EDITORIAL COMMENT

The Hazards of Comparative Effectiveness When We Cannot Effectively Compare*

Richard A. Lange, MD, L. David Hillis, MD
San Antonio, Texas

According to the Agency for Health Care Research and Quality, “comparative effectiveness research is designed to inform health-care decisions by providing evidence on the effectiveness, benefits, and harms of different treatment options” (1). The data derived from clinical registries play an important role in assessing the effectiveness and appropriateness of various treatments, and their use is likely to grow in light of the increasing focus on assessing and improving healthcare delivery and performing comparative effectiveness studies. Such clinical registries capture data that reflect “real-world” clinical practice in large patient populations (2).

Whereas most randomized controlled trials examine treatment effects in highly selected patients who are enrolled and studied at a limited number of sites and followed for a relatively short time, observational clinical registries often include data from larger numbers of patients with more heterogeneous characteristics, many of whom are traditionally underrepresented in randomized trials (i.e., women, elderly, minorities, patients with various comorbid conditions, and so on); these subjects often are gathered from a large number of sites and are followed for a relatively long duration. Although the data from observational clinical registries are not designed to replace those that are obtained from randomized controlled trials, they often offer insights and information that are complementary to the results of such randomized trials.

The major limitation of data obtained from observational clinical registries is that differences in baseline characteristics that are often present because the patients are not assigned randomly to a particular treatment (i.e., known or unknown confounding factors) may affect the treatment choice, leading to treatment selection bias. Accordingly, several statistical tools, such as multivariate adjustments based on regression analysis, stratification by background characteristics, matching techniques, and propensity scoring, are utilized in an attempt to “balance” these baseline characteristics (e.g., biases) in the nonrandomized subjects. Although the statistical methodologies that are employed in an attempt to overcome these potential biases have become more robust, they nonetheless continue to be limited by a reliance on measured variables, which are typically traditional risk factors. In this issue of JACC: Cardiovascular Interventions, McNulty et al. (3) draw attention to the consequences of this limitation in assessing the comparative effectiveness of percutaneous coronary intervention (PCI) versus coronary artery bypass graft surgery (CABG) in patients with unprotected left main (ULM) coronary artery disease (CAD).

By carefully scrutinizing the records of 101 consecutive patients with ULM CAD who underwent nonemergent PCI at a California Kaiser Permanente Hospital, the investigators assessed how often these subjects were referred for PCI because they were thought to be ineligible for CABG and whether the conditions that made them ineligible for CABG were identified by the American College of Cardiology–National Cardiovascular Data Registry (ACC–NCDR) Cath-PCI Registry. This registry contains detailed information on patient and hospital characteristics, coronary angiographic findings, and PCI and in-hospital outcomes on >8 million subjects from >1,000 sites within the United States (roughly 60% of American cardiac catheterization laboratories); as such, it is the largest national registry of diagnostic cardiac catheterization and PCI data and, as a result, is a commonly utilized source of data for comparative and appropriate use studies (4,5).

In the study of McNulty et al. (3), 54% of the patients undergoing nonemergent PCI for ULM CAD were deemed to be ineligible for CABG. Not surprisingly, these “CABG-ineligible” patients had a greater prevalence of traditionally measured risk factors and a higher estimated CABG–associated mortality (using STS scores, NCDR Cath-PCI risk scores, and EuroScores) than those who were thought to be eligible for CABG. Importantly, most of the clinical conditions cited as reasons for CABG ineligibility—such as active malignancy, frailty, dementia, severe aortic calcification, or poor bypass targets or conduits—were not captured by the ACC–NCDR registry. One or more of these conditions were present in 78% of the CABG-ineligible patients and were independent predictors of mortality. After adjustment using the standard risk scores, patients with 1 or more of these “non–ACC–NCDR-captured” conditions were 5 to 6 times more likely to die during the 1.5 years of follow-up when compared with those without these conditions.

*Editorials published in the JACC: Cardiovascular Interventions reflect the views of the authors and do not necessarily represent the views of JACC: Cardiovascular Interventions or the American College of Cardiology.

From the Department of Medicine, University of Texas Health Science Center, San Antonio, Texas. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.
What are the clinical implications of this study? Most importantly, it highlights the potential problems of using observational registry data (which may have been intended for providing information regarding healthcare delivery processes, quality, and outcomes) for “comparative effectiveness” research. The unmeasured clinical conditions identified in the careful analysis of the 101 patients—as well as others that may not have been identified in this study—represent a substantial source of residual confounding for studies comparing modes of revascularization that rely on ACC–NCDR data for statistical adjustment of differences in patient populations. A similar analysis of a much larger population would likely reveal additional conditions that are not currently being assessed, yet they contribute to the treatment outcome. The study also points out that registry data often provide limited insight into why treatments that appear to be contrary to guidelines or appropriate use criteria are selected.

The strengths of the study are the large number of predictor variables analyzed by the investigators and the fact that it was conducted at an integrated healthcare institution (Kaiser), where decisions regarding the more appropriate mode of coronary revascularization are not influenced by economic incentives for physicians. The study has few limitations. Although only patients with ULM CAD referred for nonemergent PCI were studied, it is likely that similar treatment bias occurs in patients with multivessel CAD without left main involvement (i.e., those with malignancy, frailty, dementia, severe aortic calcification, or poor bypass targets or conduits are referred for PCI rather than CABG). Lastly, the determination and reason(s) for CABG ineligibility were assessed retrospectively, not prospectively. Nonetheless, the determination of “CABG ineligible” was based on the assessment of documentation by the treating clinicians, not on an independent assessment of eligibility for CABG by the chart reviewer.

How should the results of this study influence our current practices? Standard definitions and criteria for diagnosing “non–ACC–NCDR risk factors” (i.e., dementia, severe aortic calcification, end-stage liver disease, cachexia/frailty, and so on) need to be established. These risk factors—and those identified in the future—should be incorporated into data collection and statistical models that “correct” for baseline differences in patient populations (i.e., propensity score analysis). When comparative effectiveness studies utilize registry data, they should be scrutinized cautiously for conditions that may potentially confound the findings. Finally, the results of comparative effectiveness studies derived from observational clinical registries should be confirmed with additional registries that collect complementary data, when available. In short, basing healthcare decisions on the results of comparative effectiveness studies is hazardous when the treatment groups are not (or cannot be) effectively compared.

Reprint requests and correspondence: Dr. Richard A. Lange, Department of Medicine, University of Texas Health Science Center, 7703 Floyd Curl Drive, MC 7870, San Antonio, Texas 78229-3901. E-mail: lambera@uthscsa.edu.

REFERENCES


Key Words: comparative effectiveness • coronary bypass • left main disease • risk factors • stents.