Comparative Effectiveness and Caveat Emptor

Pamela N. Peterson, MD, MSPH; Paul D. Varosy, MD

Recent legislation has emphasized comparative effectiveness, including both the American Recovery and Reinvestment Act of 2009 and the Patient Protection and Affordable Care Act of 2010, which established the Patient Centered Outcomes Research Institute (PCORI) to foster comparative effectiveness research (CER). Defined as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care,” the intent of CER is to assist stakeholders, including patients and clinicians, with making informed decisions to improve health. Although many stakeholders are hopeful that CER will improve health and increase the value of care provided, it is not uniformly accepted that CER will achieve these important goals. In particular, the use of nonexperimental observational methods is questioned by some. Skepticism around observational CER is warranted. Indeed, observational studies have inherent limitations, including several potential sources of bias and confounding that may threaten the validity of findings. The effect of hormone replacement therapy (HRT) on cardiovascular outcomes is a notorious example. In the 1990s, high-profile publications using complex multivariable statistical modeling approaches in a variety of observational settings and datasets demonstrated lower risk of cardiovascular outcomes associated with use of HRT in women. Subsequently, randomized trials determined that HRT was actually harmful with respect to cardiovascular outcomes. With the proper perspective, however, observational studies can contribute useful evidence and have an important role in CER. In fact, there is great appeal to using observational data to efficiently answer multiple important scientific questions where costs and/or other factors make performing trials impractical.

In this issue, Reynolds and colleagues present an observational clinical comparative effectiveness analysis of catheter ablation versus antiarrhythmic therapy for the prevention of stroke or transient ischemic attack among patients with atrial fibrillation (AF). The comparative effectiveness and safety of medical antiarrhythmic therapy and catheter ablation of AF are of great importance. AF is a common and increasingly more prevalent problem that carries an increased risk of stroke. The authors used a large administrative claims dataset to identify patients with a diagnosis of AF who received an ablation procedure or antiarrhythmic therapy. In a propensity score-matched analysis, further adjusted for multiple measured confounders, ablation therapy was associated with a 40% lower risk of stroke or transient ischemic attack and no significant difference in heart failure admissions compared with antiarrhythmic drug therapy. The authors clearly and appropriately state that the advanced statistical methods used cannot account for unmeasured differences between the patient groups. We agree and believe that the observed association of AF ablation with apparently lower risk of stroke in this study should be interpreted with great caution.

A key concern in comparative effectiveness analyses using observational data are confounding by indication, or treatment selection bias. Investigators do not have control over treatment assignment, but, rather, clinicians and patients make treatment decisions. Propensity score analysis balances the observed characteristics of patients nonrandomly assigned to different treatments and reduces treatment selection bias. Many factors that influence patient selection for AF ablation, however, may not be documented in a medical chart and are certainly not available through claims data. For example, duration of atrial fibrillation, left atrial size, functional status, and symptom burden may influence the choice of treatment and also affect outcomes but were not available for analysis. It is remarkable that the majority of ablation patients could not be paired with nonablation control subjects. The match rate of 25% highlights that the patients who received ablation generally differed substantially from those receiving antiarrhythmic therapy on measured variables and suggests that confounding by indication, or treatment selection bias, is an important concern in the study interpretation.

Similar to confounding by indication, other unmeasured confounding, and the “healthy user effect” may also threaten the validity of findings in CER. When a number of measured variables are available, they can collectively represent, in part, the effect of unmeasured confounders. In this sense the propensity score, which uses available measured variables, serves as a proxy for unmeasured factors that are dependent on measured factors. However, propensity score methods cannot balance unobserved confounders that are independent of the observed confounders. To assess for potential healthy user effect, Reynolds and colleagues examined admission for...
datasets, as described above, is costly and complex. Chal-
data, but also have limitations. Creating robust clinical
able, many lack 1 or more of these components. Integrated
Although a number of observational data sources are avail-
and provider perspective; and 5) track longitudinal outcomes,
detailed clinical information for risk adjustment; 4) include
populations receiving and not receiving therapies; 3) include
lacking in granularity and nuance of the clinical information
required for the ascertainment of specific procedures, clinical
burden of disease severity and comorbidity, and clinical
outcomes. In this study, the accuracy of identifying those who
underwent ablation using claims data are uncertain because
an administrative code specific to AF ablation did not yet
exist at the time the data for this study were collected. This
results in an important possibility of misclassification and
calls into question the validity of the comparison groups.
Additionally, as discussed above, a number of clinical vari-
ables that likely influenced treatment selection and that may
be potential confounders may have been incompletely cap-
tured or not captured by claims data at all. Inability to
influence clinical decision-making; and 3) potential con-
ounding factors. In general, administrative datasets such as
the one employed in the current study are based on adminis-
trative codes (CPT and ICD9/10 codes) and, as such, are often
lacking in granularity and nuance of the clinical information
required for the ascertainment of specific procedures, clinical
burden of disease severity and comorbidity, and clinical
outcomes. In this study, the accuracy of identifying those who
underwent ablation using claims data are uncertain because
an administrative code specific to AF ablation did not yet
exist at the time the data for this study were collected. This
results in an important possibility of misclassification and
calls into question the validity of the comparison groups.
Additionally, as discussed above, a number of clinical vari-
ables that likely influenced treatment selection and that may
be potential confounders may have been incompletely cap-
tured or not captured by claims data at all. Inability to
adequately account for such factors in the analysis increases
the likelihood of bias and confounding.

This study highlights the need for high quality data to
perform CER. Databases are needed that 1) reflect routine
community-based clinical practice; 2) include broad patient
populations receiving and not receiving therapies; 3) include
detailed clinical information for risk adjustment; 4) include
information on treatment decision-making from the patient
and provider perspective; and 5) track longitudinal outcomes,
including patient-centered outcomes, such as health status.
Although a number of observational data sources are avail-
able, many lack 1 or more of these components. Integrated
health care systems such as the VA provide the most robust
data, but also have limitations. Creating robust clinical
datasets, as described above, is costly and complex. Chal-
lenges include the protection of private health information
and connectivity of health information systems, in addition to
ensuring that the data collected are valid. The investment in
creating such rich datasets, financial and otherwise, is neces-
sary for CER to achieve its intended goals.

Of critical importance is how the results of observational
studies are interpreted and used. In the case of HRT, millions
of women received therapy, in the absence of other indica-
tions, with the sole intent of reducing cardiovascular risk. The
findings of the current study of AF therapies should stimulate
further study; however, they should not be interpreted as
adequate to advocate for the use of catheter ablation of AF as
a stroke-reduction strategy. A randomized trial, CABANA,
is underway, which will provide additional data on the longitu-
dinal safety and efficacy of atrial fibrillation ablation and the
procedure’s effects on the risk of stroke.

Although this study should be interpreted with caution and
should not change practice, it does provide important insights
into the real world outcomes of patients undergoing catheter
ablation. Even with the limitations of observational data, the
authors used the most rigorous possible statistical methods
and appropriately framed the findings in the context of the
limitations of the data. Further, the finding of a relatively low
risk of hospitalizations for stroke among the patients who
underwent AF ablation is somewhat reassuring about the
safety of AF ablation, as it is currently employed in practice.

It is certain that observational studies will play an impor-
tant role in CER, particularly, if the various emports (stake-
holders) adequately consider the relevant caveats. It is critical
that researchers apply the highest quality methods and are
honest about potential limitations, exactly as Reynolds and
colleagues have done. Similarly, healthcare decision-makers
must understand the relative strengths and weaknesses of
various types of evidence. Finally, it is critical that all
stakeholders interpret results cautiously and in the context of
a larger body of evidence. Meanwhile, concerted efforts to
develop data sources that enhance the validity of observa-
tional CER will ensure that such studies provide the most
useful information to clinicians and their patients.

Sources of Funding
Funded in part by a Research Career Development Award (RCD
04-115-2 to Paul D. Varosy) from the Veterans Administration
Office of Health Services Research and Development and K08 award
from the Agency for Healthcare Research and Quality to Pamela N.
Peterson.

Disclosures
None.

References
1. Patient Protection and Affordable Care Act of 2010 (PPACA), PL
3. Institute of Medicine. Initial national priorities for comparative effec-
4. Lauer MS, Collins FS. Using science to improve the nation’s health
system: NIH’s commitment to comparative effectiveness research. JAMA
2010;303:2182–2183.
5. Djulbegovic M, Djulbegovic B. Implications of the principle of question
propagation for comparative-effectiveness and “data mining” research.
JAMA. 2011;305:298–299.

**Key Words:** Editorials | arrhythmia | atrial fibrillation | comparative effectiveness | stroke
Observational Comparative Effectiveness Research: Comparative Effectiveness and Caveat Emptor

Pamela N. Peterson and Paul D. Varosy

Circ Cardiovasc Qual Outcomes. 2012;5:150-152
doi: 10.1161/CIRCOUTCOMES.112.965574

Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 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/5/2/150

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/