

Chapter 9

Developing Patient-Reported and Relevant Outcome Measures

Kirstie L Haywood, Maarten de Wit, Sophie Staniszewska, Thomas Morel, and Sam Salek

9.1 PROM Development

This chapter will examine good practice guidance for patient-centred approaches towards PROM development. During the last decade, we have witnessed a paradigm shift in how outcomes are measured from a more clinical, physician-oriented perspective to a more patient-focused perspective, which has led to the emergence of the notion of patient-reported outcome (PRO). The concept of PRO seeks to understand how patients feel, function and live their lives in relation to health challenges and associated healthcare and is more encompassing than earlier terms, such as patient global assessment, health status, quality of life or symptom checklists. In this chapter, we argue that well-developed questionnaires, or PRO measures (PROMs), which reflect patients' perspectives, have the potential to provide valuable patient-based evidence in HTA. PROM development should engage with patients as participants (US Food and Drug 2009) and increasingly as research partners (Staniszewska et al. 2012; de Wit et al. 2013; Chap. 8) through all stages of development. This promotes patients as the determinants of the key constructs

K.L. Haywood (✉) • S. Staniszewska

Royal College of Nursing Research Institute, Department of Health Sciences, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry, CV4 7AL, UK
e-mail: k.l.haywood@warwick.ac.uk

M. de Wit

Department of Medical Humanities, VU University Medical Centre,
Amsterdam, The Netherlands

T. Morel

UCB Biopharma, Brussels, Belgium

S. Salek

School of Life and Medical Sciences, School of Pharmacy, Pharmacology and Postgraduate Medicine, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK

underpinning the PROM. This approach will support a transparent and auditable approach towards capturing patients' contributions to the measurement of relevant outcomes, thereby enhancing the face and content validity, relevance and acceptability of measures. In this chapter, we describe eight key stages in PROM development (Fig. 9.1) and reflect on how patients can participate in this process.

9.2 Key Stages in Developing a PROM

9.2.1 *Establishing the Need for a New Measure*

Developing a new PROM is a costly and time-consuming activity. Initially in relation to a HTA, efforts should be made to select a measure already available for the intended purpose, embarking upon development of a new measure only when there is an unmet need.

Systematic reviews of PROMs' availability, quality and acceptability are essential in supporting any decision to develop a PROM (Haywood et al. 2014a). If PROMs are available, one needs to establish if they are 'good' enough for the intended purpose, taking into consideration evidence of their development, relevance and acceptability as outlined above, alongside evidence of quality (Haywood et al. 2012; Terwee et al. 2007; Streiner et al. 2014) and consideration of their appropriateness for the proposed application.

9.2.2 *Identifying Key Collaborators*

From the beginning a new PROM should be developed with both the end users and intended application in mind. Key considerations comprise by whom, when and how the measure will be completed and who will receive the scores or analyses (such as HTA bodies). A team of experts is required throughout the development process including patient representatives, clinicians, clinical academics and measurement experts. However, if the new PROM is intended for use also by device manufacturers, health service or health technology developers and HTA bodies as the end users who will receive the scores for strategic and reimbursement decision-making, their representative should join the team of experts as additional stakeholders.

9.2.2.1 Core Research Team and Advisory Group

A small core research team, responsible for conducting the day-to-day research activities, should seek to include measurement experts, clinical academics, clinicians and patient research partners. A larger advisory group will include representatives

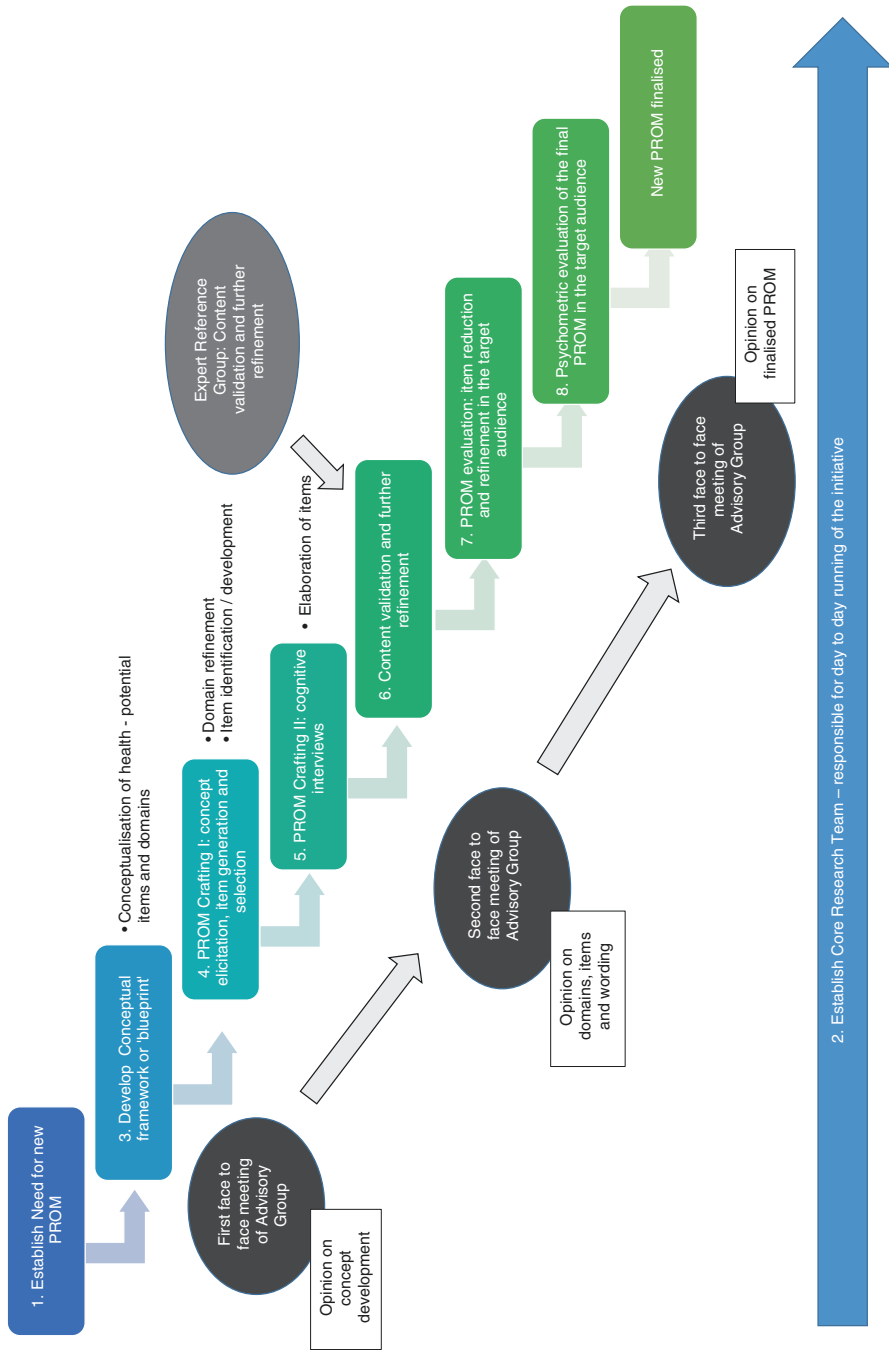


Fig. 9.1 Key stages in PROM development

from these same groups, with the addition of patient representatives, scientific organisation representatives, sponsors of the research and relevant health technology developer participants. In contrast to the core research team, the advisory group provides a more strategic oversight to the development of the PROM, commenting and contributing on each stage of the PROM development process.

9.2.2.2 Expert Reference Groups

Two external reference groups may also be established. These include (1) an expert patient reference group and (2) a professional expert group, both of whom will be utilising the measures and the information arising in their decision-making. These panels will be called upon at key stages in PROM development to comment on content, structure, format and appropriateness. An example of where these panels can play a critical role is in helping to find a resolution for the tension that may occur between the findings of the qualitative research (*Stage 9.2.4*) and the demands of the psychometric evaluation (*Stage 9.2.8*) (Gossec et al. 2014).

9.2.3 Developing a Conceptual Framework

Defining what a PROM is intended to measure is a crucial but often overlooked and poorly reported step in PROM development. Guidance has highlighted the importance of providing a clear conceptual framework of ‘what’ the PROM is intended to measure (US Food and Drug 2009; Patrick et al. 2011a).

A first step is to understand the medical, or disease, model of an illness (US Food and Drug 2009; Patrick et al. 2011a; Victorson et al. 2014), for example, the biology of the disease, associated symptoms and extent of impairment. This should underpin an appreciation of any potential patient-reported symptom and associated illness impact and hence the variables that may contribute to a developing biopsychosocial model of illness.

The conceptual framework describes the overriding concept of health underpinned by ‘hypothesized relationships among items, domains and concepts measured’ (US Food and Drug 2009, p. 9). That is, the specific questions (items) or groups of questions (domains) that should be considered for inclusion within a PROM to reflect the aspects of health (concepts) to be assessed. In effect, the conceptual framework is an ‘organising tool that summarises what has been found in the literature and discussions with experts’ (Patrick et al. 2011a, p. 971). It informs the developing topic guide for the qualitative research. Furthermore, it evolves as a consequence of findings from the qualitative research providing a ‘blueprint’ of the outcomes that really matter to patients with the target illness and hence the outcomes that should be considered for inclusion in the developing PROM (Parslow et al. 2015; Gorecki et al. 2010) (Fig. 9.2).

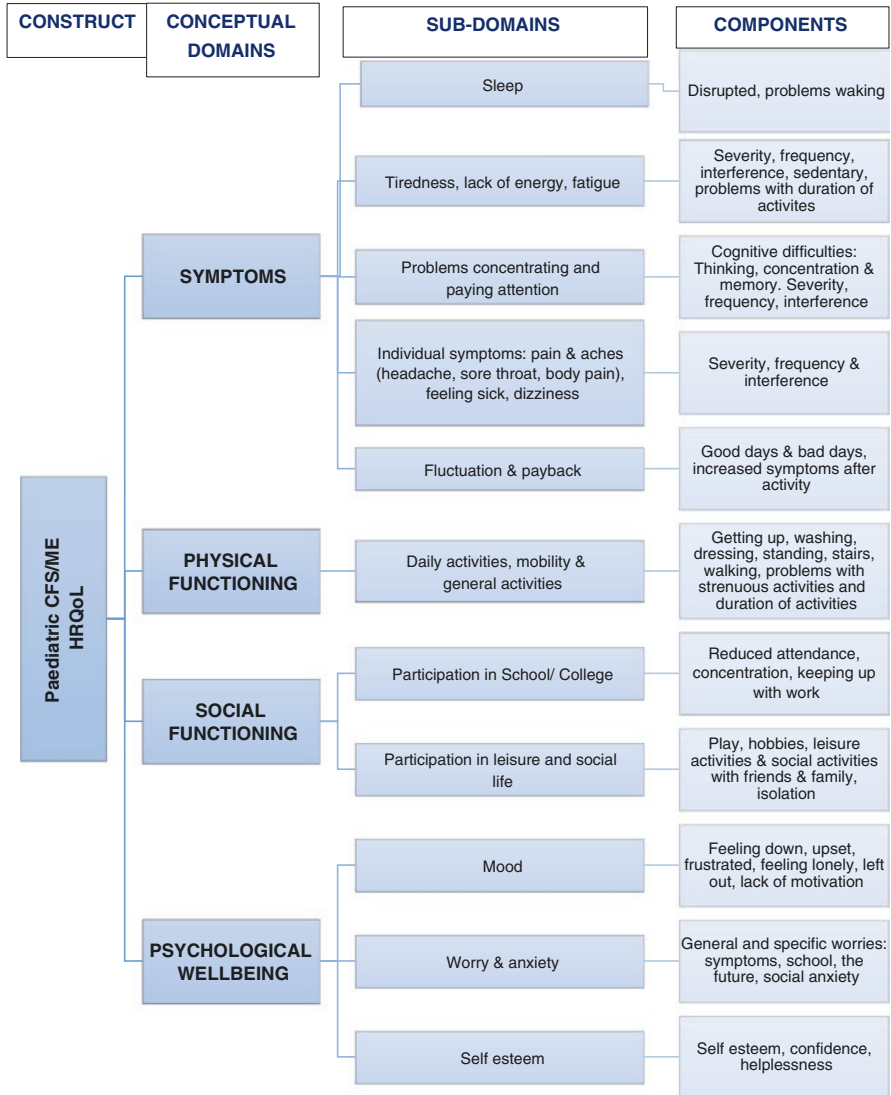


Fig. 9.2 A conceptual framework to underpin a new PROM for assessing the health-related quality of life (HRQoL) of children with chronic fatigue syndrome/myalgic encephalitis (CFS/ME) (Parslow 2016)

9.2.4 *Crafting the PROM-I: Concept Elicitation, Item Generation and Selection*

Current guidance on PROM development stipulates the importance of transparency in the data generation and analytical processes—creating a clear audit trail from concept elicitation to final items (US Food and Drug 2009; Patrick et al. 2011a).

9.2.4.1 Existing Measures

The content and focus of existing PROMs may contribute to both the developing conceptual framework and the list of potential items. For most instances of new PROM development, existing scales within the same disease or with a similar focus are available and should be reviewed.

Organisations such as the Patient-Reported Outcomes Measurement Information System (PROMIS) have established topic-specific ‘item banks’ (Health Measures 2016)—large numbers of items, or questions, derived from established measures and qualitative research with patients, but whose association has been determined by item response theory and hence form the basis of computer adaptive testing (CAT) approaches to PROM administration (Reeve et al. 2007). Such item banks can play a useful contribution to the development of new measures. For example, development of the Headache Impact Test (HIT) group of measures was informed by an item bank founded on several established migraine/headache measures and clinical judgement (Bjorner et al. 2003). Initial testing revised and reformatted the items and response formats to produce the CAT-HIT which has access to 54 items within the HIT item bank; a short-form, standardised version includes just six items—the HIT-6 (Kosinski et al. 2003).

9.2.4.2 Existing Literature

Systematic reviews and meta-syntheses of the qualitative literature can further assist in understanding the lived experience of patients and identifying relevant outcomes, contributing to the developing conceptual framework and item pool (e.g. Parslow et al. 2016).

9.2.4.3 Experts: Defining the Sample

The extent to which participants are representative of the target population and condition—considering variations in gender, age, disease severity and presentation—is essential to concept elicitation and item generation, ensuring content relevance and validity. For example, development of the EASi-QoL for Ankylosing Spondylitis (AS) included qualitative data generated from in-depth interviews with 29 patients and a UK survey of 462 patients (Haywood et al. 2010). Respondents identified the most important areas of their life affected by AS, ensuring that priorities and values representative of the wide spectrum of AS presentation and a broad socio-demographic mix contributed to concept elicitation and item generation.

Driven by changing global regulatory systems and HTA, it is increasingly recognised that for PROM data to have greater universality and relevance to a wide range of cultures, patients from different cultures and settings should be involved in item generation and selection. The result of such participation seeks to avoid culture-specific words or phrases and concepts that would be difficult to reproduce

cross-culturally. Models of PROM development which build in universality and translatability from the start are increasingly observed. For example, development of the PsAID questionnaire included 12 patient research partners from 12 European countries who were active through all stages of PROM development; all were fluent in English and had personal experience of psoriatic arthritis (Gossec et al. 2014). Moreover, the domain selection and external validation of the developing measure were further supported by an international cross-sectional study of 140 patients from ten countries who were invited to rank the domains in order of importance. Whilst it may not always be possible to achieve such integration, developers should be cognisant of the importance of these issues.

Increasingly social media is utilised to contribute to item generation and further development of PROMs. For example, an online forum of members of the hyperhidrosis patient organisation contributed to the generation of items for the Hyperhidrosis Quality of Life questionnaire (HidroQOL) (Kamudoni et al. 2015). Added benefits of this approach include the large number of international contributors, adding to the universality of the approach.

9.2.4.4 Qualitative Research

Rigorous qualitative research which seeks to better understand patients' perspectives and experiences is essential for concept elicitation and item generation so that PROMs are comprehensive and relevant to the target population (Brédart et al. 2014). A range of qualitative methods including semi-structured interviews, focus group discussions and modified Delphi surveys (Haywood et al. 2010; Gossec et al. 2014, Bartlett et al. 2012) can be used. However, this information is often poorly reported by developers (Patrick et al. 2011a). Recent guidance has highlighted the importance of transparency in both the qualitative approach and methods of data collection (US Food and Drug 2009; Patrick et al. 2011a). Where, historically, such qualitative exploration and analysis have been undertaken by academics or clinical researchers, patients are increasingly involved in this process as patient research partners (Gossec et al. 2014; Chap. 8).

9.2.4.5 Analysis of Qualitative Data: Quality Assurance in PROM Development

Data analysis seeks to refine the large amount of qualitative data into a long list of items that reflects the evolving conceptual framework in a manner that is transparent and meaningful and which ultimately supports the allocation of scores to enable quantification of the target construct. The data analysis should be both inductive—discovering new patterns and themes—and deductive, that is, regarding the evolving conceptual framework (Patrick et al. 2011a).

The analysis consists of several steps. First, the accuracy of the transcribed audio recordings should be checked to ensure preservation of the integrity of the generated

data (Patrick et al. 2011a; Golics et al. 2014). Data analysis seeks to use words and phrases generated by participants to craft the evolving concepts, themes and sub-themes of the conceptual framework. Several trained researchers, or coders, should be involved in this process—working independently in the first instance, before discussing the developing themes to identify areas of consistency, inconsistency and concept saturation, a process which is repeated throughout data analysis. The transparent illustration of developing themes and codes, for example, on a thematic map, may assist with communicating data pattern conceptualisation. The thematic prevalence of a concept, that is, the number of patients expressing a concept, can also assist with item selection. For example, potential items were selected for the Family Reported Outcome Measure (FROM-16), a population-derived measure of the impact of illness on the partner or family members of patients, if mentioned by more than 5% of interviewees (Golics et al. 2014). Recent examples of PROM development have highlighted where patient partners, trained in qualitative data analysis, have actively collaborated with experienced coders in this process (Chap. 8).

Guidance suggests that the process of documenting concept saturation should be specified within the study protocol (US Food and Drug 2009; Patrick et al. 2011a). To demonstrate that concept saturation has been achieved, first attention must be paid to the representativeness of the population. Once this has been satisfactorily achieved, good practice supports the continuation of interviews with some additional 10–20% of patients before confirming saturation (Golics et al. 2014; Salek et al. 2016).

The use of computer-assisted qualitative data analysis software programmes, for example, NVivo, facilitates the data management, the assessment of between-coder reliability and the documentation of concept saturation and aids quality assurance audits (Patrick et al. 2011a). Data analysis creates a model for the data that makes the data understandable by the research team in the next stage.

9.2.4.6 Item Crafting: Generation and Selection

Once the analysis is complete, the core research team seeks to further refine the conceptual framework, developing domains and subdomains from the defined themes and subthemes and crafting specific questions, or ‘items’, with which to populate an initial long-form version of the developing PROM. Item crafting seeks to convert long, transcribed text into comprehensible, jargon-free, easy-to-read, specific and universal statements which link the essence of the patients’ experience with the content of the developing PROM. The target concept and purpose of measurement must be closely adhered to during this process (Patrick et al. 2011b); clearly specified item selection criteria can assist in guiding the appropriateness of developing items.

The large amounts of data generated at this stage often results in too many potential themes and associated items. The process of item selection is an iterative one, during which multiple viewpoints should be considered and integrated—including the qualitative data, the multidisciplinary team, patient research partners and meth-

odological experts. An important challenge is to avoid losing the patient perspective, and strategies to ensure that the patient voice is retained should be considered. For example, involving patients in the prioritisation of the most important themes can assist in the process of refining the conceptual model and shortlisting items (Gossec et al. 2014).

9.2.4.6.1 Recall Period

The appropriateness of the recall period, that is, the timeframe against which a specified concept is considered, requires special attention. A range of variables including the target population, objectives and frequency of assessment and the content and frequency of an event may influence the appropriateness of the recall period. Commonly used recall periods include ‘current time’ and shorter periods such as the ‘past week’. For example, if the PROM is used in research scenarios such as clinical trials, a recall period which captures an individual’s experience ‘at the present time’ could be more appropriate.

9.2.4.6.2 Response Options and Scaling

The ability to communicate the subjective, qualitative experiences of the patient as an objective, numerical value is a central tenet of PROM development. Selection of an appropriate numerical scale with which to capture the patient experience is a crucial step. A large number of response scales are available, including categorical and adjectival, Likert-type, numerical rating and visual analogue scales (Streiner et al. 2014; Patrick et al. 2011b).

The appropriate number of response options in a scale is driven by a balance between accuracy and practicality. The greater the number of options will enhance the ability of the patient to communicate their experience, thus enhancing precision and discriminant validity, whilst also increasing reliability and responsiveness (Streiner et al. 2014). However, a smaller number of response options improve practicality: good practice supports the adoption of between five and seven responses (Streiner et al. 2014). The interval between each response option needs to be logical and ‘equal’ so that there is a gradual progression from one end of the scale to the other. Whilst there are other schools of thought that challenge this approach (e.g. Andrich 2011), this continues to be a common practice as an initial attempt for scaling of a newly developed PROM.

In arriving at the final score, for most PROMs, a simple summation of item scores is often described. Dependent on the context in which the PROM will be used. For example, at an individual or aggregate level, the final score can be represented either as the actual score or as a percentage. For PROMs which may be utilised within a routine practice setting, a further driver when considering the appropriateness and acceptability of response scales is the ability to score the final PROM and provide timely, interpretable and meaningful data to both clinicians and patients.

9.2.4.6.3 Mode of Administration

Patient self-completion is the preferred format for PROM administration and is a crucial consideration at the start of PROM development. However, there are instances—such as for patients with cognitive impairment or for young children—where proxy completion, such as by a caregiver, is essential (Haywood et al. 2014b).

9.2.4.6.4 Engaging with Experts

PROM development is an iterative process which requires several stages of drafting, evaluation and further refinement (Patrick et al. 2011b). The potential suitability of developing items and item stems, suitability of phraseology, recall period(s) and response scales should be explored with members of the advisory group. Insight from patients, experienced clinicians and measurement experts will help to refine the items—seeking to group, merge, order or delete items and endorse or refine domain development. This process will result in a long-form PROM suitable for cognitive interviewing.

9.2.5 *Crafting the PROM-II: Cognitive Interviews*

This stage represents the last opportunity for significant revision to the PROM (Patrick et al. 2011b). The focus of the cognitive interviews is to verify the relevance, acceptability, comprehension and comprehensiveness of the new PROM with participants' representative of the target population (Brédart et al. 2014; Patrick et al. 2011b; Hay et al. 2014). Four stages of cognitive processing should underpin the interviewing process: comprehension, the process of making sense of the question and developing a response; memory retrieval, the process of relevant information to enable a response; judgement, the process to determine if memory retrieval is accurate and complete; and response mapping, the process by which an appropriate response option is selected (Tourangeau 1984; Patrick et al. 2011b; Gorecki et al. 2012; Hay et al. 2014).

The two most commonly used interview techniques include 'thinking aloud', where respondents express aloud their thought processes whilst answering the question, and often followed by 'verbal probing', where respondents are invited to retrospectively paraphrase or rephrase items (Christodoulou et al. 2008; Brédart et al. 2014). Most authors describe several rounds of semi-structured interviews during which the patient completes either a subset of items or the full PROM (Haywood et al. 2010; Gorecki et al. 2010; Hay et al. 2014)—with both the patient and interviewer highlighting items or aspects of completion which are judged to be difficult or confusing, warranting further exploration. During this process, interviewers should pay careful attention to both verbal and non-verbal respondent clues. Whilst there is no standard approach for using cognitive interview data

for PROM modification (Christodoulou et al. 2008; Gorecki et al. 2012), good practice supports the exploration of results from each round with ‘experts’ (Haywood et al. 2010; Gorecki et al. 2012; Hay et al. 2014), for example, the core research team or advisory group. Where significant revisions are made, subsequent interview rounds will be required. A summary report of the interviewing process should highlight changes made to the PROM. The number of interviewees per round varies, with total sample size estimates ranging from 7 (Leidy and Vernon 2008) to more than 100 with three rounds of interviewing (Hay et al. 2014). The goal is to achieve consensus from a group of patients that the PROM is appropriate.

The ability of patients with different literacy levels to accurately and adequately complete the PROM is a key consideration at this stage of development (Streiner et al. 2014; Petkovic et al. 2015). Sophisticated software—using readability formulas such as Flesch reading ease, FOG and FORECAST—is available with which to evaluate PROM readability (e.g. Zraick and Atcherson 2012), providing a useful adjunct to the cognitive interviewing process.

9.2.6 Content Validation and Further Refinement

Further exploration of the content validity of the developing measure seeks to ascertain that the focus and emphasis of the measure is fit for purpose (Patrick et al. 2011b; Rothman et al. 2009). Developers have adopted different approaches in seeking to establish PROM content validity. For example, the developers of the HidroQoL (Kamudoni et al. 2014) and FROM-16 (Golics et al. 2014) utilised modified nominal groups. First, copies of the developing PROM and a content validation questionnaire were sent to two expert panels—one formed of patients and the second of clinicians. Participants were asked to rate the PROM for language clarity, completeness, relevance and appropriateness of response scale using a 4-point Likert scale for agreement. These groups then met separately to discuss the results and reach consensus on proposed refinements. Agreement between panel members was reported both quantitatively and qualitatively, supporting the process of content validation and informing PROM refinement.

This process results in the final long-form version of the PROM which will be evaluated in the target population.

9.2.7 PROM Evaluation: Item Reduction and Refinement in the Target Population

Item reduction is an important next step in refining the long-form PROM (Streiner et al. 2014). A preliminary psychometric evaluation should be undertaken using both traditional psychometrics (classical test theory) (US Food and Drug 2009;

Streiner et al. 2014) and modern psychometric methods such as Rasch measurement theory (Hobart and Cano 2009) or item response theory (Streiner et al. 2014; Reeve et al. 2007). The importance of this step is to realise a set of items contributing to the measurement of the concept of interest and to elucidate on the internal structure of the new measure.

9.2.7.1 Sample and Sample Size

The initial evaluation should be undertaken in a large, representative population of patients with the target condition. Purposive sampling should be undertaken to ensure that patients representative of key disease features, severity levels and socio-demographic variables are included.

Sample size guidance for ‘new’ summated scales suggests a minimum of five to ten subjects per item (Blazeby et al. 2002). For example, for a new measure with several potential domains, the longest of which includes ten potential items, 100 patients will be required. The subject to item ratio is a frequently used method to determine a required sample size to perform exploratory and confirmatory factor analysis (E/CFA). However, guidance for sample size calculations for performing EFA ranges from 2 to 20 subjects per item, depending on the nature of the data (i.e. the stronger the data, the smaller the required sample size). Recent guidance from COSMIN¹ supports a more conservative maximum number of seven subjects per item or an absolute minimum total of 100 subjects (Terwee et al. 2012). Modern psychometrics requires consideration of the impact of sample size on item fit statistics which, when using polytomous data, are highly sensitive to sample sizes (Streiner et al. 2014). In general, as large a sample size as possible is ideal (Streiner et al. 2014), with a sample size of up to 250 recommended to produce a statistically stable measure.

9.2.7.2 Analyses: Traditional and Modern

Traditional analyses should seek to establish preliminary evidence in support of the acceptability, data quality (scaling assumptions) and internal structure of the measure. Modern psychometrics contribute to this understanding, with the addition of a further exploration of scale targeting, item response, item fit and response bias to further guide PROM refinement and identification of items with poor psychometric properties which are considered for removal. These analyses and comparisons between both approaches are further elucidated by Gorecki et al. (2013) (Table 1, p 4–5). This will result in the final version of the PROM for which final, further psychometric evaluation in the target population is required.

¹ [Consensus-based standards for the selection of health measurement instruments](#)

9.2.8 *Psychometric Evaluation of the Final PROM in the Target Population*

Finally, a comprehensive psychometric evaluation of the final version PROM is required in a large, independent and representative population to confirm evidence of quality, relevance and acceptability. The precision of the PROM depends on the quality of the psychometric evaluation and the evidence of measurement properties. Psychometric evaluations should include the following.

9.2.8.1 Reliability (Internal Consistency; Test-Retest; Measurement Error)

Evaluation of reliability considers the degree of measurement error and is central to the measurement process (Streiner et al. 2014). For example, poor reliability may obscure the correlation of a measure with other measures in the assessment of convergent validity. Similarly, a measure's ability to detect change over time, responsiveness, is equally effected by poor measurement reliability. For multi-item PROMs, both the internal consistency (inter-item correlations, item-partial total correlations and Cronbach's alpha coefficient) and test-retest reliability should be evaluated. Measurement reliability is affected by the target population and setting in which it is completed and hence should be re-established each time a measure is put to new use.

9.2.8.2 Validity (Internal Analyses and Analyses Against External Criteria)

Evaluation of measurement validity seeks to establish evidence in support of the proposed measurement construct. Although delineation is made between different types of validity (content, criterion and construct), a unified perspective considers all forms of validity to be encompassed by construct validity (Streiner et al. 2014). Construct validity relates to the extent to which theoretically derived hypotheses relating to the construct being measured by a PROM are supported by empirical evidence. As there is no single 'ultimate test' for construct validity (Streiner et al. 2014), its assessment involves testing for various hypotheses relating to the relationship between the underlying variable and the items of the PROM in different situations. Therefore, assessing PROM validity requires the testing of a number of clearly specified hypotheses (Terwee et al. 2012; Mokkink et al. 2010).

9.2.8.3 Responsiveness (Criterion or Construct-Based Assessment)

The assessment of responsiveness, also referred to as longitudinal validity, requires an external measure as a criterion for determining whether the patient's condition has changed, improved or deteriorated (Streiner et al. 2014). Establishing evidence

of PROM responsiveness requires not only showing that a PROM can capture statistically significant changes (changes beyond chance) but more importantly that it can capture minimal changes considered important by the patient (Mokkink et al. 2010). The hypotheses to be considered when testing the new PROM include:

1. If the new PROM can capture change in the group of patients experiencing minimal but important change in the condition.
2. If the magnitude of change in patients with minimal improvement in their condition is greater than those with no change in their condition.
3. If change will be greater over the longer period in those patients receiving active treatment.

9.2.8.4 Interpretability

The qualitative meaning of PROM scores is not intuitively apparent (de Vet et al. 2006); the credibility and usefulness of such data are dependent on interpretative guidance and its appropriate use. The cross-sectional comparison of between group ‘differences’—also referred to as ‘minimal important difference’ (MID)—in scores for clearly defined groups can facilitate score interpretation, for example, comparing score differences between the general population and patients with inflammatory rheumatic disease (Salaffi et al. 2009) or between groups categorised according to mild, moderate or severe levels of impact of a condition (Hongbo et al. 2005).

However, interpretation of change scores is crucial to understanding if an individual’s health has improved or deteriorated to an extent that warrants a change in treatment. Two values are important in this context (de Vet et al. 2006): (1) the smallest detectable change (SDC), a change that is greater than measurement error, and (2) the minimal important change (MIC), ‘the smallest difference in score ... which patients perceive as beneficial’ (Jaeschke et al. 1989, p. 408). Consensus is lacking on the most appropriate evaluation of MIC, but both anchor-based—which adopts an external anchor which specifies ‘minimal important’—and distribution-based approaches are described (Crosby et al. 2003). Recent guidance emphasises the importance of understanding meaningful change at the individual level (i.e. the responder), recommending estimation of a ‘responder definition’ based on an empirically derived minimally important change (MIC) (US Food and Drug 2009).

In addition, evidence which supports MID and MIC interpretation adds to the robustness of the measure and its utility both at individual and aggregate level. For example, HTA appraisal of PROM data for a new product compared with ‘standard of care,’ where MIDs are used to demonstrate a between group difference which is important to patients, would be important evidence to facilitate reimbursement recommendations in favour or against the product.

9.2.8.5 Acceptability and Feasibility

Evidence for practical properties including acceptability (relevance and respondent burden) and practicality (completion time, cost, etc.) should also be documented.

9.3 Concluding Remarks

Well-developed PROMs seek to ensure that research and decision-making better capture patient-derived evidence about how they feel, function and live their lives, often aiming to provide a standardised, relevant and acceptable assessment of this experience. Good practice guidance recommends the use of both generic and disease-specific measures in HTA evaluations. However, for many patients, generic measures such as the EuroQoL EQ-5D may lack relevance (Haywood et al. 2016). In recent years, approaches that support the ‘mapping’ of scores from disease-specific PROMs into utility values for the purpose of economic appraisal and HTA evaluations have been developed (Longworth and Rowen 2013). This has the advantage of moving away from utilising a generic measure alongside a disease-specific measure, as has been a common practice. However, HTA appraisal should use PROMs to assist in their decision-making for reimbursement and not just to inform economic appraisal. Although the quality and quantity of life is built in to cost-effectiveness analyses, it does not entirely reflect the impact of the health technology on what patients can and cannot do. HTA should be more cognisant of the value of PROMs in their own right—that is, in isolation from their use in economic appraisal. The selection of well-developed, patient-derived PROMs, developed in a way that reflects the key stages discussed in this chapter, will support this and can provide high-quality, robust patient-based evidence to contribute to HTA.

Acknowledgements The authors would like to acknowledge comments received from Dr. Stefan Cano whilst revising the original transcript. All views expressed and any errors are entirely the responsibility of the authors.

All authors declare no conflict of interest.

References

- Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:571–85.
- Bartlett SJ, Hewlett S, Bingham III CO, Woodworth TG, Alten R, Pohl C, OMERACT RA Flare Working Group, et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis.* 2012;71:1855–60.
- Bredart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop patient-reported outcome (PRO) measures for clinical research: eliciting patients’ experience. *Health Qual Life Outcomes.* 2014;5:12–5.
- Bjorner JB, Kosinski M, Ware Jr JE. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). *Qual Life Res.* 2003;12:913–33.
- Blazeby J, Sprangers MA, Cull A, Groenvold M, Bottomley A. EORTC quality of life group: guidelines for developing questionnaire modules. 3rd ed. Brussels: EORTC Quality of Life Group Publication ; 2002.2-930064-24-2
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395–407.

- Christodoulou C, Junghaenel DU, DeWalt DA, Rothrock N, Stone AA. Cognitive interviewing in the evaluation of fatigue items: results from the patient-reported outcomes measurement information system (PROMIS). *Qual Life Res.* 2008;17:1239–46.
- de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes.* 2006;22:54.
- de Wit M, Abma T, Koelwijn-van Loon M, Collins S, Kirwan J. Involving patient research partners has a significant impact on outcomes research: a responsive evaluation of the international OMERACT conferences. *BMJ Open.* 2013;3(5):e002241.
- Golics CJ, Basra MK, Finlay AY, Salek S. The development and validation of the Family Reported Outcome Measure (FROM-16)[®] to assess the impact of disease on the partner or family member. *Qual Life Res.* 2014;23:317–26.
- Gorecki C, Lamping DL, Brown JM, Madill A, Firth J, Nixon J. Development of a conceptual framework of health-related quality of life in pressure ulcers: a patient-focused approach. *Int J Nurs Stud.* 2010;47:1525–34.
- Gorecki C, Lamping DL, Nixon J, Brown JM, Cano S. Applying mixed methods to pretest the Pressure Ulcer Quality of Life (PU-QOL) instrument. *Qual Life Res.* 2012;21:441–51.
- Gorecki C, Brown JM, Cano S, Lamping DL, Briggs M, Coleman S, et al. Development and validation of a new patient-reported outcome measure for patients with pressure ulcers: the PU-QOL instrument. *Health Qual Life Outcomes.* 2013;11:95.
- Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scriver R, EULAR PsAID Taskforce, et al. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, a 13-country EULAR initiative. *Ann Rheum Dis.* 2014;73:1012–9.
- Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, NCI PRO-CTCAE Study Group, et al. Cognitive interviewing of the US National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *Qual Life Res.* 2014;23:257–69.
- Haywood KL, Garratt AM, Jordan K, Healey EL, Packham JC. Evaluation of ankylosing spondylitis quality of life (EASI-QoL): reliability and validity of a new patient-reported outcome measure. *J Rheumatol.* 2010;37:2100–9.
- Haywood KL, Staniszewska S, Chapman S. Quality and acceptability of patient reported outcome measures in chronic fatigue syndrome/Myalgic encephalitis (CFS/ME): a structured review. *Qual Life Res.* 2012;21:35–52.
- Haywood KL, Collins S, Crawley E. Assessing severity of illness and outcomes of treatment in children with Chronic Fatigue Syndrome/Myalgic Encephalitis (CFS/ME): a systematic review of patient-reported outcome measures. *Child Care Health Dev.* 2014a;40:806–24.
- Haywood KL, Whitehead L, Perkins GD. The psychosocial outcomes of cardiac arrest: relevant and robust patient-centred assessment is essential. *Resuscitation.* 2014b;85:718–9. doi:10.1016/j.resuscitation.2014.03.305.
- Haywood KL, Wilson R, Staniszewska S, Salek S. Using PROMs in healthcare: who should be in the driving seat—policy makers, health professionals, methodologists or patients? *Patient.* 2016;9(6):495–8.
- Health Measures. Applications in research. 2016. <http://www.nihpromis.org/researchers/researchershome>. Accessed 21 Dec 2016.
- Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess.* 2009;13:1–177.
- Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: what do dermatology life quality index scores mean? *J Invest Dermatol.* 2005;125:659–6.
- Jaeschke R, Singer J, Guyatt G. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–15.
- Kamudoni P, Mueller B, Salek MS. The development and validation of a disease-specific quality of life measure in hyperhidrosis: the hyperhidrosis quality of life index (HidroQOL[®]). *Qual Life Res.* 2015;24:1017–27.

- Kosinski M, Bayliss MS, Bjorner JB, Ware Jr JE, Garber WH, Batenhorst A, et al. A six-item short-form survey for measuring headache impact: the HIT-6. *Qual Life Res.* 2003; 12:963–74.
- Leidy N, Vernon M. Perspectives on patient-reported outcomes. Content validity and qualitative research in a changing clinical trial environment. *PharmacoEconomics.* 2008;26:363–70.
- Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health.* 2013;16:202–10.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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.
- Parslow R, Patel A, Beasant L, Haywood KL, Johnson D, Crawley E. What matters to children with CFS/ME? A conceptual model as the first stage in developing a PROM. *Arch Dis Child.* 2015;100:1141–7. doi:10.1136/archdischild-2015-308831. Epub 2015 Oct 9
- Parslow R, Harris S, Broughton J, Alattas A, Crawley E, Haywood K, et al. Children's experiences of Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME): A systematic review and meta-ethnography of qualitative studies. *BMJ Open.* 2016;7(1):e012633.
- Parslow RM. (2016). Developing a Patient Reported Outcome Measure (PROM) for Children with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME). Thesis submitted to the University of Bristol.
- Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, 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.* 2011a;14:967–77.
- Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, 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.* 2011b;14:978–88.
- Petkovic J, Epstein J, Buchbinder R, Welch V, Rader T, Lyddiatt A, et al. Toward ensuring health equity: readability and cultural equivalence of OMERACT patient-reported outcome measures. *J Rheumatol.* 2015;42:2448–59.
- Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, PROMIS Cooperative Group, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care.* 2007;45(5 Suppl 1):S22–31.
- Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. 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.
- Salaffi F, Carotti M, Gasparini S, Intorcchia M, Grassi W. The health-related quality of life in rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis: a comparison with a selected sample of healthy people. *Health Qual Life Outcomes.* 2009;7:25.
- Salek S, Kamudoni P, Oliva E, Ionova T. Quality of life issues important to patients with haematological malignancies. *Value Health.* 2016;18:A709.
- Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures: evolution not revolution. *Patient.* 2012;5:79–87.
- Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use.* 5th ed. UK: Oxford University Press; 2014.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–7.

- Tourangeau R. Cognitive science and survey methods. In: Jabine T, Straf M, Tanur J, & Tourangeau R (Eds.). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press. 1984. pp. 73–100.
- US Food and Drug Administration guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Rockville: Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research; 2009. <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>. Accessed 30 Jan 2016
- Victorson DE, Cella D, Grund H, Judson MA. A conceptual model of health-related quality of life in sarcoidosis. *Qual Life Res*. 2014;23:89–101.
- Zraick RI, Atcherson SR. Readability of patient-reported outcome questionnaires for use with persons with dysphonia. *J Voice*. 2012;26:635–41.