Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using “Real-World Data”

Demissie Alemayehu, PhD; Riaz Ali, MPP; Jose Ma. J. Alvir, DrPH; Joseph C. Cappelleri, PhD; Mark J. Cziraky, PharmD; Byron Jones, PhD; Jack Mardekian, PhD; C. Daniel Mullins, PhD; Eleanor M. Perfetto, PhD; Robert J. Sanchez; Prasun Subedi, PhD; and Richard J. Willke, PhD
Table of Contents

Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using “Real-World Data”

Demissie Alemayehu, PhD; Riaz Ali, MPP; Jose Ma. J. Alvir, DrPH; Joseph C. Cappelleri, PhD; Mark J. Cziraky, PharmD; Byron Jones, PhD; Jack Mardekian, PhD; C. Daniel Mullins, PhD; Eleanor M. Perfetto, PhD; Robert J. Sanchez; Prasun Subedi, PhD; and Richard J. Willke, PhD

S3  Introduction
    C. Daniel Mullins, PhD, and Robert J. Sanchez, PhD

S5  Something Old, Something New, Something Borrowed… Comparative Effectiveness Research: A Policy Perspective
    Prasun Subedi, PhD; Eleanor M. Perfetto, PhD; and Riaz Ali, BA, MPP

S10 “Ten Commandments” for Conducting Comparative Effectiveness Research Using “Real-World Data”
    Richard J. Willke, PhD, and C. Daniel Mullins, PhD

S16 Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research
    Demissie Alemayehu, PhD, and Jack Mardekian, PhD

S22 Statistical Issues with the Analysis of Nonrandomized Studies in Comparative Effectiveness Research
    Demissie Alemayehu, PhD; Jose Ma. J. Alvir, DrPH; Byron Jones, PhD; and Richard J. Willke, PhD

S27 Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research
    Demissie Alemayehu, PhD; Robert J. Sanchez, PhD; and Joseph C. Cappelleri, PhD

S34 Developing a Collaborative Study Protocol for Combining Payer-Specific Data and Clinical Trials for CER
    Robert J. Sanchez, PhD; Jack Mardekian, PhD; Mark J. Cziraky, PharmD; and C. Daniel Mullins, PhD
Demissie Alemayehu, PhD, is currently Executive Director at Pfizer, where he is Statistics Head for Global Outcomes Research as well as Bone and Endocrinology Disease Areas. He received his PhD in statistics from the University of California at Berkeley, and was elected a fellow of the American Statistical Association in 2002. Dr. Alemayehu has held academic appointments for more than 20 years at major universities, including Columbia University in the City of New York and Western Michigan University. He has been active in professional societies, has held numerous elected and appointed positions with the American Statistical Association and the International Biometric Society, Eastern North American Region, and has served on the editorial boards of major journals, including the Journal of the American Statistical Association and the Journal of Nonparametric Statistics. Dr. Alemayehu has authored or co-authored several publications in the statistical and medical literature on topics ranging from the asymptotic theory of bootstrap methods to goodness-of-fit tests.

Riaz Ali, BA, MMP, is Director, Avalere Health. He provides clients with research, analytic, and project management support on evidence-based medicine and health care quality improvement issues. Prior to joining Avalere Health, Riaz completed a fellowship at the Office of Congressman Bobby Jindal (LA-01) as well as terms in the policy group at Pharmaceutical Research and Manufacturers of America and Jeffery J. Kimbell & Associates, a Washington-based government strategy firm. Prior to these graduate school fellowships, Riaz was Senior Strategist at Wunderman NY, a Young & Rubicam Inc.-owned marketing services company, where he developed communications strategies for product and brand launches for pharmaceutical clients. Riaz holds a BA in Political Science from Columbia University and an MMP in Health Policy from Georgetown University.

Jose Mar. J. Alviar, DrPH, was born in Manila, Philippines, where he graduated magna cum laude with a BA in Sociology from the University of the Philippines and was given the Most Outstanding Graduate award. Dr. Alviar earned his MPH and DrPH degrees from Columbia University. After a long career in academic research at Hillside Hospital and NYU, he joined Pfizer in 2004. At Pfizer, he serves as Senior Director, Statistical Scientist, partnering with Outcomes Research and Market Access colleagues. Jose has served as Industry Representative in the Medicare Evidence Development & Coverage Advisory Committee and is currently one of the editors of the Biopharmaceutical Report of the American Statistical Association Biopharmaceutical Section.

Joseph C. Cappelleri, PhD, earned his MS in statistics from the City University of New York, PhD in psychometrics from Cornell University, and MPH in epidemiology from Harvard University. In June 1996, Dr. Cappelleri joined Pfizer as a biostatistician collaborating with Outcomes Research and is a Senior Director in biostatistics at Pfizer. He is also an adjunct professor of medicine at Tufts Medical Center and an adjunct professor of statistics at the University of Connecticut. He has delivered numerous conference presentations and has published extensively on clinical and methodological clinical topics, including regression discontinuity designs, meta-analysis and health measurement scales. Dr. Cappelleri has also been instrumental in developing and validating a number of patient-reported outcomes for different diseases and conditions. He is a Fellow of the American Statistical Association.

Mark J. Cziraky, PharmD, is co-founder and Vice President, Industry Sponsored Research, at HealthCore. He received Fellowship status in the American Heart Association in 2000 and the National Lipid Association in 2007. Dr. Cziraky earned his Bachelor and Doctorate of Pharmacy degrees from the Philadelphia College of Pharmacy and Science. He completed his residency training in Ambulatory Care Pharmacy Practice at Blue Cross and Blue Shield of Delaware. He served from 2005-2010 as a member of the Cardiology Expert Committee for the United States Pharmacopeia. He is currently a member of the Board of Trustees for the Institution of Safe Medication Practice and Boards of Directors for the American Pharmacists Association Foundation, the National Lipid Association, and the North East Lipid Association. Dr. Cziraky also holds an Adjunct Associate Professor appointment at the University of Delaware School of Nursing, Philadelphia College of Pharmacy and Science, and a Clinical Faculty Position at The University of Florida College of Pharmacy.

Byron Jones, PhD, joined the pharmaceutical industry in 2000. Prior to that, he was a university professor and consultant to the pharmaceutical industry. Dr. Jones is a Fellow of both the American Statistical Association and the Royal Statistical Society. He has co-authored 4 books and has more than 100 publications in peer-reviewed journals. He was a founding Editor-in-Chief of the Journal Pharmaceutical Statistics and previously was the Regional Editor of Biopharmaceutical Statistics. Dr. Jones still retains close ties with Academia and holds Honorary Professorships at University College London and at the London School of Hygiene and Tropical Medicine.

Jack Mardekin, PhD, is Senior Director, Statistics at Pfizer, Inc., where he supports outcomes research projects focusing on retrospective databases and the statistical methods used in their analyses. Dr. Mardekin holds an MS in Statistics from Rutgers University and a PhD in Applied Statistics from the University of Wyoming. He is a Visiting Part-Time Lecturer at Rutgers University.

C. Daniel Mullins, PhD, is a Professor within the Pharmaceutical Health Services Research Department at the University of Maryland School of Pharmacy. He received his BS in economics from MIT and his MA and PhD in economics from Duke University. His research and teaching focus on pharmacoeconomics, comparative effectiveness research, and health disparities research. He has received funding as a Principal Investigator from the NIA and NHLBI and was the Shared Resources Core Director for the NIH-sponsored University of Maryland Center for Health Disparities Research, Training, and Outreach. In addition to his work on federal grants, Professor Mullins has designed many cost-effectiveness analyses and budget impact models for both the pharmaceutical and insurance industries. He is co-Editor-in-Chief for Value in Health and is author/co-author of more than 150 peer-reviewed articles.

Eleanor M. Perfetto, PhD, is Senior Director, Reimbursement & Regulatory Affairs, Federal Government Relations at Pfizer. Dr. Perfetto holds BS and MS degrees in pharmacy from the University of Rhode Island, and a PhD from the University of North Carolina School of Public Health. She currently serves as a Pharmacy Quality Alliance (PQA) board member and co-chairs the Research Coordinating Council. She serves on the Center for Medical Technology Policy and Health Industry Forum Boards of Advisors and was appointed to the Centers for Medicare and Medicaid Services Medicare Evidence Development & Coverage Advisory Committee. She served on the National Quality Forum Steering Committee for a Framework for Measurement, Evaluation, and Reporting of Healthcare Acquired Conditions. Dr. Perfetto served for 6 years on the Drug Information Association Board of Directors and is a past President. Dr. Perfetto is also an adjunct faculty member at the University of Maryland and University of Rhode Island schools of pharmacy. Prior to Pfizer, she served in the U.S. Public Health Service as senior pharmacoepidemiologist at the Agency for Health Care Policy & Research.

Robert J. Sanchez, PhD, is Director, U.S. Health Economics and Outcomes Research at Pfizer, where he supports both brand and payer channels. Dr Sanchez holds a pharmacy and a Masters degree from the University of Texas at Austin and a PhD from the University of Wisconsin. Prior to joining Pfizer, he was Director of Drug Information at a mail order pharmacy.

Prasun Subedi, PhD, is a Director in Pfizer's Worldwide Policy group, where he focuses on matters related to U.S. health care reform. Previously, Dr. Prasun worked in Pfizer's Specialty Care Business Unit, focusing on pricing and reimbursement issues. He received his PhD in pharmaceutical health services research from the University of Maryland in 2008.

Richard J. Willie, PhD, is Head of Global Health Economics and Outcomes Research, Global Market Access, Primary Care Business Unit, at Pfizer, based in New York City and Peapack, New Jersey. He first joined one of Pfizer's legacy companies, Upjohn, in 1991, and received a PhD in economics from Johns Hopkins University in 1982, concentrating in econometrics and labor economics. Dr. Willie served on the ISPOR Board of Directors (2007-09), was Chair of the ISPOR Institutional Council in 2010, and was co-chair of the ISPOR Good Research Practices Task Force on Cost-Effectiveness Analysis in Randomized Clinical Trials in 2003-2005. He was a member of the Health Outcomes Committee of PhRMA from 1998-2009, having been its chair from 2002-2004.
ABSTRACT

BACKGROUND: The Patient Protection and Affordable Care Act brought considerable attention to comparative effectiveness research (CER).

OBJECTIVES: To (a) suggest best practices for conducting and reporting CER using “real-world data” (RWD), (b) describe some of the data and infrastructure requirements for conducting CER using RWD, (c) identify statistical challenges with the analysis of nonrandomized studies and suggest appropriate techniques to address those challenges, (d) recognize the value of patient-reported outcomes in CER, (e) encourage the incorporation of observational data into randomized controlled studies, and (f) highlight the importance of incorporating payers in industry-sponsored research.

SUMMARY: The first article in this supplement, “Something old, something new…” provides a policy perspective on the recent evolution of CER. It reviews the historical context, discusses the “promise and fear” of CER, and then describes the new role of the Patient-Centered Outcomes Research Institute (PCORI) in defining and sponsoring CER. The second paper, “Ten Commandments,” proposes a series of tenets for planning, conducting, and reporting CER done with RWD. Oriented for basic-to-intermediate researchers, it combines standard scientific research principles with considerations specific to nonrandomized, RWD studies. The third article, “Infrastructure Requirements,” points out that effective use of secondary data requires addressing major methodological and infrastructural issues, including development of analytical tools to readily access and analyze data, formulation of guidelines to enhance quality and transparency, establishment of data standards, and creation of data warehouses that respect the privacy and confidentiality of patients. It identifies gaps that must be filled to address the underlying issues, with emphasis on data standards, data quality assurance, data warehouses, computing environment, and protection of privacy and confidentiality. The fourth paper, “Statistical Issues,” discusses how the validity of analytic results from observational studies is adversely impacted by biases that may be introduced due to lack of randomization. It reviews some of the methodological challenges that arise in the analysis of data from nonrandomized studies, with particular emphasis on the limitations of traditional approaches and potential solutions from recent methodological developments. The fifth paper, “Considerations on the Use of Patient Reported Outcomes (PROs),” describes how PRO data can play a critical role in guiding patients, health care providers, payers, and policy makers in making informed decisions regarding patient-centered treatment from among alternative options and technologies and have been noted as such by PCORI. However, collection and interpretation of such data within the context of CER have not yet been fully established. It discusses some challenges with including PROs in CER initiatives, provides a framework for their effective use, and proposes several areas for future research. Lastly, “Developing a Collaborative Study Protocol…” indicates that there is the potential, the desire, and the capability for payers to be involved in CER studies, combining elements of their own observational data with prospective studies. It describes a case example of a payer, a pharmaceutical company, and a research organization collaborating on a prospective study to examine the effect of prior authorization for pregabalin on health care costs to the payer.

CONCLUSION: Researchers at Pfizer routinely conduct CER-type studies. In this supplement, we have proposed some approaches that we believe are useful in developing certain kinds of evidence and have described some of our experiences. Our experiences also make us acutely aware of the limitations of approaches and data sources that have been used for CER studies and suggest that there is a need to further develop methods that are most useful for answering CER questions.

J Manag Care Pharm. 2011;17(9-a):S3-S4

Copyright © 2011, Academy of Managed Care Pharmacy. All rights reserved.
Introduction

Because payers collect data ranging from pharmacy claims to more advanced approaches of data collection, such as electronic medical records (EMR), it is important to consider infrastructure needs from an organizational viewpoint when it comes to using secondary databases for conducting research. The third article in this supplement discusses the infrastructure required for using these sources of data to conduct CER trials. The article discusses not only the required infrastructure, but also touches on issues like data standards, quality assurance, and patient privacy protection.

While RCTs are considered the gold standard for providing efficacy claims, their ability to provide information on drugs’ real world value is often limited, particularly from a payer perspective. While data from RCTs satisfies the regulatory requirements for safety and efficacy, their strict inclusion/exclusion criteria may make them less generalizable to populations often covered by third-party payers. For example, an RCT evaluating the efficacy of a pain medication may require subjects to be free of all other medications used to control pain and allow only limited use of rescue medications. Furthermore, the protocol may exclude patients with past failures to certain pain medications as well those with certain comorbidities. The characteristics of patients meeting the inclusion and exclusion criteria do not fully reflect the general population.

For these reasons, observational studies have gained more attention to fill the evidence gaps that remain after traditional explanatory trials have been completed. However, while observational studies are more generalizable to the real world, they are fraught with issues of their own. Therefore, the fourth article in this supplement helps to bring those issues to light and offers some suggestions on how to handle those issues using valid and reliable statistical methods. There are many definitions of CER; however, they all have one common objective…to help people make more informed decisions about health care. As CER results are intended to be relevant to a broad array of individuals, it is not surprising that patient-reported outcomes (PROs) from a variety of patients are becoming incorporated into CER trials. The fifth article in this supplement, therefore, reviews the challenges associated with the use of PROs in CER and provides a framework for their effective use in such trials. Finally, the last article in this supplement discusses the importance of CER studies and offers a collaborative approach for conducting such studies that will better equip patients, physicians, and payers to make more informed decisions about which health care resources are most appropriate for specific clinical conditions and patients.

The supplement describes up-to-date research techniques and policies related to CER and represents a guide by which Pfizer conducts similar research. This supplement therefore provides the reader with an understanding on one company’s approach to ensuring that their scientific investigations within the umbrella of CER studies follow strict guidelines to ensure credible application of CER for evidence generation and use of our medicines.

DISCLOSURES

This supplement was sponsored by Pfizer, Inc. Alemayehu, Alvir, Cappelleri, Jones, Mardekian, Perfetto, Sanchez, Subedi, and Willke are employees of Pfizer, Inc. Mullins reported receipt of consulting income, speaker’s fees, grant support, and compensation for travel expenses from Pfizer, Inc. and serves on Pfizer advisory boards; he also received compensation for his contributions to the manuscripts in this supplement. Cziraky is an employee of HealthCore, which has received research grants and has consulting relationships with Pfizer and other pharmaceutical manufacturers; he did not receive separate compensation for his contribution to this manuscript. Ali is an employee of Avalere Health, which receives consulting income from Pfizer and other health care organizations. Ali and Avalere Health did not receive specific consulting fees from Pfizer for his contributions to this manuscript.

REFERENCES

Something Old, Something New, Something Borrowed…
Comparative Effectiveness Research: A Policy Perspective

Prasun Subedi, PhD; Eleanor M. Perfetto, PhD; and Riaz Ali, BA, MPP

In the nearly 18 months that have elapsed following enactment of the Patient Protection and Affordable Care Act (PPACA), health care scholars have debated a key portion of the legislation that has the potential to dramatically impact health care. Section 6301 of the PPACA outlines how the federal government will play an increasingly direct role in shaping comparative effectiveness research (CER). The inclusion of CER in the PPACA marked a critical step in the advancement of health services research, given that the legislation called for significant federal investment in CER with the ultimate goal of improving the efficiency of the health care system. Here, we briefly review the policy history of CER as it relates to prior efforts in health services research and health technology assessment, discuss the promises and fears associated with CER as outlined in the PPACA, and highlight the important role that the new Patient-Centered Outcomes Research Institute (PCORI) will likely play in advancing both CER methods and evidence generation.

Historical Policy Context
It is important to recognize that current efforts around CER represent the latest in a series of evolutionary steps that were borne out of the constructs of “health technology assessment” (HTA), “effectiveness research,” and “evidence-based medicine,” among others. Conceptually, each of these constructs serves a distinct purpose (e.g., evidence generation and analysis versus application in decision making), and was advanced toward discrete objectives (e.g., understanding if something works versus is something worth doing or paying for). These differences belied a common thread—integration of clinical evidence about an intervention or service into decision-making.

Figure 1 demonstrates a chronology of this activity—dating back several decades—demonstrating the overall effort of infusing evidence into health care decisions. In the United States, these early activities can be traced back to the Congressional Office of Technology Assessment (OTA, 1972-1995), an agency tasked with oversight of various scientific and technical issues. Over 2 decades, the OTA issued a series of reports on a variety of process- and disease-oriented health care matters. Although defunded in the mid-1990s as part of government consolidation reforms, the reviews conducted by the OTA set an example for similar processes that have subsequently been developed and implemented by both public and private entities in the United States and other countries. Prior efforts toward evidence integration conducted in the United States have proven useful in broadening knowledge, but federally-funded agencies have faced significant challenges in sustaining their efforts, largely due to political opposition borne out of perceptions that this work was primarily focused on reducing costs, and that its implementation would lead to rationing of health care.

In light of a legacy of challenges and lingering conflation of these distinct concepts, the federal government has again sought to wade carefully—but concertedly—in this space. Recently it has focused its efforts on organizing investment on clinical comparative effectiveness. Table 1 includes a selection of recent definitions of CER advanced by the Agency for Healthcare Research and Quality (AHRQ), which runs the Effective Healthcare Program; the Federal Coordinating Council, which offered a definition to help orient the $1.1 billion in funding through the American Recovery and Reinvestment Act; the Institute of Medicine; and the definition of CER provided in the PPACA. These definitions share several common elements: (a) focus on clinical effectiveness; (b) inform a wide range of decision-makers (i.e., patients, providers, and policymakers); and (c) focus on a broad set of interventions and services. More recently, the PCORI released a draft definition of patient-centered outcomes research (PCOR) for public comment. While the PCOR definition is similar to many of the CER definitions discussed above, its clear focus on preferences and needs marks an important shift towards the patient—and this shift may have important implications for evidence generation and dissemination.

Given that CER-like efforts have been percolating in the United States for the better part of 4 decades, an inevitable question arises: why was CER so clearly singled out as a critical component of the 2010 health care reform legislation? One factor is an acute awareness of the rising costs in health care and growing questions about whether those have translated to meaningful improvement in quality of care. As an example, recent technological advancements have brought about many novel diagnostic and treatment paradigms, but every incremental innovation seemingly brings with it additional questions regarding how the new technology can best be used in the context of existing treatments to improve outcomes (i.e., advancing the quality of care), while not overburdening increasingly constrained resources (e.g., without significantly increasing the cost of care). The majority of the prior CER-like efforts were designed to answer similar questions about quality and cost; however, the PPACA’s clear focus on this research is also motivated by the fact that “in the next decade, the United States must absorb 32 million currently uninsured people into the health care system, while simultaneously improving the quality of care and slowing cost increases.” While there is significant theoretical promise that comparative studies will provide evidence both to improve quality and to decrease cost, there are also many potential fears regarding the inappropriate
The Promise and Fear of CER

The real "promise" of CER lies in its potential to generate "more and better evidence on what works best."7 There have been significant technological advances in recent years, yet the evidence of effectiveness of health care interventions as a whole is suboptimal. This is partly because much of the published evidence on health care interventions is defined and driven by randomized controlled trials (RCTs) designed to answer specific regulatory questions. Although RCTs are praised for their capacity to identify causal relationships between treatments and health outcomes, it is important to note that these studies are not designed to answer more intricate questions regarding how a new therapy should be considered for use in the context of existing treatment options. RCTs are typically conducted in carefully selected patient populations, under highly controlled settings, with placebo comparators and often provide only aggregated averaged results—largely ignoring variation in treatment response by patient characteristics. By definition, CER studies, with their objective of "comparing health outcomes and the clinical effectiveness of 2 or more medical treatments, services, or items,"1 have the potential to significantly add to the evidence base used in the treatment selection process. In the short term, this evidence is likely to be derived from observational studies conducted using the myriad real-world data sources developed in recent years; in the longer term, it may be that RCT designs are increasingly adapted to provide more relevant evidence for comparative questions.

CER studies that generate evidence through an evaluation of the spectrum of health care interventions and services, that reflect true patient choices for a given clinical situation, will improve patient and physician decision making. With the appropriate evidence, CER has the potential to lead to...
better, more patient-relevant treatment decisions—allowing the “right” treatment to be delivered to the “right” patient in the “right” setting. Achieving this alignment would likely yield significant downstream effects. Specifically, CER evidence, if rigorously produced and effectively transmitted, represents a significant opportunity to reduce current variations in the quality of care, which in turn would serve to improve outcomes and reduce currently observed health care disparities.

Ultimately, CER evidence also has the potential to represent an important step forward in the progression of personalized medicine—the development of “treatment regimens based on the molecular biology of individuals or their diseases”—which to date has been a promising, albeit elusive, goal.

While the promises of CER are great, so are the “fears” regarding the impact such research may have on how health care is practiced and financed—both for the public and private health insurance markets. In the months leading up to the passage of PPACA, there were significant concerns that government-sponsored health care studies would ultimately lead to homogenized (i.e., “one-size fits all”) treatment recommendations that ignored patient heterogeneity. These anxieties were coupled with worries that such treatment recommendations would inevitably be used to justify cost control efforts, leading to indiscriminate coverage restrictions that could potentially devastating impacts on patients themselves (e.g., “death panels”). Moreover, some have argued that such blunt restrictions would force a reduction of investment in health care innovation, which has advanced clinical paradigms and generated economic value in the United States for many years.

In the end, many of these fears seem to have been heard by Congressional officials and were addressed in the final language of the health care reform legislation. PPACA established the PCORI, and the legislation directed that the public-private institute should seek to “advance the quality and relevance of evidence concerning the manner in which diseases, disorders, and other health conditions can effectively and appropriately be prevented, diagnosed, treated, monitored, and managed,” to inform “patients, clinicians, purchasers and policy makers in making informed health decisions.” This evidence- and stakeholder-focused language, taken in the context of the PPACA’s mandate that the PCORI board (a) establish a expert advisory panel on rare diseases, and (b) provide “support and resources to help patient and consumer representatives effectively participate” in its activities, suggests a clear sensitivity to the fears discussed above.

In addition, the legislation mandates that PCORI not make clinical, coverage, or reimbursement recommendations on the basis of the evidence generated at its direction. The Secretary of Health and Human Services is granted the authority under the establishing legislation to use CER findings from PCORI-sponsored work, in conjunction with other evidence, in coverage determinations, but there are specific caveats that coverage decisions based on CER must be developed in a transparent and iterative manner (where iterative refers to the public
Role of PCORI in Advancing CER Infrastructure and Methods

A critical aspect of PCORI’s remit, especially in the short run, will focus on developing both the infrastructure to support substantive CER studies, as well as the methodological standards by which the research it funds should be carried out. Significant initial investments in these 2 areas have already been made as part of the CER funding allocated through the American Reinvestment and Recovery Act. From an infrastructure perspective, much work is needed to advance health information technology (HIT) from its current disaggregated state to a point where data sources such as electronic medical records, clinical and claims databases, and patient registries can be appropriately combined for use in CER. PCORI can and likely will play a critical role in ensuring that these and other data from routine, clinical encounters can be appropriately utilized in conjunction with clinical trials (randomized and pragmatic) to serve as a rich source of raw data. Alemayehu and Mardelkian provide additional thoughts on the infrastructure requirements needed for secondary data sources to be optimally utilized for CER in a separate article in this supplement (pages S16-S21).

As these data sources are developed, it will be equally important for PCORI to establish a clear methodological framework so that the CER studies it funds can have maximum scientific validity and broad acceptance by the end-user(s). Toward this end, the legislation requires PCORI to establish a Methods Committee, which will seek the advice of experts in biostatistics, health services research, and epidemiologists (among other disciplines) to advise and assist PCORI on developing best practices for conducting CER. As previously discussed above, a variety of organizations and entities have focused on CER-like efforts in the past. As a result, there exists a fairly substantial foundation of methods and standards from which PCORI can begin its work. However, it is critical that the PCORI board have adequate expert advice with which to understand, interpret, and potentially adopt existing methods standards, as well as develop new guidance. As a potential first step in this process, several articles in this supplement may provide PCORI and other interested stakeholders with “food for thought” in terms of methods development. Alemayehu et al. provide their insights on statistical issues related to the analysis of nonrandomized studies (pages S22-S26); Sanchez et al. outline how hybrid studies may be used to answer important CER questions (pages S34-S37); and Alemayehu et al. provide a framework that outlines how patient-reported outcomes may be optimally included in CER studies (pages S27-S33).

Conclusions

Although the core concepts behind CER have been in development under different labels for a number of years, it is clear that the health care reform effort, as outlined in PPACA, has raised the awareness of CER to demonstrably higher levels. The formation of PCORI, and its clear opportunity to advance both the infrastructure and methods of CER, demonstrates that the federal government recognizes the important role CER can play in addressing both quality and cost considerations. Through clear and open dialogue and engagement with all stakeholders—including patients, providers, insurers, academics, and industry—PCORI can leverage the wealth of existing resources to build an initial framework from which it can advance and promote the appropriate use of CER to improve upon the overall value of health care in the United States.


The use of “real-world data” (RWD), defined as “data used for decision-making that are not collected in conventional RCTs” (randomized controlled trials)1 to inform comparative effectiveness research (CER) questions holds tremendous promise, which can be realized only if such research is conducted by strictly—religiously, one might say—following good research practices.2,3,4 The well-recognized potential for biases associated with analysis of nonrandomized data, as well as the increasing accessibility of these data and their potential for being data-mined, might lead some to view them as “forbidden fruit” for informing medical decisions.5,6 In fact, some may argue that RWD CER results based on nonrandomized data a priori compromises their credibility. Others may argue that clinical trials that target only regulators rather than post-regularity decision makers, including patients, consumers, payers, prescribers, and policy makers are similarly, albeit differently, flawed because they are less informative for medical decision making than pragmatic clinical trials that address patient, prescriber, and payer concerns. In both randomized trials and studies using data with nonrandom assignment, the virtues of RWD CER results are more likely to be valued by appropriately skeptical audiences if decision makers are confident that the work has been conducted and reported with a dedication to high standards.

In this spirit of devotion to good research practices for CER using RWD, we offer “ten commandments” for conducting and reporting CER based on analysis of RWD, without any claim of having received them from on high. The purpose of this article is to provide the beginning-to-intermediate practitioner or decision maker with a concise list of practices that are crucial to the proper execution of this kind of work. It is not meant to replace the growing literature which, in many cases, more extensively reviews important technical aspects of RWD analysis, and we strongly recommend that readers also review other guidance documents and Task Force reports, such as those published by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). However, we believe there is merit in a brief overview of some key tenets of the RWD research process, from planning, to analysis, to reporting, that combine general good research practices with considerations specifically relevant to CER with RWD. Before beginning, we strongly urge those who conduct RWD studies to involve those who are part of the RWD data generation and decision-making processes when designing CER studies. This will maximize the usefulness of the RWD CER results.

I. Design your study to address the 3 central pragmatic features of CER, all oriented to informing a specific treatment choice: active comparators; relevant patient populations; and outcomes that are meaningful to patients, prescribers, payers and policy makers.

CER is intended to improve the evidence base for making decisions that impact the health of “real world” patients. Thus, CER studies should make comparisons—directly or indirectly—of the drug or medical technology being studied to other medical technologies that are commonly used or recommended to treat the targeted indication. The comparators should be selected from among those most frequently prescribed as well as those recommended in clinical practice guidelines. CER must be pragmatic in nature, reflecting a reasonable cross section of patients who are likely candidates for the comparators being studied.7 Study outcomes, including the measurement, frequency, and timing of reporting outcomes, must be meaningful to patients and their providers, as well as payers and policy makers who affect access to drugs and other medical technologies. In order to be meaningful, outcomes must be relevant and important to patients; however, in a CER study, it also must be the case that outcomes vary across comparators and patients.8 That is, there must be a plausible causal relationship between the treatment and the meaningful outcomes and a recognition that the relationship may vary across subgroups.

As with all components of CER study design, analysis, and interpretation, stakeholder engagement can help to assure that the study is appropriately designed to be maximally relevant and informative for decision making. When CER is conducted with a particular payer or subgroup of patients in mind, the comparators, patient population, and outcomes should reflect that perspective.

II. Develop your research question such that all benefits and harms relevant to the treatment decision for the product relative to the comparator are considered. The research question must be well-defined a priori and targeted to provide a clear answer for a specific audience. Choose a research design (e.g., case-control, cohort) and a corresponding data-set (right population, right variables, large enough sample) that are likely to be able to answer and are suitable for your research question).

Both the blessing and the curse of large RWD sets are the many research questions that can be addressed with them, making them ideal for exploratory data analysis. However, when the goal is to present evidence on a question as outlined in Commandment I, especially for decision-making purposes, one’s work must be free of any suspicion that the bulls-eye was painted around the arrow. Just as a conventional RCT starts with a research question, with the subsequent protocol...
and data collection designed specifically and parsimoniously to answer a pre-specified question, a CER study using RWD must start with a clear objective, which is usually best framed as a research question. Sometimes a research question begins as a very specific one, either to replicate or extend previous research. More often it begins rather broadly (e.g., “Does medication X result in better outcomes than medication Y in treatment of Z?”). If a broad CER question is being posed, then all benefits and harms of both products relevant to the treatment decision should be included in the analysis.

Before the analysis begins, a number of more specific conditions need to be imposed to clarify the question—which patients, with which characteristics, over what timeframe, under what definition of medication use, etc.? Those conditions should be based on what questions prior studies explored or left unanswered, or which specific coverage or treatment decisions require better information. Before a specific research question can be finalized, potential data sources should be reviewed for their feasibility (e.g., presence of the right variables, enough patients, and proper time frame) for answering the research question; even a very good data source may additionally delimit the research question. Finally, the specific research question, as well as the nature of disease and its treatment process, the prevalence of the outcomes of interest, the data source, and other factors will help determine the most appropriate research design (e.g., case-control, cohort, case-crossover, etc.). In the end, the research question should lead to an analytic framework that directly can test the hypothesis present in the question in a scientifically rigorous and informative manner. If the research question and relevant hypothesis cannot be tested satisfactorily with the RWD available, the research should likely not be done.

III. Investigate your data sources to understand the “real-world” process by which the data are generated. Describe the limitations of the data, as well as how patients are selected into or exposed to treatment, and when appropriate, describe potential concerns (e.g., classification bias, immortal time bias, adherence concerns, etc.) and how they are addressed.

Whoever said “what you don't know can't hurt you” never worked with RWD. Data are an inherently imperfect representation of the underlying characteristics they are meant to measure, even when collected following a strict protocol. Considering the highly variable conditions under which RWD are collected, recorded, transmitted, merged, etc., it’s best to ask yourself, and possibly others, questions about any datum important to your study. Are data complete and, if not, are data missing at random or is there a systemic bias in under-reporting that could impact the results of the study? Why are different diagnosis codes for a condition used? Do those codes vary by location or other factors, and if so, for transparent reasons? Are the patients or physicians not representative of typical practice in some way? Could their choices of treatments be limited by external factors, such as formulary restrictions or insurance provisions? What drugs used by patients may be missing from the data, and why? Incorrectly attributing exposure to treatment is called “classification bias.” Given test result data (e.g., blood pressure) under what conditions were those data collected, and why?

What can you do? Thoroughly review any underlying data manuals and/or questionnaires when they are available. While staying blinded to outcomes by treatment group, examine not only the descriptive statistics of key variables but also the distributions and lots of cross-tabulations. Consider consulting with a practicing physician, pharmacist, or a billing department employee to test your assumptions about your data. In addition, when constructing any outcome or control variables, be careful not to introduce any biases. For example, in a time-to-event analysis, introducing any time period during which the outcome could not have occurred will create “immortal time” bias. When categorizing patients based upon treatment, perform sensitivity analysis to see whether using different codes or time periods for exposure would affect how patients are categorized. When constructing total costs, be cognizant of systematic reasons why certain costs may be missing and exacerbate differences between treatment groups. In the end, it’s never possible to find or adjust for all the imperfections in one’s data, but doing the due diligence needed to be reasonably confident that the data are fit for the research task at hand is a fundamental responsibility of any empirical researcher. As in Commandment II, if the data are not deemed to be fit for the task, the research should not be continued; if the question is sufficiently important, a prospective study may be necessary.

IV. Write a full statistical analysis plan a priori that reflects current knowledge about comparator products and the evidence gap to be addressed; document any changes made along the way.

A pre-specified, well-written statistical analysis plan for a CER study provides benefits that are similar to those achieved by a pre-specified analysis plan for a conventional RCT. Having a roadmap provides a predetermined course for conducting the analysis and prevents deviations that otherwise could unintentionally change the validity or overall intent or direction of the study. It also avoids post hoc or selective reporting that tends to reduce the value and believability of results in the eyes of many decision makers. In fact, excessive post hoc analysis almost guarantees that certain results will appear to be statistically significant by chance rather than by true causation. Thus, pre-specified analysis plans enhance the credibility, efficiency, reliability, validity, and transparency of CER studies.

The statistical analysis plan should reflect a scientifically rigorous and clinically meaningful approach to answering the
study question. There should be specific aims and testable hypotheses that are directly related to the overall study objective and research question that are relevant for the comparator therapies being assessed. The analytic approach should be informed by what is known about the disease or condition being studied as well as the comparators being evaluated (see Commandment II). The statistical analysis plan should identify pre-specified subgroup analyses, specific codes (e.g., International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] and Current Procedural Terminology [CPT] codes) for inclusion/exclusion criteria, and the general approach for both descriptive and multivariable analyses (see also Commandments VI and VII).

As with conventional RCTs, there may be necessary deviations from the original statistical analysis plan because new evidence emerges from outside the trial or because of unexpected findings during the implementation of the analysis that require additional exploration. It also may be possible that the original statistical analysis plan failed to address a particular element appropriately. In all cases in which an amended analysis plan is required, it is important to report transparently not only what part of the statistical analysis plan was altered but why it was changed.

V. Carefully review univariate statistics for patient characteristics, outcomes, and control variables and how they differ across comparators. Investigate thoroughly the nature and degree of missing data (attrition, nonresponse, noncoverage, etc.) or miscoding, including anything that may affect treatment group identification.

In following Commandment III, you should have investigated some of these same issues in order to ensure that it was feasible to answer your research question with your data. Commandment V concerns the data analysis needed to inform not only yourself but also your audience about the nature of the data, its strengths and weaknesses, and its potential biases. The analysis begins with a thorough review of each relevant variable—outcome or control—and how it is distributed across comparison groups and across other relevant treatment subgroups. By identifying any fundamental imbalances, this descriptive analysis should inform and support any subsequent stratified or multivariate analyses. While this analysis cannot reasonably include “all possible” cross-tabulations, it should follow a logical process that ensures review of potentially important bivariate relationships, such as outcomes by disease severity across treatment groups.

A key aspect of this univariate review is attention to missing data. When control variables are missing, one should examine differences in outcomes across treatment groups for those with such variables present versus missing, in order to understand the biases that may be introduced by excluding observations with control variables missing. In cases where it appears that control variables are missing at random, it may be feasible to impute them. Several techniques to handle missing data exist (e.g., listwise deletion, pairwise deletion, or multiple imputation), and the reader is encouraged to carefully consider the pros and cons of each method.12,13,14 Missing outcome variables or completely missing observations are generally more problematic, but methods are available to at least partly manage those problems.15,16,17,18,19,20,21 Sometimes missing data can lead to poor treatment group identification, called classification bias (see Commandment III). The key task at this stage of the analysis is to analyze and report the extent of the missing data as well as any information about why it occurred that can guide subsequent analysis.

VI. Control for observed confounders and other effect modifiers (explanatory variables) in a systematic and unbiased fashion and pay particular attention to how these may vary across comparator treatments; be wary of their correlation with the treatment variable. Choose 1 or more methods to address unobserved confounders (also known as selection bias); none is perfect and comparisons of different methods can be informative.

In RCTs, both effect modifiers (factors which affect outcome but not treatment choice) and confounders (factors affecting both outcome and treatment choice) are randomly, and in large trials, generally equally distributed between treatment groups, making explicit controls for these factors unnecessary to estimate unbiased average treatment effects. Nevertheless, a pre-specified multivariate analysis controlling for patient characteristics that affect treatment outcome can reduce residual variance, result in a smaller confidence interval on the treatment effect estimate, and using interactions, potentially identify treatment-effect heterogeneity.

Outside of RCTs, treatment groups are rarely balanced on observed characteristics and the potential for confounding of outcomes by unobserved factors is high. Physicians and patients commonly make choices about treatments based on factors that also affect treatment outcomes (e.g., patients who are more severely ill [in ways sometimes not observed] are often treated more aggressively, making the more aggressive treatment a priori biased towards having worse outcomes). Treatment effect estimates that don’t both control for observed factors and consider unobserved factors are likely to be significantly biased. The literature on these issues is vast and distributed across statistical, econometric, epidemiological, psychological, and other disciplines. An overview of methods in this area, such as propensity score matching, stratification, instrumental variables, and others, as well as an extensive set of references, is found later in this supplement (Alemayehu et al., pages S22-S26).

Concerns around use of these methods can be grouped into 2 points. First, there are many choices of methods, including
selection of control variables, for a given problem, and each one may yield a different treatment-effect estimate. To avoid the temptation or appearance of picking a method post hoc that gives a “desirable” answer, the methods need to be clearly specified a priori. Second, one cannot know that a given method is going to give the “right” answer; none of the known methods can fully adjust for unobserved influences. Comparing the results of several methods, via a sensitivity analysis or simulation, given a sense of the strengths of each one, can provide insight into the robustness, or lack thereof, of conclusions around the comparative effectiveness estimates obtained.22

VII. Choose a statistical technique and functional form for your estimation that is most appropriate to the outcomes of interest (time to event, linear regression, 2-part model, general estimating equations, etc.) across therapies as well as the relationship between treatment, confounders, and outcomes.

There frequently may be more than 1 analytic approach and multiple ways in which a regression equation can be specified to answer a particular CER question; however, there usually is one that is preferable based upon the study perspective, the conceptual design, or the data-generating process. While it may sometimes seem that no matter what you select, peer reviewers prefer an alternative statistical approach, it is important to remember that part of the responsibility of conducting a study is describing the pre-specified methods and defending why the specific statistical technique and functional form were selected. There is both a science and an art to conducting CER, and the best research balances the 2 considerations. The art of CER requires that the analytic approach is informed by clinical practice and patient decision making so that the regression results provide meaningful and interpretable output. The science of statistical analysis provides guidance for assuring that one can draw conclusions from the results because the statistical technique is appropriate and the functional form has been informed by model specification tests. It is equally important to pre-specify such alternative model specifications in the analysis plan and follow up the analysis with statistical testing of alternative functional forms. Specification testing provides critical information for determining whether there are interaction effects (e.g., whether the treatment effect varies by age or other observable patient characteristics), whether higher-order terms are required (e.g., whether variables are related in linear or nonlinear ways), and whether variables should be continuous or categorical. At the same time, it is always important to review the final regression approach and results for clinical plausibility.

VIII. Report univariate and multivariate results in an unbiased and complete fashion such that the benefits and risks of all comparators reflect “fair balance.”

No study can answer all important questions, nor should one report every single data run, yet every study must provide objective and balanced reporting of the most relevant results regarding benefits and risks of all comparators included within the analysis. To achieve this balance, it is important to consider the viewpoints of decision makers, who are interested in comparisons of all clinically relevant benefits and potential side effects of treatments. This list should be informed by what is known or suspected about all treatment options included within the study. While all pre-specified outcomes should be provided in tabular form, it may be appropriate to highlight only those benefits and risks that are statistically different between the comparator treatments; however, in other cases, it may be important to comment on the fact that there is not a difference in key clinical outcomes. The reporting should include sufficient detail on the methods and results, including those from any alternative statistical approaches used, to provide the reader with a reasonably complete picture of the analyses performed; an online appendix can be useful for this purpose.

Objective reporting of outcomes requires that all benefits of all comparator treatments are given equal weight. Unfavorable outcomes should not be downplayed or “explained away.” It is acceptable to translate the clinical importance of both positive and negative impacts of therapies on health so long as this, too, is done in with fair balance.

IX. Do not “over-interpret” results in the Discussion or Conclusion sections; remain objective in describing differences in outcomes across comparators.

The Discussion section should interpret CER results for key stakeholders and decision makers and place the study’s results in context with prior knowledge and publications. Authors should comment on why comparative effectiveness results seem plausible, how the magnitudes of relative benefits and harms compare with those reported in prior studies, and whether observed differences are clinically and statistically significant. Although the Discussion may be somewhat subjective in nature, the interpretation of results should reflect an objective evaluation of what an unbiased individual could reasonably conclude from the study design and results. Authors should be careful to accurately reflect whether causal inference or correlation has been established and should avoid generalization of comparative benefits or harms beyond the study population and time frame.

Similar to regulators examining claims, payers and journal editors express strong criticism of manufactured-sponsored CER studies that appear partial in selecting which results are highlighted in the Discussion and Conclusion sections. Guidance on transparency in reporting can be found in the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) statement.23 The focus of the entire
paper, including the Discussion and Conclusion sections, is driven by the research question (see Commandment II). Therefore, the importance of following all commandments simultaneously is essential.

X. Know and follow any external requirements (e.g., from ethics committees, federal or local governments), as well as any internal organizational protocols or SOPs for RWD study conduct and reporting.

Use of RWD data for CER studies is increasingly considered to impose 2 ethical obligations on the researcher—to use the data in sanctioned research and to report the results. Conditions for use of individual patient data are set by the owner of the data. In some cases, the research proposal must be reviewed by either the data owner or an ethics committee before the research can be carried out. Their intent is generally to protect patient confidentiality and ensure that the research is conducted along whatever lines were agreed to by the patients when their data were collected. Once the research is complete, it may be necessary to post results of safety-related or effectiveness-related outcomes in the spirit that it would be unethical to withhold potentially relevant information about treatments from the public. The state of Maine has required that RWD CER studies examining safety and effectiveness outcomes of drugs be posted in a way similar to posting of clinical trials.

Some institutions and companies have created their own standard operating procedures to provide both information and processes for employees to follow that help them comply with these obligations. For example, an institution may require that study protocols for both randomized trials and observational studies be posted on www.ClinicalTrials.gov and study results be posted in the ClinicalTrials.gov Results Database; we would encourage this practice even if it is not required. Researchers should ask data providers and their own institutions about such requirements before engaging in CER studies using RWD.

■ Summary

Decision makers want RWD studies and CER that provide meaningful evidence about the benefits and harms of alternative treatments. At the same time, they remain skeptical when RWD studies are not appropriately designed to answer relevant questions in a scientifically rigorous and transparent manner. While our proposed “ten commandments” cannot guarantee that studies are free from bias or other flaws—they can only address the devils you know—they nevertheless can serve as a useful checklist for improving the systematic use of principles that are aimed at achieving the goals of developing credible and germane CER studies using RWD.

Authors

RICHARD J. WILLKE, PhD, is Head, Global Health Economics & Outcomes Research, Global Market Access, Primary Care, Pfizer Inc., New York, New York. C. DANIEL MULLINS, PhD, is Professor, Pharmaceutical Health Services Research Department, University of Maryland School of Pharmacy, Baltimore, Maryland.

DISCLOSURES

This supplement was funded by Pfizer, Inc. Willke is a Pfizer employee. Mullins reported financial and other relationships with Pfizer that include receipt of grants, consulting fees or honoraria, support for travel, consulting fees for participation in review activities such as data monitoring boards, payment for writing or reviewing the manuscript, advisory board membership, payment for lectures including service on speakers bureaus, and payment for development of educational presentations. Willke and Mullins contributed equally to the concept and design, and writing and revision of the manuscript.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the helpful comments of a number of colleagues, including Demissie Alemayehu, PhD, Jens Grueger, PhD, Frank Jen, PhD, Jack Mardelkian, PhD, Andreas Pleil, PhD, and Robert J. Sanchez, PhD.

REFERENCES


“Ten Commandments” for Conducting Comparative Effectiveness Research Using “Real-World Data”


Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research

Demissie Alemayehu, PhD, and Jack Mardekian, PhD

The growing interest in comparative effectiveness research (CER) has re-ignited the debate about the inadequacy of data from randomized controlled trials (RCTs) to address patient-centered decision making. Despite their well-known internal validity and use as the gold standard for regulatory decision making, the limitations of RCTs are widely recognized. In addition to their lack of statistical power due to inadequate sample size to address certain research hypotheses, practical and ethical considerations may preclude their viability. A case in point is the ethical dilemma in conducting an RCT to establish whether a diet high in fat content may be a risk factor for dementia, which might produce useful public health information but would not be acceptable in terms of protection of human subjects. Frequently, RCTs provide substantial information regarding the efficacy of drugs and other medical interventions, yet leave large gaps in evidence that would be relevant for medical decision making. Even when RCTs are carried out with this intent, they may not necessarily reflect “real-world” experience and, therefore, may not provide sufficient evidence to guide patient-centered care.

The substantial investment in CER and the broad objective implied in the American Recovery and Reinvestment Act of 2009 have necessitated the need to seek alternative sources of data to meet the emerging health care questions. The stated requirements include an “…assessment of a comprehensive array of health-related outcomes for diverse patient populations and sub-groups” as well as “…a wide range of interventions;” and “[d]evelopment, expansion, and use of a variety of data sources and methods to assess comparative effectiveness.”

Contending with the changes in health care policy and delivery clearly requires doing things differently, as discussed at a recent workshop sponsored by the Institute of Medicine. The workshop summary highlighted the dependence on clinical trials as the “sole source of evidence on the constantly accelerating flow of diagnostic and treatment challenges is unfeasible.” The need for a “learning healthcare system” with “real-time learning from the clinical experience and seamless application of the lessons in the care process” was emphasized.

Secondary data, such as registries and retrospective databases, can be used to complement RCTs, since they are less costly and can be used to incorporate real world experience to answer questions. Further, important questions, such as adherence, treatment patterns, and burden of disease, can be answered in retrospective analyses of databases. However, effective use of secondary data requires addressing major methodological and infrastructural issues that may be related to, but often go beyond, those encountered with most RCT-based work. Infrastructural issues, such as tools to efficiently access and correctly analyze the data, need to be developed for effective use of such data sources. Guidelines need to be formulated and data standards established using RCTs as a role model. Data warehouses are also required to be established that respect the privacy and confidentiality of patients.

In this paper, we discuss the infrastructural requirements for secondary data utilization in CER, and identify gaps that must be filled to address the underlying issues, with emphasis on data standards, data quality assurance, data warehouses, computing environment, and protection of privacy and confidentiality.

Secondary Data Sources and Associated Challenges

Secondary data can be generated from registries, chart reviews, electronic health records (EHR), administrative claims databases, or national surveys such as the National Health and Nutrition Examination Survey (NHANES) and the Behavioral Risk Factor Surveillance System (BRFSS). There may be linkage between distinct sources (e.g., U.S. Renal Data System [USRDS]) registry of end-stage renal disease patients in which claims data from Medicare patients are linked).

In other cases, the source may involve alternative designs (e.g., a combination of RCTs and nonrandomized studies). For instance, the United Kingdom’s General Practice Research Database (GPRD) has been developing the capability to run real-world primary care clinical trials through recruitment at point of care by general practitioners in their system. Patients are recruited at point of care. Software from GPRD’s information technology (IT) system informs the doctor when a patient satisfies inclusion and exclusion criteria for a particular study protocol, prompts the doctor for other needed information as well as patient consent, and provides the randomization (if there is one) to an assigned drug group. Drugs are given open label. Patient follow-up can be according to the treating physician’s standard care or can be a specific prompted return to the doctor. All data, including safety and patient reported outcomes, are recorded in the standard EHR data downloaded on a regular basis by GPRD.

Secondary data sources offer several benefits. First, they are rooted in real life, and if analyzed using appropriate statistical methods can document the effectiveness of drugs in everyday use under a wide spectrum of clinical practice. The patient population is diverse, mimicking the real world. Further, this patient diversity allows the rapid identification of large numbers of patients in a cost-effective manner, and permits comprehensive, long-term safety follow-ups. The latter is particularly important when the focus is on rare diseases, atypical therapy responses, or uncommon clinical questions.
However, observational studies also have inherent methodological and infrastructural limitations. Statistical methods have been developed to mitigate the limitations,\(^7\)\(^{10}\) and guidance documents have been generated to upgrade the analysis and reporting of data from secondary sources.\(^{11}\)\(^{13}\)

From the operational perspective, the infrastructural limitations of secondary data are considerable. In general, all relevant data may not be available. For example, reasons for therapeutic substitution may not be known, actual low-density lipoprotein cholesterol (LDL-C) levels for statin users may not be tracked, or diagnoses associated with a medication prescription may not be recorded. Further, claims data are generally built for billing and record-keeping purposes, and not for research. Therefore, the potential for error occurs at many points along the record keeping process.\(^14\) The implication for researchers is that both systematic and random error can occur in the identification of treatment exposure and outcome. In addition, there is currently no simple approach to link the health information of patients from separate data sources. A case in point is the inability of insurers to readily link laboratory results with patient information from separate pharmacy plans. In addition to logistical constraints, the risk for re-identification of patients increases as the amount of information increases.

### Infrastructural Requirements

#### 3.1 Data Standards

Lack of standardized data limits the analyst’s ability to efficiently process data, implement standard statistical packages, integrate analysis results, and report results with transparency. Progress in defining standards in secondary data is slow relative to the progress made in defining standards in the clinical trial world, which is largely due to the efforts of the Clinical Data Interchange Standards Consortium (CDISC).\(^5\)\(^6\) CDISC, founded almost 10 years ago, is “a global, open, multidisciplinary, nonprofit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata.” Standards established by CDISC are intentionally “vendor-neutral, platform-independent, and freely available,” and seek to optimize workflow from protocol authoring to final study reports and regulatory submission.

The extension of CDISC standards to secondary data is reasonable. CDISC’s Healthcare Link project, which started in 2005, is an initiative that specifically focuses on the mission of “interoperability between health care and clinical research.” This effort has established the capability “to collect relevant data from the EHR for critical secondary uses such as safety reporting (and bio-surveillance), clinical research, and disease registries.”\(^9\)\(^6\)

One example of standardization in secondary data used by major U.S. providers of administrative claims databases is the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), which is designed to code and classify diagnoses from inpatient and outpatient records. Prescription drugs and insulin products are coded using the National Drug Code (NDC) scheme that is maintained by the U.S. Food and Drug Administration (FDA).

Although the coding of diseases and drugs is highly standardized, more effort is needed in defining diseases through the specification of codes. Analyses of claims databases vary in their application of coding. For example, a patient with fibromyalgia may be identified as having either 1 or 2 medical claims with diagnosis code ICD-9-CM 729.1. Another definition may include a requirement that in addition to ICD-9-CM 729.1, the patient has filled at least 1 prescription for a drug indicated for fibromyalgia during a defined time period. In CER, standardized definitions of common diseases and conditions enable the comparison of results across studies.

The announcement by Google to retire Google Health in 2012 underscores the fact that creating a standardized infrastructure for needed health information is not an easy problem to solve. Google established Google Health in 2006 as a personal health information centralization service. The service allowed Google users to merge potentially separate health records into 1 centralized profile either manually or through partnered health services. Lack of widespread adoption was the reason provided by Google for abandoning the project.\(^17\)

#### 3.2 Computing Environment

A reliable and efficient computing infrastructure, including hardware, software, and support staff, is fundamental to the success of performing CER with secondary data sources.\(^18\) The computing environment must address data acquisition, storage, and integration in addition to housing analytical tools for data mining and analysis. The data need to be dynamically maintained over time with updated data and links to other data sources in the presence of increasing numbers of users. Understanding adherence and treatment patterns for patients on new drugs and treatments as they become available is an important capability. Therefore, it is important for suppliers of administrative claims databases to be able to provide adjudicated claims data in a timely fashion.

Data warehouses consisting of high-quality clinical trial data, administrative claims databases, and registries from various sponsors including industry, federal health agencies, and health care providers are possible and need to be established, maintained, and updated easily with new studies in a timely manner. The data do not necessarily need to be aggregated in a single warehouse and can remain in their existing secure environments using recent advances in database structure and high-speed computing to link across data sources. The U.S. Department of Health and Human Services (DHHS) is creating a “multi-payer claims database” that would combine claims data into a distributed warehouse from a range of public and private payers.\(^19\) The FDA is creating a similar infrastructure
for its Sentinel System, which will enable FDA to monitor the safety of drugs and other medical products with the assistance of a wide array of collaborating institutions from a range of academic medical centers, health care systems, and health insurance companies.20

The FDA sanctioned a clinical trial data repository known as Janus to enable FDA and the pharmaceutical industry to look retrospectively at clinical trial data and also prospectively to design future clinical trials.21 Janus is a highly structured data warehouse of clinical trial data based on the CDISC Study Data Tabulation Model and is characterized by containing information on large cohorts of patients.

Administrative claims databases are highly structured data warehouses and typically contain information on large cohorts of patients followed over long periods of time. More generally, secondary data sources are characterized by large numbers of records that require extensive data processing during analyses. For example, medication records may need to be sorted by patient and by prescription fill date in order to merge with outpatient visit records that also must be sorted by patient and diagnosis date. Even in the presence of highly structured data, other aspects of the data, such as the timing of office visits to a physician, may require processing large numbers of patient records for analyses. A retrospective database study including millions of patient visits, for example, may require summarizing cardiovascular events that occur at 3 and 6 months after the start of a drug therapy. The schedule of visits according to usual care practices necessitates establishing visit windows and extensive data processing to classify the cardiovascular events into the defined windows for analysis. In contrast, an RCT protocol visit schedule is aligned with its objectives, with visits occurring at periodic intervals to enable analysis of outcomes at pre-specified time points.

A data warehouse of secondary data sources needs to be accessible by all users, many of whom are performing intensive computations at the same time. Users should be able to generate extracts containing different types of secondary data, such as claims data or EHR or both, for further analyses both quickly and easily rather than having to rely on a small group that has extract responsibility. Adequate disk space can be an issue. A typical extract for a retrospective database study that involves 100,000 patients and their pharmacy claims, medical procedures, and clinical diagnoses that is generated by 1 analyst might be as large as 30 gigabytes.

Cloud computing has become a viable option for secondary data sources in health care in the past few years. The National Institute of Standards and Technology (NIST)22 defines cloud computing as a “model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” One example company developing this technology to provide a secure cloud-computing platform that is specialized for the health care industry is Explorys.23 This technology’s target is to perform queries on a data repository that consists of over 10 million patients and billions of clinical events in a Google-like manner, efficiently and quickly.

Data warehouses demand high performance analytical tools, methods, and best practices for data visualization, simulations, and analyzing data. Analytical tools should be flexible to be used with various secondary data sources with little or no modification. A software tool for querying data from a database vendor should be able to be used on data obtained from a payer without major effort.

An example of the direction of analytical software development for CER toward user access to data to generate queries and perform basic analyses is the selection of Thomson Reuters by the U.S. DHHS. The company was selected “to develop a secure, interactive tool that will enable researchers to perform comparative effectiveness studies without the need for professional computer programming.”24 While completing the project, Thomson Reuters will develop a pilot system linking multiple health care data sources. The company will test the pilot system by conducting 2 high-priority analyses on care delivery options for selected medical conditions.

Medical dictionary diagnosis and procedure coding browsers, such as EncoderPro25 from Ingenix and drug product browsers, should be used to establish common definitions of diseases and outcomes through diagnoses, procedures, and drugs. It is not uncommon for RCTs to have centralized adverse event coding so that investigator terminology is coded consistently from study to study, which is especially important for regulatory submissions. Tools to track projects can help so that similar research questions can be answered efficiently and in a consistent manner.

One example of analytical tool development is software evolved from the application of Classification and Regression Trees. Secondary data sources are useful to identify individualized patient subgroups where outcomes are optimal. Analytical software that identifies these subgroups has been developed and relies not only on access but also computing power.26

3.3 Good Practices and Quality Assurance. It is important to establish internal and external processes to ensure quality, efficiency, and transparency. Protocols for studies involving secondary data sources should be registered and study results should be posted in a similar fashion as RCTs on the U.S. National Institutes of Health registry and results database (ClinicalTrials.gov) of federally and privately supported clinical trials conducted in the United States and around the world.27 Increased transparency should reduce the potential bias of study sponsors and improve the acceptance of results from studies involving secondary data sources.
Quality guidelines exist to offer direction on good practices and assuring quality when using secondary data sources. Examples include the recommendations of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR)\textsuperscript{7-9} and the International Society of Pharmacoepidemiology,\textsuperscript{11} as well as the STrengthening the Reporting of Observational studies in Epidemiology (STROBE) statement,\textsuperscript{11} the recently published Good Research for Comparative Effectiveness (GRACE) Principles,\textsuperscript{28} and numerous other resources for evaluating nonrandomized studies of comparative effectiveness.\textsuperscript{29-31}

**Protection of Patient Privacy**

A framework to improve the infrastructure to collect and share secondary data should have a provision to address the privacy concerns of patients in a transparent and credible way, and in accordance with current applicable laws such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA, Title II).\textsuperscript{32} Ambiguity in this regard will limit the voluntary and active participation of patients, and also discourage health care providers to contribute data toward this effort. Researchers, patient advocacy groups, and legislators should work together to ensure that there is a viable consensus among the various stakeholders. The need for researchers to get access to critical data should be carefully weighed against patients’ rights to privacy.

In the study of rare diseases, which tend to affect vulnerable populations, particular attention should be paid to relevant policies and requirements relating to patient privacy. When the rare disease under study is associated with special prognoses or visible phenotypes, de-identification of data alone may not be adequate to guarantee anonymity. In these circumstances, processes should be in place to prevent re-identification of patients when combining data from alternative sources.

One major reason for inadequate participation by patients and excessive concern for privacy may be lack of awareness on the part of patients and providers about the underlying purpose of the CER initiative. Therefore, it would be worthwhile to make efforts to educate the public about the scientific merit of the CER initiative, and the implications for health care utilization. In this regard, institutions, such as the Patient-Centered Outcomes Research Institute, can play a constructive role by publicizing the overarching goals of CER vis-à-vis patients’ need for privacy.

**Discussion**

The lofty goal of CER to promote high-quality health care to patients can be achieved mainly through the acquisition of reliable scientific information that helps health care providers, patients, and policymakers to determine the most optimal strategy for health care delivery. This in turn is predicated on the establishment of reliable infrastructure for data access, analysis and reporting, particularly when the sources of information are nonrandomized studies.

In this paper, we considered relevant gaps that must be filled to address the issues, with particular emphasis on data standards, data quality assurance, data warehouses, software requirements, and protection of privacy and confidentiality. There are needs to develop tools to readily access and correctly analyze the data; satisfy requirements to formulate guidelines to enhance quality and transparency; establish data standards using RCTs as a role model; and create data warehouses that respect the privacy and confidentiality of patients. Further, the infrastructure should leverage cutting-edge technology and permit implementation of state-of-the-art data analytical tools.

Given the scope of the problem, strong collaboration among stakeholders is critical to address the issues effectively and efficiently. This may involve establishment of processes to link and share databases, and the harmonization of hardware and software to facilitate the exchange of information among various health care entities. The collaboration may also need to involve the creation of a framework to overcome logistical impediments as well as proprietary constraints to access of information for effective systematic reviews and analyses of randomized and nonrandomized studies involving RCTs and secondary data sources. In this respect, the Observational Medical Outcomes Partnership (OMOP), which draws on the resources of the pharmaceutical industry, academic institutions, nonprofit organizations, the FDA, and other federal agencies, may serve as a model of a viable public-private collaboration.\textsuperscript{33}

**Conclusions**

Secondary data, such as registries and retrospective databases, are often considered to complement randomized clinical trials since they are less costly and can be used to incorporate real-world experience to answer important health care questions. Analyses of secondary data provide a relatively efficient means for addressing hypotheses regarding adherence, treatment patterns, and burden of disease. However, effective use of secondary data requires addressing major methodological and infrastructural issues, including development of analytical tools to readily access and analyze data, formulation of guidelines to enhance quality and transparency, establishment of data standards, and creation of data warehouses that respect the privacy and confidentiality of patients. This paper described infrastructural requirements for secondary data utilization in the context of comparative effectiveness research and identified gaps that must be filled to address the underlying issues, with emphasis on data standards, data quality assurance, data warehouses, computing environment, and protection of privacy and confidentiality.
Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research

DISCLOSURES
This supplement was funded by Pfizer, Inc. Alemayehu and Mardekian are Pfizer employees.
Alemayehu and Mardekian contributed equally to writing and revision of the manuscript.

ACKNOWLEDGEMENTS
Margaret McDonald, PhD; C. Daniel Mullins, PhD; Robert J. Sanchez, PhD; and Richard J. Wilke, PhD; provided valuable input to various draft versions of the manuscript.

REFERENCES

Authors
DEMISSIE ALEMAYEHU, PhD, is Executive Director, OR and Disease Area Statistics Head, and JACK MARDEKIAN, PhD, is Senior Director, OR Statistical Scientist, Pfizer Inc, New York, New York.
Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research


Observational studies are used to inform health care policy and decision making when comparable data from randomized controlled trials (RCTs) are inadequate or unavailable due to ethical reasons, practical considerations, and other logistical issues. The need for evidence from observational studies is particularly relevant in comparative effectiveness research (CER) given the large evidence gaps that exist regarding the comparative effectiveness and value of a broad array of treatments. Furthermore, CER may require data from different sources, including RCTs, nonrandomized studies, and systematic reviews.3

It is generally accepted that RCTs are the gold standard for generating evidence pertaining to the benefits and risks of medical treatments. A major advantage of RCTs is that by design, the experimenter is able to control for selection bias. The assignment of study subjects through a random mechanism ensures comparability of the treatment groups with respect to both known and unknown confounding factors. This implies that any difference between groups before randomization is attributable to chance alone. The latter in turns permits the application of standard inferential procedures to draw conclusion about treatment efficacy in the trial population.2 However, results of RCTs often do not provide evidence of comparative effectiveness because clinically important active comparators were not selected by the study designers or because of other design limitations.

Even when there are comparative effectiveness data from RCTs, the data may be inadequate to address all relevant decisions. The conditions under which the trials are conducted may not reflect the real-world setting or important subpopulations. Under these circumstances, it may be necessary to rely on nonrandomized studies to inform medical decision making.

Use of observational studies, however, requires a careful consideration of important conceptual and practical issues. From a design perspective, the absence of random assignment of subjects to treatments almost always introduces selection bias that confounds the relationship between treatments and outcomes. More specifically, in the absence of randomization, study subjects use treatments dictated by factors, other than chance, that have the potential to confound outcomes. This problem results in imbalances with regard to known and unknown confounding factors that may influence the outcome of interest. For measured covariates, there are statistical approaches to mitigate the bias introduced by the imbalances. However, the problem is more challenging for important covariates that may not exist in the dataset. Thus, the standard inferential procedures are likely to lead to invalid conclusions, if applied uncritically to such data.3

With the growing awareness of the importance of data from nonrandomized studies in making critical health care decisions, considerable progress has been made in recent years in establishing guidance for best practices in the design, analysis, and reporting of observational studies.4-9 In this paper, we consider some of the major statistical issues that arise in the analysis of data from observational studies, with particular reference to the limitations of existing approaches, and recent methodological developments aimed at addressing bias introduced by unmeasured or latent confounders.

**Bias in Nonrandomized Studies**

There are several ways in which bias may arise in nonrandomized studies. Bias can arise as a consequence of systematic measurement error or misclassification of subjects on 1 or more of the explanatory or response variables. Another important type of bias is one that is intrinsic to observational studies, often referred to as selection or channeling bias. Since assignment to treatment is not random, the channeling of individuals into treatments results in imbalance with respect to relevant attributes. From a methodological perspective, the bias that results from imbalance of known and unknown risk factors is of particular interest, and will be the focus of the next 2 sections.

In the absence of randomization, differences in apparent treatment effects may be attributable to pretreatment differences in risk factors among subjects receiving the intervention groups being studied. For overt biases emanating from known covariates, there are established methodological approaches aimed at removing bias through appropriate matching and regression analysis. When the bias is hidden (i.e., caused by risk factors that have not been measured), the problem is generally complex, and the analytical procedures are not as well developed.

Although there has been considerable methodological progress in addressing both overt and hidden biases in observational studies, all the available techniques have certain limitations that require careful assessment to ensure the validity of the results for particular applications. In the next section, we review some of the commonly used approaches and highlight their limitations and other relevant features. It is essential for each investigator to carefully and thoroughly assess the potential biases in each proposed study and tailor the methods or combination of methods to best address these biases, while recognizing the general limitations of observational research relative to RCTs.

**Traditional Analytical Approaches**

In this section, we consider adjustment techniques, including
matching, stratification, and analysis of covariance, generally employed for overt biases, and instrumental variable procedures that are typically used for hidden biases. The emphasis will be on nontechnical aspects of the procedures, without delving into their mathematical formulations (see Johnson et al. for a review of such techniques).

Methods for Overt Bias

Matching. A common approach to adjust for overt biases is matching, which involves comparing each individual in the treated group with 1 or more subjects in a comparison cohort with respect to observed covariates that are known to confound the relationship between treatment and outcomes. When performed properly (e.g., with appropriate and adequate matching criteria), the procedure has the dual advantage of improving the precision of estimators as well as reducing the overt bias.

Propensity Score. One way of achieving balance among the treated and comparison groups with regard to the distributions of observed covariates is through propensity score analysis, which involves quantifying the conditional probability, given the covariates, that a subject receives the treatment rather than the control. It has been long established that when interest is in balancing treatment groups on all observed covariates, it is sufficient to balance on the propensity scores. The propensity score is particularly useful when the number of covariates is large and matching is not practical. However, matching or adjusting for propensity scores does not solve the problem of hidden biases. Further, the validity of the propensity score matching is heavily dependent on the adequacy of the model used to estimate the scores. It is, therefore, necessary to check whether balance has been achieved in the distributions of observed covariates, and to update the model, as appropriate, through inclusion of interaction or other higher order terms in the logit model. In a recent study, Basu et al. showed that in moderate sample sizes, balancing on estimated propensity scores may fail to balance higher-order moments and covariates among covariates and that the usual inverse-probability weighting in propensity scores may be sensitive to misspecification of the model for estimating propensity scores.

Implementation of propensity score methods in the medical literature has been a subject of some scrutiny that can be illuminating—in a “what not to do” sense—for those planning to use such methods; see Weitzen et al. and Austin for critical reviews. D’Agostino provides a useful tutorial and some basic SAS (SAS Institute Inc., Cary, NC) code for creating propensity scores. Basu provides an interesting overview and empirical comparison of 7 different methods of creating propensity scores.

Stratification. Stratification attempts to create balance between control and study drug subjects by matching subjects as groups rather than pairs. Stratification may be achieved based on 1 or more known covariates. When there are several covariates, suitable cut-off points (e.g., quintiles) of a propensity score may be employed to define strata. Optimal stratification strategies are available to ensure that subjects in a given stratum are as similar as possible. In general, stratification is known to reduce bias and enhance precision of estimates considerably. However, the value of stratification is reduced by the often arbitrary way strata are defined. The available approaches to determine an optimal stratification are not commonly used in routine applications.

Model-Based Approaches. An alternative to matched sampling and stratification is use of suitable models, such as analysis of covariance, to estimate treatment effects adjusting for observed covariates and/or propensity scores. The performance of model-based adjustments is, of course, dependent on the accuracy of the model and validity of model assumptions. In fact, when there is significant departure from model assumptions, the procedure may increase bias rather than reducing it. Accordingly, a combination of matching and model-based adjustments may be preferred.

Methods for Hidden Biases

Without randomization, hidden biases might result from imbalances between treatment groups with respect to important covariates that were not observed by the investigator. Such hidden biases are likely to distort the conclusions of observational studies. While traditional propensity scoring can only condition on observed confounders and cannot deal with unobserved confounding, traditional instrumental variable methods can not only condition on observed confounders but also average over unobserved confounders, thereby addressing hidden selection biases in observational data. Below, we discuss some of the measures that may be taken to mitigate consequences of hidden biases.

Instrumental Variables. A method that is borrowed from econometrics is instrumental variable analysis, which involves identifying 1 or more variables (instruments) that are highly correlated with treatment but are unassociated with other confounders and have no direct effect on the response variable. In RCTs, an obvious instrument is the randomization mechanism. In observational studies, common instruments include prescriber preference and the distance a patient has to travel to a hospital or site of care. Suppose \( E[Y|Z=z] \) is the average value of the response \( Y \) for all subjects with values for an instrument \( Z=z \). A measure of the effect of treatment \( X \) on \( Y \) may be given by:

\[
\beta = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}
\]

In RCTs where \( Z \) is an indicator of random assignments,
a Wald estimator of $\beta$ corresponds to an intention-to-treat (ITT) estimator, while when $Z$ is an instrumental variable in observational studies, it corresponds to the instrumental variable estimator. A common approach to instrumental variable estimation involves 2-stage least squares, in which 1 model (generally probit or ordinary least squares [OLS] regression) is specified for the treatment assignment process that depends on the instrument and potential confounding variables, and a second for the outcome that includes the predicted probability of treatment from the first stage and the additional covariates that are included in $Y$.27

A major drawback of instrumental variable techniques is that suitable variables frequently are not available. Even when such variables are available, it is often difficult to assess the validity of the underlying assumptions. For example, if the instrument is weakly correlated with treatment, the resulting treatment effect estimate may be biased.28,29 In addition, the estimators may be inefficient relative to OLS when the instrument is redundant.30 For further discussion about instrumental variable techniques, see references 25 and 31-33.

One should note that even with the successful implementation of instrumental variable methodology, the interpretation of the results is limited to what is called the local average treatment effect.31 This local average could apply to a small proportion of the study population, so-called marginal patients who are defined as the subset of patients whose treatment choices vary with the instrument. In the case where the instrumental variable is a binary indicator of distance from a hospital offering a particular treatment or procedure, this local average treatment effect pertains only to the comparison between patients who received the treatment because they lived relatively close to a hospital offering the treatment and those who lived further away but would have received the treatment had they lived close by. If one were to use a different instrumental variable, the resulting treatment effect would be different because it would apply to a different group of marginal patients.

Sensitivity Analysis. A general approach to assessing the impact of unobserved confounders involves sensitivity analyses that attempt to quantify the degree to which hidden bias would explain any observed association. For a discussion of alternative methods of sensitivity analysis, see references 34-36.

Pattern Specificity. Pattern specificity is a technique employed to detect hidden biases or to reduce sensitivity to hidden biases, and is based on the fact that observational studies are variable in terms of their sensitivity to hidden bias. Typically, latent biases tend to leave “visible traces” in observed data37 and the approach involves distinguishing real treatment effects from hidden biases.37-40

Recent Developments and Future Directions

Individualization in CER. Basu discusses the need to individualize comparative effectiveness research.43 Although a rich array of biomarkers is usually required to generate individual-level treatment effects, Basu proposes 2 methods that can be used to learn about treatment effect heterogeneity even in the absence of such biomarkers.43,44 Both methods estimate treatment effect heterogeneity conditional on individual level confounders, some of which are observed in the data and the remaining are unobserved. The first is a method of local instrumental variable (LIV) that addresses limitations of traditional instrumental variable approaches.42,44 LIV methods attempt to leverage this selection and allow for unobserved confounders to be moderate treatment effects. Therefore, they can be used to estimate marginal treatment effects that are conditional on both observed and unobserved confounders. Such marginal treatment effects can also be estimated using a second method that uses latent factors to proxy for the unobserved confounding.41 The data requirements are different for the alternative methods, and usually careful nonparametric identification is required to make sure that the methods are estimating the relevant parameters.

Bias Adjustment through Prior Event Ratio. Tannen et al. introduced a technique they dubbed prior event rate ratio (PERR) to adjust for hidden confounders in the analysis of data from electronic medical record databases.45 The adjustment involves knowledge of event rates in the 2 groups prior to initiation of the interventions. While the technique worked reasonably well to identify and reduce the effects of unmeasured confounding when applied to cardiovascular outcomes considered in the study, the procedure requires strong assumptions about constant temporal effects, absence of confounder-by-treatment interaction, and nonterminal events as outcome. However, these issues are present to some degree in other estimators, and the PERR technique can provide a useful alternative approach in CER estimation.

Bayesian Inference for Observational Data. Despite the growing body of literature on the role of Bayesian statistics in the analysis of observational studies, the potential is not fully realized among practitioners. The application may range from sensitivity analysis for unmeasured confounding in observational studies46 to covariate adjustment based on a Bayesian propensity score.47 Additional information may be found in references 48-50.

Meta-Analysis of Observational Studies. In addition to the known issues with the synthesis of data from RCTs, meta-analysis of observational studies requires a careful assessment of problems peculiar to such studies.51-52 Accordingly there have been efforts to establish good practices for the
reporting of meta-analyses of observational studies. Central to the proposed guidelines is the need to have a strategy for addressing potential confounding in the primary studies.

Discussion

Well-conducted observational studies are useful for CER. When RCTs are inadequate for decision making, observational databases can provide relevant information from the real-world setting in a timely manner. However, effective use of data from nonrandomized studies requires overcoming significant conceptual and technical issues. In this paper, we highlighted some of the available statistical methods that can be used to mitigate the effects of overt and hidden biases, with emphasis on limitations of the approaches and opportunities for further research. A major issue with the analysis of observational data is the preservation of privacy. Accordingly, there are laws and regulations such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA, Title II) that govern the transmission and use of such data. For pooling de-identified data on the patients from alternative sources, probabilistic record linkage or similar machine learning techniques may be used. The techniques typically are computationally intensive and involve identification of similar groups of records, relative to predefined criteria, and then evaluating the likelihood that the records belong to the same patient. As an integral part of the methodological consideration, parallel efforts must also be exerted to enhance other aspects of the studies, including sound design, pre-specification of analytical strategy, high-quality data, and appropriate reporting of results.

Conclusions

When RCTs are inadequate or unavailable, observational studies may play useful roles in addressing major health care questions. However, the validity of analytic results from observational studies is adversely impacted by biases that may be introduced due to lack of randomization. In this paper, we reviewed some of the methodological challenges that arise in the analysis of data from nonrandomized studies, with particular emphasis on the limitations of traditional approaches and potential solutions from recent methodological developments.

REFERENCES


ACKNOWLEDGEMENT

The authors are grateful to C. Daniel Mullins, PhD, for his constructive comments on a draft version of the manuscript.

DISCLOSURES

This supplement was funded by Pfizer. Alemayehu, Alvir, and Willke are Pfizer employees. Jones was a Pfizer employee during the production of the manuscript.

The 4 authors contributed equally to writing and revision of the manuscript.

Authors

DEMISSIE ALEMAYEHU, PhD, is Executive Director, OR and Disease Area Statistics Head; JOSE MA. J. ALVIR, DrPH, is Senior Director, OR Statistics; BYRON JONES, PhD, is currently Biometrical Fellow at Novartis, formerly Senior Director, Statistical Research Consulting Center; and RICHARD J. WILLKE, PhD, is Head, Global Health Economics & Outcomes Research, Global Market Access, Primary Care, Pfizer Inc., New York, New York.
Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research

Demissie Alemayehu, PhD; Robert J. Sanchez, PhD; and Joseph C. Cappelleri, PhD

Comparative effectiveness research (CER) involves studies that generate evidence through an evaluation of the spectrum of health care interventions and services that reflect patient choices for a given clinical situation, with the intent of improving patient and physician decision-making. In this paradigm, CER can be defined as a rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such studies may compare similar treatments, such as competing drugs, or they may analyze very different approaches, like surgery and drug therapy. To date, the areas of emphasis in CER have primarily been on clinical endpoints, with extensive work in mixed and indirect treatment comparisons, use of Bayesian approaches, simulated treatment comparisons, real-world data use, and therapeutic index determination.

Despite the potential role of patient-reported outcomes (PRO) data in CER, a central role for PRO data has not yet been fully established in CER because of the challenges associated with the collection and interpretation of such data within and across studies. A PRO is any report on the status of a patient’s health condition that comes directly from the patient. PROs can help to identify areas (e.g., functioning, well-being, symptomatology, and satisfaction) that are most important to the patient’s health status. Thus, data generated by a PRO instrument can provide evidence of a treatment benefit or harm to decision makers if presented in the context of the value to patients as determined from a PRO and how this translates to improved health-related quality of life. PROs can help to identify areas (e.g., functioning, well-being, symptomatology, and satisfaction) that are most important to patients in a specific disease area and allow for frequent and complementary information to other clinical endpoints for use in CER. For example, in oncology, the interpretation of progression-free survival may be made more meaningful to decision makers if presented in the context of the value to patients as determined from a PRO and how this translates to improved health-related quality of life.

Role of PROs in CER

As stated by the Institute of Medicine (IOM), a primary purpose of CER is “…to assist consumers, clinicians, purchasers, and policy makers to make the informed decisions that will improve health care at both the individual and population levels.” By definition, PROs are measurements of a patient’s health status that come directly from the patient, without any interpretation of the patient’s responses by a physician or anyone else. Therefore, utilization of PRO data meets the criteria for IOM’s stated purpose of CER. For example, since 2009, the National Health Service (NHS) has required that all providers of NHS-funded care collect PRO measures (PROMs) for certain conditions to measure quality from the patient’s perspective. The PROMs can then be used to help patients and general practitioners exercise choice.

Within the realm of PRO research, there are numerous validated instruments that appropriately and accurately measure different domains of health from the perspective of the patient. The choice of a PRO instrument is contingent on the research question and the population under study and can either be generic or disease-specific. A partial list of a variety of common PRO instruments is described elsewhere. Briefly, they include generic instruments, such as the Sickness Impact Profile, Nottingham Health Profile, Medical Outcomes 36-item Short Form, and EuroQol; while disease-specific instruments include such instruments as the European Organisation for Research and Treatment of Cancer QLQ-C30 and its disease or treatment-specific modules, Functional Assessment of Cancer Therapy (General) and its specific disease or treatment-specific modules, and Rotterdam Symptom Checklist. The most commonly used and cited instruments in clinical practice include the Medical Outcomes 36-item Short Form and the Dartmouth Primary Care Cooperative Information Project (COOP) Charts, both of which are generic instruments, and the Sexual Health Inventory for Men, a disease-specific instrument.

PRO instruments typically capture concepts related to how a patient feels or functions and help establish the burden of illness and impact of treatment on one or more aspects of the patient’s health status. Thus, data generated by a PRO instrument can provide evidence of a treatment benefit or harm from the patient’s perspective and can provide supplementary and complementary information to other clinical endpoints for use in CER. For example, in oncology, the interpretation of progression-free survival may be made more meaningful to decision makers if presented in the context of the value to patients as determined from a PRO and how this translates to improved health-related quality of life. More generally, PROs can help to identify areas (e.g., functioning, well-being, symptomatology, and satisfaction) that are most important to patients in a specific disease area and allow for frequent and
longitudinal assessments on several self-reported aspects pertinent to the disease and treatment. Regardless of the instrument chosen, PROs have the potential to play a critical role in CER directly and contribute to the patient's role in the decision-making process.

One natural question is, when are PROs worth the time and cost to collect in CER? The answer depends on providing a sufficient background to justify the resources required for an investigation of PROs in CER. The rationale should provide answers to questions such as, how exactly might the results from PROs affect the clinical management of patients in a given specific clinical situation? And, how will the PRO results be used in determining the benefits and harm of different treatments? The justification should include a motivation for the particular aspects that the PROs are measuring and how these aspects relate to the disease, treatment, and impact on patient and physician decision making.

In CER, a major objective is the establishment of the relative effectiveness of a range of treatment options. In this regard, the PRO instrument selected should be sensitive enough to differentiate among competing interventions of interest. In addition, use of PRO measures in CER may play a critical role in the assessment of heterogeneity of treatment effects. For example, baseline PRO values may provide useful information about subgroup differences in ways not captured by other baseline clinical variables. In the context of CER, acknowledgment of these considerations (among others) is essential for making optimal treatment choices for individuals and patient subgroups.

**Considerations for Use of PRO Data in CER**

For effective integration of PROs in a CER initiative, it is essential to establish a robust conceptual, analytical, and operational framework that addresses issues pertinent to such data. In this section, we outline a few points for consideration, including standardization of instruments, meta-analytic issues peculiar to PROs, and communication and reporting of results.

**Standardization of Instruments for a Given Therapeutic Area.** Different interventions often use different specific instruments, and this generally poses analytical and conceptual challenges when it is necessary to synthesize available data for comparative purposes. Effective use of PROs in CER will, therefore, presuppose establishment of standard instruments and criteria for a specific therapeutic area. This in turn entails addressing significant operational, theoretical, and methodological issues.

From an operational standpoint, if the CER goal involves inclusion of a PRO component in a trial, it is essential to integrate the PRO protocol into the initial overall plan for the trial, and to determine which PRO concept is important to assess in a particular therapeutic area. When assessing patient benefit, IQWiG, for example, applies criteria that are important to patients by consulting with patient representatives in order to establish patient-relevant outcomes. In fact, as part of its responsibilities and objectives, IQWiG stipulates that results important for patients need to be assessed when evaluating the benefits of interventions (www.iqwig.de). However, it is not clear how to determine which endpoints are of importance and their hierarchy of importance to patients, as well as the extent of how newer endpoints add new information different from traditional areas of symptoms, function, health-related quality of life, and satisfaction with care. This has led some researchers to consider use of interpretative phenomenological analysis, analytic hierarchy process, and conjoint analysis to aid prioritizing patient outcomes based on patient preferences.

Certainly, to advance the use of PRO data in CER there needs to be globally accepted measures that can be used within a therapeutic area rather than individually developed PROs for specific therapy. The Critical Path PRO Consortium (http://www.c-path.org/) is leading the way in endeavoring to develop standardized signs and symptoms measures across a wide range of diseases, such as Alzheimer's disease, oncology, and depression. Standardized measures will allow easier comparison across therapies, especially if meta-analyses are to be utilized. Also, a standardized measure for a given construct within a therapeutic area will make interpretation of what a change or difference is between treatments more understandable. At the same time, it is important to encourage the development of new PRO instruments and to enhance and improve existing PRO instruments as research and new evidence evolve.

From a theoretical perspective, if a new PRO instrument needs to be created for a given therapeutic area that is the focus of a CER platform, then a robust and theory-based conceptual framework for the PRO must be established, linking the desired outcome to the concept of interest and subsequently linking that concept to the specific symptoms or latent variable being measured. In the process, considerable input must be obtained from patients, as is customary, using focus groups and cognitive interviews to establish face and content validity and ensuring that the instrument covers what patients consider important outcomes. Additionally, exploratory factor analysis and confirmatory factor analysis should also be conducted to examine the factor structure of which items go with what domains (construct validity). In accordance with standard procedures in instrument development, psychometric methods should be applied to test reliability, validity, and responsiveness of the PRO measure. For PROs intended to be used in the real-world setting, it is also important to keep PROs short and simple since, unlike in routine clinical trial settings, study nurses and monitors are not available to ensure proper completion of PRO instruments. Further, to effectively address the objectives of CER, the PROs need to be sensitive enough to distinguish among alternative treatment options and to enable...
assessment of heterogeneity of treatment effects.

From a methodological perspective, item response theory (IRT) using computerized adaptive testing (CAT) is another approach to PRO standardization in CER. For example, PROMIS (Patient-Reported-Outcomes Measurement Information System) is a National Institute of Health (NIH) Roadmap network project (information available at: http://www.nihpromis.org/default) intended to standardize PROs and to improve their reliability, validity, and precision for chronic diseases. This large-scale initiative also aims to provide definitive new instruments that will exceed the capabilities of classic instruments and enable improved outcome measurement for research. IRT models allow the reduction and improvement of items according to a single (unidimensional) concept. Item banking uses IRT methodology and models to develop item banks from large pools of items from many available questionnaires. CAT provides a model-driven algorithm and software to iteratively select the most informative remaining item in a domain until a desired degree of precision is obtained. Through these approaches, the number of patients required for a study may be reduced while holding statistical power constant. These PROMIS tools are expected to improve precision and enable assessments that are specifically tailored to the individual patient level, which should broaden the appeal of PROs in CER.

If the CER analytic plan involves use of an existing instrument for diverse population groups, appropriate modifications should be considered to ensure that it is valid in the populations being studied. Once a therapeutic area-specific PRO is established, the standardization should include a determination of how much of a response should be meaningful. In particular, the amount of change that will be considered a clinically meaningful response should be defined, and consistent approaches should be employed to compare patients receiving alternative treatments for the therapeutic area of CER interest.

Another methodological consideration is the mode of administration of PROs used in CER. With the recent advancement in technology, there has been considerable interest to adapt paper PROs into electronic format (ePROs), due to the many advantages of ePROs including less administrative burden, higher patient acceptance, avoidance of secondary data entry errors, easier implementation of skip patterns, and more accurate and complete data. For purposes of registration studies, the U.S. Food and Drug Administration (FDA) has given guidance on the use of PROs in clinical studies and has raised specific issues about the comparability of paper PROs versus ePROs, with particular reference to minimization of measurement error within a study. In the context of CER, the emphasis is on standardization of the mode of administration. If the decision is made to use an ePRO in a CER study then it is necessary that all sites (and patients) have access to a computer to minimize the combination of paper and ePRO use.

Synthesis of Data from the Literature. The wide scope of CER requires synthesis of data from alternative sources. In the context of clinical endpoints, much work has been done to extend traditional meta-analytic techniques to address CER needs. When data are not available from head-to-head comparative trials involving PROs, the feasibility of network meta-analytic techniques would need to be explored. Network meta-analysis is a statistical technique that combines trials involving different sets of treatments, using a network of evidence, within a single analysis. This integrated and unified analysis may provide a defendable, digestible answer to a question relevant to a decision maker.

The multiplicity of endpoints, discussed below, and differences in outcome measures may pose additional obstacles in extending the available methods to the analysis of PRO data for use in CER. While Bayesian procedures are often proposed as a viable alternative in general, their use with PROs has not been extensively studied. A central issue with pooled data analysis of aggregate (study-level) data, of course, is the assessment and handling of study-level heterogeneity. Given the nature of PRO instruments, the problem may be even more important with PRO studies than traditional clinical endpoint synthesis. Specifically, cultural, geographic and other socio-economic variables may contribute to lack of consistency of PRO results across sources of information, subgroups and other categories, especially if data from pragmatic trials are to be used. Despite the unique challenge presented by PROs, the usual approaches should still be applied to investigate the presence of heterogeneity and to mitigate any potential bias. As a matter of good practice, subgroup definitions and sensitivity analyses should be preplanned, and appropriate statistical procedures for heterogeneity be performed when applicable. If relevant study-level information is available, modeling techniques (e.g., meta-regression) may be used to adjust for imbalance in potential confounders, while recognizing the limitations of such approaches (e.g., the ecological fallacy with meta-regression). It is generally advisable to assess the consistency of results by performing sensitivity analyses. For example, a cumulative meta-analysis, which shows how the summary effect and variance shift as studies are added to the analysis, can be part of a sensitivity analysis. However, the most informative data, when available, involve the meta-analysis of individual patient data from all the available studies addressing the same question.

When synthesizing data from studies in which different scales are used for the same disease and treatment comparison, each study’s treatment effect can be converted into a standardized mean difference so that the combined treatment effect is expressed in terms of standard deviation units. According to an arbitrary but commonly used interpretation of effect size by Cohen, such standardized mean effect sizes of 0.2, 0.5, and 0.8, for example, indicate small, moderate, and large effect sizes,
respectively. However, this approach loses the ability to draw inferences on the original scale of measurement and may lose its appeal for CER, where standardization and interpretation of instruments are key considerations.

**Communication of PRO Data in CER.** The Patient Protection and Affordable Care Act in the United States, which authorized the formation of the Patient-Centered Outcomes Research Institute (PCORI), includes a key provision relating to the reporting of CER results. More specifically, PCORI is mandated with the dissemination of CER “research findings with respect to the relative health outcomes, clinical effectiveness, and appropriateness of the medical treatments, services, and items.” In addition, PCORI “…shall ensure the findings are conveyed in a manner comprehensible and useful to patients and providers in making health care decisions; discuss considerations specific to certain subpopulations, risk factors, and co-morbidities, as appropriate.”

In the light of the above provision, the dissemination of PRO results should be executed to address the needs of the various stakeholders, which include patients, payers, policy makers, and other health care providers. It is imperative that the end user of health care—the patient—be well-informed about the health state that different treatment options yield. For example, as mentioned previously, results of PRO data should state major findings relating to symptom-free days, percentage of persons experiencing improvements, percentage of persons experiencing a loss of function, and the length of time required to experience an important change. This dissemination will ultimately lead to better and more informed decision making that results in the appropriate use of health care resources and dollars.

Thus, PROs are directly wedded to PCORI’s mission on patient-centered outcomes research that is designed to inform health care decisions by providing evidence on the benefits and harms of different treatment options for different patients. This research recognizes that the patient’s voice should be heard in the health care decision-making process. PCORI research is charged with being responsive to the preferences, values, and experiences of patients in making health care decisions, as well as with highlighting the impact that diseases and conditions can have on daily life. Patient-reported outcomes are often relevant in studying a variety of conditions—such as pain, erectile dysfunction, fatigue, migraine, mental functioning, physical functioning, and depression—that cannot be assessed adequately without a patient’s evaluation and whose key questions require patient input on the impact of a disease or a treatment (after all, who knows better than the patient herself?) It is this broad and indispensable application of PROs that make them a critical part of CER.

**General Issues with PRO Data Analysis**

Effective incorporation of PRO data in CER, however, would require a thorough understanding and surmounting of inherent conceptual and methodological challenges, including establishment and use of standardized instruments, reliability and validity testing of new instruments, and handling of such technical, conceptual, and operational issues as multiplicity of endpoints, missing values, and definitions of a clinically important difference and responder criteria.

In PRO data analysis, multiple endpoints are naturally of interest as a consequence of the intrinsic design features of the instruments used to generate the data. In CER, multiple endpoints pose additional problems, since interpretation of results may be complex when the goal is to compare a range of treatment options. From a statistical perspective, the multiplicity issue is of particular relevance since multiple testing can result in inflation of false positive rates (i.e., falsely concluding statistical significance) and can incite problems with result interpretation. The available approaches generally depend on research objectives, endpoints, decision rules, and other factors. In addition to standard statistical techniques (e.g., step-down, step-up, and other gatekeeping procedures), other approaches for PRO analysis in CER may include suitable definitions of composite endpoints when a PRO measure includes multiple domains. Although the latter is intuitively appealing, it also has its own drawbacks, since it implicitly assumes that individual components are of similar importance.

Another aspect of using PRO data in the real-world or CER setting is the greater likelihood, compared with clinical trials, that a subject will not answer all questions in a given instrument. It therefore becomes important to examine the data for missing values. While missing data problems are not unique to PROs, missing data may arise in several ways. For example, observations may be missing for an entire patient, an entire domain, or for specific items within domains. What is more important than the missing values is the pattern of the missing data. If the data that are missing are random, then techniques can be employed to correct the problem (e.g., multiple imputation). Conversely, if the missing data are nonrandom, the generalizability and perhaps the validity of the results can be in question. In this case, appropriate techniques should be used to determine if the missing data are random or nonrandom.

In CER, it is essential to know how to interpret scores on a PRO so that they have meaning and clinical importance. The PRO has to be readily interpretable to the patient, as well as to health care providers, policy makers, and payers. Traditional approaches to determining a clinically important difference (CID)—anchor and distribution based approaches—should be supplemented and taken a step further to relate the CID to other relevant parameters, such as symptom-free days, percentage of persons experiencing improvements, percentage of persons experiencing a loss of function, and the length of time required to experience an important change.

Several strategies have been proposed for the interpretation of scores from PROs. Among the more recent ones are
responder analysis and a cumulative distribution function.11 Other approaches intended to enrich interpretation of PROs have been published.48-53 In the context of CER, a preferred approach is to use a measure of effect that facilitates the pooling of information from disparate instruments and studies.

**Discussion**

With the establishment of PCORI, CER activities should take on a patient-centered focus. The relevant literature on generating and translating PROs is growing, and new areas are being explored and tested to establish a solid methodological and analytical framework for effective use of PRO data to influence health care decision making and formulary coverage. Although the focus within CER heretofore has tended to be on traditional clinical endpoints, there is a realization that PROs as specialized clinical endpoints also have a unique place in CER. Given the fact that the patient is at the center of all treatment and policy decisions affected by CER initiatives, PROs are expected to be an integral part of CER strategic initiatives in the near future.

To ensure that PROs play an effective complementary role to traditional clinical endpoints in CER, it is essential to understand the issues that are inherent in such data and to put in place processes to guide researchers and other stakeholders. In particular, standardization of PRO instruments should be given primary focus, as well as consideration of optimizing implementation to address potential issues with missing data. Further work on multiple testing (and its accompanying risk of false-positive findings) and how best to address it, is also necessary. Existing statistical approaches employed in the synthesis of available clinical information for use in CER should be adapted to the analysis of PRO data, and new techniques should be explored to tackle problems that are particular to PROs. Lastly, an effective CER strategy should also address the communication of PRO results to relevant stakeholders with clarity, transparency, and fair balance.

**Conclusions**

PRO data can play a critical role in guiding patients, health care providers, payers, and policy makers in making informed decisions regarding patient-centered treatment from among alternative options and technologies and have been noted as such by the newly formed PCORI. However, collection and interpretation of such data within the context of CER has not yet been fully established. In this paper, we discussed some challenges with including PROs in CER initiatives, provided a framework for their effective use, and proposed several areas for future research.
Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research


Developing a Collaborative Study Protocol for Combining Payer-Specific Data and Clinical Trials for CER

Robert J. Sanchez, PhD; Jack Mardekian, PhD; Mark J. Cziraky, PharmD; and C. Daniel Mullins, PhD

The demand for comparative effectiveness research (CER) by health care providers and payers represents new opportunities for the U.S. government, research organizations, and pharmaceutical companies to generate “meaningful evidence” for use in medical decision making.1 CER studies conducted with a payer perspective should develop questions, select outcomes, and utilize data that are applicable to the payers themselves for use with their formulary and reimbursement decision-making processes. CER studies for prescribers should be designed and implemented to inform evidence-based therapeutic guidelines, providing actionable information from their everyday practice use. The challenge is how to conduct CER studies that satisfy the simultaneous requirements of scientific rigor and applicability to the respective decision makers. One solution is to address the demand for “real-world” data (RWD) by involving decision makers and other key stakeholders early on in the development of the research designs and implementation of study protocols when conducting CER studies. RWD have been defined “as data used for decision-making that are not collected in conventional RCTs” (randomized controlled trials);2 therefore, the ability to gather input from the payer is essential to ensure collected endpoints are applicable to the decision makers themselves.

RCTs are considered the “gold standard” for providing evidence about a product’s efficacy and are the basis for supporting formulary decision making. While the internal validity of RCTs is well known and established, the controlled protocols of RCTs may not have the desired level of external validity for a managed care organization’s (MCO) population. Consequently, health care decision makers are examining other sources of data to supplement RCTs for their health care coverage policies. Health care providers and payers use available evidence from both RCTs and RWD sources to decide whether a particular drug product offers tangible clinical benefits and value compared with existing therapies. Improving medical outcomes and providing positive impact on health care expenditures are shared goals of providers, payers, and the pharmaceutical industry.3

Developing CER Studies to Inform Payer Decision Making

Payers are interested in CER results and evidence-based value assessments of comparator therapies to use in their coverage decision-making processes. Some have proposed that CER involving systematic reviews of effectiveness evidence could improve the coverage and reimbursement processes.4 However, now more than ever, there is a need for better evidence generation rather than just better synthesis of existing evidence, which raises the question of how more meaningful evidence could be generated and how the decision makers could be involved in the identification of evidence gaps, design of study protocols, and implementation of CER studies, particularly those that propose to use RWD. It also is important to determine when additional studies, and related designs, are needed; value of information analysis, which examines the value of generating new evidence for decision making,5 can assist in that process since there is a need to prioritize in addition to grading the quality of the evidence.6

Stakeholder engagement in CER is encouraged by the Agency for Healthcare Research and Quality (AHRQ). The selection of stakeholders and processes for engagement will continue to evolve. Stakeholder engagement will no doubt involve patients and physicians, yet when it comes to coverage and formulary decisions, it is clear that payers and other health care stakeholders have an interest in participating in the research design and conduct. In fact, a recent article that reports on key informant interviews from major U.S. payers documents their willingness to be involved in studies that address the value of drug therapies.7

A Case Study in Neuropathic Pain

The remainder of this paper describes a collaborative effort between a payer, a research organization (HealthCore), and a drug manufacturer-sponsor (Pfizer) to develop a study protocol that combines elements of an RCT with RWD sources to answer mutually aligned research questions. These types of collaborative research studies can never replace clinical trials done for regulatory approval and labeling; however, in the post-regulatory environment, they may provide supplemental evidence that is valued by some payers. The example of the collaborative development of a study protocol highlighted in this paper is from an ongoing study. Pfizer is currently working with a large MCO and a research organization, HealthCore, to examine the relationship of its medication utilization strategy for pregabalin to utilization and expenditures. Medication utilization strategies, such as prior authorization (PA) and step therapy, are effective tools used by payers to control medication costs or to control access to medications in which the potential for harm may outweigh the benefits. With respect to the former, studies of the impact of PA and step therapy on medical and/or total cost of care (pharmacy and medical cost) have shown mixed results with respect to overall savings.8-17

Recently, Pfizer-sponsored retrospective studies examining the association of a pregabalin PA on the total cost of care in a Medicaid and a commercial population were presented to the MCO.16-17 Because the MCO did not believe that the studied population was representative of its beneficiaries, Pfizer and
Developing a Collaborative Study Protocol for Combining Payer-Specific Data and Clinical Trials for CER

the MCO agreed to undertake a prospective study to answer the question of whether the PA on pregabalin would affect costs; the study uses the plan's beneficiaries and the physicians who treat the MCO's patients with painful diabetic peripheral neuropathy (pDPN) or fibromyalgia (FM).

The MCO's PA for pregabalin is paper-based and requires the physician to fax the PA form to the MCO. The specific requirements for a pregabalin approval include (a) a certification of a diagnosis of FM, pDPN, postherpetic neuralgia (PHN), or epilepsy; (b) confirmation of pharmacy benefit eligibility and; (c) for patients with these diagnoses other than epilepsy, a trial of at least 180 days on a formulary agent approved for treating pain (e.g., tricyclic antidepressants, cyclobenzaprine, fluoxetine, trazodone). Pfizer, HealthCore, and the MCO agreed to study the effect of PA under “real world” conditions, using a hybrid between an RCT and an observational study, with randomization at the physician level. All parties also mutually agreed on endpoints consisting of health care costs and patient-reported outcomes (PROs).

Process for Developing the Study Protocol. Before an appropriate study design was identified, a process was mutually developed to ensure that Pfizer and the MCO had equal decision-making authority and contribution into the research design with the research organization serving as operational hub of the project. A core study team of 10 researchers; 2 from the MCO, 3 from Pfizer, 4 from HealthCore, and 1 independent statistician was formed. Because the proposed study would most likely use a nontraditional study design, a scientific advisory board composed of 5 members, including 1 external methodologist and 2 clinical experts, as well as 1 contributor each from the MCO (medical director), and Pfizer (senior health economist) was established to help guide and advise the study design. In order to ensure parity in decision making, all organizations contributed to and agreed to the selection of the scientific advisory board members. A study outline was prepared when there was agreement on the framework for the study in order to obtain internal agreement within each organization to proceed with the study and to obtain necessary funding within Pfizer for the research conduct. The study protocol was written and endorsed by all participating collaborators. It is known as the ExPAND (Examination of Pregabalin Access for Treatment of Indicated Pain Disorders) study and is posted on www.clinicaltrials.gov as NCT01280747. Results will also be posted once the study is presented according to the approved institutional review board.

Study Design. Much like the prior retrospective claims database studies, the objective of this study was to determine the impact of a PA on pregabalin, not a direct comparison of treatment effects of specific medications. It was clear to the research team that a study design was needed that would be feasible and test the impact of a PA on pregabalin. While a traditional RCT was preferred, this study design seemed unlikely since blinding and randomization to a group were not feasible. We also considered a pragmatic clinical trial (PCT), a type of RWD which aims at exploring a hypothesis and study design to inform decision making. While a PCT study design seemed most appropriate, the team wanted to go beyond the traditional definition of a PCT, which generally does not include aspects of retrospective data collection. Therefore, the collaborative research team proposed an observational PCT and also brought in retrospective data elements into the study (e.g., administrative claims for visits and charges) to better inform the payer in an economic decision. The retrospective component of the study was necessary to assess disease-related health care utilization and cost as well as total all-cause cost of care. The study design included a cluster randomization at the physician level in an attempt to reduce confounding, and endpoints were to be evaluated mainly through observational follow-up. The final study design was agreed upon by study team members at the MCO, HealthCore, and Pfizer and endorsed by the scientific advisory board. The study will enroll 2,280 patients from 228 physicians (i.e., 10 patients per physician) across the 14 states where the health plans have membership. The physicians will be randomized on a 1:1 basis to usual care (PA policies in place) or expanded access (non-PA group). Although all patients for the 114 physicians in the non-PA group can receive pregabalin without restriction (i.e., regardless of prior use of formulary medication and regardless of diagnosis), the 10 patients selected for each physician will be required to have a diagnosis of either FM or pDPN.

Physician and Patient Recruitment and Randomization. The retrospective elements in this study are utilized to inform aspects of the study including the primary endpoint, cost to treat FM and pDPN, and the identification of physicians treating FM or pDPN patients. Participating physicians are randomized to 1 of the 2 study arms, usual care or expanded access. The usual care group will continue to have a PA on pregabalin while the expanded access group will have no PA on pregabalin. Following the design of a PCT, the inclusion criteria were established to increase external validity. Therefore, all patients aged 18 years or older with a diagnosis of either FM or pDPN are considered eligible for the study if they (a) are newly prescribed treatment for their either FM or pDPN or (b) a change in existing treatment is needed due to lack of effectiveness on their current treatment as determined by the physician. Choice of treatment for either disease state is at the discretion of the physician and patient. Patients enrolling in the study are consented according to the approved institutional review board.
(IRB) protocol and followed for 6 months; however, following
the pragmatic study design, patients will see the physician
under routine care, and patient visits are not mandated beyond
the baseline visit except for the end-of-study visit. Additionally,
patients are not compensated for office visit care, nor are they
compensated for the cost of prescription medications.

All patients in both the PA and non-PA groups will meet
the PA criterion of a diagnosis of either FM or pDPN. The
difference between the groups is that physicians in the non-PA
group will be able to prescribe pregabalin without restrictions,
if deemed appropriate, whereas physicians in the PA group will
be required to (a) complete and fax the PA approval form and
(b) document a trial of 180 days on a formulary agent (e.g.,
tricyclic antidepressants, cyclobenzaprine, fluoxetine, trazo-
done), to obtain coverage for a pregabalin should the physician
prescribe pregabalin.

Measured Outcomes and Reporting of Assessment. All
patients will be evaluated on 2 primary endpoints: pain-related
patient-reported outcomes (numeric rating scale [NRS]) and
all-cause health care resource costs (from administrative claims
records). There are also a number of secondary outcomes
measured including the Brief Pain Inventory, Fibromyalgia
Impact Questionnaire (FM patients only), Work Productivity
and Activity Impairment Questionnaire, and the Patient Global
Impression of Change.19-22 Patients complete the instruments
at baseline, month 1, month 3, and month 6. However, as
mentioned above, patients are not required to have office vis-
its at the above time periods. As a result, subjects, given a
binder with all the PRO instruments and will be instructed to
mail (return postage provided) the PRO instruments directly
to HealthCore. Alternatively, if patients have a scheduled visit
within a 2-week time period of the schedule above, they will be
asked to bring the instruments with them to the visit.

Database Development. All prospectively generated study
data will be collected using electronic records (eCRFs) and will
reside in a Health Insurance Portability and Accountability Act
(HIPAA)-compliant secure database. A data management plan
will be developed with cleaning and validation instructions
consistent with both traditional clinical trial and real world
data.

Study Limitations. All CER studies have limitations and
potential biases. The current study was designed to limit these
biases while attempting to balance internal and external valid-
ity; nonetheless, biases remain, and publication and dissemina-
tion of the study results will need to address these biases. The
non-PA group will have the entire restriction lifted while the
PA group will continue to have the PA in place for pregabalin.
While patients in this study may meet the MCO’s criteria for
pregabalin, it is hypothesized that many physicians in the
PA group will not prescribe pregabalin due to the process of
getting the medication. Furthermore, the fact that the study
focuses only on patients with FM or pDPN reduces the ability
to fully assess the potential cost implications of a PA program
on pregabalin since the drug may be prescribed for patients
who do not meet the labeled indications.

Benefits to Participating Organizations. Manufacturers and
payers have a mutual interest in conducting CER studies that
inform coverage and reimbursement decisions. The current
study provides benefits to both Pfizer and the participating
MCO. As a participating partner, the MCO benefits through
its ability to conduct a CER study on its own enrollee popu-
lation with financial support from Pfizer. Historically, many
pharmacoeconomic studies were designed by the sponsoring
manufacturer, and the majority of “input” from the MCO was
the use of its administrative claims. In contrast, the current
study integrates the MCO as an equal partner in the study
design and conduct. Furthermore, there is a prospective data
capture component to expand outcomes to include patient-
centered outcomes using validated instruments. As a sponsor,
Pfizer benefits from the assurance that the study will produce
“meaningful” evidence since the MCO participated in the
design and execution of the study, as well as demonstrate its
leadership in collaborative CER design and conduct. Another
benefit is the insight the pharmaceutical sponsor gains on the
MCO decision-making process regarding a payer’s require-
ments to establish PA, step-therapy edits, and other utilization
control tools that are used routinely by MCOs. Finally, from an
“internal management” perspective, CER researchers at Pfizer
were able to provide exposure to their clinical trial specialist
colleagues at Pfizer, whose focus is primarily on regulatory
approval, to key post-approval research requirements which
are being requested by many payers. Thus, the clinical trials
group at Pfizer obtains first-hand knowledge of the potential
benefits of RWD sources to assess effectiveness.

Conclusion and Next Steps
The increasing demand for CER studies and evidence of com-
parative clinical benefits and value likely will be addressed
through continued development of novel approaches to CER
studies that involve decision maker participation. Moving for-
ward, CER protocols that are jointly designed and conducted
by manufacturers and payers likely will attempt to combine
the best concepts from clinical trials and analysis of RWD.
This effort will require scientifically rigorous investigations
that produce meaningful evidence in an efficient manner. The
results will supplement prior evidence from RCTs and provide
additional information for payers to potentially aid in coverage
determination. There no doubt will be a variety of case stud-
ies, such as the one described in this article. These early CER
endeavors will provide insights for enhancing CER methods
and the entire evidence generation process. The pDPN and FM
study described in this article is expected to be completed in
mid-2012. The study team along with the scientific advisory board will work to determine an appropriate venue to disseminate the results, which will shed light not only on the specific research being addressed but also on the approach to conducting collaborative CER studies.

**Authors**

ROBERT J. SANCHEZ, PhD, is Director, U.S. Health Economics and Outcomes Research, and JACK MARDEKIAN, PhD, is Senior Director, OR Statistical Scientist, Pfizer Inc, New York, New York. MARK J. CZIRAKY, PharmD, is Vice President, HealthCore, Inc., Wilmington, Delaware. C. DANIEL MULLINS, PhD, is Professor, Pharmaceutical Health Services Research Department, University of Maryland School of Pharmacy, Baltimore, Maryland.

**DISCLOSURES**

This supplement was funded by Pfizer. Sanchez and Mardekian are Pfizer employees. Cziraky reported that HealthCore receives funding for research projects from pharmaceutical companies including Pfizer. Mullins reported financial and other relationships with Pfizer that include receipt of grants, consulting fees or honoraria, support for travel, consulting fees for participation in review activities such as data monitoring boards, payment for writing or reviewing the manuscript, advisory board membership, payment for lectures including service on speakers bureaus, and payment for development of educational presentations.

The 4 authors contributed equally to writing and revision of the manuscript.

**ACKNOWLEDGEMENTS**

The authors would like to acknowledge the helpful comments from Shelly Stanley, MS, RD; Richard J. Willke, PhD; and Zhanna Jumadilova, MD, MBA.

**REFERENCES**


