Developing and Implementing Performance Outcome Assessments: Evidentiary, Methodological, and Operational Considerations

Marriott Marquis • Washington, DC
December 7 & 8, 2016

Discussion Guide

In recent years, there has been increasing focus on making the health care system more patient-centered. This approach has extended into nearly every aspect of the health care spectrum, including the development and regulatory approval of new drugs. A key focus of these efforts has been on advancing the science of patient input: identifying rigorous and systematic approaches to incorporating the patient voice into drug development, developing patient-centered outcomes, and applying innovative drug development tools to capture data on those outcomes. An important part of this advancement has been the increased development and implementation of valid and reliable clinical outcome assessments (COAs). COAs measure an outcome that describes or reflects how a patient feels, functions, or survives and determines whether or not a drug provides clinical benefit. The US Food and Drug Administration (FDA) published guidance for one type of COA—patient-reported outcomes—in 2009, and while many of the principles outlined in that guidance can generally be applied to other types of COAs, this is not true for all of the principles as described. Performance outcomes (PerfOs)—which measure a patient’s ability to perform tasks according to instruction—present particular challenges, owing in part to a relative lack of scientific guidance and regulatory experience with these measures.

To support further progress in this area, and under a cooperative agreement with FDA, the Duke-Margolis Center for Health Policy is convening this expert workshop in order to advance the conceptual and methodological considerations for using PerfOs in the regulatory setting. Specifically, this workshop will engage expert thought leaders and stakeholders to 1) discuss conceptual and methodological approaches to determining whether a PerfO is fit-for-purpose (including the evaluation of its validity), 2) explore the extent to which different types of PerfOs (both physical and cognitive) require different considerations or approaches, and 3) discuss challenges and possible methodological approaches to identify meaningful within-patient change when utilizing a PerfO.

Clinical Outcome Assessments in Drug Development

COAs are generally divided into four broad categories according to how the assessment is conducted and reported: patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), observer-reported outcomes (ObsROs), and performance outcomes (PerfOs). A description of each of the four types of clinical outcome assessment is provided in Table 1 below.

Table 1: Clinical outcome assessment definitions

<table>
<thead>
<tr>
<th>COA type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-Reported Outcomes (PROs)</td>
<td>A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient’s health condition without amendment or interpretation of the patient’s response by a clinician or anyone else.</td>
</tr>
</tbody>
</table>

* Fit-for-purpose: well-defined and reliable within a stated context of use
Clinician-Reported Outcomes (ClinROs)

A measurement based on a report that comes from a trained healthcare professional after observation of a patient’s health condition. Most ClinRO measures involve a clinical judgment or interpretation of the observable signs, behaviors, or other manifestations related to a disease or condition. ClinRO measures cannot directly assess symptoms that are known only to the patient.

Observer-Reported Outcomes (ObsROs)

A measurement based on a report of observable signs, events, or behaviors related to a patient’s health condition by someone other than the patient or a healthcare professional. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life and are particularly useful for patients who cannot report for themselves (e.g., infants or individuals who are cognitively impaired). An ObsRO measure does not include medical judgment or interpretation.

Performance Outcomes (PerfOs)

A measurement based on a task(s) performed by a patient according to instructions that are administered by a health care professional. Performance outcomes require patient cooperation and motivation.

COAs may be further subcategorized in a variety of ways, such as by the concepts being measured, the method and mode of administration, data collection, and analysis, or whether the measure is specific to a particular disease, population, or concept, or is ‘generic’ (i.e. applies across diseases, populations, or concepts). Most trials submitted to FDA to support registration of a therapy utilize some type of COA—and many use multiple COAs—either to support their primary endpoint, as part of the secondary endpoint that provides supportive information on the drug’s effects, or as an exploratory endpoint that can inform future research.\(^3\) The selection of a particular COA for use in a trial necessarily depends on the measurement concept (e.g., pain intensity or frequency) and the context of use in which the assessment will be applied (i.e., key study aspects that can impact the adequacy of the measurement, such as the disease, patient sub-population, method of administration, number and timing of assessments, etc.).

**Opportunities and Challenges in the Development and Use of PerfOs in Drug Trials**

PerfOs may assess several different aspects of patient function in order to determine clinical benefit, including physical (e.g., grip strength or gait speed), cognitive (such as memory recall), or sensory function, typically through tasks completed by the patient. These assessments may consist of just one task or may involve completing a series of tasks in order to assess a broader range of function. The patient’s performance on these tasks is then quantified and reported using defined procedures. Because PerfOs are typically evaluated in a health care setting but generally do not rely on clinical judgment, they may be particularly attractive for use in multicenter trials for standardizing assessment where improving (or mitigating the decline of) function is the focus of treatment.\(^4\) The use of a PerfO in a trial can also help to eliminate errors related to personal recall or the influence of a patient’s perception of their ability versus their actual level of ability to perform a particular task.

To date, much of the focus around the development of valid, reliable, and patient-centered COAs has been on PROs, which are a more direct method of capturing data on how patients feel and function. However, PerfOs have been used to support regulatory drug approvals in a number of different disease areas, and there is increasing interest from industry, academia, and health care providers to develop and implement PerfOs in the drug development process.\(^5\)
Although there is currently no specific FDA guidance regarding the creation and implementation of PerfOs, the agency released final guidance on the use of PROs to support labeling claims in 2009. This guidance outlines the steps necessary to develop and evaluate a patient-reported outcome assessment, and many of the principles outlined in the document can provide useful direction for PerfO development as well. However, not all aspects of the PRO guidance are applicable to PerfOs, and PerfOs present additional development and implementation challenges not encountered with PROs.

**Measurement Properties**

The degree to which a PerfO (or any COA instrument) is suitable and valid for a particular context of use depends in large part on its measurement properties, which include its reliability, validity, and sensitivity to change. Content validity refers to the evidence showing that a PerfO measures its concept of interest, is comprehensive in what it measures, and is the appropriate type of COA for the context in which it is used. However, there remain a number of unanswered questions regarding how best to determine the content validity of a PerfO, and what level of evidence is necessary to support that validity. It can also be challenging to determine for whom the PerfO is valid, as a given PerfO may be better suited for some populations more than others (e.g., adults vs. children vs. elderly). Additionally, the considerations for evaluating key measurement properties may vary depending upon whether the PerfO is measuring cognitive or physical function.

PerfOs are unique in consisting of an actual demonstration of tasks performed by the patient in a standardized manner that is generally less subject to clinical or patient judgment and cultural effects (e.g., timed test of walking speed) and, as a result, can provide evidence that complements other COA types. They therefore often have good reliability, and can provide more direct evidence of a patient’s ability to function in a way other COAs cannot. This can be particularly relevant and useful for assessments of cognitive function where another type of COA, such as a PRO, is not feasible or would not provide a reliable assessment of the patient’s ability. However, PerfOs may be subject to a “learning effect” (i.e., a patient may improve their performance on a task owing to motor learning or motivation to improve on previous efforts, rather than due to treatment benefit), which can impact the reliability of the assessment and potentially the interpretability of the results.

**Defining and Interpreting Meaningful Within-Patient Change**

One of the key challenges in developing COA instruments and implementing them within clinical trials is determining what level or degree of change measured by the COA is considered clinically important and meaningful, rather than simply statistically different. The methods for interpreting score changes on COAs have evolved over time, but there is no established consensus on the ideal approach and FDA does not specify or require a particular methodology for the purposes of its review.

While PerfOs are often sensitive to small changes in outcomes, their sensitivity can also make it difficult to interpret those changes. This issue is compounded by the fact that PerfOs often do not directly measure the activities and health outcomes that are most important to patients, but are instead measuring simpler and more narrowly defined abilities that are hypothesized to be related to more meaningful or complex activities and functions. For example, a PerfO may measure leg strength as a component of walking ability or short-term memory as a component of cognitive decline. In many cases, it may be more challenging to determine a meaningful level of change when the relationship between the measure and the intended treatment benefit is more indirect.
Implementing PerfOs in Multinational Trials

Many of the barriers to COA development and utilization relate to challenges in conducting clinical trials across multinational settings, as such trials recruit diverse populations speaking multiple languages and living in very different settings. Obtaining culturally appropriate and accurate translations of an instrument can be a significant burden, which is magnified each time that the study protocol is amended during the course of the trial. It can also be challenging to ensure that both administrators and patients are able to understand and interpret what the PerfO is asking them to do, and that the task or tasks are administered and performed consistently across trial sites. Additionally, PerfOs may require extra steps to ensure that the equipment used in the execution of tasks is adequately standardized across trial sites.

These logistical challenges are complicated by the fact that the specific activities being measured by a PerfO may be less relevant or meaningful across regions, cultures, and populations, which can impact the interpretation of the resulting data.

Regulatory Efforts to Support the Development and Use of PerfOs

In addition to finalizing its guidance on PROs, FDA has taken a number of steps in recent years to better incorporate the patient’s voice into the drug development and regulatory decision-making process. Under the fifth authorization of the Prescription Drug User Fee Act (PDUFA) in 2012, the agency committed to developing a new program known as the Patient-Focused Drug Development Initiative. Through this initiative, the agency aims to systematically gather patient input and perspectives on a number of specific disease areas and the available treatment options for those conditions, which will in turn be used to inform its standardized benefit-risk assessment framework. FDA has signaled its ongoing commitment to this effort in its recently released set of draft commitments for the upcoming reauthorization of PDUFA in 2017. Under these commitments, the agency will develop a series of guidance documents that will focus on approaches and methods for translating these initial patient-focused drug development meetings into COA tools that can be used to collect meaningful patient and caregiver input.

FDA has also developed a formal, voluntary process by which drug development tools—which include biomarkers and animal models as well as COAs—may be designated as “qualified” by the agency. This designation indicates that a given tool can be relied upon to have a specific interpretation and application in the drug development and regulatory review process, provided that it is used within its defined context of use. In theory, this allows the tool to be used across multiple drug development platforms without the need to gather additional data to support its use, thus mitigating the financial and opportunity costs associated with developing valid and reliable new tools.

The qualification process for COA instruments is overseen by the Clinical Outcome Assessments Staff (formerly the Study Endpoints and Labeling Development staff), who serve as a cross-divisional resource to promote the development and implementation of patient-focused endpoint measures in medical product development to describe clinical benefit in labeling. FDA released final guidance on this process in January 2014, providing drug developers with a framework for how best to engage the agency and navigate the qualification process, including what types of evidence are required to support qualification.

FDA has also developed supporting communication tools that provide further information on how COA developers might approach the development and qualification process (see Appendix 1). The Roadmap to Patient-Focused Outcome Measurement outlines how developers can identify the concept of interest, define the specific context of use in which the instrument will be used, and identify the
appropriate COA, while the Wheel and Spokes Diagram provides an overview of the major steps involved in the COA qualification process. There are currently 33 COAs undergoing qualification (of which four are PerfOs), and two COAs (both PROs) have achieved qualification status.\(^\text{19,20}\)

**The COA Compendium**

More recently, the agency published a pilot COA Compendium, which is intended as a communication resource for researchers and drug developers who may wish to develop and use COAs in clinical trials.\(^\text{21}\) The Compendium provides a description of how certain COAs have been previously used to measure a patient’s experience and to support labeling claims, and includes a list of qualified COAs, COAs currently in the qualification process, and labeled COAs from a subset of new molecular entities approved between 2003 and 2014.\(^\text{22,23}\) The most common type of COA listed in the Compendium is PROs, reflecting the relative amount of effort directed at PRO development compared to other kinds of COAs to date.

The Compendium is designed to be a starting point for early drug development utilizing COAs, and the agency intends for the COA Compendium to facilitate communication and collaboration amongst COA developers to fulfill unmet measurement needs as well as to provide clarity and transparency regarding how COAs can be used in clinical trials.\(^\text{24}\) Moving forward, the agency plans to modify the Compendium based on public comments, expand the scope of information contained within it, and develop guidance to support its use by external audiences.

**Meeting Objectives and Structure**

The purpose of this two-day workshop is to explore the ongoing challenges to developing, implementing, and interpreting PerfOs in the drug development process, highlight potential solutions to these challenges, and identify priority areas that will require additional work and consensus-building. Using a hybrid approach of case studies and group discussion, these sessions will explore conceptual and methodological approaches to determining whether a PerfO is fit-for-purpose (including evaluating its validity), the extent to which different types of PerfOs (physical and cognitive) require different considerations/approaches for selecting methods and modalities for data collection, and approaches to interpreting meaningful within-patient change. Discussion will also encompass the special considerations that would apply to the use of a PerfO in heterogeneous populations and trials with small sample sizes.

**DAY ONE**

**Session I: Determining Whether a Performance Outcome Assessment Fit-for Purpose:**

**Evidentiary Standards**

**Objective:** Using two case studies—one physical and one cognitive functional outcome (see Appendix 3)—this session will tackle questions related to building the context of use for a PerfO, including the evidence necessary to support that the outcome assessment is representative of, or a requisite component of, daily life functioning and is appropriate for differing populations. Discussion will also encompass special considerations for heterogeneous and small populations

**Questions to address:**

- What information should be provided (or what level of evidence is needed) to support that the performance outcome is representative of daily life functioning or a clinically meaningful component of a patient’s normal function disorder?
- What information should be provided to support that the performance outcome assessment is appropriate for populations with differing:
  - Cultures, nationalities, and languages
  - Socioeconomic background
  - Age, race, and gender
  - Educational attainment/literacy

**Session II: Determining Whether a Performance Outcome Assessment is Fit-for-Purpose: Methodological and Stakeholder Considerations**

**Objective:** Drawing on both case studies and using a breakout group structure, this session will focus on the role of various stakeholders in the development and validation of a PerfO, as well as different methods to ensure both patients and administrators can understand and interpret the PerfO tasks as required.

**Questions to address:**
- What type of methods should be utilized to ensure patients are able to understand and interpret how to complete the task(s) in the performance outcome assessment?
  - To what extent can changes in a patient’s approach impact the outcome observed (e.g., cognitive strategies, practice effect with repeated measurements)?
- What type of methods should be utilized to ensure administrators are able to understand and interpret how to administer the task(s) in the performance outcome assessment?
  - How do you establish a standardized process for administration of a task in a performance outcome assessment (e.g., training procedures and manuals, appropriate prompting)?

**DAY TWO**

**Session I: Implementing PerfOs in Clinical Trials: Operational Considerations**

**Objective:** Implementing COAs within multi-national clinical trials presents several operational challenges, including selecting the optimal methods and modalities for collecting and evaluating data. Utilizing the same breakout group structure as Day One, this session will explore strategies for addressing those challenges and highlight areas that would benefit from additional research and/or regulatory clarity.

**Questions to address:**
- What methods should be considered to collect data for performance outcome assessments (smartphone/tablet, video, etc.)?
- What is the role for use of modalities such as videoconferencing and documentation of performance using videotaping?
- What is the evidence required to demonstrate that the assessment is valid across multiple (or additional) platforms?
- What is the appropriate role for central raters vs. on-site raters or self-administered techniques?

**Session II: Interpreting Meaningful Within-Patient Change**

**Objective:** One of the key challenges in developing PerfO instruments and implementing them within clinical trials relates to determining the degree of change that is considered important and meaningful to patients, physicians, and regulators. Through panelist remarks and open moderated discussion, this
session will explore these challenges and identify possible methodological approaches to capturing meaningful within-patient change.

Questions to address:

- How do we establish meaningful within-patient change for a performance outcome assessment?
- How do we select methods and anchors for establishing meaningful within-patient change for performance outcome assessments?
- How do we evaluate the need for normative data (e.g., age equivalents scores: vs. chronological age in pediatric and geriatric patients)?

Session III: Major Takeaways and Next Steps

Objective: With input from the assembled experts, review the major themes of the workshop’s discussion and identify and prioritize for action the major questions that will require further input. This may help to guide/inform the agenda for a future workshop or other collaboration.
Appendix 1: Roadmap to Patient-Focused Outcome Measurement in Clinical Trials

Roadmap to

PATIENT-FOCUSED OUTCOME MEASUREMENT
in Clinical Trials

1. Understanding the Disease or Condition
   A. Natural history of the disease or condition
   B. Patient subpopulations
   C. Health care environment
   D. Patient/caregiver perspectives

2. Conceptualizing Treatment Benefit
   A. Identify concept(s) of interest (COI) for meaningful treatment benefit
   B. Define context of use (COU)
   C. Select clinical outcome assessment (COA) type

3. Selecting/Developing the Outcome Measure
   A. Search for existing COA measuring COI in COU
   B. Begin COA development
   C. Complete COA development
Appendix 2: Wheel and Spokes Diagram: Qualification of Clinical Outcome Assessments