Chapter 12 DISCUSSION

An enormous array of patient-reported health instruments have been reported in the literature. At least 1275 different instruments were identified a few years ago and the number will have grown substantially since then (Garratt et al., 2002). This makes the task of selecting instruments for any given purpose challenging. The reviews reported here have used standard and fairly widely accepted criteria to assess instruments, basically focused on measurement properties and evidence of practical feasibility to reach, wherever possible, specific recommendations. As emphasised in the Introduction, whilst such criteria are widely accepted and to a large extent can be expressed in operational terms, how a review should weigh up the sum of evidence available has been less extensively discussed and agreed. To take an example, how should a review assess the overall merits of an instrument if there is extensive positive evidence supporting the reliability and responsiveness of an instrument but a small amount of weak evidence indicating poor response rates? There are problems of weighing up contrasting positive and negative features of performance of instruments on different criteria, weighing the volume versus quality of evidence, possible biases against reporting negative evidence, the likelihood of more long-established instruments accumulating more favourable evidence and so on. In the end considerable judgement is unavoidably involved in assessing overall performance of one instrument with another.

Despite such unresolved difficulties in the assessment of health instruments, for the most part it was possible to reach reasonably confident recommendations, with some caveats, for each of the conditions. In some reviews, for example, diabetes, there did not seem to be sufficient evidence to highlight one diabetes-specific instrument over others. It is not surprising that for each of the six specific conditions, the largest amount and best quality of supportive evidence was found for SF-36 as a generic health instrument. It is generally the most widely used and most extensively examined of instruments, to capture broad aspects of health in general populations as well as in studies of specific health problems.

It is frequently recommended that a generic instrument should be used in combination with a disease- or condition-specific instrument when assessing the health problems of individuals with a particular condition, so that both broad general features of health and rather more specific problems are equally captured in assessment. It is assumed here that that is the optimal strategy for group-based uses in the NHS whenever it is feasible and especially when change over time is an important issue. Generally the evidence is that disease-specific instruments are more sensitive to changes within patient groups with specifically identified health problems (Wiebe et al., 2003). Most of the recommendations made in this report are compatible with such a strategy. No very long disease-specific instruments have been recommended, so that being combined in a battery with, for example, SF-36 is feasible. However in a small number of instances it does not make sense to use some specific combinations of generic and disease-specific instruments, for example where the content of the disease-specific instrument has been partly or largely derived from a generic instrument and there would be resulting repetition if used in a battery. An example is discussed in the chapter on epilepsy.

It may be an important goal to assess utilities or values regarding health, for example in the context of an economic evaluation. The reviews of long term conditions presented here have found encouraging evidence for instruments such as EQ-5D and HUI, designed and developed to assess utilities. Instruments such as EQ-5D are also quite short and hence appropriate to be used in combination with disease-specific instruments. Most recently, it has been argued that an alternative strategy is to use disease-specific instruments but to derive appropriate utilities for the various health states described in them. There is nothing inherently difficult in such a 'hybrid' solution and examples of utility-based disease-specific instruments have begun to appear (Wasserman et al., 2005; Casey et al., 2006). However there was little evidence of such work in the conditions specifically reviewed here.

The reviews have tried to give as much attention as possible to whether instruments are feasible to use on a routine or regular basis, for example taking account of evidence response rates, respondent burden and acceptability. Unfortunately these aspects of instruments are not remotely as commonly assessed and reported as are traditional measurement properties. Typically the only direct evidence available are matters such as the number of questionnaire items, any apparent complexity of format, and whether or not a trained interviewer is required for instrument administration. The instruments recommended in this report are for the most part suitable for respondents to self-complete without supervision, for example in a postal survey. This is usually the most cost effective form of administration whilst potentially sacrificing some of the more favourable response rate that can be obtained if a questionnaire is either administered by an interviewer or under supervision. This is supported in a study by Duncan et al., (2005) where it is reported that telephone administration was twice as expensive as postal administration. However, no actual costs were reported.

The simplest and most quantitative expression of an instrument's overall performance in terms of feasibility and acceptability, as stated in the individual reviews, is particularly poorly reported and operational definitions of response rates in any case vary. Reports of response rates in research studies do not necessarily generalise to the routine context. It was not possible to assess relative performance of instruments of instruments in terms of response rates. Nevertheless the general literature provides some useful general guidance on factors that may impact on response rates for patient reported health instruments.

A systematic review of randomised controlled trials to improve response rates to questionnaires across a broad spectrum of contexts including health, shorter length of questionnaire improved response rate (Edwards et al., 2002); A subsequent report from this group showed that the greatest improvement in response rates were achieved by very short questionnaires (Edwards et al., 2004). A second, independent systematic review of trials to improve response rates to questionnaires in health research confirms the importance of shorter questionnaires (Nakesh et al., 2006). A review of response rates to patient satisfaction surveys showed modest improvements to responses by shorter questionnaires (Sitzia and Wood (1998).

Two trials illustrate potential implications for patient-reported health instruments. A randomised trial compared response rates in patients with stroke to (the shorter) EQ-

5D compared with SF-36 and obtained a 5% better response rate with EQ-5D (Dorman et al., 1997). Similarly Iglesias and Torgerson (2000), in a survey of women aged 70 years and older conducted through general practices, randomised respondents to receive a questionnaire that included the EQ-5D and (a longer) SF-12. A difference of 9% was observed in favour of the EQ-5D.

Other aspects of the overall design and conduct of a survey may make a contribution to the response rate, over and above specific instruments selected (McColl et al., 2001). Personally conducted interviews may not necessarily result in higher response rates but do tend to have less missing data (Smeeth et al., 2001). Personalised letters, reminder letters, stamped return envelopes and telephone contact may also positively impact on reponse rates (Edwards et al. 2002; Nakesh et al., 2006.).

It should also be noted that characteristics of respondents have an important effect on response rates. For example, older respondents do appear to have difficulties with questionnaires such as SF-36, especially if they have additional physical or psychological morbidity and impairments (Malison 1998; Parker et al., 2006). Overall it is difficult, from the available evidence to separate out evidence of instrument-specific effects on response rates from the many broader determinants.

Even in contexts where it is possible to automate collection, processing and analysis of information gained from patient-reported health instruments, for example by use of computer interface, the costs of analysis and interpretation of information to relevant audiences should not be neglected. There is some evidence supporting electronic administration of the SF-36 in terms of patient preference and shorter completion times but no details were provided of costs, acceptability to staff and training needed (Caro et al., 2001). In a study by Bendsten et al., (2003 (Sweden)), the feasibility of implementing a computerised system for collecting and analysing patient responses to the SF-36 with patients with COPD reported the thoughts and attitudes among physicians of the utility of the results. Patients completed the SF-36 prior to consultation and the physician reviewed the results following the interview. The physicians rated the patient's health status and then compared the patient's assessment with their own. While there was correlation between physician and patient's responses, the physicians reported that the information from the SF36 did not provide any new information or lead to further clinican decisions. The physicians in this study embraced the concept of incorporating patient's perspectives of their health at an individual level, but there was more interest in incorporating such measurement for the purposes of group evaluation and quality improvement initiatives.

Unlike traditional measurement performance which is reported in quantitative data, feasibility and patient acceptability has been explored qualitatively with focus groups with physicians and patients. Several barriers have been reported and McHorney et al., (2002) illustrates potential and actual barriers to the completion of the SF-36 and a disease-specific instrument at home prior to an appointment at an asthma clinic. Patients reported a lack of feedback from physicians about their responses but they did prefer completing the instruments at home. Interestingly, patients in this study expressed preference for instruments tailored to their specific problems and there was resistance to the inclusion of mental health questions. Patients though could see the benefit of collecting such information from a physician and healthcare provider perspective and also for their own benefit increasing self-awareness of their health in

general. The physicians in this study could see the benefits of using such measurement particularly with people with chronic illness promoting a more holistic view of the patient and the information provided extra protection for documented consultation. Further benefits perceived by the physicians related to the potential marketing value in terms of patient satisfaction with services and for evaluation of services. Again there was preference for disease-specific instruments. Also, the. physicians in this study were concerned about several issues echoed in other studies regarding the economical impact of implementing such assessment as well as the organisational aspects in terms of staff training, real time analysis of data and score interpretation.

It has been suggested that the use of patient-reported health instruments is a useful strategy for improving collaboration between patient and providers and as a screening tool. Espallargues et al., (2000) suggests though that using patient-reported instruments as a screening tool is minimally collaborative. The use of instruments as a strategy to improve communication suggests that it increases topics discussed and has a positive effect on provider behaviours. However, there is little evidence to suggest this method results in better patient reported-health overtime (Espallargues et al., 2000).

Greenhalgh (1998) describes criteria for assessment of instruments for use in clinical practice. Whilst there is still emphasis on measurement performance, it is suggested that there may be a trade-off between psychometric measurement standards and practical aspects of feasibility and clinical utility. For example, some instruments are lengthy and take 15-20 minutes to complete. Practically this may not be feasible in a clinical setting. Patients may require assistance or find completing questionnaires too burdensome. Shorted instruments which have been developed from longer parent versions may be more acceptable to both patients and clinicans with the added benefit of less time for completion and analysis. The SF-12 is an example of a generic instrument and the MiniAQLQ asthma-specific. Greenhalgh (1998) points out that feasibility and clinical utility should have higher priority than traditional quantitative measurement properties when selecting an instrument for use in clinical practice.

The evaluation of services is an important feature of quality management in the NHS. A whole systems approach to care is being advocated including not only the measurement of patient-reported health outcomes, but the use of services, satisfaction with care and user involvement in defining quality. In a study by Steinwachs et al., (1994), as part of a quality improvement programme, evaluation of the feasibility of implementing an outcomes management system for patients undergoing cardiac angiography and patients with asthma across 13 different types of organisations was examined. Outcomes management in this study was defined as a systematic approach to collecting information on the impact of medical care on patients' health outcomes. Feasibility was defined as successful collection of outcomes data; is the information collected reliable, valid and discriminative; and will the information provide useful predictors of outcomes to improve the overall quality of care. Response rates were higher for physicians than patients in this study and no differences were observed between different organisations. There was a positive relationship between response rates and data collection and staffing levels especially where there was committed personnel allocated to collecte data with a specified protocol to increase responses

from patients. This study also found that those patients who were satisfied with their care also had a positive change in their health status.

Whilst many of the instruments included in this review for people with chronic conditions are suitable for self-completion individualised components/domains within instruments may prove difficult for some patients (Garratt 2000).

Several barriers have been asserted with reference to successful implementation of patient-reported health measurement. Complex scoring systems advocated by the developers may be an important barrier and specific training is required. Jenkinson (2000) argues that for the London Handicap Scale simple summation yields almost identical information to the complex weighed scheme suggested by the developers.

The greater challenge is to ensure that information gathered by this methodology is made meaningful to those audiences. Although increasing effort is put into interpretability of health status scores, it remains one major barrier to wide-spread up-take. To date evidence demonstrating that information from patient-reported outcomes can make a difference to individual clinical practice and health outcomes is not persuasive (Gilbody et al., 2002; Greenhalgh et al., 2005). It is encouraging though that there is evidence from several exploratory studies that both physicians and patients embrace the concept of measurement. Of great interest would be the somewhat different question of whether such information can be used to make a difference at the system level when used to assess quality and performance of services. Such evidence is even more lacking.

For PHI's to be implemented successfully there needs to be a cultural acceptance and strong clinical leadership; financial resources; specific training for staff; and time allocated.

More generally, it needs to be emphasised that users, clinicians, patients, managers and service providers will have legitimate and valuable views about the relative importance and appropriateness of instruments for any given application. In addition to the evidence assembled in reviews such as those reported here, a full appraisal of the relative value of instruments for a given task must take account of stake-holders judgements of the fit between instruments and specific intended uses. This element of judgement about appropriateness and relevance to a given context cannot be assessed in the same way as formal measurement properties. 'Appropriateness' is one of the key criteria emphasised by Fitzpatrick and colleagues (1998). This criterion has to rely on users' judgements of the degree of fit of the content of an instrument to a specific intended application; something that, a priori, cannot be determined by reviews of formal measurement properties.

This report has also assessed instruments to involve individuals with long term conditions in assessing the quality of their care. Although the field of involving patients and users in assessing quality of care (measures of patient satisfaction, patient experience etc) is as long-standing as that of health status measurement, it has not evolved in the same way. Few if any instruments have emerged to dominate and there is fairly constant flux of instruments. It is not surprising that this should be the main pattern also to emerge from the current review, focusing on instruments for use with patients with long term conditions. A likely reason for the absence of instruments

emerging with clearly superior measurement properties is that instruments are required to address patients' experiences of diverse, specific contexts and services (which in turn constantly change their forms). Instruments end up being developed for specific dedicated purposes and cannot be applied to different settings. The specific questions and concerns of those who commission such work may also vary enormously.

Nevertheless the review did provide important, useful insights for how patients and users with long term conditions might be involved in assessment of service quality. There are commonalities in the domains and topics important to individuals with long-term conditions across the more promising instruments found in the review. Also it is clear that such experiences can be addressed via standard self-complete survey instruments.

Somewhat similar observations may be made about carer impact. Although no instruments emerged clearly to dominate assessments of relative performance, it is clear that there are some possibly promising instruments. There is also some convergence in terms of the range of domains and topics of concern to carers of individuals with long term conditions.

Overall it is hoped that this report provides a clear and encouraging body of evidence to increase the contribution that patients, service users, carers and the public can make to the evaluation of services in the UK.

REFERENCES

Preben Bendtsen, Matti Leijon, Ann Sofie Somme, Margareta Kristenson. Measuring health-related quality of life in patients with chronic obstructive pulmonary disease in a routine hospital setting: Feasibility and perceived value *Health and Quality of Life Outcomes* 2003, **1:**5

Caro JJ, Caro I, Caro J, Wouters F, Juniper EF. Does electronic implementation of questionnaires used in asthma alter responses compared to paper implementation? *Quality of Life Research* 2001; **10**:683-91.

Casey R, Tarride JE, Keresteci MA, Torrance GW. The Erectile Function Visual Analog Scale (EF-VAS): a disease-specific utility instrument for the assessment of erectile function. *Can J Urol.* 2006;**13**:3016-25

Dorman P, Slattery J, Farrell B et al., A randomised comparison of the EuroQol and Short-Form 36 after stroke. *BMJ* 1997; 315-461.

Duncan P, Reker D, Kwon S, Lai S, Studenski S, Perera S et al.,. Measuring stroke impact with the Stroke Impact Scale: telephone versus mail administration in veterans with Stroke. *Medical-Care* 2005; **43**: 507-15

Edwards P, Roberts I, Clarke M, et al., Increasing response rates to postal questionnaires: a systematic review *BMJ* 2002; 324: 1183-5.

Edwards P, Roberts I Sandercock P, Frost C, Follow-up by mail in clinical trials: does questionnaire length matter? *Controlled Clinical Trials* 2004; 25: 31-52.

Espallargues M, Valderas JM, Alsonso J. Provision of Feedback on Perceived Health Status to Health Care Professionals. *Medical Care***.** 38(8):877-878, August 2000; 175-183

Garratt AM, Hutchinson A, Russell IT. Patient-assessed measures of health outcome in asthma: a comparison of four approaches. *Respiratory Medicine* 2000; **94**:597-606.

Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ*. 2002;**324**:1417

Gilbody S, House A, Sheldon T, Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome – a systematic review *Psychological Medicine* 2002: **32**: 1345-1356

Greenhalgh J, Long AF, Brettle AJ, Grant MJ.Reviewing and selecting outcome measures for use in routine practice.*J Eval Clin Pract*. 1998 Nov;4(4):339-50.

Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med*. 2005; **60:833-**43

Iglesias C, Torgeron D, Does length of questionnaire matter? A randomised trial of response rates to a mailed questionnaire *J Health Serv Res & Policy* 2000; 5: 219-21.

Jenkinson CP, Mant JW, Carter J, Wade DT, Winner S. The London Handicap Scale: a re-evaluation of its validity using standard scoring and simple summation. *Journal of Neurology, Neurosurgery and Psychiatry* 2000; **68**:365-7.

McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, Thomas R, Harvey E, Garratt A, Bond J. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess.* 2001;**5**(31):1-256

Mallinson S The Short-Form 36 and older people: some problems encountered when using postal administration. *J Epidem Comm Health* 1998; 52: 324-28.

McHorney CA, Bricker ED. A qualitative study of patients' and physicians' views about practice-based functional health assessment. *Medical Care*, 2002;40(11):1113-1125..

Nakesh R, Hutton J, Jorsted-Stein E, et al., Maximising response to postal questionnaires--a systematic review of randomised trials in health *research BMC Med Res Methodolog* 2006; 1471-2281.

Parker S, Bechinger-Engish D, Jagger C et al., Factors affecting completion of the SF-36 in older people. *Age Ageing* 2006; 35: 376-81.

Smeeth L, Fletcher A, Stirling S, et al., Randomised comparison of three methods of administering a screeing questionnaire to elderly people. *BMJ* 2001; 323: 1403-6.

Sitzia J and Wood N, Response rate in patient satisfaction research: an analysis of 210 published studies. *International Journal for Quality in Health Care* 1998; 10:311-17

Steinwachs DM, Wu AW, Skinner EA. How will outcomes management work? *Health Affairs* 1994.
Wasserman J, Aday LA, Begley CE, Ahn C, Lairson DR. Measuring health state preferences for hemophilia: development of a disease-specific utility instrument. *Haemophilia.* 2005;**11**:49-57

Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol*. 2003;56:52-60