VALUE IN HEALTH 16 (2013) 554–563

Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Patient-Reported Outcomes

Examining Item Content and Structure in Health Status and Health Outcomes Instruments: Toward the Development of a Grammar for Better Understanding of the Concepts Being Measured

Pennifer Erickson, PhD[1,*], *Richard J. Willke, PhD*[2]

[1]*OLGA, State College, PA, USA;* [2]*Pfizer, Inc., New York, NY, USA*

ABSTRACT

Objectives: Health outcomes instruments assess diverse health concepts. Although item-level concepts are considered fundamental elements, the field lacks structures for evaluating and organizing them for decision making. This article proposes a grammar using item stems, response options, and recall periods to systematically identify item-level concepts. The grammar uses "core concept," "evaluative component," and "recall period" as intuitive terms for communicating with stakeholders. Better characterization of concepts is necessary for classifying instrument content and linking it to treatment benefit. **Methods:** Items in 2 generic and 21 disease-specific instruments were evaluated to develop and illustrate the use of the grammar. Concepts were assigned International Classification of Functioning, Disability and Health codes for exploring the value that the grammar and a classification system add to the understanding of content across instruments. **Results:** The 23 instruments include many core concepts; emotional function is the only concept assessed in all instruments. Concepts in disease-specific instruments show obvious patterns; for example, arthritis instruments focus on physical function. The majority of instruments used the same response options across all items, with five-point scales being the most common. Most instruments used one recall period for all items. Shorter recall periods were used for conditions associated with "flares," such as chronic obstructive pulmonary disease and "skin disease." Every diagnosis, however, showed variation across instruments in the recall period used. **Conclusions:** This analysis indicates the proposed grammar's potential for discerning the conceptual content within and between health outcomes instruments and illustrates its value for improving communication between stakeholders and for making decisions related to treatment benefit.

Keywords: Core concepts, International Classification of Functioning, Disability, and Health (ICF), patient-reported outcomes, taxonomy.

Introduction

Instruments that measure health from an individual's perspective assess a wide range of concepts of health and well-being and are used for evaluating treatment benefit throughout the world [1–4]. The development of these instruments can be traced to generic measures of health that emerged in the 1970s [5–8] and were based on the 1948 World Health Organization's definition of health as "physical, mental, and social well-being" [9]. This tripartite classification was expanded by Ware [10] to include disease, personal functioning, psychological distress/well-being, general health perceptions, and social/role functioning. Subsequently, Wilson and Cleary [11] modified the Ware model to include a medical focus and external factors. Patrick and Chiang [12] expanded this model to include more detail on both the health concepts and factors due to the environment and personal and lifestyle factors. These conceptual models have been used as the basis for the development of numerous generic and disease-specific instruments ranging in complexity from multi-item scales that assess a single concept, for example, depression, to scales that measure multiple concepts by using multiple scales that may, or may not, be aggregated to form an overall outcome measure [13–15]. In the past decade, generic instruments have become increasingly widely used to measure health status, for example, in population surveys by the Medical Care Expenditure Survey and the Medicare Health Outcomes Survey [16,17]. During the same time, specific measures have become routinely used to assess health and treatment outcomes in clinical research studies including those used for new drug development [18].

Approximately one third of all new drugs approved by the U.S. Food and Drug Administration in the period 1997 to 2003 and one fourth in the period 2006 to 2010 included labeling claims based on at least one patient-reported outcome (PRO) instrument [18,19]. This information was used in the process of generating

*Address correspondence to: Pennifer Erickson, OLGA: The On-Line Guide to Quality-of-Life Assessment, 1316 Deerfield Drive, State College, PA 16803, USA.

E-mail: pae6@psu.edu.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.01.008

a Food and Drug Administration Guidance on PROs and a European Medicines Agency Reflection Paper [1,2]. The positions put forth in these regulatory documents are now being expanded to include information reported by clinicians and by patients' observers. Together these reported assessments are referred to as clinical outcomes assessments (COAs) [20].

As a result of the growing importance of obtaining individuals' input into health care decision making, the number and diversity of instruments have increased exponentially in the past 30 years. Common features, however, exist across instruments. At the item level, COAs include core concepts that are considered to be fundamental elements of the broader concepts that may be represented by multi-item scales.

Together with the core concept, other aspects of item wording are also important to fully understand item content, namely, the recall period, the response option, and the verb-adverb of the question ("how often are," "how troubled by," "how severe are"); these elements, which comprise the "grammar" of COAs, convey the precise way that the core concept can be systematically evaluated within a given item. Better characterization of the choice set, not only for the core concepts but also for the other aspects of the grammar, would lead to a more complete classification of the conceptual content of an item in a COA and how it measures a specific treatment benefit.

Two recent initiatives that aim to provide a structure for classifying core concepts within COAs are the Patient Reported Outcomes Measurement Information System and the International Classification of Functioning, Disability, and Health (ICF) [21,22]. Both of these add lower-level concepts, for example, activities of daily living and instrumental activities of daily living, to the physical concept. Because both the Patient Reported Outcomes Measurement Information System and the ICF were developed for uses broader than determining treatment efficacy, each lacks a structure that relates a health concept to its role in evaluating treatment outcome, for example, whether or not a concept is appropriate for making an explicit statement of treatment benefit as might appear in a label claim.

To aid the understanding of COA concepts and their contribution to the drug-approval process and decision making more broadly, Erickson et al. [23] proposed the PRO Concept Taxonomy. This hierarchical structure has four levels: family, compound, singular, and low-level singular concepts. While concepts as well as their names and definitions have yet to be specified, family-level concepts can be thought of as corresponding to those in the 1948 World Health Organization definition and its subsequent modifications. Concepts at each level correspond to the importance placed on them by those who are directly involved with the patient's response to treatment. Some item-level concepts, such as shampooing hair, are unlikely to be of sufficient importance for labeling and would be considered low-level singular concepts. Other item-level concepts, such as pain severity, may be singular concepts if they can be used for claiming treatment benefit. Multi-item scales form singular or higher-level concepts. A hierarchical structure, such as that in the PRO Concept Taxonomy, allows investigators to identify patient-relevant concepts that are appropriate for the intended claim of treatment benefit at the time of trial design, thus increasing the likelihood that the trial results will support the desired labeling.

While other taxonomies have been proposed for use with PRO measures [24–26], these efforts focus on practical characteristics of instruments, for example, respondent burden. These systems are similar to the PRO Instrument Hierarchy that was developed as a companion structure to the PRO Concept Taxonomy. Structures that categorize instruments on the basis of their measurement features, rather than conceptual content, are also important tools that can improve communication among researchers, sponsors, and other consumers of COA information.

There is certainly an awareness of the importance of not only the core concepts but also other aspects of the grammar of COA items [27,28]. The final FDA PRO guidance devotes specific attention to recall periods and response options. There has been less effort to classify these aspects, however, than there has been to organize the concept taxonomies. Nor has there been much effort to analyze the impact of differences in recall and response formats on the measurement and interpretation of otherwise similar core concepts. Whether or not variations in these other components of the grammar will make significant differences in the overall content and psychometrics of an item is an empirical, case-by-case, question, but it cannot be assumed that they will not matter. Better delineation of the structure and usage of this grammar would be a useful step in addressing these considerations. As this brief review indicates, the key for having meaningful discussions about the use and interpretation of COAs is to have a common language for understanding an instrument's content and potential for decision making. This language needs to be compatible with concepts measured by existing measures and also with those in instruments being developed now and in the future.

This article evaluates the syntax and language of items to develop a descriptive, item-level, grammar of COAs that can guide users in identifying item-level concepts and understanding each concept's role in evaluating and making a statement of treatment benefit. For comparison, analysis, and communication, core concepts are best understood as part of cohesive concept taxonomy. In addition to introducing the grammar, this article suggests the value of using both the grammar and taxonomy for evaluating instrument selection within the context of a given application. Similarities of concepts within levels of the taxonomy indicate confirmatory information; differences suggest disparate concepts that need further investigation. The grammar is designed to be applicable to existing and new and generic and disease-specific instruments. Items in a sample of PRO instruments are compared and contrasted by using the grammar to identify similarities and differences in the core concepts across instruments. The ICF was used to explore the value that the proposed grammar adds to the evaluation and classification of the conceptual content of various instruments. The ICF was selected because it has 1) a well-defined hierarchical structure similar to that in the PRO Concept Taxonomy; 2) clearly stated concept definitions that enable the matching of core concepts in items in the sampled instruments to levels in the hierarchy; and 3) a coding system that indicates the location of the concept in the taxonomy, in terms of its conceptual content and its potential role in decision making. The longer the ICF code, the more detailed the concept, the narrower the concept, and the more restricted the concept's role in decision making. Thus, a subtext is to evaluate the ICF's potential for populating a PRO Concept Taxonomy as a step toward developing a standardized terminology for improving communication among the various stakeholders involved in evaluating treatment benefit.

Methods

To systematically evaluate the conceptual content of items in COA instruments, as well as to provide considerations for new instrument development, we propose a grammar based on the item-stem wording, recall period, and response option. This grammar takes into account the role that modifying phrases play in understanding the item's core concept. Table 1 identifies the primary components in this descriptive tool and gives a brief definition and statement of purpose for each. In addition, grammar components are expressed in terms chosen to make them more intuitive to those unfamiliar with terms used primarily by COA researchers.

All items have three components—*core concept, evaluation, and recall period*—with the first two each having a subcomponent.