlated with treatment but uncorrelated with outcomes. This generally turns out to be a difficult task—leading to the conclusion that the researcher can sometimes make the bias problem worse by trying to repair it with instrumental variables versus acknowledging the problem and leaving it uncorrected [5].

LARGE SIMPLE TRIALS
There will always be a role for RCTs in regulatory decision making for product approvals but what is the potential for using randomization in real world effectiveness studies? The Institute of Medicine recently held a workshop on Large Simple Trials and Knowledge Generation in a Learning Health System [6]. As this workshop highlighted, the rapid growth in adoption of electronic systems by health care providers, particularly electronic medical record systems, offers the potential to substantially reduce the cost of data collection. For example, the systems can often be very helpful in identifying potential study subjects. Randomization itself is a simple matter. Thus, if the outcome of interest is captured in the data, the only drawback to randomization versus purely observational studies is typically the need for IRB approval of the intervention (if it involves direct patient care) as well as the elapsed time waiting for the data to accumulate post-randomization. Of course, there will still be studies where there is a need for primary data collection in the form of clinical samples or patient reported outcomes. Nevertheless, conducting studies with the aid of a beginning data infrastructure is potentially a great research accelerator and cost reducer.

Virtually all scientists would agree that, everything else equal, randomized study designs are stronger than comparable observational study designs. But it does not follow that results from observational studies are destined to be unreliable. Good observational studies are really not that difficult to identify. To have confidence in the results from an observational study three things are needed: 1) strong measurement on all of the relevant variables necessary for a particular research question; 2) pre/post data for both a treatment and control group; and 3) appropriate statistical methodology. When all of these elements are present, many of the challenges that arise in observational studies can be dealt with effectively.

Health care stakeholders need evidence to make decisions that range from coverage and reimbursement of new drugs to point-of-care clinical decisions on the part of physicians. Meta analyses of well-designed, large, randomized clinical trials provide the highest level of evidence for the questions they are intended to answer. But they are not very informative about how a therapy may work in real world patient populations that are usually very different from those studied in registration trials. Moreover, we simply do not have the time or the money to answer all questions with clinical trials. By necessity, observational research will be the source of much of the evidence with respect to the effectiveness, safety, and value of alternative therapies. As the quality of observational data continues to improve, so too will the quality of observational studies themselves. Ongoing comparisons of results from observational and randomized studies evaluating comparable treatment effects in similar patient populations will continue to be helpful in informing where we can productively rely upon evidence from observational studies versus where we should use randomization to address the remaining challenging issues of confounding.

REFERENCES