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. PRO is an umbrella term that includes a whole host of subjective outcomes, such as pain, fatigue, depression, aspects of well-being (e.g., physical, functional, psychological), treatment satisfaction, health-related quality of life, and physical symptoms, such as nausea and vomiting.

In the traditional clinical research domain, there have been great advances with regard to the recognition of the role of PROs, as evidenced also by the recent publications of guidance documents by regulatory agencies. In different parts of the world, agencies or government bodies like the Institute for Quality and Efficiency in Healthcare (IQWiG) in Germany, the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, the National Institute of Health and Clinical Excellence (NICE) in the United Kingdom, and the Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada have long histories of using PROs. While there are ongoing initiatives aimed at selecting preferred PRO instruments that would support validity and comparability of PRO measures and results, the use of PROs for CER is less defined than it is for regulatory approval.

In this paper we discuss the role of PROs in CER, review the challenges associated with the inclusion of PROs in CER initiatives, provide a framework for their effective utilization, and propose several areas for future research.

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 when determining the benefits and harm of the 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 a 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 defensible, 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.\textsuperscript{11} Another approach is a content-based interpretation that uses a representative item, along with its response categories, internal to the measure itself to understand the meaning of different scores on that measure.\textsuperscript{20,45-47} Other approaches intended to enrich interpretation of PROs have been published \textsuperscript{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.

\section*{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.

\section*{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.

\begin{center}
\begin{tabular}{|c|c|}
\hline
\textbf{Authors} & \\
DEMISSE ALEMAYEHU, PhD, is Executive Director, OR and Disease Area Statistics Head; ROBERT J. SANCHEZ, PhD, is Director, U.S. Health Economics and Outcomes Research; and JOSEPH C. CAPPELLERI, PhD, is Senior Director, OR Statistical Scientist, Pfizer, Inc., New York, New York. & \\
\hline
\end{tabular}
\end{center}

\section*{DISCLOSURES}

Alemayehu, Sanchez, and Cappelleri are Pfizer employees. The 3 authors contributed equally to writing and revising the manuscript.

\section*{ACKNOWLEDGMENTS}

The authors would like to thank Tracey Gerthoffer, PhD, RPh; C. Daniel Mullins, PhD; Tara Symonds, PhD; and Richard J. Wilke, PhD; for substantial input in the development of the manuscript.

\section*{REFERENCES}


www.amcp.org    Vol. 17, No. 9-a    November/December 2011    JMCP    Supplement to Journal of Managed Care Pharmacy    S31
Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research


Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research


