Patient-reported outcome measures: an overview

Keith A Meadows
Keith A Meadows, Founder & Director, Health Outcomes Insights
Email: keith@healthoutinsights.com

Over the past few years there has been a fundamental shift in focus to give greater emphasis to the involvement of the patient in the care they receive and which is reflected in a number of policy and national initiatives (Darzi, 2008; Department of Health, 2010). Patients also want to be involved in the decision making process (Guadagnoli and Ward,1998).

Against this background, both nationally and internationally, the assessment of outcomes based on the patient's perspective using patient-reported outcome measures (PROMs), are increasingly accompanying the traditional clinical ways of measuring health and the effects of treatment on the patient (Devlin and Appleby, 2010). Furthermore, for more than a decade PROMs have played and continue to have an important role in national and international academic and clinical research, resulting in the development of many thousands of reliable and valid instruments to measure the patient's self-reported health.

At a national level, from 1st April 2009 all licensed providers of NHS-funded elective procedures including hip and knee replacements, groin hernia surgery or varicose vein surgery, have been required to implement the requirement to collect PROMs contained in the Standard NHS Contract for Acute Services (Browne, 2010). And in 2010, the UK Department of Health in collaboration with the University of Oxford initiated a pilot study to determine if PROMs are an acceptable way of collecting health information for a number of long-term conditions (LTCs) in primary care.

This article provides an overview of some of the key issues relating to the selection and use of PROMs starting with a description of the different health constructs they purport to measure, key issues in the selection of an appropriate PROM through to an overview of some of the methodological approaches to interpreting PROM data and maximizing the benefits of their use.

ABSTRACT

With the increasing prominence of the patients' involvement in the care they receive, the assessment of outcomes based on the patient's perspective using patient-reported outcome measures (PROMs), are increasingly accompanying the traditional clinical ways of measuring health and the effects of treatment on the patient. This article provides an overview as to what PROMs are and the different health constructs they are purported to measure. Differences between generic and disease-specific, multidimensional and index scored PROMs are also described. Factors relating to the choice of a PROM are discussed with a particular focus on the importance of developing a measurement strategy and endpoint model to ensure the appropriate PROM is selected to measure the desired outcome. Examples of the application of PROMs are given together with some of the methodological approaches to interpreting PROM data. Finally, issues on maximizing the benefits of using PROMs are briefly discussed.

KEY WORDS

Validity • Reliability • Quality of life • Health-related quality of life • Health status • Wellbeing • Patient satisfaction

What are proms and what do they measure?

What is a PROM?

PROM's are the tools we use to gain insight from the perspective of the patient into how aspects of their health and the impact the disease and its treatment are perceived to be having on their lifestyle and subsequently their quality of life (QoL). They are typically self-completed questionnaires, which can be completed by a patient or individual about themselves, or by others on their behalf.

But why should patients' perspective be considered? Don't the clinical outcomes inform of the effectiveness of care provided or the benefits of a new prescribed drug or treatment?

Clinical outcomes such as an HbA1c for a patient with diabetes or the peak flow rate of an asthma sufferer are useful outcomes in terms of treatment effectiveness. However, as indicators of the impact the disease might be having on the patient's QoL, for example, because of the restrictions on their social activities as a consequence of a fear of going into a hypoglycaemic coma, clinical outcomes alone do not provide the full picture.

Access to the patients' perspective through the use of PROMs can impact on a wide range of aspects related to the delivery of effective health care including, identifying those issues faced by patients and their families living with an illness and how this knowledge might impact on treatment decisions and adherence and provide a better understanding as to how health-practitioners can affect outcome.

What do PROMs measure?

A PROM should be designed to provide information around a given construct which must be made explicit by the instruments' authors. Common constructs measured by PROMs include, health status, health-related quality of life (HRQoL), QoL, wellbeing, treatment satisfaction. symptoms and functioning.

Health status

Measures of health-status such as the SF-36 (Ware and Sherbourne, 1992) and EQ-5D (EuroQol Group, 1990) which are in common use, focus on the quality of health including, the biological and physiological dysfunctioning, symptoms as well as the physical e.g. the ability to walk a block or climb a flight of stairs, as well as the psychological and social functional impairments. Measures of health status are however, not synonymous with measure of QoL or HRQoL.

QoL - In defining QoL there is the general consensus that it is based on the individual's subjective evaluation of the psychological, physical and social aspects of their life, which is changing over time as a result of different influences such as treatment (Speight et al, 2009). QoL is what the patient says it is (Joyce, 1994). It has been argued to be the extent to which the treatment, management or living with a disease or condition is perceived as impacting on the individual's life where issues such as family life, work and school are, in addition to health, an important priority. (Mc Gee et al, 1991).

HRQoL

Often referred to as the degree to which the treatment and the disease as perceived by the individual to impact on those aspects of their life - in addition to health – which are considered important (Speight et al, 2009)

Wellbeing

Measures of psychological wellbeing focus on those aspects of psychological illness including, anxiety, depression, coping, positive wellbeing and adjustment and a sense of control and self-esteem. Typical measures of wellbeing include the Beck Depression Inventory (BDI) (Beck et al, 1961) and the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983). Although related to the patient's QoL psychological wellbeing is not synonymous with it.

Patient satisfaction

PROMs developed to gauge the patient's satisfaction with treatment focus on the patient's subjective appraisal of their experiences of treatment including, side effects, efficacy and convenience.

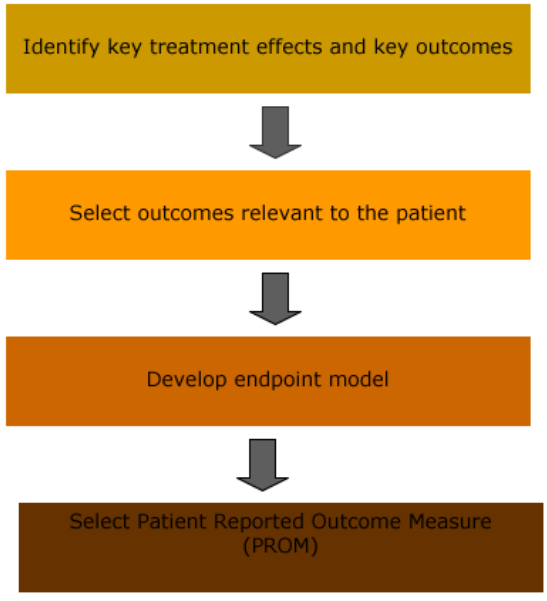
Symptoms and functioning

Measures of symptoms can concentrate either on a range of impairments or a specific impairment such as depression or pain. PROMs assessing functioning focus on activities such as personal care and activities of daily living.

Selecting the appropriate PROM

Selecting the most appropriate PROM is of course the most critical aspect of the study design and in the absence of some universally agreed definition as to what the measured health concept is and its relationship to the objectives of the study, choosing the appropriate PROM can be problematic. It is not uncommon that the choice of an outcome measure such as QoL is based on whether the instrument has been used in previous studies or the name of the instrument appears to be appropriate for its intended use (Polonsky, 2000). The golden rule for selecting the appropriate PROM is 'think about what you are going to do before doing it'.

Figure 1. Key stages in the development of a measurement strategy



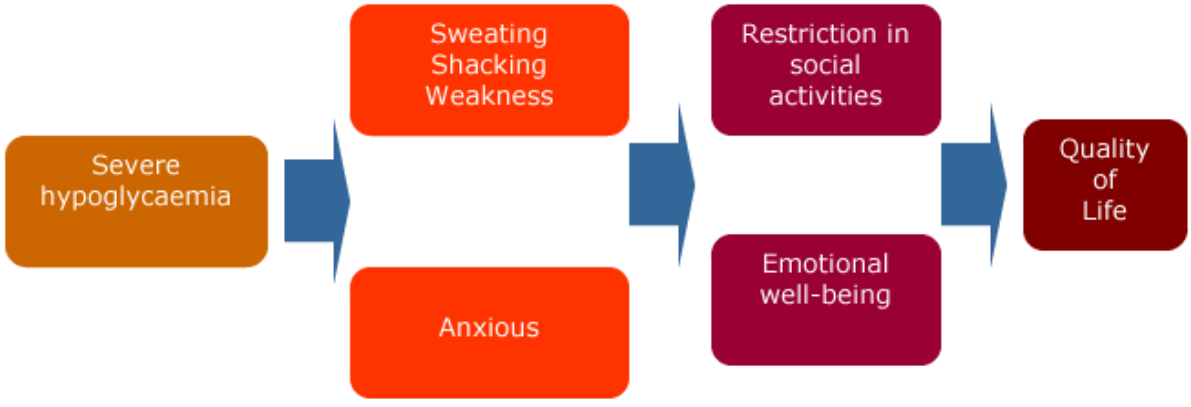
An effective way to establish the link between the measured outcome such as the patient's wellbeing following an intervention programme is to develop a measurement strategy, which requires a clear understanding of the disease and the outcomes relevant to the disease area and patient (*Figure 1*).

Central to the measurement strategy is the endpoint model which provides the rationale for the measurement model, by making explicit the associations among the different health outcomes, patient and domains to be measured by the PROM. *Figure 2* depicts a simple endpoint model showing the relationship between the physio/biological outcomes of severe hypoglycaemia and their predicted impact on those domains relevant to the patient such as restrictions in social activities and emotional wellbeing and subsequent impact on QoL. Based on the understanding of the disease and the expected treatment effects, *Figure 3* illustrates a more focused model showing the specific link between the primary clinical endpoint - of a reduction in the risk of hypoglycaemia - and the subsequent impact on patients' QoL as the secondary endpoint to be measured by the PROM. Once the relevant endpoint(s) (outcome (s)) has been identified, the appropriate PROM can then be selected or developed.

Generic and disease-specific PROMs

PROMs can be generally categorized as generic and disease or condition-specific, with each having their own strengths and weaknesses.

Figure 2. Endpoint model example diabetes



The generic PROM measures health concepts that are of relevance to a wide range of patient groups and the general population and as such can be used for comparison across different conditions as well as with healthy populations. Owing however, to the generic nature of their content, they will most likely include items that are not particularly relevant to many patient groups. For example, questions relating to physical mobility such as ability to walk a block, climb one flight of stairs or reach for items on a shelf will not be relevant to adolescents with diabetes whose concerns are more likely to be focused on dealing with the implications and complexities of the disease at a time of changing self-perception, sexual and physical development. Also generic measures are more likely to exclude content that is of particular relevance to a specific disease group. For example, both for patients and clinicians issues around diet management and the enjoyment of food are key factors in the management of diabetes (Schlundt et al, 1996). Yet content specifically focusing on these important aspects are unlikely to be included in a generic measure, making them of less relevance to the respondent and less sensitive to detecting change in outcome.

In contrast, disease-specific or condition-specific PROMs have been developed to capture those elements of health and QoL of relevance to a specific patient group. Often disease-specific PROMs are developed using qualitative research methodologies such as in-depth interviews and focus groups with people with the specific disease or condition, thus, ensuring that patients are only asked questions which are meaningful and acceptable to them. Disease-specific measures are also more likely to be of greater clinical relevance as well as being more responsive to detecting clinically important changes resulting from treatment intervention. For example, the Diabetes Health Profile (DHP) contains content specifically focusing on diabetes-specific areas which have been identified by patient groups as important including, emotional distress; barriers to activity including social activities and eating problems (Meadows et al, 1996; Meadows et al, 2002).

However, unlike generic measures, owing to the specific content, disease-specific PROMs cannot be used for making comparisons with other conditions.

Multidimensional or unidimensional?

The manner in which data from PROMs can be derived depends on the rationale behind its development and manner of scoring. For example, questionnaires that measures a single construct, for example, pain are described as unidimensional and the items (questions) are added to yield a single overall score.

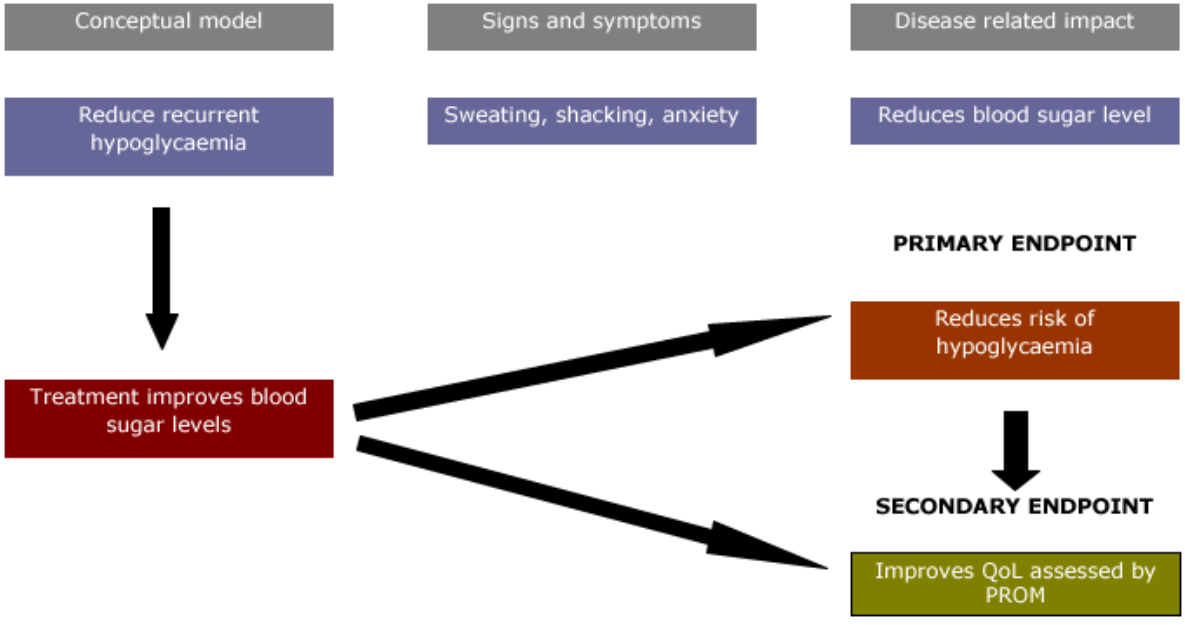
In contrast a multi-dimensional questionnaire is used to provide a profile of scores. For example the Short Form SF-36 which is a generic multidimensional PROM comprising 36 questions scored to produce a score for each of the eight different areas ranging from physical functioning, role limitations because of physical health, bodily pain through to role limitations resulting from emotional problems and mental health (Ware and Sherbourne, 1992). In addition, to providing a profile of scores, scoring algorithms enable the eight areas to be summated to provide two separate summary scores for mental and physical health respectively.

However, there needs to be a cautionary note in scoring both unidimensional and multidimensional PROMs. When a PROM has been developed specifically as a multidimensional scale, it can never be assumed that an overall score can be derived simply by adding each of the items without evidence that the scale can be treated as unidimensional, for example, through confirmatory factor analysis (Brown, 2006). Similarly, a PROM providing an overall score cannot be assumed to be measuring a single construct unless there is sound psychometric evidence to support this assumption. A PROM must always be scored in accordance with the author's recommendations and any deviation from this must be supported by evidence and approval from the instrument's author.

Reliability and validity

A PROM is only of value if it has been well designed and based on a sound theoretical underpinning or conceptual

Figure 3. Endpoint model showing relationship between treatment effects and outcomes



model of what it is purported to measure (*Figure 4*). The content of a PROM should be fully representative of the construct measured (content validity) and be derived from patient input to ensure its relevance to the study population.

Content of a PROM is arguably its most important quality, as without satisfactory content validity, no amount of statistical manipulation will improve its measurement qualities. It must also demonstrate evidence of having been developed with scientific rigour and satisfactory psychometric properties including, reliability, validity and sensitivity to detect change. A poorly designed PROM is unlikely to meet the required psychometric criteria and detect changes in the measured construct as a consequence of treatment. Without a theoretical or conceptual framework and patient input, content is likely to be of less relevance to the patient, leading to annoyance and alienation and failure to answer the questionnaire and most importantly, provide less than valid data. A more detailed discussion of the psychometric properties of PROMs is beyond the scope of this article and the interested reader is referred elsewhere. (Loewenthal, 1996; Nunnally, 1994).

Last but not least, other practical considerations in selecting the appropriate PROM include, its acceptability to patients, length and relevance to the particular patient group or study. Is it easy for patients to complete? How long does it take to complete? How easy is it to score? Does it come with user manual and scoring instructions?

Applications

PROMs can be applied to obtain data from the patient's perspective in a range of areas. Data derived from a PROM can guide clinicians in making decisions about different clinical inputs and for monitoring the outcomes of specific interventions. PROMs can also provide a baseline assessment of the health status, QoL, patient satisfaction or wellbeing etc., of a specific population to identify need and the delivery of effective care as well as aid communication between patient and doctor. For example, the doctor can discuss with the patient why a question was answered in a particular way or why the patient's score has improved or declined since the previous visit. PROMs can also be routinely administered in clinical settings for audit and quality assurance such as in the assessment of the effectiveness of different procedures.

As already highlighted, PROMs are playing an increasingly significant role in clinical research such as in randomized control Phase III trials as secondary endpoints to provide 'added value' to the primary biomedical outcomes. Commercially, data derived from PROMs can for example, be used to support product differentiation, provide intelligence for the targeting of drug development and marketing and in some cases support labelling claims.

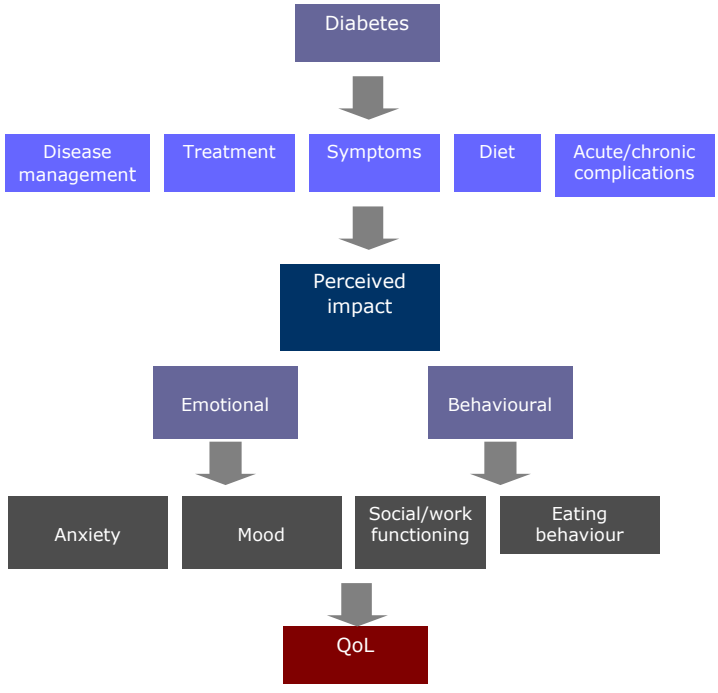
Interpreting prom data

Interpretation of data derived from a PROM can be challenging, particularly with regard to understanding the meaning of what a change or difference in a score means clinically.

Interpretation of PROM data is linked with both the objectives of the study and the constructs measured by the PROM and as a consequence should be a key factor in the development of any measurement strategy.

Understandably, optimising the patient's health is of primary importance for the clinician and therefore, is more likely to look for improvements in health status. However, what might be of specific interest to the clinician in terms of outcome may not always correspond to

Figure 4. Conceptual model for Diabetes Health Profile



what is of relevance to the patient. Remembering also that PROMs are assessing outcomes from the perspective of the patient, any change in the score may or may not be related to a clinical outcome. As an example, patients with diabetes who have changed to a different insulin treatment considered more convenient – but, in all other respects is the same – are unlikely to report any improvement in health status but, may well report that their QoL has been enhanced as a consequence of the greater flexibility in their lives.

Despite the many thousands of PROM developed and their application, there is often lacking an understanding as to what a PROM score represents and what is a meaningful change in score (Osoba and King, 1995). Should the differences be big or small and what are the implications for clinical practice and health policy?

When large samples (macro level) of patients are studied, differences in PROM scores between patient groups might be numerically small but, highly statistically significant. However, data obtained at the macro level are difficult to apply at the individual patient (micro) level (Cella et al, 2002). Furthermore, interpretability is also something that cannot be established from a one-off study but is based on a body of evidence developed over time through a variety of studies, and perspectives (Osoba and King, 1995).

A detailed discussion of the various approaches to interpreting PROM data is beyond the scope of this paper. However, there are a number of approaches that that can aid the interpretation of PROM data and which are summarised below (Osoba and King, 1985).

- Minimal Important Difference (MID) - the smallest difference in a score that is considered to be worthwhile or important
- Known groups – the mean scores underlying particular clinical groups or clinical indicators which give rise to them and which can be used as a clinically-based benchmark to compare other groups
- Normative and reference groups - mean scores underlying particular clinical groups or clinical indicators which give rise to them and which can be used as a clinically-based benchmark to compare other groups
- Statistical significance - the statistical significance in the probability of treatment (A) is better than treatment (B)
- Effect size – a way of quantifying the difference between two groups that has many advantages over the use of tests of statistical significance alone and emphasises the size of the difference rather than confounding this with sample size.

Other approaches include, reference to the PROMs content when interpreting its score as well as comparing them with known clinical parameters such as days in hospital and illness severity (Osoba and King, 1995), the proportion of patients whose PROM score improve or get worse after intervention (Detmar et al, 2002) and cross-instrument calibration using advanced statistical analyses such as Item Response Theory (IRT) (Reeve and Fayer, 1995).

What still needs to be done to get the most out of PROMS?

Despite the increasing use of PROMs across a range of clinical settings and research, there are a number of important issues that remain to be addressed before we can maximize the benefits from their use, such as in routine care. These include for example, the need to ensure instruments, data collection and analysis is highly credible. There must also be demonstrable evidence that PROMs have been developed with scientific rigour with proven reliability, validity and sensitivity to detect change where change is present as well as the need to make PROM data meaningful to those who need to be informed including, nurses, clinicians, commissioners, providers and policy makers. Data must be relevant, affordable and practical to collect and not affect the delivery of care in the process. There is also a need to know more about how best PROMs can be embedded into the decision-making process and combine the derived data with other clinical information as well as explore ways to provide feedback as insightful information to enable sound clinical decision making.

Conclusion

By providing a standardized way of quantifying the patient's perspective on the impact of illness and its treatment on life, PROMs are an increasingly important adjunct to the delivery of appropriate patient care and treatment efficacy. This article has attempted to provide an overview of what PROMs are and what they measure as well issues in their selection and interpretability.

PROMs assess a wide array of health constructs, including, health status, HRQoL, QoL, wellbeing and satisfaction, etc. However, it is not uncommon to find in the literature that these different health constructs are being used interchangeably or considered as being synonymous. This is a false assumption and potential users of PROMs need to be aware of the conceptual differences underlying each construct.

Choosing the appropriate PROM needs to be based on an explicit measurement model and an articulation as to what the measured endpoint is, e.g. health status, QoL, HRQoL etc. combined with evidence to support its reliability, validity and sensitivity to detect change. As part of the measurement strategy, consideration must be given to whether the outcome will be assessed as a unidimensional or multidimensional construct using a disease– specific or generic instrument.

The interpretation of data derived from PROMs along with reliability, validity and responsiveness, is an essential attribute of the instrument. Nevertheless, it can be challenging particularly with regard to the understanding of what a meaningful difference in score is. Although there are a number of different approaches to aid interpretation of PROM scores, ranging from content analysis through to the more complex methodologies, scores must be meaningful to those most that need to be well informed.

PROMs are an important adjunct to the traditional clinical indicators to understanding the impact of health on the patient and the delivery of quality care. Nevertheless, issues remain before the benefits of their use across a range of settings can be maximized. Least not the need to ensure instruments, data collection and analysis is highly credible and that the information they provide is interpretable and insightful to enable sound clinical decision making. **BJCN**

KEY POINTS

- PROMs measure outcomes from the patient's perspective.
- PROMs increasingly accompany the measurement of traditional clinical outcomes.
- PROMs measure a range of constructs including health status, quality of life; health-related quality of life; wellbeing and symptoms.
- PROMs should be selected on the basis of explicit measurement strategy and endpoint model.

Beck A, Ward C, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. *Arch Gen Psychiatry* **4:** 561-71

Browne J, Jamieson L, Lewsey J et al (2010). *Patient Reported Outcome Measures (PROMs) in Elective Surgery.* DH, London. http://tinyurl.com/5s6xrfj (Accessed 12 February 2011)

Brown TA (2006) *Confirmatory factor analysis for applied research*. Guilford Press, New York

Cella D, Eton DJ, Lai JS, Peterman AH, Merkel DE (2002) Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. *Journal of Pain & Symptom Management* **24:** 547-61

Darzi A (2008). *High Quality Care For All: NHS next stage review final report. Cm 7432*. The Stationery Office, London. http://tinyurl.com/6hymms4 (Accessed 22 February 2011)

Department of Health (2010) Equity and excellence: Liberating the NHS, White paper. DH, London

Detmar SB, Muller MJ, Schornagel JH, Wever LD and Aaronson (2002) Health-related quality-of-life assessments and patient-physician communication: a randomised control trial. *Journal of the American Medical Association* **288:** 3027-34

Devlin NJ, Appleby J (2010) *Putting health outcomes at the heart of NHS decision making.* King Fund, London

EuroQol Group (1990) EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy 1990* **16:** 199-208

Guadagnoli E, Ward P (1998) Patient Participation in Decision Making. *Soc Sci Med* **47**(3)**:** 329-39

Loewenthal KM (1996) *An Introduction to Psychological Tests and Scales.* UCL Press Limited, London

McGee HM, O'Boyle CA, Hicky A, O'Malley K, Joyce CR (1991) Assessing the quality of life of the individual: the SEIQoL with a healthy and gastroenterology unit population. *Psychol Med* **21:** 749-59

Meadows KA, Steen N, McColl et al (1996) The diabetes Health profile (DHP): a new instrument for assessing the psychosocial profile of insulin requiring patients - development and psychometric evaluation. *Qual Life Res* **5:** 242-54

Meadows KA, Abrams, Sandaek A (2000) Adaptation of the Diabetes Health Profile (DHP-1) for use with patients with Type 2 diabetes mellitus: psychometric evaluation and cross-cultural comparison. *Diabetic Medicine* **17:** 572-80

Nunnally JC (1994) *Psychometric theory*. McGraw Hill, New York

Osoba D, King M (1995) *Meaningful differences. In: Fayers P, Hays R, eds. assessing quality of life in clinical trials*. Oxford University Press, Oxford: 243-258

Polonsky WH (2000) Understanding and assessing diabetes-specific quality of life. *Diabetes Spectr* **13:** 36

Reeve BB, Fayer P (1995) *Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers P, Hays R, eds. assessing quality of life in clinical trials.* Oxford University Press, Oxford: 55-73

Schlundt D, Pichert JW, Gregory B, Davis D (1996) 'Eating and diabetes: a patient-centred approach'. In: Rubin RR (ed.) *Practical Psychology for Diabetes Clinicians: How to deal with Key Behavioural Issues Faced by Patients and Health Care Teams.* American Diabetes Association, Alexandria: 63-72

Speight J, Reaney MD, Barnard KD (2009) Not all roads lead to Rome – a review of quality of life measurement in adults with diabetes. *Diabetic Medicine* **26:** 315-27

Ware JE, Sherbourne CD (1992) The MOS 36-item Short Form Health Survey. I Conceptual framework and item selection. *Med Care* **30:** 473-83

Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. *Acta Psychiatr Scand* **67:** 361-70