

## 1.1 Theoretical basis of instruments

Section IV.A of the document establishes the importance of developing a conceptual measurement framework for identifying and justifying the PRO measurement constructs used as clinical trial endpoints. Section IV.B.1 emphasises the importance of referring to that conceptual framework in identifying appropriate methods / sources for item generation. However, only passing reference is made to consideration of the theoretical approach used (line 311).

PRO's that assess a non-observable variable (latent trait) such as depression, health-related quality of life (HRQL) or quality of life (QoL) must be based on a clear theoretical model of the construct assessed. It is inappropriate for researchers to justify the use of certain domains in, for example, a HRQL measure simply because these are commonly used domains. There must be a clear theoretical model underpinning their selection. For example, a PRO assessing functioning might use as its theoretical basis the World Health Organisation's classification of disability/activities<sup>1</sup> or a measure of QoL might be based on the needs-based model of QoL.<sup>2</sup> Unless researchers state explicitly their theoretical model it is not possible for the FDA to determine whether the definition is reasonable and, hence, whether the questionnaire is measuring what it is claimed to be measuring. Furthermore, without such information it is not possible to establish that an instrument has construct validity.<sup>3</sup>

It is recommended that the guidance clearly emphasises the need for researchers to state the theoretical framework used as the basis of all PRO questionnaires used as clinical trial endpoints; whether these already exist or are newly developed.

## 1.2 Fundamental measurement properties

The quality of most existing PRO scales has been assessed using classical psychometric properties (reliability, validity and responsiveness). Assessment of fundamental measurement properties is a relatively new area for PROs. However, item response theory (predominantly Rasch analysis<sup>4</sup>) is now commonly used to address some of the fundamental measurement attributes of instruments; most notably, level of measurement and unidimensionality.

Assessing the level of measurement of the instrument (ordinal, interval, ratio) is crucial to determining which statistical tests can be used on data derived from instrument use. Most existing PRO instruments are at the ordinal level suggesting that parametric statistics should not be used to assess data.

Any scale from which a summary (total) score is produced must be unidimensional if it is to provide valid change scores. This applies to both single-scale instruments and to individual scales within multidimensional instruments. Internal consistency coefficients (Cronbach's alpha) are not sufficient to indicate unidimensionality as it has been shown that this statistic merely indicates the degree of interrelation between the items. Different scales can be added together and still have relatively high internal consistency.<sup>5</sup> Factor

analytic methods have also been used to assess the dimensional structure of scales. However, these are parametric methods requiring interval level data and so their use may not always be scientifically valid. In addition, it has been known for some time that factor analysis, particularly with dichotomous data, produces spurious factors. Rasch analysis is now considered to be the most efficient means of establishing unidimensionality.<sup>6</sup>

It is recommended that the FDA include the need to demonstrate level of measurement and evidence of unidimensionality for all PRO instruments used as clinical trial endpoints. Where evidence of unidimensionality is lacking, scale summary scores must be treated with extreme caution. Total scores for profile measures should be treated with extreme caution without evidence that the total score is unidimensional – which is unlikely to be the case.

## 2. Additional Comments

**Line 31-32.** The definition of the term “patient-reported outcomes” provided by the FDA currently excludes other potential client groups who could not be described as “patients”. We agree that the term PRO should incorporate only those measurements where information is provided directly by the subject, rather than by expert raters or by other proxy assessments. However, there may be occasions where the subject of direct interest is not the patient but another party such as a caregiver (for example, where the patient is a child or is mentally impaired) or wider family group. The FDA should clarify whether they would consider data arising from the impact on, for example, caregiver fatigue. If such data are considered acceptable it may be appropriate to broaden the definition of “patient” reported outcome and potentially modify the terminology to “self-reported outcome”.

**Line 33-34.** The FDA asserts that PROs “can be used to measure the impact of an intervention on one or more aspects of patients’ health status”. The term “health status” should not be used in this context as it represents a well-defined concept in its own right that is often used interchangeably with the term “health-related quality of life”.<sup>7,8,9</sup>

**Line 134-137.** In many instances it is inappropriate to suggest that the most appropriate means of assessing validity is to compare patient-reported results to those of expert assessors. The patient’s view of the impact of their condition, particularly in terms of non-observable variables, is likely to differ from that of clinical or other experts.

**Lines 495-497.** It would be useful to comment on what the FDA would consider to be an appropriate (or inappropriate) time frame for assessment of test-retest. For example, a time-interval of 24 hours would be considered inappropriate by most researchers as recall factors are likely to influence instrument completion. A time interval of 2 weeks has been recommended and is most commonly adopted.<sup>6</sup>

**Lines 499-516.** The guidance should emphasise that assessment of construct validity is only feasible where the PRO is based on a clear underlying model of the construct assessed.

### Summary

We welcome the FDA's draft document in that it suggests raising the standards required of PRO measurement. Current instruments are largely incapable of showing significant differences between alternative treatments.

We do have a concern that the measurement of symptoms and functioning is over-emphasised at the cost of issues that may be of more direct concern to patients such as psychological distress and ability to meet fundamental needs. It would be unfortunate if PROs simply duplicate clinical data collected in trials. It is acknowledged that assessing less objective outcomes presents a greater challenge but this does not mean that they should be ignored. The document does emphasise the perspective of the patient rather than that of the clinician.

As this is a guidance document it must be sensitive to the notion that the science is not static but evolving. Document wording should allow for such developments in the science.

## References

1. World Health Organisation. The International Classification of Impairments, Disabilities and Handicaps. Geneva: WHO, 1980.
2. World Health Organisation. International Classification of Functioning and Disability: ICFIDH-2. Stationery Office Books. November 1999
3. Streiner D, Norman G. Health Measurement Scales. Oxford: OUP, 1989.
4. Rasch G. Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980.
5. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 1993;78:98-104.
6. Doward LC, Meads DM, Thorsen H. Requirements for Quality of Life instruments in clinical research. *Value in Health*. 2004, 7 Issue s1: S13-S16.
7. Doward LC, McKenna SP. Defining Patient-Reported Outcomes. *Value in Health* 2004, 7 Issue s1: S4-S8.
8. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Annals of Internal Medicine*. 1993; 118: 622-629.
9. Patrick DL, Bergner M. Measurement of Health Status in the 1990's. *Annual Review of Public Health*. 1990; 11:165-83.