VALUE IN HEALTH 15 (2012) 217–230

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

ISPOR TASK FORCE REPORT

Prospective Observational Studies to Assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report

Marc L. Berger, MD[1,], Nancy Dreyer, PhD, MPH[2], Fred Anderson, PhD[3], Adrian Towse, MA[4], Art Sedrakyan, MD, PhD[5], Sharon-Lise Normand, PhD[6]*

[1]*OptumInsight, Life Sciences, New York, NY, USA;* [2]*Outcome, Cambridge, MA, USA;* [3]*Center for Outcomes Research, University of Massachusetts Medical School, Worcester, MA, USA;* [4]*Office of Health Economics, London, UK;* [5]*Comparative Effectiveness Program at HSS and NYP, Weill Cornell Medical College, New York, NY, USA;* [6]*Harvard Medical School, Department of Health Care Policy, Boston, MA, USA*

ABSTRACT

Objective: In both the United States and Europe there has been an increased interest in using comparative effectiveness research of interventions to inform health policy decisions. Prospective observational studies will undoubtedly be conducted with increased frequency to assess the comparative effectiveness of different treatments, including as a tool for "coverage with evidence development," "risk-sharing contracting," or key element in a "learning health-care system." The principle alternatives for comparative effectiveness research include retrospective observational studies, prospective observational studies, randomized clinical trials, and naturalistic ("pragmatic") randomized clinical trials. **Methods:** This report details the recommendations of a Good Research Practice Task Force on Prospective Observational Studies for comparative effectiveness research. Key issues discussed include how to decide when to do a prospective observational study in light of its advantages and disadvantages with respect to alternatives, and the report summarizes the challenges and approaches to the appropriate design, analysis, and execution of prospective observational studies to make them most valuable and relevant to health-care decision makers. **Recommendations:** The task force emphasizes the need for precision and clarity in specifying the key policy questions to be addressed and that studies should be designed with a goal of drawing causal inferences whenever possible. If a study is being performed to support a policy decision, then it should be designed as hypothesis testing—this requires drafting a protocol *as if* subjects were to be randomized and that investigators clearly state the purpose or main hypotheses, define the treatment groups and outcomes, identify all measured and unmeasured confounders, and specify the primary analyses and required sample size. Separate from analytic and statistical approaches, study design choices may strengthen the ability to address potential biases and confounding in prospective observational studies. The use of inception cohorts, new user designs, multiple comparator groups, matching designs, and assessment of outcomes thought not to be impacted by the therapies being compared are several strategies that should be given strong consideration recognizing that there may be feasibility constraints. The reasoning behind all study design and analytic choices should be transparent and explained in study protocol. Execution of prospective observational studies is as important as their design and analysis in ensuring that results are valuable and relevant, especially capturing the target population of interest, having reasonably complete and nondifferential follow-up. Similar to the concept of the importance of declaring a prespecified hypothesis, we believe that the credibility of many prospective observational studies would be enhanced by their registration on appropriate publicly accessible sites (e.g., clinicaltrials.gov and encepp.eu) in advance of their execution.

Keywords: comparative effectiveness, prospective observational studies.

Introduction

Context and background

In both the United States and Europe there has been an increased interest in comparative (or relative) effectiveness of interventions to inform health policy decisions. In the United States, the American Reinvestment and Recovery Act established a federal coordinating council for comparative effectiveness research (CER). This council defined CER as the "conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real world' settings" [1]. It noted that the purpose of this research is to inform patients, providers, and decision makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances. To provide this information, CER must assess a comprehensive array of health-related outcomes for diverse patient populations. Interventions may include medications, procedures, medical and assistive devices and technologies, behavioral change strategies, and delivery system interventions. Furthermore, it noted that CER necessitates the development, expansion,

* *Address correspondence to:* Marc L. Berger, OptumInsight, Life Sciences, 301 West 118th Street, Apartment PH 2-C, New York, NY 10026, USA.

E-mail: Marc.Berger@Optum.com.

1098-3015/$36.00 – see front matter

doi:10.1016/j.jval.2011.12.010

Table 1 – Categories of intervention effects.

Term	Efficacy	Relative efficacy	Effectiveness	Relative effectiveness
Definition: *Extent to which*	An intervention does more good than harm under *ideal* circumstances	An intervention does more good than harm, under *ideal* circumstances, compared with one or more alternative interventions	An intervention does more good than harm when provided under the *usual* circumstances of health-care practice	An intervention does more good than harm compared with one or more intervention alternatives for achieving the desired results when provided under the *usual* circumstances of health-care practice
Key features	Randomization versus placebo; select patients; high-volume centers	Randomization versus active control; or use of indirect comparisons of trials versus placebos or active comparators	Observational study; heterogeneous patient population; typical treatment environment; comparison typically made to other treatments	Observational study of several competing interventions; or randomized naturalistic pragmatic clinical trial

and use of a variety of data sources and methods to assess comparative effectiveness.

The American Reinvestment and Recovery Act provided $1.1 billion in funding to the U.S. Secretary of Health and Human Services, the National Institutes of Health, and the Agency for Healthcare Research and Quality to promote CER. At the request of Congress, the Institute of Medicine developed a list of 100 priority topics for CER, most of which involved processes of care rather than specific therapies. Subsequently, U.S. Health Care Reform legislation—the Patient Protection and Affordable Care Act—created a new entity, the Patient-Centered Outcomes Research Institute, to identify national research priorities for CER, appoint advisory panels on research design, facilitate public comment, and disseminate CER findings, as well as to work to improve the science and methods of CER through developing and updating standards on internal validity, generalizability, and timeliness.

In Europe, the European Network for Health Technology Assessment initiative was launched in 2006 following the request of European Union member states in the High Level Group on Health Services with a work program focusing on a pan-European "core model" for health technology assessment in Europe, with initial reports on diagnostics and medical and surgical interventions. The 2011 European Network for Health Technology Assessment work program includes research on pharmaceuticals and other technologies, reflecting a recent focus in Europe on the relative effectiveness of pharmaceuticals. The Pharmaceutical Forum was developed in 2005 to bring the European Commission, member states, representatives of the European Parliament, and a wide range of stakeholders together to examine challenges relating to providing information to patients on pharmaceuticals, pricing, reimbursement policy, and relative effectiveness assessment. In its 2008 report [2], the forum adopted working definitions of efficacy, relative efficacy, effectiveness, and relative effectiveness. These are shown in Table 1 along with this task force's update of the key features.

The report noted that the aim of a relative effectiveness assessment is to compare health-care interventions in practice to classify them according to their practical additional therapeutic value. It acknowledged that differences between the objectives and priorities of different national health care systems may create differences in the way in which health-care interventions will be evaluated relative to one another and differences in relative effectiveness valued. In a survey of 27 member states in 2007, however, the forum found that little distinction is currently made in member state assessments between efficacy and effectiveness. Member states mostly relied on relative efficacy data to inform their health technology assessments and felt that there was inadequate effectiveness data available.

Generating evidence about new pharmaceuticals, including biological entities, is increasingly being seen as an activity that occurs throughout the entire product life cycle rather than prelaunch for a one-off "at-launch" review. Drug regulatory authorities are exploring both early access and provisional access schemes in which some studies about effectiveness and safety are conducted postlaunch. Similarly, health technology assessment and pricing and reimbursement bodies are experimenting with "coverage with evidence development" including risk sharing that involves collection of additional data postlisting. At the same time, concerns about safety have led to augmented postlaunch pharmacovigilance requirements. For most of these initiatives, prospective observational studies have been the vehicle for evidence collection.

Like pharmaceuticals, medical devices demand scrutiny across their total life cycle, albeit a life cycle that is typically much shorter than that of drugs. There is a growing debate about future evidence requirements for medical devices in both the United States [3] and Europe. Safety and effectiveness evidence for medical devices, along with novel surgical procedures and diagnostics, has typically involved observational studies.

The ISPOR Board of Directors approved on May 16, 2010, the formation of the Prospective Observational Clinical Studies Good Research Practices Task Force to develop good research practices for prospective observational clinical studies that focus on the effectiveness and/or comparative effectiveness of health-care interventions. Researchers, experienced in biostatistics and outcomes research working in academia, government health organizations, contract research organizations, and hospitals from the United States and the United Kingdom, were invited to join the Task Force Leadership Group. The task force met about once a month to develop the topics to be addressed and outlined and to prepare the first draft report. A face-to-face meeting was held on March 23, 2011, to debate and finalize any contentious issues in the draft report. The draft report was presented for comment at the ISPOR 13th European Congress in Prague, Czech Republic, in October 2010 and the ISPOR 16th International Meeting in Baltimore, MD, USA, in May 2011. The draft report was sent for comment to the Task Force Reviewer Group (82 invited and self-selected individuals interested in this topic) on October 12, 2011. Comments were then considered. The final draft report was sent for comment to the ISPOR membership via the ISPOR eBulletin October 2011. Collectively, there were 11 written comments. All written comments are published at the ISPOR Web site. All comments (many of which are substantive and constructive) were considered, and once consensus was reached by all authors of the article, the final report was submitted to *Value in Health*.