EUROPEAN JOURNAL OF CANCER 46 (2010) 3149–3157

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Review

Choosing patient-reported outcome measures for cancer clinical research – Practical principles and an algorithm to assist non-specialist researchers

T. Luckett [a,b,*], *M.T. King* [a]

[a] *Psycho-oncology Co-operative Research Group (PoCoG), University of Sydney, Camperdown, NSW 2006, Australia*
[b] *Improving Palliative Care through Clinical Trials (ImPaCCT) New South Wales (NSW), South Western Sydney Clinical School, Faculty of Medicine, University of NSW, Wetherill Park, NSW 2164, Australia*

ARTICLE INFO

Article history:
Received 5 May 2010
Received in revised form 23 July 2010
Accepted 3 August 2010
Available online 23 September 2010

Keywords:
Cancer
Quality of life
Patient-reported outcomes
Psychometric properties

ABSTRACT

Aim: The purpose of this article is to give practical advice to researchers wishing to choose measures of quality of life and other patient-reported outcomes (PROs) for cancer clinical research.

Method: Readers are guided through the process of selecting a patient-reported outcome measure (PROM) by means of six principles, illustrated with examples.

Results: PROM selection should always be undertaken with consideration of specific objectives, samples, treatments and available resources. Guiding principles include: (1) always consider PROMs early in the design process within the context of other methodological decisions; (2) choose a primary PROM that is as proximal to the cancer and/or its treatment as will add to knowledge and inform practice; (3) identify candidate PROMs primarily on the grounds of scaling and content; (4) appraise the reliability, validity and 'track records' of candidate PROMs in studies similar to that planned; (5) look ahead to practical concerns; and (6) take a minimalist approach to ad hoc items.

Conclusion: The principles and algorithms presented in this article will assist cancer clinical researchers who lack specialist expertise in patient-reported outcome measurement to make appropriate choices when selecting PROMs for their next study.

1. Introduction

Patient-reported outcome measures (PROMs) are routinely used to assess quality of life (QOL) and other patient-reported outcomes (PROs) in cancer clinical research. Available PROMs range from those that assess severity of individual symptoms (e.g. nausea, xerostomia) to multi-dimensional QOL questionnaires that evaluate the impacts of disease and treatment across every aspect of life. Selection of the optimal set of PROMs for a particular study can be a daunting and difficult task for researchers who lack specialist knowledge. The current article offers principles and an algorithm to guide researchers through the steps needed to make the best choice. These principles are illustrated with examples based on the authors' experience in cancer clinical research.

* *Corresponding author:* Contact address: ImPaCCT NSW, Department of Palliative Care, Braeside Hospital, Locked Bag 82, Wetherill Park, NSW 2164, Australia. Tel.: +61 2 9616 8654; fax: +61 2 9616 8657.
E-mail address: t.luckett@unsw.edu.au (T. Luckett).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2010.08.002

2. Principles of choosing a PROM

Principle 1. Selection of PROMs should be considered early during study design rather than left to the end. Like all aspects of design and methodology, choice of PROMs should be driven by specific research objectives, samples, treatments and available resources[4–7] with the aim of creating optimal conditions for the effect of interest to be identified should it occur. Choice of PROMs and other decisions should be reciprocal and mutually informative. As subsequent principles will illustrate, the content, scaling and psychometric properties of candidate PROMs are of mutual relevance to sample eligibility criteria, sample size calculation, bias control, selection of informative time-points and planning of analysis and interpretation.

Online resources are informative regarding the range of available PROMs. Run by the Mapi Research Trust, the Patient Reported Outcome and Quality of Life Instruments Database (PROQOLID; www.proqolid.org) provides comprehensive information about PROMs for use in a number of health populations, including cancer patients. Australasian researchers can also search the free-to-access Psycho-oncology outcomes Database (PoD) run by the Psycho-oncology Co-operative Research Group (PoCoG; www.pocog.org.au), which focuses on PROMs with applications in cancer research.

Wherever available, local experts in measuring and analysing PROs should be consulted early in the design process.

Principle 2. For the primary outcome, choose as 'proximal' a PRO as will add to knowledge and inform practice

Proximal versus distal PROs

In the most widely cited model of PRO measurement, Wilson and Cleary[8] locate different 'levels' of PROs along a continuum with regard to their 'proximal' (symptoms) versus 'distal' (overall QOL) relationship to the disease or treatment involved.

The model indicates that more distal PROs are subject to greater mediation by personal and environmental characteristics than are more proximal PROs. The most distal outcome, overall QOL, is affected not only by health status but also by non-medical factors (e.g. bereavement, financial stress). Intermediate levels of PROM are available as health-related or disease-specific QOL measures, which limit assessment to the impacts of health in general or a particular condition.

Box 1: Commonly used response options with examples

Visual analogue scales (VAS)

E.g., GLQ-8 : [1]'Please mark these lines with a "I" to indicate how severe each of the following problems has been since your last appointment.'

Feeling anxious or depressed

None_____ Worst I can imagine

Ordered categorical scales

E.g., EORTC QLQ-C30 : [2]'Please answer all of the questions yourself by circling the number that best applies to you.'

'How would you rate your overall quality of life during the past week?'

1	2	3	4	5	6	7
Very poor						Excellent

	Not at all	A little	Quite a bit	Very much
Did pain interfere with your daily activities?	1	2	3	4

Likert scales

E.g., SF-36v2 : [3]'How TRUE or FALSE is each of the following statements for you?'

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false
My health is excellent					

However, these are subject to psychological adaptation (or 'response shift') in response to changing health. Response shift can be of three kinds.[9] In the first kind, patients recalibrate their concept of 'good' and 'bad' health as a result of new experiences. For example, a patient's idea of 'worst pain imaginable' may change following a particularly painful medical procedure. In the second type of response shift, patients revise their values relating to component dimensions of QOL, for example by prioritising existential concerns over physical functioning at the end of life. In the third type of response shift, patients reconceptualise QOL altogether, for example, by trying to focus only on the positive aspects of life and ignoring their declining health. The most proximal outcome, symptom severity, is arguably the least subject to mediation by personal and environmental factors because the evaluative requirement is minimal; in this sense, it is the most 'objective' of subjective outcomes.

Advantages of proximal measures

The PRO relationships put forward in Wilson and Cleary's model have been broadly supported by structural equation modelling.[10] Important implications include:

1. The more distal the measure, the smaller the effects of disease and treatment will be (i.e. there will be a reduced 'signal to noise' ratio).
2. Because overall QOL is often good despite disease, it may be difficult to show an improvement due to treatment (i.e. a 'ceiling effect' will occur).
3. Successful treatment may have greater distal effects when symptoms are severe. However, this may be offset by the greater response shift that is likely to occur when change in health status is dramatic.

(adapted from Brenner et al.[11])

As a result, proximal measures should be more sensitive to differences and responsive to change than those that are more distal. In general, this hypothesis has been supported by studies comparing disease-specific QOL measures with generic QOL measures across a range of health conditions.[12–14]

The practical implication for choice of PROMs is that a primary measure should be as proximal to the disease or treatment as will yield important new evidence of impact. Secondary measures can focus either on other proximal outcomes or on more distal outcomes where the impact of disease or treatment is expected to be large and/or external influences can be controlled for. In many cases, identifying proximal effects will be adequate to contribute to knowledge, inform practice[15] and secure approval of new treatments,[16] especially where there is pre-existing evidence for a causal relationship between the proximal outcome and QOL more generally, as in the case of nausea and vomiting.[17] Symptom measures were among the first PROMs to be routinely used in cancer clinical trials (by the National Cancer Institute of Canada) and continue to be broadly supported by clinicians.

Advantages of distal measures

An exception to Principle #2 occurs when the primary objective is to compare the effects of one disease with others and/or the general population in order to better understand its relative burden. In these cases, generic measures of QOL may be more appropriate than disease-specific measures because they enable comparison without bias towards the proximal effects associated with one disease in particular.

Combining proximal and distal measures

Where practical considerations such as patient burden are not prohibitive, the best solution may be to use a suite of PROMs that will build a comprehensive picture of impacts across both proximal and distal levels. This approach enables analysis of relationships between proximal and distal effects that may be especially informative to the development of interventions. For this reason, it is often useful to include a single item on which respondents are asked to rate their overall QOL. Several QOL instruments include a global item, including the most widely used cancer-specific measures, the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core-30 (EORTC QLQ-C30)[2] (see Box 1) and the Functional Assessment of Cancer and Therapy – General measure (FACT-G).[18]

Reliance on a global QOL item – a cautionary note

Occasionally, there may be good reason to rely on a global QOL rating as the primary outcome (see Example 1 in Box 2). Global rating scales are convenient, reduce patient and staff burden, produce results that are easy to analyse and can in some instances be as valid as those generated using a large number of questions.[19] However, limitations in interpretability and reliability caution against reliance on global QOL ratings alone. A key feature of overall QOL questions is that they are open to interpretation and, when used alone, provide no information about what QOL refers to in each respondent's case. This is a strength inasmuch as it allows each patient to focus on what is most important to them, even if their conceptualisation is different from that of other patients.[20] However, it also poses methodological problems in that:

(a) Global ratings are especially vulnerable to the most problematic kinds of response shift, wherein patients revise priorities among the various dimensions of QOL or even revise their conceptualisation of QOL altogether.
(b) Apparent changes in overall QOL may have been unduly biased by changes in one particular dimension. An example of this occurs when patients start chemotherapy and experience a boost in emotional wellbeing because they are 'doing something' about their cancer whilst simultaneously experiencing worsening symptoms and functioning.[21]

Taken together, the strengths and weaknesses of global QOL items recommend the combined use of these with more detailed assessment of dimensions. This approach will enable patterns of responses to be compared in ways that enable testing for response shift and bias, as well as providing insights into which dimensions have the strongest relationships to overall QOL. For a more detailed discussion of the

relative merits of global measures versus multiple sub-indices, see Sloan and colleagues.[22]

Box 2: 'The importance of considering context when selecting PROMs'

Example 1. When overall QOL is the best choice for a primary outcome measure.

Dr. Jones was developing a randomised phase II study of drug X plus drug Y versus drug X alone for management of symptoms. The primary hypothesis was that adding drug X would have a sparing effect and enable the dose of drug Y (which has nasty side-effects) to be reduced without reducing efficacy. Investigator experience and a review of the literature indicated debilitating but largely distinct side-effects for the two drugs, and it was agreed that the primary PRO should be the difference in mean net benefit reported by patients in each arm. The potential for various benefits and side-effects to be weighted differently by different patients was highlighted as a key consideration. No suitable, validated index could be found that covered all the symptoms, side-effects and concerns, and limited time, funds and eligible patients precluded the possibility of developing and validating a new measure. As a result, the team decided to use the single item global QOL scale from the EORTC QLQ-C30 as the primary PROM. An agreement was reached to administer this scale together with the rest of the QLQ-C30 and several separate scales assessing additional symptoms and side-effects, with a view to describing each of the symptoms and side-effects singly and exploring the variance in overall QOL explained by each separately and by the set together.

Example 2. Social functioning versus social wellbeing.

The social scales of the EORTC QLQ-C30 and FACT-G measure different aspects of social QOL. Correlations have been consistently low and a review of items reveals that the QLQ-C30 assesses impact on social activities and family (*social functioning*) while the FACT-G assesses closeness with friends and partner, social support, family acceptance of illness and satisfaction with family communication and sex life (*social wellbeing*). Where social QOL is an outcome of importance, these differences in content become key considerations when choosing between the two measures. *Social functioning* is more likely to be affected by medical treatments than is *social wellbeing*, and may be more relevant earlier in the patient's cancer journey. In contrast, *social wellbeing* is more likely to respond to psychosocial interventions, such as those aimed at improving social support and intimate communication, and to be more relevant towards the end of life.

Principle 3. Identify candidate PROMs primarily on the grounds of scaling and content

The first questions to ask when narrowing the pool of candidate PROMs are:

1. Which items and aggregate scales offer the best coverage of the impacts of interest?
2. Which aspects of score distributions will be most meaningful to consider, and how amenable to related analyses are the scalings of each candidate PROM?

Compare the relationship between content and scaling in each candidate PROM

All PROMs comprise one or more scales, which in turn consist of one or more items. Most scales are designed to be uni-dimensional (i.e. assess only one 'construct'). When a PROM is developed, the uni-dimensionality of each scale and the way the scales relate to each other – the PROM's 'internal structure' – should be justified on theoretical grounds and supported by evidence from factor analysis, Rasch analysis and/or multi-trait scaling. Frequently, PROM internal structure is developed so that some or all of the scales can be combined to give a summary or overall scale.

At a minimum, comprehensive evaluation of QOL requires assessment of physical, emotional and social wellbeing, as well as some evaluation of how disease and/or treatment impact everyday occupations (e.g. housework, paid work) and roles (e.g. as parents or employees).[23] The extent to which these 'core' dimensions, experience of symptoms and additional disease- or treatment-specific issues (e.g. impact on sexuality) can be aggregated to give an overall index of QOL is controversial.

The distinction between content and scaling is exemplified by the two most widely used cancer-specific QOL measures, the EORTC QLQ-C30 and FACT-G. Both offer scales for the core dimensions and an overall index. But while the FACT-G aggregates items on pain, fatigue, nausea and insomnia into its core dimensions, the QLQ-C30 maintains separate scales for each. Furthermore, while the FACT-G offers an overall score via summation of all twenty-seven items, an overall score on the QLQ-C30 is generated by averaging responses to just two items (global health status and QOL). The consequences of these different approaches are that:

(a) The QLQ-C30 provides more specific information about the effects of disease and treatment that will be useful where symptoms, cognitive functioning or financial impact are outcomes of special interest. However, this results in 15 scales compared to the FACT-G's five, which gives rise to problems of multiple hypothesis testing unless clear hypotheses are made regarding which scales are considered the primary and the secondary PROMs.
(b) According to classical test theory, FACT-G subscale and overall scores should be more sensitive and responsive than their QLQ-C30 counterparts due to their larger number of items; more items provide a finer grading and therefore greater precision (reduced 'noise'). However, this advantage may be countered if only some items in a composite scale are affected while others are not; each issue is implicitly weighted according to the number of relevant items.

The issue of weighting is pertinent when considering any PROM that combines more than one dimension of a given

construct into the same scale. The most important examples are the so-called 'symptom indices' that include assessment of various symptoms within an overall scale of symptom burden (e.g. the MD Anderson Symptom Inventory (MDASI)[24]). Where the research objective is to compare two interventions, care is needed to ensure that one of these is not unduly disadvantaged by PROM scaling. This requires painstaking review of individual items and how these are aggregated with reference to both expected benefits and adverse effects. Similar names for scales in two instruments do not necessarily imply they measure similar constructs (see Box 2, Example 2).

Think ahead to analysis and interpretation

Analysis plans should normally be concerned with the clinical importance of differences or changes in PROs as well as with statistical significance. Relevant jargon includes 'minimal clinically important difference' (MCID),[25] 'minimally important difference' (MID)[26] and 'subjectively significant difference' (SSD).[27] Definitions of all the three terms require that the difference in the target dimension should be perceptible to patients. An MCID also requires the difference to be sufficient to mandate a change in patient management. MCIDs, MIDs and SSDs need to be estimated for each PROM via anchor-based or (acceptable for MCIDs) distribution-based approaches (see Yost and Eton[28] for the FACIT approach). Using PROMs for which these differences have been estimated will facilitate analysis and interpretation of results. Because MCIDs/MIDs/SSDs are nearly always expressed as number of points difference (or change) in means, mean difference between groups or the proportions of samples who improve, remain stable or deteriorate, are the common foci for analysis. Where no estimates are available, common rules of thumb are 0.5 standard deviation (SD)[29] or 10% of the scale.[30] Power and sample size calculations therefore often require an estimate of the scale(s) SD, which can usually be found in papers reporting results from previous studies that have used the PROM, ideally in samples similar to that of the planned research. When reviewing distribution of scores in previous studies, researchers should take note of any risk of ceiling or floor effects with reference to their own samples and objectives. When using 10% of the scale as a rule-of-thumb, scales with more items typically have an advantage (because they typically have a smaller SD); attention should be paid to the relative contributions made to scores by the number of items versus the number of response options (see below). Since the primary PROM scale and its analysis will form the basis for sample size calculation, researchers should be mindful of any limitations in the pool of eligible patients or resources when making these decisions. Where the sample is expected to be small, extra care is needed to ensure that the PROM focuses precisely on the issues of interest (i.e. obtains a strong 'signal'), provides comprehensive coverage of the range anticipated and avoids ceiling and floor effects and has high reliability or precision (i.e. low 'noise').

Principle 4. Appraise the reliability, validity and 'track records' of candidate PROMs in studies similar to the one planned

Much emphasis has been placed on the importance of appraising reliability and validity when choosing PROMs.[31] It should never be said that a PROM is reliable or valid in its own right; rather, evidence for these properties grows with the body of evidence for performance in various populations and clinical settings. Readers are referred to the COSMIN checklist for step-by-step guidance with reviewing evidence for validity and reliability[32] and previous commentaries on the properties of: generic QOL measures;[33] measures for assessing impacts of cancer treatments;[34] cancer symptom indices;[35] cancer pain measures;[36] disease-specific QOL measures for head and neck cancer,[37] gastrointestinal cancers,[38] prostate cancer[39] and gynaecological cancers;[40] measures of distress in women with breast cancer;[41] paediatric QOL measures;[42] end of life QOL measures;[43] QOL measures for long-term survivors[44] and QOL measures for caregivers.[45]

Researchers are encouraged to look beyond articles that focus on reliability and validity to consider the full catalogue of studies in which each candidate PROM has been used. A PROM's 'track record' offers a practical indication of its psychometric properties and is quick and easy to appraise. The researcher reviews all studies where a PROM has been used and identifies any where it has demonstrated an effect. These 'successes' are then cross-referenced to the planned study with regard to samples, treatments and variables relating to design and methodology to see whether similar performance can be reasonably expected. Importantly, absence of effect should not be taken as evidence of PROM failure. A lack of 'real' effect or limitations in design (e.g. small sample size) or method (e.g. poor choice of time-points) may equally have been to blame.

Readers might infer from this advice that an efficient way to identify candidate PROMs is to review studies similar to the one planned to see which measures have been used. While a review of this kind may be helpful, care must be taken not to bypass the preeminent principles described above. Widespread use does not, by itself, recommend a PROM, because measures may become popular for reasons other than superior performance. That said, it is advantageous to be able to compare results between studies, especially those with similar objectives and samples.

Principle 5. Look ahead to practical considerations such as patient and staff burden, the preferred method of administration, the cost of PROM use, and the availability of translated versions and guidelines concerned with scoring and interpretation. Patient and staff burden are especially important considerations in PRO-based research because of the problems caused by missing data.[46] Variables relevant to burden include not only number of items but also how cognitively challenging questions are and the ways in which patients are asked to respond (see Box 1 for commonly used response options). Visual analogue scales (VAS), in particular, rule out the potential for PROMs to be administered via interview either by telephone or when patients are too ill to self-complete; they also take longer to score than do dichotomous or ordered categorical scales.

Where there is an expectation that numerous participants will be too ill to complete PROMs by the end of a study, it is worthwhile considering a measure specifically designed for rating by a caregiver or clinic staff. The Spitzer Quality of Life Index (QLI)[47] has been designed for this purpose and can be used in tandem with patient-reported measures during earlier study time-points in order to calibrate the scores from each.[48]

Changing the method of administration (e.g. self-administered to telephone interview) may affect PROM measurement

properties. Ideally, then, a candidate PROM should have established reliability and validity in a clinical context and via a method of administration similar to those of the planned research. Administration via computer touch-screen technology is becoming increasingly popular (see section on future directions, below). While the initial costs of equipment and programming may be substantial, computer administration can reduce research staff costs over the course of a project, improve the completeness of data and reduce administrative errors.

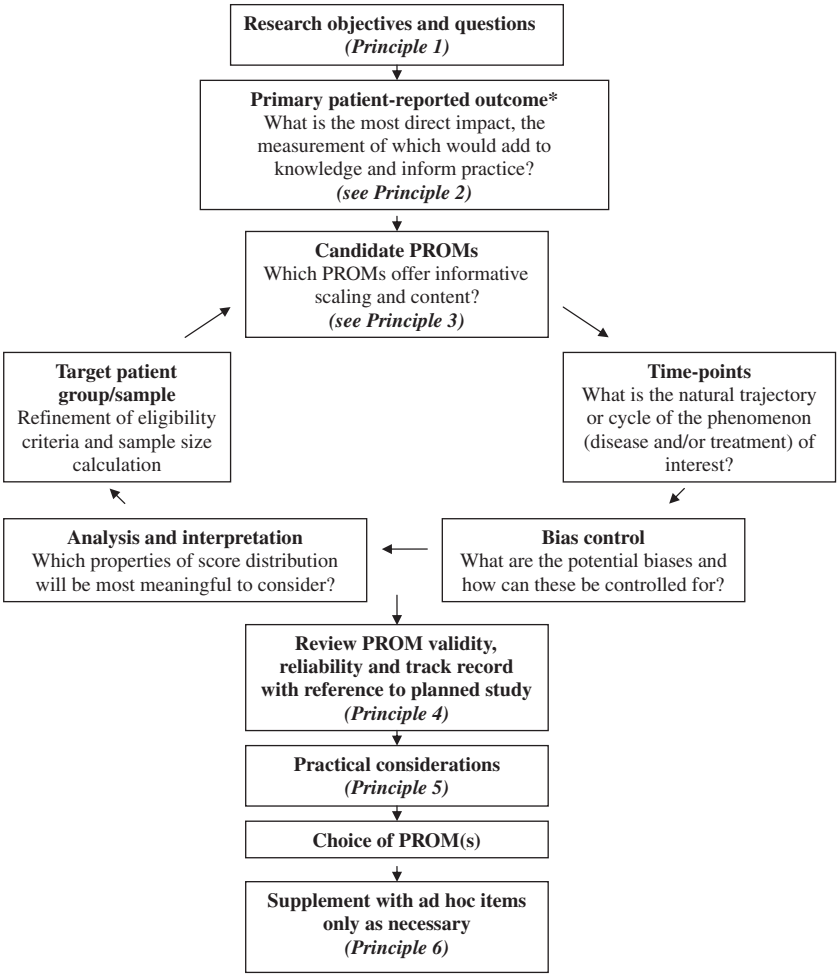
There is significant variation in the costs associated with the use of PROMs themselves and supportive information and services. For example, while both the EORTC and FACIT offer their PROMs free for research not funded by industry, FACIT charges for translations and guidelines on scoring, analysis and interpretation.

Principle 6. Take a minimalist approach to ad hoc items

Where content is similar, PROMs with proven psychometric properties are always preferable to ad hoc measures of the researcher's own invention. Use of ad hoc measures not only raises concerns about reliability and validity but also limits comparison between studies. Occasionally, however, the research objectives may be such that no validated PROM, scale or even items are informative. Given the reciprocity between choice of PROMs and other methodological decisions, it may be possible to adjust the protocol so that an existing measure will serve, although the fundamental objectives of the study should not, of course, be compromised. Whilst not ideal, minor adjustment of a PROM's recall period (e.g. from 'past week' to 'past 24 hours') may also be justifiable where this is the sole barrier to a PROM's suitability (e.g. for episodic symptoms); a few PROMs (e.g. the MDASI[24]) even offer alternative versions.

Where ad hoc items must be generated, these should preferably be used to supplement the existing PROMs rather than constitute the sole outcome measures. Items and response options should be formatted along similar lines to those of validated PROMs being used in the study or else of widely used PROMs.[49] Finally, such items should be appended to the end of standard PROMs to avoid potential measurement artefacts caused by disrupting the order of the items in the established measures.

A minimalist approach should also be taken to scaling of ad hoc items. Ad hoc items should never be incorporated into the scales of validated PROMs; they must be scored and



* The same process should be repeated for each secondary and tertiary patient-reported outcome. In these cases, the question is 'what is the next most important impact, the measurement of which would add to knowledge and inform practice?'

Fig. 1 – Algorithm for choosing patient-reported outcome measures (PROMs).

reported separately. A score will be easier to interpret *prima facie* if it comes from a single item than from an ad hoc scale with no evidence of dimensionality. Where scales are devised, at least minimal evidence for reliability and/or validity should be reported, preferably from pilot data rather than the sample for which results are being reported. Readers are referred elsewhere for guidance on specific properties and related analyses.[50,51] However, it is worth emphasising that best-practice development and validation of PROMs is a resource-intensive process that often takes several years (see the EORTC's protocol for an example of good practice[52]).

An algorithm summarising principles 1–6 is presented in Fig. 1.

A case study illustrating the authors' use of principles in selecting a PROM is presented in Box 3.

Box 3: Case study illustrating use of principles in choosing patient-reported outcome measures

Quality of life after surgery for non-small cell lung cancer

This case study illustrates how we applied these principles in a QOL substudy of a randomised controlled trial of positron emission tomography (PET) in the management of Stages I and II non-small cell lung cancer.[53] We did not expect QOL to be affected by PET. Rather, this trial provided the opportunity to describe for the first time the acute and long-term QOL outcomes of this patient group.[54] This was identified as an important secondary objective at trial concept stage. We considered dyspnoea, cough, pain (chest and arm/shoulder) and fatigue to be the key symptoms of interest for clinician's managing lung cancer patients. We wished to assess the impact of these symptoms on the core components of functioning (physical, emotional and social). We also wished to have a global assessment of QOL. After searching the literature for candidate measures, we selected the EORTC QLQ-C30 and a lung cancer-specific module (EORTC QLQ-LC13), as these provided coverage of all the symptoms and domains that we wished to assess and had been shown to have robust psychometric properties when previously tested.[55,56] We did not need to add any ad hoc questions. The questionnaires, scoring manual and interpretation guidelines were all available at no cost for this investigator-initiated study. Importantly, the questionnaires were available in the languages that we required (English, Chinese, Greek, Italian and Portuguese). Patients were able to self-complete and generally did not find the 43 questions overly arduous. Completion rate as a percent of surviving participants ranged from 96% preoperatively to 86% at 2 years for those without recurrence and 95% preoperatively to 71% at 2 years for those with recurrence. Mixed models provided a statistically robust summary of the QOL trajectories over time, and individual change in QOL relative to preoperative levels was calculated to show the proportion of patients with impaired or improved QOL (using the minimal clinically important difference recommended in the manual) in a way that was clinically meaningful.

3. Future directions

It should be remembered that the measurement of PROs is a dynamic field; measures may be revised or become obsolete over time. In the future, computer adaptive testing (CAT) is predicted by some as likely to replace 'static' questionnaires because it offers greater precision by means of a smaller number of items. In CAT, the first item is the same for all patients but subsequent items are selected by a computer program on the basis of previous answers, ensuring that the items are targeted at each patient's individual level of functioning. Item 'banks' for use in CAT are being developed both by the EORTC and the US National Institute of Health Patient-Reported Outcome Measurement Information System (PROMIS).[56] Provisional item banks and (in the case of PROMIS) the software to administer them are freely available online. PROMIS offers item banks for use not only in cancer but also in the general population and a range of other health conditions, raising the potential for comparability between studies with different populations. PROMIS has also made available on its website a number of static short-forms composed of the most generally informative items from each of its item banks. Given the outlay on equipment and change in practice required to routinely administer PROMs via computer, it may be that these short-forms gain popularity faster than CAT.

Many of the principles outlined in this paper for fixed content PROMs will hold for CAT-administered PROMs. So the principles and the algorithm provided in this paper should continue to help researchers who lack specialist knowledge to make appropriate PROM choices into the future.

Conflict of interest statement

None declared.

REFERENCES

1. Coates A, Glasziou P, McNeil D. On the receiving end – III. Measurement of quality of life during cancer chemotherapy. *Ann Oncol* 1990;**1**:213–7.
2. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;**85**:365–76.
3. Ware JE, Kosinski MA, Bjorner JB, et al. *User's manual for the SF-36v2 health survey*. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2007.
4. Bowling A. *Measuring disease. A review of disease specific quality of life measurement scales*. 2nd ed. Buckingham: Open University Press; 2001.
5. Robinson PG, Carr AJ, Higginson IJ. How to choose a quality of life measure. In: Carr AJ, Higginson IJ, Robinson PG, editors. *Quality of life*. London: BMJ Books; 2003. p. 88–100.
6. Gotay CC. Assessing cancer-related quality of life across a spectrum of applications. *J Natl Cancer Inst Monogr* 2004;**33**:126–33.
7. Osoba D, Aaronson N, Till JE. A practical guide for selecting quality of life measures in clinical trials and practice. In: Osoba D, editor. *Effect of cancer on quality of life*. Boca Raton, Florida: CRC Press; 1991. p. 89–104.

8. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;**273**(1):59–65.
9. Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. In: Schwartz CE, Sprangers MAG, editors. *Adaptation to changing health: response shift in quality of life research*. Washington, DC: American Psychological Association; 2002. p. 11–23.
10. Sousa KH, Kwok O-M. Putting Wilson and cleary to the test: analysis of a HRQOL conceptual model using structural equation modeling. *Quality of Life Research* 2006;**15**(4):725–37.
11. Brenner MH, Curbow B, Legro MW. The proximal–distal continuum of multiple health outcome measures: the case of cataract surgery. *Med Care* 1995;**33**(4 Suppl):AS236–44.
12. Murawski MM, Miederhoff PA. On the generalizability of statistical expressions of health related quality of life instrument responsiveness: a data synthesis. *Qual Life Res* 1998;**7**(1):11–22.
13. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003;**56**(1):52–60.
14. Krahn M, Bremner KE, Tomlinson G, et al. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Qual Life Res* 2007;**16**(3):509–22.
15. Cleeland CS. Symptom burden: multiple symptoms and their impact as patient-reported outcomes. *J Natl Cancer Inst Monogr* 2007;(37):16–21.
16. Rock EP, Kennedy DL, Furness MH, et al. Patient-reported outcomes supporting anticancer product approvals. *J Clin Oncol* 2007;**25**(32):5094–9.
17. Osoba D. Translating the science of patient-reported outcomes assessment into clinical practice. *J Natl Cancer Inst Monogr* 2007;(37):5–11.
18. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–9.
19. de Boer AGEM, van Lanschot JJB, Stalmeier PFM, et al. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual Life Res* 2004;**13**(2):311–20.
20. Youngblut JM, Casper GR. Single-item indicators in nursing research. *Res Nurs Health* 1993;**16**(6):459–65.
21. Cella D. What do global quality-of-life questions really measure? Insights from Hobday et al and the "do something" rule [comment]. *J Clin Oncol* 2003;**21**(16):3178–9 [author reply 9].
22. Sloan JA, Aaronson N, Cappelleri JC, et althe Clinical Significance Consensus Meeting Group. Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proc* 2002;**77**(5):479–87.
23. Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd ed. Chichester, UK: John Wiley and Sons Ltd.; 2007.
24. Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the MD Anderson Symptom Inventory. *Cancer* 2000;**89**(7):1634–46.
25. Jaeschke RJ, Singer J, Guyatt G. Measurement of health status: ascertaining the minimally clinically important difference. *Control Clin Trials* 1989;**10**:407–15.
26. Schunemann HJ, Guyatt GH. Commentary – goodbye M(C)ID! Hello MID, where do you come from? [comment]. *Health Serv Res* 2005;**40**(2):593–7.
27. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998;**16**(1):139–44.
28. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Professions* 2005;**28**(2):172–91.
29. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;**41**:582–92.
30. Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. *Eur J Cancer* 2005;**41**(2):280–7.
31. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Qual Life Res* 1993;**2**:441–9.
32. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol* 2006;**6**(2). doi:10.1186/471-2288-6-2.
33. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;**17**(1):13–35.
34. Barry MJ, Dancey JE. Instruments to measure the specific health impact of surgery, radiation, and chemotherapy on cancer patients. In: Lipscomb J, Gotay CC, Snyder C, editors. *Outcomes assessment in cancer: measures, methods, and applications*. Cambridge: Cambridge University Press; 2005. p. 201–15.
35. Kirkova J, Davis MP, Walsh D, et al. Cancer symptom assessment instruments: a systematic review. *J Clin Oncol* 2006;**24**(9):1459–73 [erratum appears in *J Clin Oncol* 2006;**24**(18):2973].
36. Jensen MP. The validity and reliability of cancer pain measures. *J Pain* 2003;**4**(1):2–21.
37. Ringash J, Bezjak A. A structured review of quality of life instruments for head and neck cancer patients. *Head Neck* 2001;**23**(3):201–13.
38. Pallis AG, Mouzas IA. Instruments for quality of life assessment in patients with gastrointestinal cancer. *Anticancer Res* 2004;**24**(3b):2117–21.
39. Litwin MS, Talcott JA. Measuring quality of life in prostate cancer: Progress and challenges. In: Lipscomb J, Gotay CC, Snyder C, editors. *Outcomes assessment in cancer: measurement, methods, and applications*. Cambridge: Cambridge University Press; 2005. p. 126–59.
40. Luckett T, King MT, Butow PN, Friedlander M, Paris T. Assessing health-related quality of life in gynecological oncology: a systematic review of questionnaires and their ability to detect clinically important differences and change. *Int J Gynecol Cancer* 2010;**20**(4):664–84.
41. Love A. *The identification of psychological distress in women with breast cancer*. Sydney: National Breast Cancer Centre; 2004. <http://www.nbocc.org.au/resources/documents/IPD_The_identification_of_psychological_distress_in_women_with_breast_cancer.pdf>.
42. Clarke S-A, Eiser C. The measurement of health-related quality of life (QOL) in paediatric clinical trials: a systematic review. *Health Qual Life Outcomes* 2004;**2**:66.
43. Mularski RA, Dy SM, Shugarman LR, et al. A systematic review of measures of end-of-life care and its outcomes. *Health Serv Res* 2007;**42**(5):1848–70.
44. Pearce NJM, Sanson-Fisher R, Campbell HS. Measuring quality of life in cancer survivors: a methodological review of existing scales. *Psychooncology* 2008;**17**:629–40.
45. Edwards B, Ung L. Quality of life instruments for caregivers of patients with cancer: a review of their psychometric properties. *Cancer Nurs* 2002;**25**(5):342–9.
46. Bernhard J, Gelber RD. Workshop on missing data in quality of life research in cancer clinical trials: practical and methodological issues. *Stat Med* 1998;**17**(5–7):511–796.
47. Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: a concise QL-index (QLI) for use by physicians. *J Chronic Dis* 1981;**34**:585–97.

48. Simes RJ, Greatorex V, Gebski VJ. Practical approaches to minimize problems with missing quality of life data. *Stat Med* 1998;**17**(5–7):725–37.
49. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). *Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims, December*; 2009. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
50. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use*. 3rd ed. Oxford: Oxford University Press; 2003.
51. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;**11**:193–205.
52. Blazeby J, Sprangers M, Cull A, Groenvold M, Bottomley A. *Guidelines for developing questionnaire modules*. Brussels: EORTC Quality of Life Group; 2002.
53. Viney RC, Boyer MJ, King MT, et al. A randomised controlled trial of the role of positron emission tomography in the management of Stage I and II non-small cell lung cancer. *J Clin Oncol* 2004;**22**(12):2357–62.
54. Kenny PM, King MT, Viney RC, et al. Quality of life and survival in the two years after surgery for non-small cell lung cancer. *J Clin Oncol* 2008;**26**(2):233–41.
55. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC core quality of life questionnaire (QLQ-C30) for use in lung cancer clinical trials. *Eur J Cancer*. 1994;**30A**:635–42.
56. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years [comment]. *Med Care* 2007;**45**(5 Suppl 1):S3–S11.