Qual Life Res (2008) 17:1125–1135
DOI 10.1007/s11136-008-9396-4

# Patient reported outcome measures: a model-based classification system for research and clinical practice

**Jose M. Valderas · Jordi Alonso**

Accepted: 6 September 2008 / Published online: 3 October 2008


**Abstract**
*Purpose* The umbrella term Patient Reported Outcomes (PRO) has been successfully proposed for instruments measuring perceived health outcomes, but its relationship to current conceptual models remains to be established. Our aim was to develop a classification system for PRO measures based on a valid conceptual model.
*Methods* We reviewed models and classification schemes of health outcomes and integrated them in a common conceptual framework, based on the models by Wilson and Cleary and the International Classification of Functioning (ICF). We developed a cross-classification system based on the minimum common set of consistent concepts identified in previous classifications, and specified categories based on the WHO International Classifications (ICD-10, and ICF). We exemplified the use of the classification system with selected PRO instruments.
*Results* We identified three guiding concepts: (1) construct (the measurement object); (2) population (based on age, gender, condition, and culture); and (3) measurement model (dimensionality, metric, and adaptability). The application of the system to selected PRO measures demonstrated the feasibility of its use, and showed that most of them actually assess more than one construct.
*Conclusion* This classification system of PRO measures, based on a valid integrated conceptual model, should allow the classification of most currently used instruments and may facilitate a more adequate selection and application of these instruments.

**Keywords** Classification · Construct · Measurement · Patient reported outcomes · Quality of life

J. M. Valderas (✉)
National School for Primary Care Research,
University of Manchester, Oxford Road,
Williamson Building, Manchester M13 9PL, UK
e-mail: jose.valderas@manchester.ac.uk

J. M. Valderas · J. Alonso
Health Services Research Unit. Institut Municipal d'Investigació Mèdica (IMIM-hospital del Mar), Barcelona, Spain

J. M. Valderas · J. Alonso
CIBERESP Spanish Network of Epidemiology and Public Health, http://www.ciberesp.es/

## Introduction

In the past decades, research focusing on health outcomes measurement has experienced an enormous expansion [1]. Health-related quality of life (HRQL) is amongst the most important of these outcomes. A recent systematic review identified 1,275 different instruments measuring HRQL and other related outcomes by the year 2000 [2].

The definition of HRQL and related concepts like health status and perceived health, among others, has been disputed and elusive, resulting in no single concept being universally adopted [3–5]. In a recent review of 68 different HRQL models, Taillefer et al. [6] observed that about 4 out of 10 models did not provide a clear definition of the concept. When definitions were provided, they differed significantly in their content.

Fostering simplification, the Food and Drug Administration (FDA) has hence proposed the umbrella term patient reported outcomes (PRO): "a measurement of any aspect of a patient's health status that comes directly from the patient (i.e., without the interpretation of the patient's responses by a physician or anyone else)" [5, 7]. The term is not new in the field [8], and it is appealing. Rather than

overcoming the conceptual problems related to the conceptualization of the constructs being measured, this approach simply avoids them by focusing on the source of information rather than on the content. At the same time, it emphasizes the genuine importance of the individual's own perspective when making the evaluation.

Several rigorous and useful classifications of health outcome measures have been published previously [1, 3, 7, 9–20]. Many of them claim to have been specifically devised for HRQOL. A current Guidance for industry by the Food and Drug Administration has been the first to address such a classification from the unique perspective of PRO. Nevertheless, concerns have been raised about its limited focus (clinical trials) [21] and utility, due to the number of criteria (by far the longest) and their nature (including as such the numbers of items or the frequency of administration, not well supported in the literature) [7]. Its most significant limitation, similar to that of other previous attempts in the literature, is the lack of an explicit link to any valid model of health outcomes, whether supported by empirical evidence of validity or not [4, 22].

A classification system linked to a conceptual model would represent a substantial improvement for identifying a candidate pool of PRO instruments for a given purpose, since it would facilitate a comprehensive view of measures (including the identification of areas where there are a number of measures, and areas where there is a lack of them). It would also facilitate the selection of PRO measures to be used in research, management, and, eventually, in clinical practice, if standard guidelines were provided along with the classification system [23].

In this paper, we present the development of a classification system of PRO instruments, based on a valid conceptual model of health outcomes, and we apply it to the most commonly used instruments. We also discuss the added value of our approach for a broad range of health professional users.

## Methods

We aimed to develop a simple classification system based on the minimum possible set of relevant criteria. We reviewed previous classifications of PRO measures and identified different areas of classification. Starting with selected previous classifications [1, 3, 7, 11], we applied a snow ball technique to identify additional references [24] (Table 1). Three concepts were consistently pointed out as important, even if using different wording: the construct (or measurement object), the population to be assessed (range and characteristics of the people to whom the instrument should be applied), and the measurement model (Table 1). These concepts are the independent non-hierarchical principles (axes) in our classification system, and are instrumental in answering key questions in the measurement process (what? whose? and how?) (Table 2a). Within each axis, categories were established which characterize each instrument in relation to that particular axis (cross-classification) [26].

We then applied the classification to a selection of the most evaluated and used PROs, as identified in a previous systematic review [2]. In the absence of a 'gold standard', the identification of the constructs measured by an instrument is performed on the basis of the review of its content (content validity), further supported with evidence of its relationship with other related variables (construct validity) [27]. The assessment of the content of each instrument was based on the content of each item, the minimal units that form all PROs. Every item is a stimulus, in the form of a question, task or individual component in a scenario, which the individual is given in order to elicit a response [28]. It can thus be considered as an operational definition of the intended measurement object, and this was the basis for our classification of constructs of the selected instruments, using the definitions provided in the next section.

One of the authors (J.M.V.) classified all of the instruments using the previously defined axes and categories, and the final classification for each instrument was agreed by consensus among the authors (J.M.V., J.A.). To exemplify the use of the classification system across categories not covered by these instruments, we further exemplified its use with other selected instruments.

## Results

### The classification system and its rationale

#### *Construct*

Construct is the range of characteristics (traits and states) measured by the instrument, its measurement object. Our classification of constructs relies on the model proposed by Wilson and Cleary [29], a well-established bio-psychosocial model for health outcomes [30] (Fig. 1). Sullivan et al. [31] tested the model in Dutch community-dwelling elders and did not find it completely satisfactory. In more recent years, though, strong empirical evidence has been obtained supporting its validity in a variety of contexts, including both the general population aged over 65 [32] as well as patients living with HIV [33] or suffering from coronary heart disease [34, 35], and, most importantly, with very different types of measures (including the SF-36 Health Survey, the Nottingham Health Profile, Health Assessment Questionnaire-Disability Index (HAQ-DI), and

**Table 1** Criteria used for the classification of patient reported outcome measures

| Criteria | Source | Original wording |
|---|---|---|
| Construct | Bowling [19] | Concept (functional ability, health status, psychological well-being, social networks, life satisfaction) |
| | Tully et al. [3] | Underlying concept (functional status, health status, general health perceptions, health related quality of life, quality of life) |
| | Patrick et al. [25] | Concepts (symptoms, functional status, health perceptions, spiritual, disadvantage/opportunity, resilience, environmental) |
| | McDowell et al. [11] | Aspects of health (physical disability, social health, psychological well-being, depression, mental status testing, pain measurement, general health status/quality of life) |
| | FDA Guidance [7] | Concepts measured (overall health status, symptoms/signs, functional status, health perceptions, satisfaction with treatment or preference for treatment, adherence to medical treatment) |
| | Cella et al. [20] | Health domains (global health, physical health, mental health, social health) |
| Population | Guyatt et al. [15] | Generic versus specific instruments |
| | McKeigan et al. [17] | Scope |
| | Patrick et al. [18] | Range of populations and concepts |
| | Tully et al. [3] | Breadth of content |
| | FDA Guidance [7] | Intended measurement or condition |
| Measurement | Patrick et al. [1] | Type of scores produced, weighting system |
| | McKeigan et al. [17] | Aggregation of scores |
| | FDA Guidance [7] | Types of scores, weighting of items or concepts |
| | Patel et al. [62] | Standardized versus patient-generated |
| Others | Kirshner et al. [13] | Purpose |
| | Patrick et al. [18] | Mode of collection |
| | Osoba [12] | Levels of decision making |
| | FDA Guidance [7] | Intended use of measure, number of items, mode of data collection, timing and frequency of administration, response options |

the MacNew Heart Disease Quality of Life Questionnaire, among others) [32–35].

While the model of Wilson and Cleary is the foundation of our classification proposal, this model can, to a considerable extent, be integrated with the theoretical model underpinning the International Classification of Functioning, Disability and Health (ICF). This ambitious classification system of health states has been proposed by the World Health Organization, and is based on a sociological perspective of health that considers disability along the whole functioning continuum [36, 37]. Both models have been conceived independently, and still they share significant characteristics. They both differentiate health related variables and contextual factors, further splitting the latter into environmental and individual characteristics (Fig. 1). Biological and physiological variables in the model by Wilson and Cleary correspond to the structure component of the ICF, and functional variables in the first model equally correspond to the activities and participation components of the latter [38]. Current and previous successful mapping of responses to PRO measures onto the international classification of functioning system support the validity of this integration [39, 40].

Based on the model by Wilson and Cleary, we differentiated and defined the following concepts: symptom status, functional status, health perceptions, and health related quality of life (Box 1). Although the original formulation of the model considered the more general concept of overall quality of life, all empirical evidence has been obtained for health related quality of life [33–35], and this was the construct finally included. In addition, the original model also considered biological or physiological variables, but the patient is then usually not the preferred source of information, and so this category has not been included in this classification system for patient reported outcome instruments.

We also considered some other health related constructs that were not specified in Wilson and Cleary model nor in

**Table 2a** A classification system for patient reported outcome measures: axis and categories

| Axis | Categories |
|---|---|
| A. Construct | A.1. Symptoms (see Table 2b) |
| | A.2. Functional status (see Table 2c) |
| | A.3. Health perceptions |
| | A.4. Health related quality of life |
| | A.5. Other health related constructs |
| | Satisfaction with care |
| | Disadvantage |
| | Resilience |
| | Environmental |
| B. Population | B.1. Age |
| | (a) All ages |
| | (b) Children |
| | (c) Adolescents |
| | (d) Adults |
| | (e) Seniors |
| | B.2. Gender |
| | (a) All genders |
| | (b) Female |
| | (c) Male |
| | B.3. Disease (see Table 2b) |
| | B.4. Culture (country and language dyads) |
| C. Measurement | C.1. Metric |
| | (a) Psychometric |
| | (b) Econometric |
| | (c) Clinimetric |
| | (d) Other metrics |
| | C.2. Dimensionality |
| | (a) Index |
| | (b) Profile |
| | (c) Index and profile |
| | C.3. Adaptability |
| | (a) Completely standardized |
| | (b) Partially individualized |
| | (c) Completely individualized |

the ICF classification system, most notably satisfaction with health care (the extent of an individual's experience with health care compared to his/her expectations) [43]. Although the construct satisfaction with care may have been used less extensively, its well-described nature as a health outcome and its widespread use support its inclusion in the classification system [44], as is also the case for resilience (ability to cope or withstand stress and illness) [18, 43]. We have therefore included an additional category for "Other Health Related Constructs" in our model (Fig. 1, Table 2a)

Previous classification systems have relied on ad hoc constructed lists for symptoms and diseases. In order to achieve better standardization, we propose to classify symptoms relying on the implied codes in the International Classification for Diseases ICD-10, 2nd version [45] (Table 2b). Similarly, functional status can be specified according to ICF chapters (Table 2c) [46].

*Population*

The population of a PRO measure is the universe of persons for which the instrument is suited. It is defined in terms of age and gender, presenting diseases (if any) and culture (Table 1), all of them undisputedly relevant to the characterization of the patient from a clinical, epidemiological, and organizational point of view. Further, the important health differences in subpopulations defined according to these criteria underlie the rationale for the development of subpopulation specific instruments [47]. We will not address here the concept of culture [9], but for the purposes of this classification only, we conceptualize culture as the dyad of language and country of the population for which the instrument has been devised.

*Measurement model*

Two issues are of utmost relevance to the measurements elicited by a PRO instrument (and therefore, to their interpretation): the theoretical model that sustains the metric of the instrument, and the level of aggregation of the score (dimensionality).

Metric refers to the method used to assign numeric values to the responses given by the individuals and the construction of the scores. Three broad groups of instruments can be distinguished: *psychometric*, *econometric*, and *clinimetric* [1, 3]. Scoring algorithms of psychometric instruments are broadly based on the sum of item responses for each scale, either weighted or not [48]. The main differences between psychometric and clinimetric instruments arises from the methods used in the scale development, with the former building upon theoretical models (i.e., sample domain theory) and using of sophisticated statistical methods [46], and the latter focusing almost exclusively on clinical relevance [49, 50]. These instruments are best suited for ordering individuals along a continuum for their comparison in clinical trials, monitoring or, to a lesser extent, screening and/or diagnosing of patients.

Econometric measures have come about due to the need to assess and value health states as separate entities. They aim to obtain values based on health state preferences (of patients, populations, experts, etc.), using methods from the field of econometrics, based on decision theory. These preferences are known as 'utilities', and the measuring instruments are called utility or preference-based measures. Utilities can be associated with an appropriate time interval

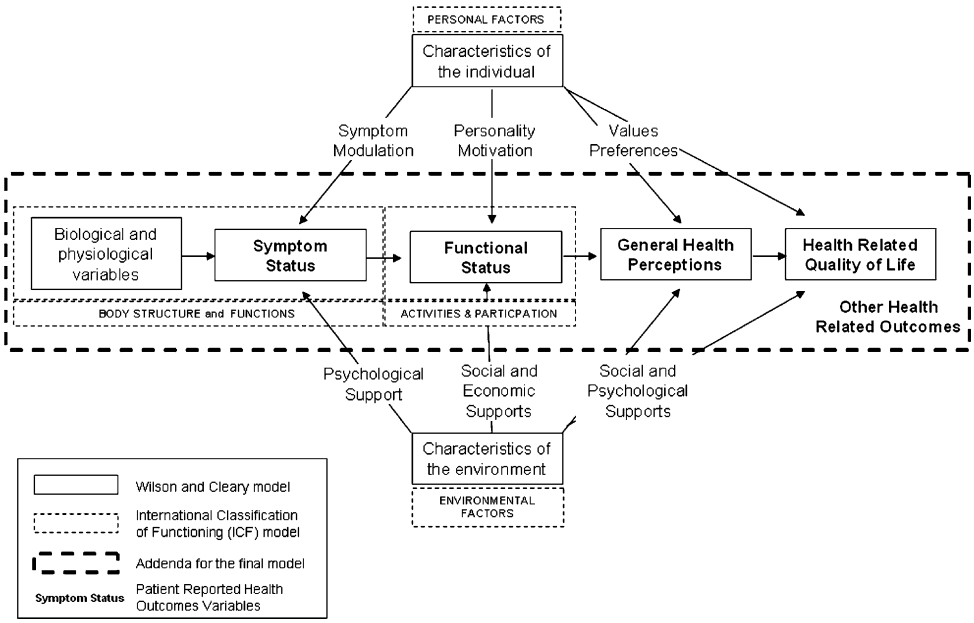


**Fig. 1** An integrated model for health outcomes. Modified from [29] and [46]

**Box 1** Definitions of relevant terms used in the classification system [29]

- *Symptom status*: patient's perception of an abnormal physical, emotional or cognitive state

E.g.: "How much bodily pain have you had during the past 2 weeks? [Response options: None, Very mild, Mild, Moderate, Severe, Very severe]" (MOS Short Form SF-36); "I'm tired all the time [Response options: Yes, No]" (Nottingham Health Profile); "I often act irritable toward my work associates, for example, snap at them, give sharp answers, criticize easily [Response options: Yes, No]" (Sickness Impact Profile)

- *Functional status*: ability of the individual to perform tasks

E.g.: "Does your health limit you in these activities? If so, how much? … Lifting or carrying groceries [Response options: Yes, limited a lot; Yes limited a little; No, not limited at all]" (MOS Short Form SF-36); "I am unable to wash or dress myself [Response options: Yes, No]" (EuroQol EQ-5D); "I stop often when traveling because of health problems [Response options: Yes, No]" (Sickness Impact Profile)

- *Health perceptions* subjective integration of all the information related to symptom status and functional status

E.g.: "I seem to get ill more easily than other people [Response options: Definitely true, Mostly true, Not sure, Mostly false, definitely false]" (MOS Short Form SF-36); "How would you rate your overall health during the past week [Response options: 1 (Very poor), 2, 3, 4, 5, 6, 7 (Excellent)]" (EORTC QLQ C-30); "Have you recently felt that you are ill? [Response options: Not at all; No more than usual; Rather more than usual; Much more than usual" (General Health Questionnaire)

- *Health related quality of life*: aspects of quality of life[a] that relate specifically to a person's health

E.g.: "How much time during the last 2 weeks … have you been a happy person? [Response options: All of the time Most of the time Some of the time A little of the time None of the time]" (MOS Short Form SF-36); "How would you rate your overall quality of life during the past week [Response options: 1 (Very poor), 2, 3, 4, 5, 6, 7 (Excellent)]" (EORTC QLQ C-30); "Considering all the ways that your arthritis affects you, rate how you are doing? [Response options: visual analogue scale ranging from 0 (Very well), to 100 (Very poor)]" (Health Assessment Questionnaire)

[a] Quality of life has been defined as a construct equivalent to subjective well being, and comprising cognitive judgment, positive affect, and negative affect [36, 41]; more generally, it has been defined as the standard of living, or degree of happiness, comfort, etc., enjoyed by an individual [42]

in order to calculate the quality-adjusted life years (QALY) index [1]. Here, the proposal becomes an important aid to the use of PRO measure, since it has been clearly pointed out that some interpretation uses (e.g., cost-effectiveness analyses) are challenged when psychometric instruments are used [51]. Since other approaches might be also possible [47], a category for 'other metrics' was included in the proposal.

Dimensionality, on the other hand, refers to the number of scores produced for each individual. When the information of the instrument can be summarized in a single value we refer to an Index. Beyond obvious cases of instruments consisting of a single item (indicators), such as a single question concerning self-perception of general health [53] or a visual analogue pain scale, all unidimensional instruments produce index type scores. Many disease specific psychometric questionnaires, such as the Beck Depression Inventory, as well as most econometric measures are index-type measures. When more than one score is needed we refer to a profile. These categories are

**Table 2b** Specific categories for the construct "symptoms" and for the category "disease" in the construct "populations"

| ICD–10 Title | Codes |
|---|---|
| (a) Certain infections and parasitic diseases and related symptoms | A00-B99 |
| (b) Neoplasms and related symptoms | C00-D48 |
| (c) Diseases of and symptoms related to the blood and blood-forming organs and certain disorders involving the immune mechanism | D50-D89 |
| (d) Endocrine, nutritional and metabolic diseases and related symptoms | E00-E90 |
| (e) Mental and behavioural disorders and related symptoms | F00-F99 |
| (f) Diseases of and symptoms related to the nervous system | G00-G99 |
| (g) Diseases of and symptoms related to the eye and adnexa | H00-H59 |
| (h) Diseases of and symptoms related to the ear and mastoid process | H60-H95 |
| (i) Diseases of and symptoms related to the circulatory system | I00-I99 |
| (j) Diseases of and symptoms related to the respiratory system | J00-J99 |
| (k) Diseases of and symptoms related to the digestive system | K00-K93 |
| (l) Diseases of and symptoms related to the skin and subcutaneous tissue | L00-L99 |
| (m) Diseases of and symptoms related to the musculoskeletal system and connective tissue | M00-M99 |
| (n) Diseases of and symptoms related to the genitourinary system | N00-N99 |
| (o) Pregnancy, childbirth and puerperium and related symptoms | O00-O99 |
| (p) Certain conditions originating in the perinatal period and related symptoms | P00-P96 |
| (q) Congenital malformations, deformations and chromosomal abnormalities and related symptoms | Q00-Q99 |
| (r) Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified | R00-R99 |
| (s) Injury, poisoning and certain other consequences of external causes | S00-T98 |
| (t) External causes of morbidity and mortality | V01-Y98 |
| For the construct "population" only: (u) All diseases | A00-Y98 |

**Table 2c** Specific categories for the construct "functional status"

| International classification of functioning, disability and health chapter | Codes |
|---|---|
| (a) Learning and applying knowledge | d110-d199 |
| (b) General tasks and demands | d210-d299 |
| (c) Communication | d310-d399 |
| (d) Mobility | d410-d499 |
| (e) Self-care | d510-d599 |
| (f) Domestic life | d610-d699 |
| (g) Interpersonal interactions and relationships | d710-d799 |
| (h) Major life areas | d810-d899 |
| (i) Community, social and civic life | d910-d999 |

not exclusive: instruments can produce only index scores, others elicit profile scores only, and finally yet others can produce both of them.

At the beginning of the outcomes research movement, collections of different instruments, called batteries, were very popular [1]. They are not considered in the classification system because the focus is on individual instruments rather than their eventual combinations.

Adaptability is the third concept relevant to the measurement methods. By this, we refer to the extent to which the instrument can be tailored to the specific circumstances and preferences of each individual [51]. Most of the PRO questionnaires are completely standardized: they include explicitly formulated questions and predefined response options. There are obvious advantages inherent in such a high degree of standardization which explain the success of this type of instrument, most notably the simplification of procedures, the reliability of the estimates, and the comparability of the results obtained.

This approach, though, has been criticized for not taking the perspective of each individual patient into consideration, but rather that of the "average" patient or some other abstract subject [54, 55]. More flexible instruments have been developed, usually referred to as "patient-generated", "patient-centered", or "individualized" measures. In these instruments, domains and/or weights are not fixed. Each individual subject elicits them, indicating, for example, which activities or problems they would like to select for assessment. These measures might offer clear advantages over standardized instruments in clinical settings, where patient-centeredness is more an issue than standardization. Some instruments include a mix of both approaches and can be conceptualized as partially individualized.

Item banking and Computer Adaptive Testing (CAT) procedures allow reaching a high precision of measurement

with shorter formats rather than increasing the sensitivity to preferences of the individual, and they should be classified as standardized [56].

### Application of the classification system to frequently used instruments

In order to exemplify the use and applicability of the classification system we applied it to the ten most evaluated PRO measures [2]. We were able to classify all the instruments across all the axes and categories. All of them measured the constructs *Symptoms* and *Functional Status*. Most of them actually measured at least one additional construct, usually Health Perceptions, and two measured four different constructs. We present some examples of this analysis in Box 1.

Eight instruments were applicable to all adults, and two to adults with certain diseases only, but none of them was designed to measure reported outcomes in children. The majority were psychometric instruments (seven) and all of them were completely standardized (Table 3). For the purpose of exemplification, Table 3 also presents five additional instruments covering categories not applicable to the first 10.

## Discussion

### What is really new about our proposed PRO classification system

Firstly, previous classifications schemes considered either a simple list of examples of what constructs were implied without further elaboration [5, 20] or a list of the different features of the instruments without consideration to their underlying relationships [7]. Furthermore, previous schemes were not explicitly based on current conceptual models of health outcomes [22]. We have used an explicit methodology for the development of the classification system, and we have relied on a conceptual model that has proved valid and useful [37].

Secondly, this is a simplified classification system. While our proposal is based on attributes consistently used in the literature (the measurement object, the target population, and the measurement model) we do not consider it necessary to differentiate PRO measures by characteristics which do not fundamentally affect the nature of the instrument, such as different administration modes, weighting procedures, or use of full and reduced versions, among others [7].

Thirdly, it is important to note that ours is only a descriptive classification system and it does not provide any fundamental evaluation of the measurement properties of the instruments. We consider such evaluation crucial for adequate selection and interpretation of PROs. In fact, we have recently developed a standardized tool for the evaluation of such measures [55]. Classification and evaluation systems are complementary and should be used in tandem. Our approach results in at least three clear advantages over previous classifications [7, 13]: (a) less information is needed for the use of the system; (2) increased stability of the classification across different versions of the instruments [47, 48]; and (3) it can be applied from the very beginning in the development of the instrument.

Fourthly, previous attempts reduced differences between measures to their generic or specific nature, not considering any further systematic approach as to what is the intended population or to what disease or symptom it should be applied [7, 13]. Our proposal takes full advantage of the worldwide endorsement of the International Classification of Diseases ICD-10 of the WHO both for the construct of *Symptoms* as well as for instruments that are 'disease specific', and the International Classification of Functioning for the construct *Functional Status*. In light of our observation that most instruments (including those that focus on specific diseases) measure more than one construct, the traditional division in generic and specific measures seems very imprecise. Furthermore, all the criteria defining the populations are relevant from a clinical and a health services provision point of view [1].

### Limitations

Apart from metric properties, a number of characteristics not included here may be of interest for describing a PRO instrument, such as time for completion, availability of interpretation guidelines, or the degree of patient involvement in the generation of the items. Classification systems aim to reduce to a minimum the information needed to identify an object, and we have limited our classification system to only three fundamental characteristics. But additional information may be relevant when choosing a PRO instrument for use in clinical practice or research. Even the detail in which the fundamental characteristics are considered may seem insufficient. Potential users may also be interested in whether a given instrument considers a particular symptom (e.g., pain). Our emphasis on simplicity may have compromised the amount of information available for each instrument in the system.

For the axis "construct", our classification system relies on the health outcomes model proposed by Wilson and Cleary [29]. A number of different models have been proposed [37, 57], but they lack both the widespread use and the empirical evidence that supports the Wilson and Cleary model. However, this endorsement is contingent on available evidence. Should other models be tested and

**Table 3** Applying the classification system to the selected patient reported outcome measures

| PRO measure | A. Construct[a] | B. Population[b] | C. Measurement |
|---|---|---|---|
| MOS SF-36 | A.1. Symptoms | B.1.d. Adults | C.1.c. Profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.3. Health Perceptions | B.3.u. All diseases | C.3.a. Completely standardized |
| | A.4. Quality of Life | | |
| Sickness impact profile | A.1. Symptoms | B.1.d. Adults | C.1.a. Index |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | | B.3.u. All diseases | C.3.a. Completely standardized |
| Nottingham health profile | A.1. Symptoms | B.1.d. Adults | C.1.a. Profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.3. Health Perceptions | B.3.u. All diseases | C.3.a. Completely standardized |
| EORTC QLQ−C30 | A.1. Symptoms | B.1.d. Adults | C.1.b. Profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.3. Health Perceptions | B.3.c. Neoplasms: C00-D48 | C.3.a. Completely standardized |
| | A.4. Quality of Life | | |
| EuroQol | A.1. Symptoms | B.1.d. Adults | C.1.a. Index |
| | A.2. Functional Status | B.2.a. All genders | C.2.c. Econometric |
| | | B.3.u. All diseases | C.3.a. Completely standardized |
| Health assessment questionnaire | A.1. Symptoms | B.1.d. Adults | C.1.a. Profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.4. Quality of Life | B.3.u. All diseases | C.3.a. Completely standardized |
| Arthritis impact measurement scales | A.1. Symptoms | B.1.d. Adults | C.1.b. Profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.3. Health Perceptions | B.3.n. Diseases of the musculoskeletal | C.3.a. Completely standardized |
| | A.4. Quality of Life | System and the connective tissue (arthropathies)[c] | |
| Quality of wellbeing scale | A.1. Symptoms | B.1.d. Adults | C.1.a. Index |
| | A.2. Functional Status | B.2.a. All genders | C.2.b. Econometric |
| | | B.3.u. All diseases | C.3.a. Completely standardized |
| General health questionnaire | A.1. Symptoms | B.1.d. Adults | C.1.a. Index |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | A.3. Health Perceptions | B.3.u. All diseases | C.3.a. Completely standardized |
| | A.4. Quality of Life | | |
| Health utilities index | A.1. Symptoms | B.1.d. Adults | C.1.a. Index |
| | A.2. Functional Status | B.2.a. All genders | C.2.b. Econometric |
| | A.4. Quality of Life | B.3.u. All diseases | C.3.a. Completely standardized |
| McGill pain questionnaire[c] | A.1. Symptoms | B.1.d. Adults | C.1.b. Profile |
| | | B.2.a. All genders | C.2.a. Psychometric |
| | | B.3.u. All diseases | C.3.a. Completely standardized |

**Table 3** continued

| PRO measure | A. Construct[a] | B. Population[b] | C. Measurement |
|---|---|---|---|
| Kidscreen[c] | A.2. Functional Status | B.1.b. Children & B.1.c. Adolescents | C.1.b. Profile |
| | A.3. Health Perceptions | B.2.a. All genders | C.2.a. Psychometric |
| | A.4. Quality of Life | B.3.u. All diseases | C.3.a. Completely standardized |
| Asthma quality of life questionnaire[c] | A.1. Symptoms | B.1.d. Adults | C.1.c. Index and profile |
| | A.2. Functional Status | B.2.a. All genders | C.2.a. Psychometric |
| | | B.3.k. Diseases of the respiratory system (asthma) | C.3.b. Partially individualized |
| Schedule for the evaluation of individual quality of life (SEIQOL)[c] | A.2. Functional Status | B.1.d. Adults | C.1.a. Index |
| | A.3. Health Perceptions | B.2.a. All genders | C.2.a. Psychometric |
| | A.4. Quality of Life | B.3.u. All diseases | C.3.c. Completely individualized |
| General Practice Assessment Questionnaire (GPAQ)[c] | A.5. Other health related constructs: satisfaction with care | B.1.d. Adults | C.1.c. Profile |
| | | B.2.a. All genders | C.2.a. Psychometric |
| | | B.3.u. All diseases | C.3.a. Completely standardized |

[a] Second order categories for the constructs "Symptoms" and "Functional Status" are not included

[b] Culture dyads and second order categories for "Disease" are not included

[c] Additional instruments

proved valid, this may result in a need for considering different constructs.

The inclusion of constructs other than those included in the Wilson and Cleary model has been justified on theoretical grounds, but the available evidence supporting the model may not necessarily apply to them. Research using available techniques as structural equation modeling will be needed to confirm their inclusion [31–35]. In particular, we used a slightly modified version of the model by Wilson and Cleary, substituting the original "overall quality of life" with the more specific "health related quality of life". We did so as supported by empirical data, but one of the revised instruments was found to measure the first instead of the latter (see Box 1).

Finally, the work presented here represents only an initial step towards the development and adoption of a common classification system of PRO measures. The application of this system to a database including about 400 PRO instruments will be our next step [58]. This process will allow us to test the classifications system with a greater variety of instruments and will provide invaluable information about the generalizability of the method.

## How to use the classification system

A common framework for classifying PRO measures will serve different purposes for a broad range of health professionals. Researchers, clinicians, administrators and policy makers are all confronted with decisions based on these measures in their everyday work. All of them have now been provided with a solid guide to a better understanding of the nature of the instruments and to the interpretation of the related literature. Moreover, the system provides the basic information for identifying the candidate pool of PRO instruments available for use. Clinicians will be hereby assisted also in the selection of PRO instruments in their clinical practice [59, 60]. Administrators and policymakers are currently encouraged to integrate outcomes with existing process measures in order to get the most comprehensive view of the performance of health services [61]. Researchers, finally, will find this proposal a useful tool for the identification of areas lacking instruments and other areas where there might be a surplus of instruments [47].

Once the potential user is aware of the basic information describing what the instrument is designed to measure, a necessary next step is to compare, among the candidate instruments, the evidence that supports the robustness and adequacy of each candidate instrument. To evaluate these characteristics, a number of existing guidelines including attributes and criteria of adequacy exist [7, 35]. The use in tandem of our proposed classification and the standard evaluation guidelines should facilitate the adequate use of PRO instruments in clinical research, management, and practice.

An additional intended consequence is to foster discussion for a common definition of the constructs [3]. As a matter of fact, the application of our proposal to well-known instruments has revealed that they are much more heterogenic in their nature than had been claimed by their developers. The time has come to make the efforts

conducive to the construction and adoption of a common terminology [22].

**Acknowledgement** The authors would like to thank Dr Montse Ferrer, MD PhD MPH (Health Services Research Unit, IMIM-Hospital del Mar, Spain) and Dr Martin Roland, DM (Director of the National Primary Care Research and Development Centre, Manchester University, UK) for their thoughtful comments on a previous draft of the manuscript. We would also wish to acknowledge the contribution of two anonymous reviewers to the completeness and clarity we may have achieved in this article.

## References

1. Patrick, D. L., & Erickson, P. (1993). *Health status and health policy*. New York: Oxford University Press.
2. Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: Bibliographic study of patient assessed health outcome measures. *BMJ (Clinical Research Ed.), 324*(7351), 1417. doi:10.1136/bmj.324.7351.1417.
3. Tully, M. P., & Cantrill, J. A. (1999). Subjective outcome measurement—A primer. *Pharmacy World & Science, 21*(3), 101–109. doi:10.1023/A:1008694522700.
4. Lenderking, W. (2003). Task force report of the patient-reported outcomes harmonization group: Too much harmony, not enough melody? *Value in Health, 6*(5), 503–504. doi:10.1046/j.1524-4733.2003.65002.x.
5. Acquadro, C., Berzon, R., Dubois, D., et al. & PRO Harmonization Group. (2003). Incorporating the patient's perspective into drug development and communication: An ad hoc task force report of the patient-reported outcomes (PRO) harmonization group meeting at the Food and Drug Administration, February 16, 2001. *Value in Health, 6*(5), 522–531. doi:10.1046/j.1524-4733.2003.65309.x.
6. Taillefer, M. C., Dupuis, G., Roberge, M. A., & Le May, S. (2003). Health-related quality of life models: Systematic review of the literature. *Social Indicators Research, 64*, 293–323. doi:10.1023/A:1024740307643.
7. FDA. (2006). Draft guidance for industry on patient-reported outcome measures: Use in medicinal product development to support labeling claims. *Federal Register, 71*, 5862–5863.
8. Levine, D. M., Morlock, L. L., Mushlin, A. I., Shapiro, S., & Malitz, F. E. (1976). The role of new health practitioners in a prepaid group practice: Provider differences in process and outcomes of medical care. *Medical Care, 14*(4), 326–347. doi:10.1097/00005650-197604000-00004.
9. Schmidt, S., & Bullinger, M. (2003). Current issues in cross-cultural quality of life instrument development. *Archives of Physical Medicine and Rehabilitation, 84*(4, Suppl 2), S29–S34. doi:10.1053/apmr.2003.50244.
10. Emery, M. P., Perrier, L. L., & Acquadro, C. (2005). Patient-reported outcome and quality of life instruments database (PROQOLID): Frequently asked questions. *Health and Quality of Life Outcomes, 3*(1), 12. doi:10.1186/1477-7525-3-12.
11. McDowell, I., & Newell, C. (1996). *Measuring health. A guide to rating scales and questionnaires*. New York: Oxford University Press.
12. Osoba, D. (2002). A taxonomy of the uses of health-related quality-of-life instruments in cancer care and the clinical meaningfulness of the results. *Medical Care, 40*(6, Suppl), III31–III38. doi:10.1097/00005650-200206001-00006.
13. Kirshner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases, 38*(1), 27–36. doi:10.1016/0021-9681(85)90005-0.
14. Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care, 27*(3, Suppl), S217–S232. doi:10.1097/00005650-198903001-00018.
15. Guyatt, G. H., Veldhuyzen Van Zanten, S. J., Feeny, D. H., & Patrick, D. L. (1989). Measuring quality of life in clinical trials: A taxonomy and review. *Canadian Medical Association Journal, 140*(12), 1441–1448.
16. Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine, 118*(8), 622–629.
17. MacKeigan, L. D., & Pathak, D. S. (1992). Overview of health-related quality-of-life measures. *American Journal of Hospital Pharmacy, 49*(9), 2236–2245.
18. Patrick, D. L., & Chiang, Y. P. (2002). Measurement of health outcomes in treatment effectiveness evaluations. *Medical Care, 38*(9, Suppl II), II-14–II-25.
19. Bowling, A. (1997). *Measuring health: A review of quality of life measurement scales* (2nd ed.). Buckingham: Open University Press.
20. The Patient-Reported Outcomes Measurement Information System (PROMIS). (2007). Progress of an NIH roadmap cooperative group during its first two years. *Medical Care, 45*(5, Suppl 1), S3–S11.
21. Revicki, D. (2007). FDA draft guidance and health-outcomes research. *Lancet, 369*, 540–542. doi:10.1016/S0140-6736(07)60250-5.
22. Valderas, J. M., & Alonso, J. (2007). Linking measurement to rooted theory models in the PROMIS project. *Medical Care, 45*(10), 1008.
23. Espallargues, M., Valderas, J. M., & Alonso, J. (2000). Provision of feedback on perceived health status to health care professionals: A systematic review of its impact. *Medical Care, 38*(2), 175–186. doi:10.1097/00005650-200002000-00007.
24. Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ (Clinical Research Ed.), 331*(7524), 1064–1065. doi:10.1136/bmj.38636.593461.68.
25. [25] Patrick DL, Chiang YP. Measurement of Health Outcomes in TreatmentEffectiveness Evaluations. Medical Care 2002:38 (9, Suppl. II):II14–II25
26. Cross-classification.WordNet. Princeton University. (2006). Available at http://wordnet.princeton.edu/perl/webwn?s=cross-classification. Accessed 07 May 2008.
27. Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193–205. doi:10.1023/A:1015291021312.
28. Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). New York: Oxford University Press.
29. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. *Journal of the American Medical Association, 273*, 59–65. doi:10.1001/jama.273.1.59.
30. Engel, G. E. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136. doi:10.1126/science.847460.
31. Sullivan, M. D., Kempen, G. I., Van Sonderen, E., & Ormel, J. (2000). Models of health-related quality of life in a population of community-dwelling Dutch elderly. *Quality of Life Research, 9*(7), 801–810. doi:10.1023/A:1008987709788.
32. Orfila, F., Ferrer, M., Lamarca, R., Tebe, C., Domingo-Salvany, A., & Alonso, J. (2006). Gender differences in health-related quality of life among the elderly: The role of objective functional capacity and chronic conditions. *Social Science & Medicine, 63*(9), 2367–2380. doi:10.1016/j.socscimed.2006.06.017.

33. Sousa, K. H., & Kwok, O. I. (2006). Putting Wilson and Cleary to the test: Analysis of a HRQOL conceptual model using structural equation modeling. *Quality of Life Research, 15*, 725–737. doi:10.1007/s11136-005-3975-4.
34. Hays, R. D., Revicki, D., & Coyne, K. S. (2005). Application of structural equation modeling to health outcomes research. *Evaluation & the Health Professions, 28*(3), 295–309. doi:10.1177/0163278705278277.
35. Hofer, S., Benzer, W., Alber, H., et al. (2005). Determinants of health-related quality of life in coronary artery disease patients: A prospective study generating a structural equation model. *Psychosomatics, 46*(3), 212–223. doi:10.1176/appi.psy.46.3.212.
36. Üstün, T. B., Chatterji, S., Bickenbach, J., Kostanjsek, N., & Schneider, M. (2003). The international classification of functioning, disability and health: A new tool for understanding disability and health. *Disability and Rehabilitation, 25*(11–12), 565–571. doi:10.1080/0963828031000137063.
37. Post, M. W. M., de Witte, L. P., & Schrijvers, A. J. P. (1999). Quality of life and the ICIDH: Towards an integrated conceptual model for rehabilitation outcomes research. *Clinical Rehabilitation, 13*, 5. doi:10.1191/026921599701532072.
38. McKenna, S. P., & Doward, L. C. (2004). Integrating patient-reported outcomes. *Value in Health, 7*(Suppl 1), S9–S12. doi:10.1111/j.1524-4733.2004.7s103.x.
39. Geyh, S., Cieza, A., Kollerits, B., Grimby, G., & Stucki, G. (2007). Content comparison of health-related quality of life measures used in stroke based on the international classification of functioning, disability and health (ICF): A systematic review. *Quality of Life Research, 16*(5), 833–851. doi:10.1007/s11136-007-9174-8.
40. Cieza, A., Geyh, S., Chatterji, S., Kostanjsek, N., Ustun, B., & Stucki, G. (2005). ICF linking rules: An update based on lessons learned. *Journal of Rehabilitation Medicine, 37*(4), 212–218. doi:10.1080/16501970510040263.
41. Tennant, A. (1995). Quality of life—A measure too far? *Annals of the Rheumatic Diseases, 54*(6), 439–440. doi:10.1136/ard.54.6.439.
42. Oxford English Dictionary. (2008). Oxford: Oxford University Press. Available online at http://www.oed.com/. Accessed 7 May 2008.
43. Starfield, B. (1973). Health services research: A working model. *The New England Journal of Medicine, 289*(3), 132–136.
44. Shikiar, R., & Rentz, A. M. (2004). Satisfaction with medication. An overview of conceptual, methodologic and regulatory issues. *Value in Health, 7*, 204–215. doi:10.1111/j.1524-4733.2004.72252.x.
45. International Statistical Classification of Diseases and Related Health Problems. 10th Revision Version for 2006. Accesible at http://www.who.int/classifications/apps/icd/icd10online/. Accessed 15 May 2008.
46. International Classification of Functioning.Disability and Health. Accesible at http://www.who.int/classifications/icf/site/onlinebrowser/icf.cfm. Accessed 15 May 2008.
47. Streiner, D., & Normand, C. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). Oxford: Oxford University Press.
48. Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw Hill.
49. Wright, J. G., & Feinstein, A. R. (1992). A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *Journal of Clinical Epidemiology, 45*, 1201–1218. doi:10.1016/0895-4356(92)90161-F.
50. Ribera, A., Permanyer-Miralda, G., Alonso, J., Cascant, P., Soriano, N., & Brotons, C. (2006). Is psychometric scoring of the McNew quality of life after myocardial infarction questionnaire superior to the clinimetric scoring? A comparison of the two approaches. *Quality of Life Research, 15*(3), 357–365. doi:10.1007/s11136-005-2291-3.
51. Kaplan, R. M. (1998). Profile versus utility based measures of outcome for clinical trials. In M. J. Staquet, R. D. Hays & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials. Methods and practice*. Oxford: Oxford University Press.
52. Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health, 59*, 342–345. doi:10.1136/jech.2004.021204.
53. Carr, J. A., & Higginson, I. J. (2001). Are quality of life measures patient centered? *BMJ (Clinical Research Ed.), 322*, 1357–1360. doi:10.1136/bmj.322.7298.1357.
54. Patel, K. K., Veenstra, D. L., & Patrick, D. L. (2003). A review of selected patient-generated outcome measures and their application in clinical trials. *Value in Health, 6*(5), 595–603. doi:10.1046/j.1524-4733.2003.65236.x.
55. Gill, T. M., & Feinstein, A. R. (1994). A critical appraisal of the quality of quality of life measurements. *Journal of the American Medical Association, 272*, 619–626. doi:10.1001/jama.272.8.619.
56. Kosinski, M., Bjorner, J. B., Ware, J. E., Jr., Sullivan, E., & Straus, W. L. (2006). An evaluation of a patient-reported outcomes found computerized adaptive testing was efficient in assessing osteoarthritis impact. *Journal of Clinical Epidemiology, 59*(7), 715–723. doi:10.1016/j.jclinepi.2005.07.019.
57. Jette, A.M., & Badley, E. (2000). Conceptual issues in the measurement of work disability. In: N. Mathiowetz & G. S. Wunderlich (Eds), *Survey measurement of work disability: Summary of a workshop* (pp. 4–27). Washington DC: National Academy Press. Available online at http://books.nap.edu/html/work_disability/ch2.html. Accessed 19 July 2008.
58. Valderas, J. M., Ferrer, M., Mendívil, J., Garin, O., Rajmil, L., Herdman, M., et al. (2008). The scientific committee on "patient-reported outcomes" of the IRYSS network. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value in Health, 11*(4), 700–708. doi:10.1111/j.1524-4733.2007.00309.x.
59. Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., et al. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Quality of Life Research, 17*(2), 179–193. doi:10.1007/s11136-007-9295-0.
60. Valderas, J. M., Alonso, J., & Guyatt, G. H. (2008). Measuring patient-reported outcomes: Moving from clinical trials into clinical practice. *The Medical Journal of Australia, 189*(2), 93–94.
61. Krumholz, H. M., Normand, S. T., Spertus, J. A., Shahian, D. M., & Bradley, E. H. (2007). Measuring performance for treating heart attacks and heart failure: The case for outcomes measurement. *Health Affairs, 26*(1), 75–85. doi:10.1377/hlthaff.26.1.75.
62. Patel, K. K., Veenstra, D. L., & Patrick, D. L. (2003). A review of selected patient-generated outcome measures and their application in clinical trials. *Value Health*, *6*(5), 595–603.