Choosing Methods to Minimize Confounding in Observational Studies
Do the Ends Justify the Means?
Robert W. Yeh, MD, MSc; Laura Mauri, MD, MSc

The comparison of alternative treatments has long been a primary objective of researchers engaged in clinical investigation. However, with the commitment of $1.1 billion in support of clinical research in the American Recovery and Reinvestment Act,1 followed subsequently by the creation of the Patient Centered Outcomes Research Institute (PCORI) as part of the Affordable Care Act,2 “comparative effectiveness research” has grown to become a publicly discussed national priority.

Such comparisons have traditionally relied heavily on the randomized, clinical trial—the gold standard for comparing two treatments. However, because of limitations in generalizability, challenges with feasibility, and typically high costs of executing such trials, coupled with the increasing availability of prospectively captured clinical and administrative data, comparative effectiveness studies using observational study designs have offered an attractive alternative, harnessing the ability of various statistical approaches to overcome nonrandomized treatment allocation. It is in this context that Venkitachalam and colleagues provide a cautionary example illustrating the challenges and limitations of these approaches.

Mortality Benefit and Drug-Eluting Stents: Fact or Artifact?
The authors are motivated by a consistent discrepancy seen in results of randomized trials and observational studies comparing drug-eluting stents (DES) with bare metal stents (BMS) for percutaneous coronary intervention.3 Observational studies comparing DES and BMS have typically shown a mortality difference favoring DES, findings not corroborated in randomized trials. The authors hypothesize that the discrepancy may be rooted in the incomplete adjustment for unmeasured variables that compel physicians to select a certain stent type but may also be associated with mortality—a phenomenon referred to variously as treatment selection bias or confounding by indication.4,5 They recapitulate a new set of DES versus BMS comparisons within the EVENT Registry, by first applying commonly used multivariable regression, then propensity-score matching, and finally, instrumental variable analysis. Similar to what has been observed previously, multivariable regression and propensity score matching showed significantly lower mortality with DES compared with BMS at 1 year; however, an instrumental variable approach showed no significant difference in mortality, consistent with randomized trial results.

Reliability of Observational Data Analysis to Compare Treatments
This study raises broad questions regarding how best to identify and implement analytical methods that would allow observational data to be used to reliably compare treatment strategies, a fundamental question in the arena of comparative effectiveness research. Both multivariable regression and propensity score methods rely on the adequacy of the dataset to capture important clinical characteristics that may confound treatment comparisons.6 The determination of whether these variables have been captured is ultimately the domain of the informed clinician, who best understands the forces that compel the selection of one stent type versus another, and whether these factors are important prognostically. “Will my patient be compliant with dual antiplatelet therapy or are there barriers to obtaining medications?” “Is she expecting an upcoming surgery?” “Does she have a history of significant bleeding?” In a period encompassing a time of enthusiastic use of DES, followed by more tempered use because of concerns about stent thrombosis, these questions were among the various clinical and nonclinical factors that determined stent selection. Importantly, these factors are also likely to be prognostically important variables that were unaccounted for in EVENT and may have only indirectly been accounted for in population-based stent comparisons.7,8 Moreover, in this dataset, in which only 12% of patients received BMS and in which BMS use was associated with more than a 2-fold unadjusted increase in mortality at 1 year, there probably was sufficient prima facie evidence to have suspected that a multivariable regression or propensity score approach would be inadequate and suffer from significant residual confounding inherent to the limitations of the data available.

Instrumental Variable Analysis and Assumptions
Instrumental variables have been used by econometricians, social scientists, and clinical investigators to overcome this type of confounding. By definition, an instrumental variable

---

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA (R.W.Y.); and the Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA (L.M.).

Correspondence to Laura Mauri, MD, MSc, 75 Francis St, Boston, MA 02115. E-mail lmauri1@partners.org

(Circ Cardiovasc Qual Outcomes. 2011;4:581-583.)

© 2011 American Heart Association, Inc.

Circ Cardiovasc Qual Outcomes is available at http://circoutcomes.ahajournals.org

DOI: 10.1161/CIRCOUTCOMES.111.963538
is one that is associated with the predictor of interest—in this case, the selection of stent type—but is not expected to be associated with the outcome apart from its relationship with that predictor.9 In this study, the authors took advantage of the decline in DES use between 2004 and 2007 to develop their instrument, namely, time. To be useful, time should be a strong predictor of DES use and should not in and of itself be associated with mortality, an assumption that is generally considered an untestable hypothesis.10 Whether both of these criteria were met in the selection of this instrument is debatable. Whereas the time period was significantly associated with stent choice from a statistical standpoint, there was still a preponderance of DES use even in the later time period. So-called “weak” instruments are known to provide estimates that may be more biased and less precise than stronger ones.11

The wide confidence intervals in the instrumental variable estimates observed in this study, punctuated by the unusual finding of a lack of a statistical significant association between stent type and target lesion revascularization (TLR) despite an 80% relative risk reduction in TLR for DES compared with BMS in randomized trials,12 underscore the importance of the strength of the instrument. Furthermore, the adjusted risk difference in mortality observed by Venkitachalam et al, though not statistically significant, favored BMS. Are we to conclude that DES may not prevent restenosis and possibly increase mortality, or is this the result of violations of the assumptions of the instrumental variable approach within this dataset? Furthermore, improvements in medical treatment between 2004 and 2007, a time period in which intensive optimal medical therapy gained traction and dual antiplatelet therapy guidelines changed, could lead one to believe that the passage of time would be associated with improved mortality in this population apart from any influence of stent selection or other assessed covariates, and thus potentially mediate benefit associated with the instrument itself. Here again, the point estimate favoring lower mortality with BMS in the instrumental variable analysis hints at this possibility.

Implicit in this study is the assumption that prevention of restenosis with DES should not translate to reduced mortality. Cardiologists are likely to agree that a reduction in mortality attributable to DES as a medical device is unlikely, or at best very small in magnitude. Whereas faster vessel patency achieved with a coronary stent compared with thrombolysis has proven to reduce mortality in randomized trials in the acute setting,13 the hypothesis that late mortality can be diminished by local therapy for stable coronary disease has been disproven in randomized trials of angioplasty14 and BMS.15 DES were developed and embraced clinically after their approval in 2003, not as a means to prevent mortality but to prevent the predictable in-stent renarrowing that led to recurrent angina in about 15% of patients.16 Whereas percutaneous coronary procedures incur cost, discomfort, and a risk of morbidity, the incremental risk of mortality is small. Because stents treat only the focal stenosis and do not prevent the unpredictable rupture of nonobstructive plaques present elsewhere in the coronary tree (the more frequent culprits for subsequent events),17 improvements in local patency between BMS and DES cannot be expected to improve mortality.

Methods and Results: Do the Ends Justify the Means?

Concordance between the inferences made from a given methodological approach and our preconceived notions of truth and plausibility of the findings should not be taken, in and of itself, as validation of that approach. We would caution that each of the observational data analysis methods illustrated here is complementary—providing different inferences framed by the different assumptions and limitations of each technique within the specific dataset and treatment question. One size does not fit all. Nor should the limitations of observational research findings lead us to dismiss these methods as a whole in favor of the randomized clinical trial; broader treatment strategy comparisons require one to consider the wider population context and variation in implementation that is inherent in medical care outside of experimentation.

Venkitachalam et al are to be commended for their illustration of the strengths and weaknesses of various analytic methods. Methodological investigation of this kind will be crucial to developing a framework for the direction and interpretation of comparative effectiveness research. In randomized trials, the clinical community is well versed in the assessment of markers of quality in a randomized trial: blinding, power, and completeness of ascertainment. In observational data analysis, sharing measures of how well an applied method conforms to its required assumptions will provide the clinical community with a more informed framework to assess the quality of such studies, a more complex but certainly rich area for the study of treatment effectiveness.

Disclosures

Dr Yeh is an investigator for the Harvard Clinical Research Institute and has consulted for the Kaiser Permanente Northern California Division of Research. Dr Mauri receives institutional research support from Abbott, Boston Scientific, Cordis, Medtronic, Eli Lilly, Daiichi Sankyo, Bristol Myers Squibb, and Sanofi-Aventis and has consulted for Abbott, Cordis, and Medtronic.

References


**Key Words:** Editorials ■ comparative effectiveness ■ confounding ■ bias ■ stents ■ instrumental variables ■ propensity scores
Choosing Methods to Minimize Confounding in Observational Studies: Do the Ends Justify the Means?
Robert W. Yeh and Laura Mauri

Circ Cardiovasc Qual Outcomes. 2011;4:581-583
doi: 10.1161/CIRCOUTCOMES.111.963538
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/4/6/581

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org/subscriptions/