VALUE IN HEALTH 18 (2015) 741–752

Available online at www.sciencedirect.com

ScienceDirect

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

ISPOR TASK FORCE REPORT

Clinical Outcome Assessments: Conceptual Foundation—Report of the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes Research Task Force

Marc K. Walton, MD, PhD[1,*], *John H. Powers III, MD, FACP, FIDSA*[2], *Jeremy Hobart, PhD, FRCP*[3], *Donald Patrick, PhD, MSPH*[4], *Patrick Marquis, MD, MBA*[5], *Spiros Vamvakas, MD*[6], *Maria Isaac, MASc, MD, PhD*[6], *Elizabeth Molsen, RN*[7], *Stefan Cano, PhD, CPsychol, AFBPsS*[8], *Laurie B. Burke, RPh, MPH*[9,10]

[1]*Janssen Research and Development, Titusville, NJ, USA;* [2]*Leidos Biomedical Research in support of the Division of Clinical Research, National Institutes of Health, Bethesda, MD, USA;* [3]*Plymouth University Peninsula Schools of Medicine and Dentistry, Devon, UK;* [4]*Seattle Quality of Life Group, Department of Health Services, University of Washington, Seattle, WA, USA;* [5]*Modus Outcomes, Newton, MA, USA;* [6]*European Medicines Agency, London, UK;* [7]*International Society for Pharmacoeconomics and Outcomes Research, Lawrenceville, NJ, USA;* [8]*Modus Outcomes, Stotfold, UK;* [9]*LORA Group, LLC, Royal Oak, MD, USA;* [10]*Department of Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA*

ABSTRACT

An outcome assessment, the patient assessment used in an endpoint, is the measuring instrument that provides a rating or score (categorical or continuous) that is intended to represent some aspect of the patient's health status. Outcome assessments are used to define efficacy endpoints when developing a therapy for a disease or condition. Most efficacy endpoints are based on specified clinical assessments of patients. When clinical assessments are used as clinical trial outcomes, they are called clinical outcome assessments (COAs). COAs include any assessment that may be influenced by human choices, judgment, or motivation. COAs must be well-defined and possess adequate measurement properties to demonstrate (directly or indirectly) the benefits of a treatment. In contrast, a biomarker assessment is one that is subject to little, if any, patient motivational or rater judgmental influence. This is the first of two reports by the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes Research Task Force. This report provides foundational definitions important for an understanding of COA measurement principles. The foundation provided in this report includes what it means to demonstrate a beneficial effect, how assessments of patients relate to the objective of showing a treatment's benefit, and how these assessments are used in clinical trial endpoints. In addition, this report describes intrinsic attributes of patient assessments and clinical trial factors that can affect the properties of the measurements. These factors should be considered when developing or refining assessments. These considerations will aid investigators designing trials in their choice of using an existing assessment or developing a new outcome assessment. Although the focus of this report is on the development of a new COA to define endpoints in a clinical trial, these principles may be applied more generally. A critical element in appraising or developing a COA is to describe the treatment's intended benefit as an effect on a clearly identified aspect of how a patient feels or functions. This aspect must have importance to the patient and be part of the patient's typical life. This meaningful health aspect can be measured directly or measured indirectly when it is impractical to evaluate it directly or when it is difficult to measure. For indirect measurement, a concept of interest (COI) can be identified. The COI must be related to how a patient feels or functions. Procedures are then developed to measure the COI. The relationship of these measurements with how a patient feels or functions in the intended setting and manner of use of the COA (the context of use) could then be defined. A COA has identifiable attributes or characteristics that affect the measurement properties of the COA when used in endpoints. One of these features is whether judgment can influence the measurement, and if so, whose judgment. This attribute defines four categories of COAs: patient reported outcomes, clinician reported outcomes, observer reported outcomes, and performance outcomes. A full description as well as explanation of other important COA features is included in this report. The information in this report should aid in the development, refinement, and standardization of COAs, and, ultimately, improve their measurement properties.

Keywords: clinical outcome assessment, concept of interest, context of use, treatment benefit.

Introduction

Establishing the value of medical interventions from the perspective of multiple parties—patients, health care providers, regulators, and payers—is essential to the availability and adoption of therapies. An important element of the information establishing a therapy's value is the evidence provided from clinical trials evaluating the intervention's effect. (The term "therapy" is used

* *Address correspondence to:* Marc K. Walton, 125 Crystal Spring Drive, Ashton, MD 20861.
E-mail: mwalton9@its.jnj.com.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2015.08.006

Background to the Task Force

Since 2009, ISPOR has published eight patient reported outcome (PRO) task force reports based on addressing aspects of the development and application of PROs. These reports are consistent with the US Food and Drug Administration's (FDA's) guidance for industry, "Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims," that described how the FDA would evaluate the adequacy and appropriateness of PRO measures used as effectiveness endpoints in clinical trials.

With the FDA's evolution toward the review and qualification of clinical outcome assessments (COAs), defined as any reported assessment used to support primary or secondary endpoints to document treatment benefit, former PRO task force members submitted a proposal to focus on a specific type of COA—clinician reported outcomes (ClinROs).

In January 2013, the ISPOR Health Science Policy Council recommended the formation of a ClinRO Good Practices for Outcomes Research Task Force. The Board of Directors subsequently approved the task force.

Members and primary reviewers were selected to represent a diverse range of perspectives, including government (US FDA), academia, research organizations, and the pharmaceutical industry. The task force leadership group comprised experts in PRO and other assessment and development, psychometrics, clinical trial data collection, and regulatory affairs. In addition, the task force had international representation with members from the European Medicines Agency and reviewers.

The task force met approximately every 5 weeks by teleconference to develop an outline and discuss issues to be included in the report. In addition, task force members met in person at ISPOR international meetings. All task force members, as well as primary reviewers, reviewed many drafts of the report and provided frequent feedback in both oral and written comments.

In the course of task force deliberations, in response to specific comments and suggestions from reviewers, and a growing concern about length, it became apparent that the material would need to be covered in two task force reports to be thorough, covering the essential points, yet keep the report readable and digestible. With permission from the editors of *Value in Health*, the material has been split into two articles. This article lays the groundwork explaining the important concepts and definitions that underpin COAs. The second article delves into the specific aspects of ClinRO assessments and makes good measurement practices recommendations.

Preliminary findings and recommendations were presented four times in forum and workshop presentations at the ISPOR annual European congresses and international meetings in the period 2013 to 2015. Comments received during these presentations were addressed in subsequent drafts of the report. In addition, the draft task force report was sent out to the nearly 500-person ISPOR PRO Review Group twice.

All comments were considered, and most were substantive and constructive. The comments were discussed by the task force in a series of teleconferences and addressed as appropriate in revised drafts of the report. All written comments are published at the ISPOR Web site on the task force's Web page: http://www.ispor.org/taskForces/Clinical-Outcomes-Assessment.asp.

The task force report and Web page may also be accessed from the ISPOR homepage (www.ispor.org) via the purple Research Tools menu, ISPOR Good Practices for Outcomes Research, heading: Patient Reported & Clinician Reported Outcomes Methods and link: http://www.ispor.org/workpaper/practices_index.asp. A list of leadership group members is also available via the task force's Web page.

Once consensus was reached by all task force members, the final report was submitted to *Value in Health* in August 2015.

throughout this report to mean the intervention administered to the person, irrespective of whether the intervention is intended to improve, or prevent additional, adverse effects in patients of a disease already established, or prevent onset of a disorder not yet affecting a person.) Measurement of the beneficial effect relies on a specified assessment of study patients—evaluating how they feel, function, or survive. This evaluation is used in defining the trial's endpoint and in comparing patient groups within the study.

For many diseases, there is no agreed-upon, well-defined, and reliable method for evaluating a particular disease manifestation of interest, and in some diseases not for any important manifestation [1–11]. In other diseases, current outcome evaluation methods have known weaknesses, such as poor reliability, limited ability to detect change, inadequate interpretability of changes [12–15], or uncertain validity [16–18]. In some clinical fields, published clinical trials have used multiple outcome evaluation methods without an adequate understanding of the characteristics of any of the used methods [19,20].

This lack of understanding of outcome assessment attributes and how these affect measurement plus the dearth of good tools for outcome measurement of disease manifestations inhibits therapy development. New, improved, and better-understood methods of assessing the patient can move therapy development forward. Consequently, many investigators are seeking to develop new instruments for outcome measurement or to evaluate the quality of existing instruments. Unfortunately, developing and evaluating new instruments for clinical trial patient assessment or improving existing assessments can be difficult. It is hoped that the information in this report will serve as a basis of understanding, ultimately leading to improved existing instruments and development of new instruments with solid measurement capabilities.

This is the first of two reports by the ISPOR Clinical Outcomes Assessment – Emerging Good Practices Task Force. In this report, we delve into the attributes or characteristics that affect the measurement properties of the assessment when used in endpoints. We provide a clear understanding of what beneficial effect means, how assessments of patients relate to demonstration of a treatment's benefit, and how these assessments are used in study endpoints. In addition, this report describes intrinsic attributes of assessments and clinical trial factors that can affect the measurement properties that should be considered when developing or refining patient assessments.

These concepts are integral to the development of patient assessment tools of any type for clinical trials. The terminology, types, and distinguishing characteristics defined and discussed in this report provide a foundation upon which new outcome assessments can be developed and shown to be well defined and reliable. The explanations and recommendations in this report will also aid investigators considering inclusion of new outcome assessments in the design of clinical trials.

The second report from this task force, "Developing and Evaluating Clinician Reported Outcome (ClinRO) Assessments of Treatment Benefit – Emerging Good Measurement Practices" (J.H. Powers, D.L. Patrick, M.K. Walton, et al., unpublished report), describes principles for the development of new and evaluation of existing ClinRO assessments. Some of the concepts described in these two reports have been discussed in other publications [21–23], and in guidance from the US Food and Drug

Administration (FDA) [24]. The additional concepts presented in this report will further contribute to improving the development of good clinical assessments for use in study endpoints. Recommendations to guide the process of developing outcome assessments with good measurement properties, however, are beyond the scope of this report.

Treatment Benefit

The primary goal of clinical efficacy trials is to provide evidence that a treatment is effective. A conclusion that a therapy is effective means that there is a *treatment benefit* presumably caused by the therapy's use.

A treatment benefit is a favorable effect on a meaningful aspect of how a patient feels or functions in his or her life or on his or her survival. Two phrases in this definition deserve emphasis to ensure clarity. One is *meaningful aspect*—the effect on how a patient feels or functions should be meaningful to the patient. If the effect is not meaningful to the patient, it is not a benefit to the patient. More specifically, this means the treatment effect has a positive impact on an aspect of health affected by the disease that is an alteration in the patient's feeling or functioning. It is an aspect of health that the patient cares about and has a preference that this aspect 1) does not become worse, 2) improves, or 3) is prevented.

The second part of the definition is *in their life*. This means the treatment benefit affects an aspect that occurs in the patient's usual (typical) life. A treatment effect is not a treatment benefit if it is solely an alteration in performing a specific task that occurs only in the medical clinic and has no defined relationship with any usual activity the patient does (or would want to do) in life outside of the clinical trial setting. Establishing a well-understood relationship of the measurement with the patient's usual life is central to the conclusion that the observed effect is actually a treatment benefit.

However, if a purely clinic-based procedure provides a measurement of a simplified patient activity that has an adequately understood relationship with the patient's function in life, then a treatment's effect on that measurement can be interpreted as a treatment benefit. For example, specific differences in clinic-based pulmonary function tests in patients with asthma have been empirically shown to have a relationship with meaningful effects for similar patients in their usual life.

Patients suffering from the same disease may exhibit different symptoms or manifestations. Some manifestations, which are the most important ones to those patients exhibiting them, are not well suited to controlled clinical trial demonstration of treatment benefit in a broader patient population. Moreover, differences in a patient's life circumstances, unrelated to the disease, may influence which of these symptoms are deemed the "most important" to an individual patient.

Furthermore, a therapy typically has a varying level of effect on different disease manifestations. This means that efficacy could be readily shown on only some manifestations. These symptoms may or may not be the most important to patients. Consequently, the specific disease aspect selected for clinical trial evaluation may not be the most important for each patient in a study. In many trials, several disease aspects of varying importance to patients are evaluated to provide a more comprehensive understanding of a treatment's benefits, for example, secondary endpoints in a clinical trial. Nonetheless, the evaluated aspects of how a patient feels or functions must have some meaning (i.e., some amount of importance) to patients who have the disease.

Controlled clinical trials, for example, phase 3 trials, are designed to show a difference in the *study endpoint* results for patients who received the investigational treatment as compared with those who received a comparator treatment (often a placebo). These rigorous studies are the usual source of evidence to support a conclusion of treatment benefit. Therefore, the study endpoint difference between treatment and control patients needs to show or be confidently interpreted as indicating a meaningful effect on how patients feel, function, or survive. The importance of the treatment benefit provided by a therapy is related to both the type of benefit (a feature of the assessment) and the amount of that benefit that occurs (the size of the treatment effect, such as the magnitude of difference on a continuous scale or the percentage difference of patients achieving a particular outcome state).

In contrast, there are many measurements that are considered related to feeling or functioning for which the meaning (importance) to patients in their typical life is not self-evident and has not been adequately evaluated, for example, many cognitive function tests and in-clinic exercise tests. In diseases in which improving one or more specific aspects of feeling or functioning is the intended benefit, an inability to interpret a measurement as meaningful leads to an inability to demonstrate the therapy's treatment benefit. Methods for evaluation that describe or can be used to infer meaningful aspects of how patients feel or function are necessary for the evaluation of these disease treatments. This discussion often uses patient function to describe features of assessments or as examples. The important points and attributes, however, are applicable to assessments of both feelings and function.

This report focuses on assessments of how patients feel or function, not survival. Survival as an outcome assessment (often duration of survival) is distinctly different from feelings or functions, which can be regarded as quality of survival in contrast to duration of survival. Unlike many types of feeling or functioning, mortality has well-defined means for determination (when not restricted to cause-attributed mortality) with readily understood meaning. The meaning of death remains clear even in diseases in which some people might view certain nonfatal outcomes of the disease, for example, severe irreversible disability, as also highly undesirable. (See J.H. Powers, D.L. Patrick, M.K. Walton, et al., unpublished report, for further discussion of mortality.)

The Relationship between Patient Assessments and Study Endpoints

The primary objective of phase 3 studies is to demonstrate a treatment-related difference in patient outcomes through an analysis of patient assessment study data as specified by the endpoint description. An *outcome assessment*, the patient assessment used in an endpoint, is the measuring instrument that provides a rating or score (categorical or continuous) that is intended to represent some aspect of the patient's health status. Most, but not all, phase 3 clinical trials use some type of clinical assessment as the basis for the primary endpoint. The measurement properties of the outcome assessment include content validity, reliability, ability to detect change, and interpretability—the ability to interpret or understand the relationship between the study results and the treatment benefit. These characteristics will strongly influence the study's success.

Defining the endpoint comprises identifying a particular method for the patient assessments obtained at one or more specified times during the study and a stated statistical method for conducting analysis to provide a comparison between groups. *The assessment itself, in isolation from the other specified endpoint elements, is not the endpoint*. This is important to recognize because other aspects of an endpoint, for example, number of evaluations,

study time points, and statistical methods, will also affect interpretation of the study results.

Some patient assessments commonly used in medical care will be suitable for use as outcome assessments; some will not. A patient's medical status is evaluated multiple times in various ways for various reasons in clinical care (diagnosis, estimating prognosis, monitoring of response, etc.). Particular assessments are often suited to particular purposes. For example, often substantial medical experience led to an understanding of how the new presence or absence of a specific clinical sign or symptom before treatment aids in diagnosing a particular disease. The presence or absence of this clinical sign indicates that the patient is at risk for the disease's known effects.

In contrast, the treatment benefit after diagnosis and treatment is a reduction in or an avoidance of the disease's important effect(s). In many diseases, however, there is often insufficient medical experience to demonstrate that elimination of a particular diagnostic sign after the administration of an intervention reliably indicates that the patient will avoid a particular adverse consequence of the disease. Therefore, this particular diagnostic sign would not be a good assessment to use in an endpoint.

Identification of the Intended Treatment Benefit

As stated above, a treatment benefit is a favorable effect on a meaningful aspect of how a patient feels or functions in his or her typical life (or on survival). In many diseases, a single manifestation may impair or prevent multiple related activities that are part of the patient's normal life. This impairment is meaningful to the patient. These closely-related affected activities, as a group, can be thought of as an aspect of how a patient functions in his or her life. Identifying this grouping is a unified abstraction or conceptualization of an aspect of a person's life adversely affected by the disease.

A disease affecting arm motion, for example, multiple sclerosis or Parkinson disease, may directly diminish a patient's ability to perform many activities dependent on arm motion, such as dressing, eating, or toileting. This group of affected functional abilities might be conceptually identified by the term "upper limb–dependent personal activities" (or similar phrasing selected by the developer) to obviate the need to repeatedly list the various activities that are affected. A treatment that favorably affects this inherently meaningful aspect of a patient's life is clearly providing a treatment benefit.

As another example, patients affected with lower limb weakness may have a problem in walking from a bus stop to their office, walking around a shopping mall, a grocery store, or within their own house. These and many similar activities are a common part of a typical healthy person's life. An abstracted commonality of these activities might be named "ambulatory performance." The inability to perform any of the individual activities is meaningful to the person who cannot do so. A treatment that improved patients' ability to walk from the bus stop to their work office because of increased lower limb strength would be expected to have beneficial effects on similar activities that rely on walking as well. A treatment that improved the activities within this aspect of a person's life would be a meaningful benefit, that is, a treatment benefit on ambulation.

In some cases, an important effect of the disease can be named with less abstraction than these examples. For instance, pain is an important adverse feature of metastatic breast or prostate cancer in some patients. Palliation of pain (pain relief) would be an important treatment benefit.

For many diseases, there are several separable groups of related activities (functional abilities) affected by the disease. These functional abilities may have more commonality within the group than between groupings, for example, ambulation activities versus hand and arm–dependent activities versus cognition-based activities. Each group identifies an abstracted concept of a distinct and *meaningful aspect of health* that is affected by the disease. Improving the ability to carry out activities in one, or more than one, of these separate aspects of health would be a treatment benefit.

The meaningful health aspect that is the intended treatment benefit should be identified with phrasing that promotes clear communication during the development period of an outcome assessment to be used in a study endpoint. The identifying phrase used for the meaningful health aspect should provide clarity to the people involved during the outcome assessment development period and to those who may consider the outcome assessment for use in clinical trials, for example, investigators, study designers, and regulatory agency staff, regarding the intended benefit that would be shown with an effective therapy.

Clearly stating the treatment benefit intended to be shown with the outcome assessment is essential during the development process. Understanding the relationship between the outcome assessment's measurements and the specific intended meaningful health aspect is a critical part of good measurement practices (J.H. Powers, D.L. Patrick, M.K. Walton, et al., unpublished report). Furthermore, clear identification of the meaningful health aspect is crucial for treatment benefit communication for therapies that have gained regulatory agency approval.

It should be noted that phrasing to describe the treatment benefit in drug labeling is based on consideration of all demonstrated benefits across several endpoints after the study is completed. Nonetheless, the description is based on the known relationship of an identified meaningful health aspect with each of the outcome assessments in study endpoints.

Because diseases have multiple manifestations affecting different types of patient functioning (or feeling), for example, leg weakness, cognitive impairment, respiratory impairment, and pain, that adversely affect patients, there can be multiple health aspects that could be appropriate as separate intended treatment benefits. Each meaningful health aspect needs to be demonstrated by a distinct outcome assessment and clinical study endpoint.

In some cases, combining several separate aspects of health into an overarching single health concept will be appropriate. For example, some degenerative neurologic diseases damage the patient's ability to use both upper and lower limbs for typical activities of life and impair cognitive function. The separate groups of impaired function define separate meaningful aspects of health that a treatment might benefit. A comprehensive treatment benefit, however, might be intended, and named as an effect on "disability of disease X." Nonetheless, identification of the specific impaired functional abilities, for example, disability in personal activities using the upper limbs, disability in ambulation, and disability in cognitive activities, is needed to adequately identify the intended benefit and determine whether several outcome assessments are needed to evaluate the several components of meaningful health aspects.

If several outcome assessments are combined into a single composite endpoint, the endpoint should be formed carefully. The endpoint structure should allow analysis of the individual component assessments and how each contributed to the composite result. This will promote understanding which benefits were shown to be associated with treatment.

An important principle to keep in mind when developing therapies for multifaceted diseases is the different levels of importance to the patient of different health aspects. Some are more significant or of higher importance to the patient than are others. Moreover, not all disease manifestations are amenable to

benefit from a specific therapy. Consequently, therapy developers should consider the range of meaningful health aspects related to the disease and select for study the one(s) that the therapy may benefit *and* is of sufficient importance to the patient to warrant assessment in a clinical trial.

The Measured concept of interest (COI): A Practical Approach to Evaluating How Patients Feel and Function

In a clinical trial, the outcome assessment is the specific method to obtain a measurement used in a study endpoint. The results are analyzed leading to a conclusion regarding whether a treatment benefit on the selected meaningful health aspect has occurred. Selecting an assessment for a study endpoint warrants attention to whether the assessment directly describes or can be shown to have a relationship with (indirectly describe) a treatment benefit for the disease. Often, patient reported outcome (PRO) instruments are used to directly measure the meaningful health aspect. In many disorders, however, an appropriate PRO instrument has not been developed or directly measuring the relevant daily functioning is difficult, for example, in persons who cannot report for themselves [22,23,25].

In these circumstances, investigators interested in an evaluation of certain functional abilities may instead deconstruct the meaningful health aspect activities into simpler (more narrowly defined) bodily abilities. These abilities are thought to be relied upon when doing meaningful health aspect activities. One of these abilities may be hypothesized as both well related to many activities within the meaningful health aspect and more readily measured, for example, leg strength in impaired walking diseases or short-term memory in certain dementia disorders. The conceptualized simplified bodily activity may then be selected as an opportunity to create a practical measurement tool for studying a treatment. The conceptualized simplified bodily activity is called the concept of interest (COI) for measurement (Fig. 1).

In the example of pain as the meaningful health aspect, the full experience of pain has multiple conceptual facets or characteristics. These include intensity, duration, frequency, and quality (e.g., quality described as sharp, burning, or shooting). The pain experience arising from different disorders can differ in these characteristics. Consequently, when pain is the meaningful health aspect, investigators might select different pain characteristics as the COI. In a particular disorder, pain intensity might be selected as the pain attribute to be measured, whereas in a different disorder, pain frequency might be selected as the COI for measurement.

It is important to recognize that in all studies there must be a targeted COI for the actual measurement. Furthermore, it may or may not be the meaningful health aspect. When a PRO is the outcome assessment, the COI for measurement usually is the meaningful health aspect—the COI is identical to the meaningful health aspect. This is efficient when the meaningful health aspect can be measured directly with a valid instrument. It is also necessary if the meaningful health aspect, such as a patient's feelings, can be evaluated only directly—by the patient.

Nonetheless, in many studies, the COI for measurement is not the same as the meaningful health aspect. Moreover, frequently, there are multiple COIs that might be useful in understanding whether a treatment benefit on the selected meaningful health aspect has occurred. Investigators must select one (Fig. 2) for a measurement to form the basis of an endpoint. In these circumstances, clearly identifying the COI can aid developing the outcome assessment and establishing that it has validity for the intended use and interpretation.

The method of measurement is a procedure that produces a categorical rating or continuous score that is intended to represent the measured COI, that is, an operationalized expression of the COI. For example, pain intensity may be measured in multiple ways. A measurement tool might ask patients to 1) select a score between 0 and 10 depicting their current pain intensity, 2) score the average pain intensity over the past month, or 3) maintain a diary of the number of analgesic pills used each day.

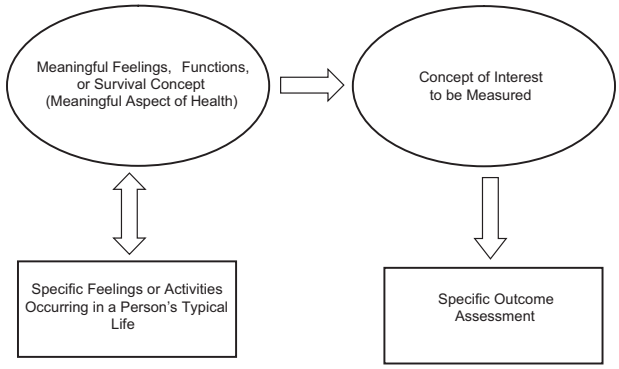

Fig. 1 – Relationship of concepts with an outcome assessment. When planning a clinical trial, investigators select a group of related feelings or activities that can occur in patient's typical lives, are adversely affected by a disease, and are expected to show benefit from the intervention. The group of feelings or functions is identified as a specific conceptualized meaningful aspect of health. If the selected group of feelings or functions is not planned for direct measurement, a concept of interest that is thought measurable and substantially influences the meaningful aspect of health is formulated. The measurement of the concept of interest is operationalized as a specific, well-defined, clinical outcome assessment. An interplay may occur between selection of the meaningful aspect of health and the concept of interest as investigators refine them to have a concept of interest for measurement that is thought both informative of the meaningful aspect of health and has a practical means to be measured (the specific clinical outcome assessment). In some cases, the affected feelings or functions affected by the disease are intended to be directly measured, and the concept of interest for measurement is identical to the meaningful health aspect.

These distinct specifically defined potential methods to measure the COI are not necessarily equivalent. They might provide different results in a clinical trial owing to measuring distinct aspects of the pain experience or having different measurement properties. For example, reliability may be different for pain scores using a 1-month recall period compared with a rating of current pain. Alternatively, the interpretation of the measurement may be not inherently clear; for example, a pill count might relate to pain relief duration from each pill or it might relate to how frequently pain intensity is beyond a patient's personal threshold. Differences in measurement properties can also affect the assessment's suitability for the specific circumstances of use. (See "Context of Use" section.)

Once the outcome assessment's interpretation is understood, evaluation of measurement properties can begin. This step is essential if developing or choosing an outcome assessment for a clinical trial to ensure that the assessment is valid and suitable for the intended use. The measurement properties and principles for evaluating outcome assessment measurement properties to determine their suitability is beyond the scope of this discussion but has been discussed by others [21–23,25–28].

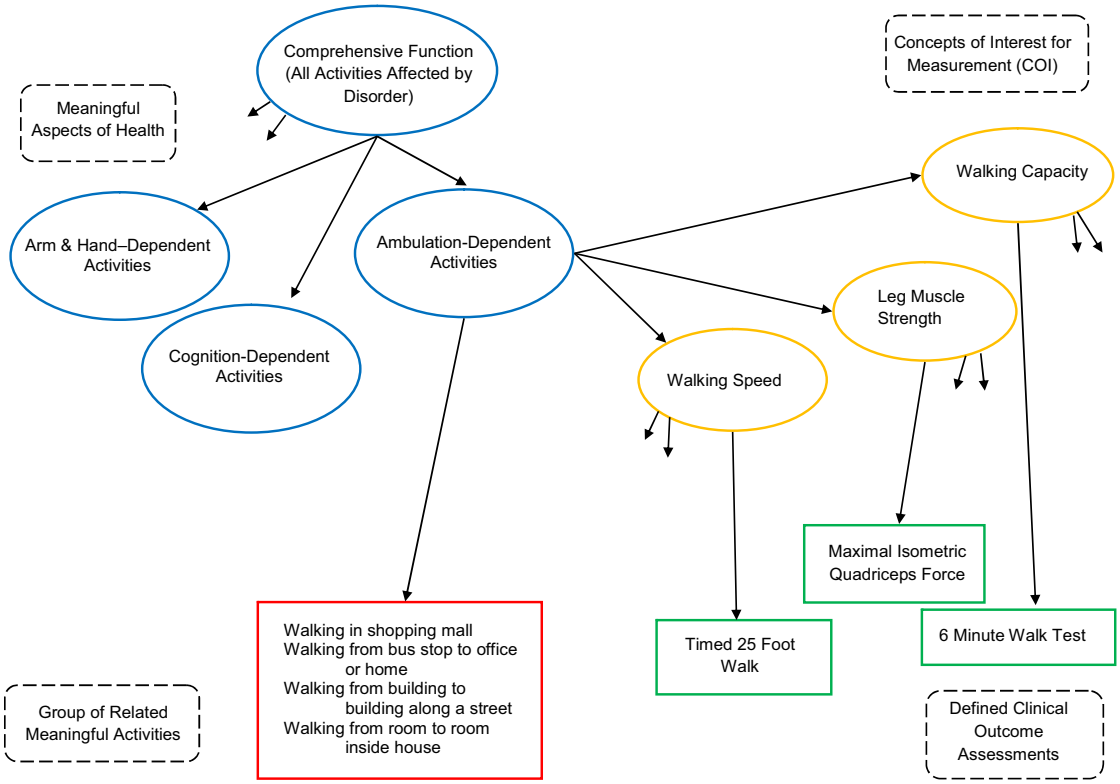

Fig. 2 – Relationships of meaningful functions with an indirect outcome assessment measure. A group of related activities meaningful to the patient (red rectangular box, Specific Meaningful Activities) is identified that would be a treatment benefit to the patient if improved by an intervention. A name is given to the meaningful health aspect that identifies the abstracted concept of the grouping (blue oval, Ambulation-Dependent Function). This specific type of function is just one from among the range of meaningful areas of feeling or functioning (blue ovals) in which patients with the disease might desire a treatment benefit. Each individual meaningful health aspect is a specific portion of the comprehensive range of all the meaningful feelings and functions affected by the disease any of which might be intended treatment benefit and evaluated in a clinical trial. In this example, Ambulation-Dependent function has been selected as the potential treatment benefit to be studied. Specific meaningful activities associated with the nonselected health aspects are not illustrated. Evaluation of the identified activities might be directly assessed in a defined outcome assessment (e.g., a PRO, not shown). The concept of interest for measurement in that case would be identical to the identified meaningful health aspect (see Fig. 1). The selected meaningful health aspect might instead be deconstructed into narrowly defined concepts of interest that are thought to be important in performing the meaningful activities (COIs, orange ovals). Procedures could be devised as specified clinical outcome assessment (COA) instruments to represent a measurement of the COI (defined COAs, green boxes). In this example, performance outcome (PerfO) tools are shown for each COI, but other PerfOs and other types of COAs could be developed for any of the COIs (see Fig. 3). The PerfO procedure, an operationalized method to measure a body action identified as the COI, is a task that is not a part of a person's usual normal life. The COA provides a score for the observed quality or quantity of performing the procedure and is used to form a study endpoint. The meaning of score or change in score to a person's typical life is not intrinsically precisely known, but is hypothesized to reflect the meaningful functional activities. The actual meaning of a specific score (or change) to the patient cannot be known from the description of the PerfO procedure alone and should be evaluated in the process of developing the instrument. Dashed outline boxes name the type of concept or category of activity illustrated in that region of the figure, as presented in Figure 1. Ovals name specific concepts: either meaningful health aspects (blue) or COIs (orange). Solid outline boxes name specific functional activities of typical daily life that are meaningful (red box group), or specific procedures each of which might be developed as a COA (green boxes) to measure the selected COI. Short arrows not leading to a concept or a COA indicate that other possibilities exist, but are not shown. COI, concept of interest; PRO, patient reported outcome. (Color illustration of figure appears online.)

The properties of an outcome assessment are critical to the ability to 1) detect a treatment effect and 2) interpret the effect as a treatment benefit. The outcome assessment's relationship with the meaningful health aspect should be understood before using the outcome assessment in an endpoint of a clinical trial. One of the important steps in ensuring that a new outcome assessment is suitable for demonstrating a treatment benefit is to confirm the hypothesis that the outcome assessment is well related to the meaningful health aspect. Differences in the outcome assessment should be reflective of changes in the meaningful health aspect. Even when the COI is identical to the meaningful health aspect, the relationship between the outcome assessment (as a specific implementation of the COI) and the meaningful health aspect should be examined.

Context of Use: Important Clinical Trial Factors

Clinical studies are conducted in a specific population with a delineated study design, set of procedures, including details on

how the outcome assessment will be used in a clinical trial endpoint. These study design features, taken together, are called the *context of use* (COU). The COU describes the outcomes assessment setting and manner of use. The importance of the COU has been highlighted by the FDA [29].

The COU should be clearly stated at the outset of development for a new outcome assessment or evaluation of an existing outcome assessment [22,23]. Outcome assessment measurement properties can be affected by factors defined in the COU. A comprehensive description of the COU will help ensure that a suitable outcome assessment is developed for the trial's specific application.

It should be noted that changes in the study design and/or other study elements can substantially affect the outcome assessment's performance characteristics [30,31]. When an outcome assessment will be applied in a changed COU, the revised COU should be clearly described. These changes can go so far as to alter the interpretation of the trial's results. Reevaluation of the outcome assessment properties in the new COU is advisable.

Important components for creating a well-defined COU may include the following:

1. Disease of study: The disease of interest should be clearly defined, for example, using diagnostic criteria. The disease's important effects on the patient should be well described to ensure that the stated meaningful health aspect is relevant for the disease. It is also important to identify whether the study participants have the disease at study enrollment or whether they are at risk of developing the disease, that is, whether the therapy is for treatment or prevention of the disease.
2. Patient subpopulations: Many diseases are heterogeneous between patients and change over time within a patient. The precise disease subpopulation intended as the target population should be identified. A wide range of factors may be important to distinguish or differentiate these patients, such as 1) the phenotype (or subtype) of the disease; 2) the patient-specific disease characteristics (severity, duration, involvement of specific portions of the body); 3) demographic characteristics of the intended patients (age, race, ethnicity); 4) history of previous treatments; 5) comorbidities; or 6) other factors that might be used as trial entry or exclusion criteria.
3. Cultural, language, or other geography-related factors: Many outcome assessments rely on questioning patients to obtain the primary input for the outcome assessment results. Thus, when multiple languages are used across a study population, it is important to determine whether the questions have the same meaning in *each* of the different languages. Some questions can invoke cultural variations, for example, differences in what factors are taken into consideration by patients in their responses. Furthermore, there can be cultural influences affecting the relationship between the COI and meaningful health aspect. This relationship can vary across cultural settings in which the outcome assessment is used. To ensure study results remain interpretable, these factors need to be considered. For more information on this topic, please see ISPOR's two PRO task force reports on translation and linguistic validation [32,33].
4. Standard concomitant care: Clinical trials commonly allow some existing therapies to be administered along with the investigational intervention when they are part of standard clinical practice and that care can affect the disease course. If the outcome assessment was developed in a time period before the use of the concomitant therapies allowed in a new clinical study, the outcome assessment's measurement properties can be affected and, therefore, be different than expected. Disparities in measurement properties between study sites can arise if these properties were evaluated with only a subset of the study-permitted choices for concomitant care and care varies across clinical sites, for example, in different global regions. Therefore, the intended choices for standard of care should be identified as part of the COU and these issues should be kept in mind when developing a new or evaluating an existing outcomes assessment.
5. Endpoint positioning: Endpoint positioning is twofold. It describes where the endpoint using the outcome assessment falls within the study objectives (as shown by the analysis plans for the study) and the regulatory role the endpoint is intended to support. An outcome assessment may be appropriate in one context and not appropriate in another. For example, it could support marketing approval decisions in one context of use, providing appropriate evidence for the central efficacy claim in a particular disease. It might, however, be appropriate only for supplementary claims of efficacy in a different COU. Some outcome assessments are not appropriate to support *any* efficacy claim. Specifying the objective and the regulatory role of an endpoint is thus an important element of the COU.
6. Manner of use within the endpoint: An outcome assessment is one of the elements of an end point. It may be incorporated into the endpoint and data analysis in various ways in different clinical trials. There may be differences in the number of times the outcome assessment is measured during the study and used in the endpoint. Alternatively, there may be differences in the time points (timing) of assessment during the study. In addition, there are several options for an endpoint analysis method, such as 1) the group average outcome assessment value at a specific time point in the study, 2) the percentage of patients who are responders at a specific time point by meeting specified outcome assessment criteria, 3) repeated measures using several time points during the study, or 4) time to event. Distinct ways of using the outcome assessment in an endpoint yield different summarizations of patient experience during the study. These distinct uses can translate to different relationships with the meaningful health aspect. The number and timing of outcome assessment measurements along with the analysis method should be specified as part of the COU.
7. Measurement setting: Outcome assessments can be obtained in various settings—in a home, an outpatient clinic, or a hospital inpatient setting, among others. These distinct settings can alter the actual measurement obtained, and thus the measurement properties of the outcome assessment. Therefore, the measurement setting should be stated as part of the COU.
8. Method of outcome assessment administration: Outcome assessments can be designed for more than one mode or method of administration [34]. There are options for 1) who administers the outcome assessment, for example, self-administration, an otherwise untrained person, or a trained professional, and 2) how it is administered, for example, visual versus auditory, face-to-face versus by telephone, and electronic versus nonelectronic. It is crucial to clearly specify the mode(s) and/or method(s) of administration of the COU.

Differences in administration can affect the data obtained, and thus alter the measurement properties of outcome assessment. This can be particularly problematic if multiple options are used within a single trial but have not been shown to provide equivalent measurement properties of the outcome assessment. If another mode or method of administration is introduced, equivalence needs to be demonstrated. For more information

on this topic, please see the two ISPOR task force reports on measurement equivalence and mixed modes data collection [34,35].

Attributes of an Outcome Assessment

Although the COU describes the setting and manner of use of an outcome assessment, it does "not" *define* the outcome assessment. A careful and complete definition of the outcome assessment will include a precise description of the procedure for obtaining the measurement. Detail is needed to evaluate, develop, and use the assessment in clinical trials. The definition determines certain distinguishing attributes (illustrated in Fig. 3) that influence the measurement properties and how the measurement properties are evaluated.

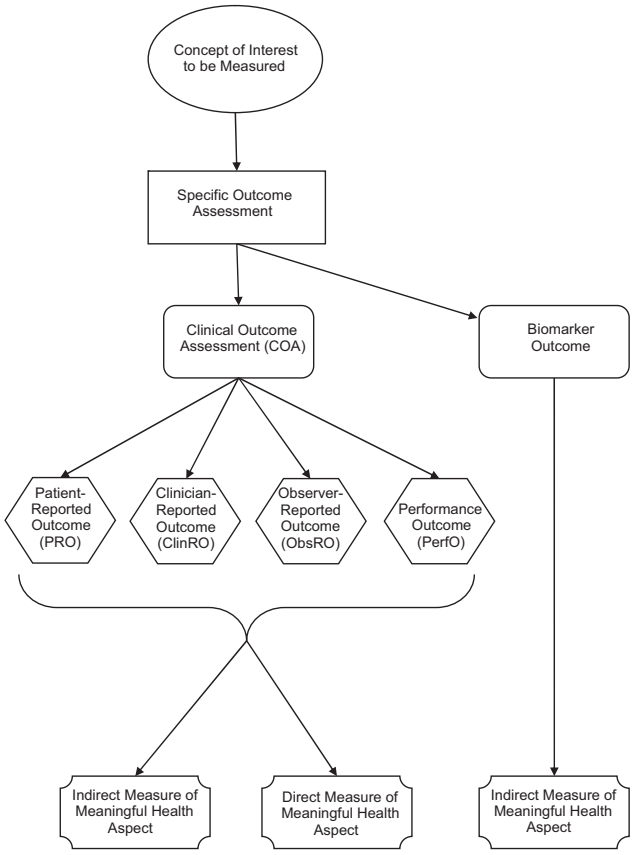

Fig. 3 – Attributes of outcome assessments. A specific outcome assessment is selected or created to operationalize measurement of the concept of interest. Outcome assessments are of two major types: clinical outcome assessments and biomarkers. Clinical outcome assessments have an attribute identifying the type of person whose judgment can influence the reported measurement. Clinical outcome assessments may be influenced by the judgment of the patient, clinician, or a nonclinician observer; they may also be a nonjudged recording of a task performed by the patient (performance outcome). Clinical outcome assessments may be directly reporting the meaningful feelings or functions selected as the potential treatment benefit, or may be reporting measurements that are thought to be indirectly informative regarding those feelings or functions (see Fig. 1). Biomarkers can only indirectly measure the meaningful aspect of health.

Important Outcome Assessment Attributes

Attribute 1: Is the outcome assessment dependent on patient's active involvement or rater's judgment?

Clinical assessments (evaluations) depend on the patient or another person, for example, a trained medical professional, spouse, caregiver, or teacher, to integrate observations, transform them into a rating, and record (report) the result (measurement). Clinical assessments are evaluations that are susceptible to variation *unrelated* to the patient's true clinical status owing to 1) variation in patient volition (which can cause rating differences for an individual patient between successive administrations or between patients) or 2) dependence on the judgment of a rater. A clinical assessment used to measure patient outcome in a clinical trial is called a clinical outcome assessment (COA).

Clinical assessments are susceptible to both sources of variation. Some patient assessments require the patient's active involvement to create a record (physical or electronic) and/or to perform an activity that is the basis of the rating or score. The process of creating the record or performing the activity is influenced by patient volition and motivation to participate in the assessment. The level of motivation and attention to the assessment process may vary over time within a patient or differ systematically between patients. Such variation can yield differences in ratings or scores unrelated to differences in the underlying medical status of interest (symptom or functional ability). These variations can potentially affect the measurement properties of the outcome assessment.

Variability is inherent in a measurement dependent on a rater, whether it is the patient (self-observation) or another person using judgment to make a rating. Raters other than the patient may apply judgment on 1) the patient's response to an inquiry, 2) observations on patient's task performance, or 3) the relative weighting among several observations in determining the rating. Rater judgement may be influenced by prior experiences and biases of the rater that can lead to scoring differences between raters unrelated to differences in patients' medical status. Wide disparities in how the rater's judgment affects the final rating can damage the COA's validity or decrease its reliability.

Although there can be patient or rater influence on measurements unrelated to the patient's medical status, COAs can be developed with assurance that they are well defined and reliable within the target COU. Such COAs can be highly informative regarding patients' medical status and reveal treatment effects with meaning to patients on a very broad range of treatment benefits.

A *biomarker* assessment, in contrast, is one that is subject to little to no patient motivational or rater judgmental influence. (A Biomarkers Definitions Working Group [36] has worked on issues involved with biomarkers used in surrogate endpoints, and defined biomarker for that discussion. The characterization of biomarker in the present discussion [focused on clinical assessments] differs from the working group's to better ensure distinguishing biomarkers and clinical assessments. This distinction used in this report is important when addressing the range of COA types and complexities that arise when developing clinical assessments for use in endpoints.) Examples include protein levels in blood or urine measured by standardized methods or an automated quantitative size measurement of a pathologic lesion visualized with magnetic resonance imaging.

It can be noted that many biomarkers are included within what are often called "clinical laboratory measurements" in clinical practice or clinical trial protocols. Nonetheless, such laboratory measurements of substances physically present in body fluids are not clinical assessments as defined in the framework described here because they are not subject to the

variations of patient motivation or rater influence. "Clinical" in the term "clinical assessment" in this framework is narrowed from the general language meaning of "concerned with or involving patients" so as to distinguish these assessments from biomarkers.

Attribute 2: Who Is the rater? Which type of rater applies judgment to form the measurement?

The rater is the person who obtains the information during the assessment procedure and applies judgment to what he or she has heard or observed to form a rating that is recorded (reported) as the measurement. Although there are many different people who may be involved in performing assessment of patients in a clinical trial, for purposes of this COA discussion, we use three rater categories: 1) patients; 2) clinicians (defined as those using professional judgment for patient assessments, e.g., investigators and nurses); and 3) nonclinician observers. Because people with different backgrounds, experience, and training are likely to have different perspectives and skills, their judgments may differ and have a large impact on the rating. There is a fourth category of clinical outcome assessment that is considered a COA that does not involve rater judgment but does rely on patients' active involvement, and thus is not a biomarker (Fig. 3).

When appropriately defined, developed, and evaluated, any of these four categories of COAs can become accepted as a well-defined and reliable COA and suitable for use in a clinical trial endpoint. Which type of COA is more advantageous for a particular COI is strongly influenced by the specific intended context of use and should be carefully considered at the outset of developing a new COA or evaluating an existing COA.

PRO assessments are those in which the patient is the rater. PRO assessments rely on a patient's direct responses to questions. These responses may be recorded by the patient in various ways such as on paper, via computerized questionnaire forms, or via interviews in which the patient's observations or reports are recorded exactly as spoken, without any interpretation (judgment) on the part of the interviewer. Because a patient's direct report can capture a wide range of feelings and functions, as well as provide a direct measurement of how a patient feels, for example, pain or low mood, there has been increasing interest in developing PRO COAs. FDA guidance on the topic [24] has highlighted the regulatory agency's interest in these instruments.

Clinician reported outcome (ClinRO) assessments are those in which a member of the investigator team with appropriate professional training is the rater. A clinician applies professional expertise and judgment to the observations of, or conversations with, the patient to arrive at a rating according to the COA's definition. For example, a ClinRO assessment may call for the clinician to interpret the patient's responses to questions in an interview, judge the quality of patient's actions, or judge findings of a physical examination. A ClinRO assessment is any COA in which the individual determining the rating must have some specific professional training to properly form a judgment. Note that although all ClinRO outcome assessments are COAs, not all COAs are ClinROs. The distinction is between clinical (attribute 1) and clinician (a specific type of reporter).

Observer reported outcome (ObsRO) assessments are those in which observations can be made, appraised, and recorded by a person other than the patient and do not require specialized professional training. The rating is nonetheless influenced by the perspective of the observer. These include COAs that are best made by a companion, for example, parent, spouse, or caregiver of the patient. The observer rater is often taught what to observe and judge to form the rating, but health care professional training is not needed. The term observer rater is used in this narrowed sense to contrast with clinical professionals who must also interact with and observe the patient when performing ClinRO assessments.

Performance outcomes (PerfOs) are a type of COA in which the patient is assessed but no rater judgment affects the measurement. It is based on the patient's performance of a defined task that is quantified in a specified way that does not rely on judgment to determine the rating. PerfOs include instruments such as the distance walked in 6 minutes and the number of pictorial symbols correctly matched to a key within a fixed amount of time. Although a clinician or observer is administering and monitoring the performance of the PerfO assessment task, this individual does not apply judgment to quantify the performance. These assessments are still categorized as COAs because a patient's motivation or volition is involved in the performance of the task.

Because PerfOs are typically evaluated in the health care clinic setting but do not rely on a rater's judgment for the score, they can often be designed to have good reliability, rendering them attractive for many multicenter clinical trials in disorders in which physical function activities are the meaningful health aspect for treatment benefit. PerfOs are usually not measuring the intended meaningful function activities, however, and are instead usually designed to measure a COI that is thought informative of the meaningful functions (Fig. 2).

Attribute 3: Is the outcome assessment directly measuring a meaningful aspect of how the patient feels or functions?

COA instruments that provide direct evidence about the meaningful health aspect are termed *direct measure COAs*. In some studies, the COA of the endpoint is directly assessing the meaningful health aspect so that observed treatment effects are inherently interpretable as showing meaningful benefit. Many, but not all, PRO instruments are intended to have this interpretability.

Indirect measure COAs are assessing a related, measurable COI (Fig. 3). The interpretation of indirect assessments is not inherently clear with regard to the intended meaningful health aspect. A well-defined relationship between the outcome assessment and the meaningful health aspect enables interpreting a favorable effect on the endpoint as a favorable effect on the meaningful health aspect (i.e., as a treatment benefit).

Recognizing whether the COA is a direct or an indirect measure of the meaningful health aspect is important for determining whether evidence to establish the relationship between the measurements of the COI and the meaningful health aspect is necessary.

The graded quality of indirectness is important to recognize. Some COAs measure abilities or actions representing a COI that is close, but not identical, to how the patient functions in typical life. Examples include some in-clinic performance instruments that simulate activities of daily living [37], or visual acuity testing. These COAs are measurements that are close to evaluating the meaningful health aspect itself. Other COAs measure COIs that are substantially unlike the patient's functioning in daily life (e.g., supine quadriceps isometric strength). Interpreting changes or differences in indirect COAs are dependent on additional evidence that clearly defines the relationship between the COA's measurement and the meaningful health aspect (see discussion of measurement principles in J.H. Powers, D.L. Patrick, M.K. Walton, et al., unpublished report).

For indirect measure COAs, the degree of indirectness between the COI and the meaningful health aspect will guide the amount and type of evidence that should be obtained during the course of developing and evaluating the COA in accordance with good measurement practices (J.H. Powers, D.L. Patrick, M.K. Walton, et al., unpublished report). Establishing the relationship

between the COA measurements of the concept of interest and the meaningful health aspect for a COA that is closely similar to a patient's functioning in typical life will generally be more straightforward than for a COA that is substantially dissimilar to a patient's typical life activities.

This consideration is also relevant to biomarkers that are generally very distant from the actual meaningful health aspect. Substantial amounts of evidence are needed to establish the relationship between a biomarker and the meaningful aspect of feeling or functioning to support use of the biomarker to demonstrate a treatment benefit. When sufficient evidence is available, however, biomarkers can be very valuable and used in surrogate endpoints, for example, blood pressure for many hypertension treatments or hemoglobin A_{1c} levels for some diabetes treatments.

Effect of Context of Use on Measurement Properties

The suitability of a COA for a particular clinical trial is dependent on its measurement properties as used in a specified endpoint for the trial. As stated previously, those measurement properties can be affected by several factors, such as the specific COU and the procedures used in performing the assessment. The COU can affect whether the outcome assessment measurement of the COI is adequately related to the intended meaningful health aspect, as well as the measurement properties of the outcome assessment such as reliability and ability to detect change.

Even within the spectrum of a single disease, the elements of the COU can alter the interpretability of an observed treatment effect. A treatment effect that is a modest fraction of the full dynamic range of a COA might be of clinical value to patients at a mildly affected stage of a disease, but not important to patients at a severely affected stage of the same disease.

The ability of the endpoint to detect change may be different for clinical studies with different design features. A COA that shows a good dynamic range for one patient subpopulation may have an unacceptable ceiling or floor limitation in a different part of the overall disease population. In a study intending to show a treatment benefit as reducing patients' decline (reduction in worsening of the disease) where slow progressive worsening is the natural disease course, a COA's reliability may be adequate to detect a treatment effect in a 1-year trial. However, it may *not* be adequate to demonstrate the treatment effect in a study of 3-month duration that is the same size as the 1-year trial.

Similarly, a COA suitable for one context of use might not be valid when a different disease is studied even if the intended meaningful health aspect does not change. For example, a COI (measured by a COA) closely linked to the meaningful health aspect in one disease may be poorly linked in another disease because other features of the second disease diminish the influence of the COI on the meaningful health aspect.

Identifying the COA

Similar to explicitly identifying the intended treatment benefit, COAs need to be clearly identified both in name and in description. Because the COA's measurement properties and suitability for a clinical trial can vary with the specific COU, the COU should also be clearly described when developing and evaluating a COA.

New COAs should be given an appropriate name, such as indicating the task performed for a PerfO assessment or the intended COI for a ClinRO assessment. Furthermore, when COAs are revised for a new context (COU), version should be clear. There are long-standing COAs in which the procedures have changed over time as different investigators use the COA in different studies. Frequently, modifications are not stated in study reports; for example, the version and the details of the COA were not described [38]. Thus, studies that appear to use the same COA may in fact have used distinctly different instruments. To avoid this confusion in evaluating or using a COA, it is important to use clear and informative names for new COAs and identify the version of an existing assessment.

For assessments that directly measure the meaningful health aspect, this may be clear from the description of the measurement procedure. Unfortunately, intrinsic clarity regarding the intended meaningful aspect of health is seldom the case for indirect assessments. Therefore, the intended meaningful health aspect should be explicitly identified for all assessments, irrespective of whether it seems implicitly clear.

Finally, it is essential to evaluate an existing COA planned for critical study endpoints to determine whether there is sufficient information to establish the relationship with the meaningful health aspect. The name itself does *not* support the conclusion that it is a good measurement of that meaningful health aspect. For example, a COA named "Disease X Disability Scale" does not alone ensure that it is a good measurement of "disability" for disease X or for any other disease. The recommended approaches are as follows: In an existing COA, evaluate the evidence to determine whether there is a well-defined relationship with the meaningful health aspect. In a new COA, name it to reflect what is actually measured and separately name the intended meaningful health aspect.

Conclusions

The benefit of medical interventions is generally demonstrated by controlled clinical trials of the intervention that show favorable effects on the efficacy endpoints. A successful clinical trial depends on many factors, among them, the availability of good outcome assessments to use in defining study endpoints.

Demonstrating benefit in a disease may require the development of a patient assessment that does not already exist or the careful reevaluation of an assessment that has not yet been shown to be well defined, reliable, and interpretable for studies of the specific disease. Development of a new assessment may also be needed to support 1) an improved endpoint that is capable of demonstrating the advantage of a new therapy over an existing one or 2) a new endpoint to evaluate treatment effect on an aspect of the disease that has not previously been evaluated.

An effect on an endpoint needs to be interpretable as a meaningful effect for patients, that is, a positive effect on how patients feel, function, or survive. A study endpoint is built on an outcome assessment (most commonly clinical assessments). It is intended to represent the meaningful health aspect when used in a well-specified context (COU). A COA measures a concept of interest that may either 1) directly measure the meaningful health aspect or 2) measure a COI thought to have a strong relationship with the meaningful health aspect and is more readily measured.

The first step to understanding the performance properties of an outcome assessment in a COU is to identify the attributes of the outcome assessment, particularly those discussed in this report. All outcome assessments can be categorized by 1) whether patient volition/motivation or rater judgment determines the rating or score; 2) the type of rater applying judgment (the patient, a clinician, an observer) or a PerfO in which no rater judgment enters into the rating; and finally, 3) whether the intended meaningful health aspect is directly assessed or indirectly assessed.

Categorizing an outcome assessment by these attributes can aid investigators in the work of developing and evaluating a COA.

Elements of these efforts that warrant careful attention include 1) specifying the detailed procedure for conducting the assessment, 2) determining how to evaluate whether the outcome assessment is well defined and reliable, and 3) refining the outcome assessment during the course of its development for a specific COU. Assessing whether a COA is well-defined and reliable and suitable for a planned clinical trial also necessitates 1) clearly and fully identifying the new COA with an appropriate name, or if an existing COA is used, include the version and 2) providing a detailed explanation of the intended meaningful health aspect, the COI measured, and the COU.

The concepts and terminology presented in this article can aid in clearly describing and effectively developing or evaluating COAs. Many of these COA features may not have been not well specified for older COAs. This does not mean that these older COAs are invalid and unsuitable for further use. Instead, investigators should recognize that the characteristics of an older COA may not be completely understood and need further evaluation. Investigators should consider this during the COA selection process. If investigators determine that substantial COA development or evaluation is needed for a future clinical trial, this effort should be initiated well in advance of finalizing the study design. Following these recommendations can lead to improved COAs and better study endpoints, enabling more informative and efficient clinical trials.

Glossary

Brief summary definitions; see text for full explanation.

Biomarker: A patient assessment that is not influenced by the patient's motivation or volition or a rater's judgment. Common types of biomarkers are biochemical measurements of blood and quantitative measurements of radiographic images.

Clinical Assessment: An assessment that is susceptible to variation unrelated to the patient's true clinical status owing to 1) variation in patient volition (which may cause rating differences for an individual patient between successive administrations or between patients), or 2) dependence on the judgment of a rater (which may differ between raters owing to differing experiences or perspective).

Clinical Outcome Assessment (COA): A clinical assessment instrument that is used as the measure of patient outcome in a clinical trial. There are four types: PRO, ClinRO, ObsRO, and PerfO.

Clinician Reported Outcome (ClinRO): A type of COA in which a member of the investigator team is the rater. The investigator's professional training is relied upon to judge what rating or score will be reported. All ClinROs are COAs, but all COAs are not ClinROs.

Concept of Interest (COI) for Measurement: The concept that the outcome assessment is intended to measure. The COI may be identical to the selected meaningful aspect of feeling or function. Frequently, however, the COI is a simplified form or component of a feeling or function that is *not* an inherently meaningful feeling or function of a patient's typical life, that is, not a complete meaningful health aspect, but thought to be indirectly well related to a meaningful health aspect.

Context of Use (COU): A description of the specifics of the study design, how the COA is used within the study, and result interpretation.

Directly Meaningful COA: A COA that directly measures the patient's actual feelings or a function(s). The feelings or function(s) must be meaningful to the patient and part of the patient's typical (normal) life. All COAs are intended to measure the COI. COAs are directly meaningful only when the COI is the same as the intended meaningful health aspect.

Feeling, Functioning, Survival: Aspects of a person's usual (typical) life that may be adversely altered by a disease.

Indirectly Meaningful COA: A COA that indirectly evaluates feelings or functions that are meaningful and are part of the patient's typical life. Indirectly meaningful COAs are, however, intended to have a good relationship with the meaningful health aspect.

Meaningful Health Aspect: An aspect of health (feelings, functions, or survival) affected by the disease that the patient cares about and has a preference that it 1) does not become worse, 2) improves, or 3) is prevented.

Observer Reported Outcome (ObsRO): A COA in which observations can be made, appraised, and recorded by a person other than the patient who does not require specialized professional training. The rating is nonetheless influenced by the perspective of the observer.

Outcome Assessment: A measuring instrument that provides a rating or score (categorical or continuous) that is intended to represent some aspect of the patient's medical status. Appropriate outcome assessments may include both COAs and biomarkers.

Patient Reported Outcome (PRO): A COA in which the report comes directly from the patient. The patients' responses to questions about their health condition are recorded without amendment or interpretation by anyone else.

Performance Outcome (PerfO): A COA in which the patient is assessed by performing a defined task that is quantified in a specified way. Although a member of the investigator team may administer the PerfO task and monitor the patient's performance, the investigator does not apply judgment to quantify the performance.

Study Endpoint: The outcome result obtained in a clinical trial and interpreted to determine whether the therapy has provided a treatment benefit. Study endpoints are composed of a specific outcome assessment, measured at specified times during the study, and analyzed according to a specified statistical method.

Treatment Benefit: A favorable effect on a meaningful aspect of how patients feel or function in their life, or on survival. It is an effect on an aspect of health affected by the disease that is an alteration in feeling or functioning, about which the patient cares that it is affected, and has a preference that it does not become worse, improves, or is prevented. The aspect of feeling or functions affected by the therapy should be what occurs in the patient's usual (typical) life.

Acknowledgement

We thank Elektra Papadopoulos and Ashley Slagle, Center for Drug Evaluation and Research, Food and Drug Administration, for their valuable comments on this manuscript as it was being written.

ISPOR member comments contribute to the high-quality consensus nature that characterizes ISPOR Good Practices for Outcomes Research reports. We are deeply appreciative of the valuable written feedback on earlier drafts of this report by the following 40 members of the ISPOR PRO Review Group: Sarah Acaster, Michael Adena, Ethan Basch, Marc Berger, G.S. Bhattacharyya, Monika Bullinger, Joe Capelleri, Lee Yee Chong, Susan Dallabrida, Slaveyko Djambazov, Celeste Elash, Francis Fatoye, João Guerra, Chad Gwaltney, Mike Hagan, Katarina Halling, Stacie Hudgens, Christine C. Huttin, Zhanna Jumadilova, Cicely Kerr, Bellinda King-Kallamanis, Kathy Lasch, Hsiao-Yi Lin, Yvonne Lis, Linda Gore Martin, Oren Meyers, Talya Miron-Shatz, Annabel Nixon, Cathy Anne Pinto, Antoine Regnault, Diana Rofail, Sarah Shingler, Ajit Singh, Lisa Strouss, Jun Su, Tara Symonds, Diane Turner-Bowker, Wendy J. Ungar, and Etta Vinik.

We are also grateful for the comments, expertise, and insight of those who provided oral comments during the four presentations of our work.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government

The views expressed in this article are the personal views of the author(s) and may not be understood nor quoted as being made on behalf of or reflecting the position of European Medicines Agency or one of its committees or working parties or any of the national agencies

Source of financial support: There are no financial sponsors of this task force and manuscript. One coauthor is required to state the source of funding for his work in general (not for this manuscript specifically). Support for J.H. Powers has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health (contract no. HHSN261200800001E).

REFERENCES

[1] Johansson GM, Hager CK. Measurement properties of the Motor Evaluation Scale for Upper Extremity in Stroke patients (MESUPES). Disabil Rehabil 2012;34:288–94.
[2] Coster WJ, Haley SM, Ludlow LH, et al. Development of an applied cognition scale to measure rehabilitation outcomes. Arch Phys Med Rehabil 2004;85:2030–5.
[3] Burk K, Schulz SR, Schulz JB. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. J Neurochem 2013;126(Suppl. 1):118–24.
[4] Howard K, Berry P, Petrillo J, et al. Development of the Shortness of Breath with Daily Activities questionnaire (SOBDA). Value Health 2012;15:1042–50.
[5] Padua L, Evoli A, Aprile I, et al. Myasthenia gravis outcome measure: development and validation of a disease-specific self-administered questionnaire. Neurol Sci 2012;23:59–68.
[6] Merkel PA, Clements PJ, Reveille JD, et al. Current status of outcome measure development for clinical trials in systemic sclerosis: report from OMERACT 6. J Rheumatol 2003;30:1630–47.
[7] Brown B, McKenna SP, Solomon M, et al. The patient-reported impact of scars measure: development and validation. Plast Reconstr Surg 2010;125:1439–49.
[8] Marquis P, Lasch KE, Delgado-Herrera L, et al. Qualitative development of a patient-reported outcome symptom measure in diarrhea-predominant irritable bowel syndrome. Clin Transl Gastroenterol 2014;5:e59.
[9] Aabenhus R, Thorsen H, Siersma V, Brodersen J. The development and validation of a multidimensional sum-scaling questionnaire to measure patient-reported outcomes in acute respiratory tract infections in primary care: the acute respiratory tract infection questionnaire. Value Health 2013;16:987–92.
[10] Drobes DJ, Thomas SE. Assessing craving for alcohol. Alcohol Res Health 1999;23:179–86.
[11] Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. J Clin Exp Neuropsychol 2012;34:11–34.
[12] Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007;6:1094–105.
[13] Paltamaa J, Sarasoja T, Leskinen E, et al. Measures of physical functioning predict self-reported performance in self-care, mobility, and domestic life in ambulatory persons with multiple sclerosis. Arch Phys Med Rehabil 2007;88:1649–57.
[14] Hobart J. Rating scales for neurologists. J Neurol Neurosurg Psychiatry 2003;74(Suppl. IV):iv22–6.
[15] Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure. Arch Dermatol 2004;140:1513–9.
[16] Terwee CB, Bouwmeester W, van Elsland SI, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Osteoarthritis Cartilage 2011;19:620–33.
[17] Terwee CB, Mokkink LB, Steultjens MPM, Dekker J. Performance based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Rheumatology 2006;45:890–902.
[18] Cano SJ, Hobart JC, Hart PE, et al. International Cooperative Ataxia Rating Scale (ICARS): appropriate for studies of Friedreich's ataxia? Mov Disord 2005;20:1585–91.
[19] Spuls PI, Lecluse LLA, Poulsen MNF, et al. How good are clinical severity and outcome measures for psoriasis? Quantitative evaluation in a systematic review. J Invest Dermatol 2010;130:933–43.
[20] Croarkin E, Janoff J, Barnes C. Evidence-based rating of upper extremity motor function tests used for people following a stroke. Phys Ther 2004;84:62–74.
[21] Rothman M, Burke L, Erickson P, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. Value Health 2009;12:1075–83.
[22] Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 1—eliciting concepts for a new PRO instrument. Value Health 2011;14:967–77.
[23] Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 2—assessing respondent understanding. Value Health 2011;14:978–88.
[24] US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064981.htm [Accessed July 20, 2014].
[25] Matza LS, Patrick D, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO Good Research Practices for the Assessment of Children and Adolescents Task Force. Value Health 2013;16:461–79.
[26] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[27] Hobart JC, Lamping DL, Thompson AJ. Evaluating neurological outcome measures: the bare essentials. J Neurol Neurosurg Psychiatry 1996;60:127–30.
[28] Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury 2011;42:236–40.
[29] US Food and Drug Administration. Clinical Outcome Assessment Qualification Program. Available from: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm. [Accessed May 13, 2015].
[30] Freeman J, Walters R, Ingram W, et al. Evaluating change in mobility in people with multiple sclerosis: relative responsiveness of four clinical measures. Mult Scler 2013;19:1632–9.
[31] Post B, Merkus MP, de Bie RMA, et al. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord 2005;20:1577–84.
[32] Wild D, Eremenco S, Mear I, et al. Multinational trials—recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient Reported Outcomes Translation & Linguistic Validation Good Research Practices Task Force Report. Value Health 2009;12:430–40.
[33] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.
[34] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force Report. Value Health 2009;12:419–29.
[35] Eremenco S, Coons SJ, Paty J, et al. PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO Mixed Modes Good Research Practices Task Force. Value Health 2014;17:501–16.
[36] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89–95.
[37] Schmitter-Edgecombe M, McAlister C, Weakley A. Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: the day-out task. Neuropsychology 2012;5:631–41.
[38] Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. Int J Stroke 2009;4:200–5.