PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES (PROMs) FOR DIABETES

Report to the Department of Health, 2009

Patient-Reported

OUTCOME

Measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH DIABETES: AN UPDATE 2009.

Elizabeth Gibbons
Ray Fitzpatrick

Patient-reported Outcome Measurement Group
Department of Public Health
University of Oxford

Copies of this report can be obtained from:

Elizabeth Gibbons
Senior Research Officer
Patient-reported Outcome Measurement Group

tel: +44 (0)1865 289402
e-mail: elizabeth.gibbons@dphpc.ox.ac.uk

Alternatively, it can be downloaded free of charge from the PROM Group website:

http://phi.uhce.ox.ac.uk/

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH DIABETES: An update 2009

CONTENTS

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH DIABETES: An update 2009

EXECUTIVE SUMMARY

Aims of the report

The aims of this report are to review the evidence of patient-reported outcome measures (PROMs) for people with diabetes and to provide recommendations to the Department of Health of PROMs for diabetes that could potentially be used on a large scale population basis, combining good measurement properties with the likelihood of modest burden to respondents in order not to jeopardise response rates. A literature review of relevant PROMs resulted in the identification of a short-list of both generic and diabetes-specific instruments which were then presented to a multidisciplinary panel for discussion. The literature review of the evidence-base and the discussions of the multi-disciplinary panel underpin final recommendations to the Department of Health.

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions (Fitzpatrick et al., 2006). The report included a review of evidence regarding PROMs for diabetes with some recommendations.

The methods of the review are described and the results of the search including sources and search terms used to identify specific published research. Details of this updated evidence are presented firstly for generic PROMs evaluated with people with diabetes, followed by diabetes-specific PROM results. A number of newly identified diabetes-specific measures are also briefly discussed. The report concludes with discussion and recommendations.

Results and short-list of PROMs for people with diabetes

The previous review reported evidence for the following PROMs:

a) SF-36
b) SF-12
c) Sickness Impact Scale
d) Health Utilities Index
e) Quality of Well-Being Scale
f) EuroQol- EQ-5D

This update identified further evidence of performance for the SF-36, SF-12, HUI, and EQ-5D. No further evidence was found for the SIP or QWBS. An additional generic PROM, the WHOQOL-BREF was identified with reporting of some evidence of performance.

Six diabetes-specific instruments were previously identified:

a) Appraisal of Diabetes Scale/ADS
b) Audit of Diabetes-Dependent Quality of Life/ADDQoL
c) Diabetes 39/D-39
d) Diabetes Health Profile/DHP
e) Diabetes Quality of Life Measure/DQOL

f) Diabetes Quality of Life Clinical Trial Questionnaire/DQLCTQ

In carrying out the current update of evidence, a further eleven diabetes-specific instruments were identified.

g) Barriers to Physical Activity in Diabetes (Type 1) BAPADI
h) Diabetes Obstacles Questionnaire (DOQ)
i) Diabetes Treatment Satisfaction Questionnaire (DTSQs, DTSQc).
j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)
k) Diabetes Symptom Checklist-revised (DSC-R)
l) Diabetes-CAT (Computerised Adaptive Testing)
m) Diabetes Impact Survey (DIS)
n) Insulin Treatment Satisfaction Questionnaire (ITSQ)
0) Diabetes Empowerment Scale (DES)
p) Diabetes Specific Quality of Life Questionnaire
q) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)

Short-listed PROMs for discussion

Based on volume of evaluations and good measurement and operational characteristics, the following were presented to a multidisciplinary panel for further consideration:

1. SF-36
2. EQ-5D
3. ADDQoL
4. DQOL
5. DHP

Recommendations

On the basis of appraisal of evidence by the PROM Group, and taking account of ratings and comments from the panel, the SF-36 and EQ-5D are considered suitable as generic measures in diabetes. Given its brevity and the fact that it yields UK-derived preferences, the **EQ-5D** is recommended for use in combination with a condition-specific PROM. The review group agreed with the views of the multi-disciplinary panel that it is difficult to see any of the reviewed instruments as ideal to address the spectrum of experiences of diabetes. Although it is difficult to make a case in favour of one of the three diabetes-specific instruments considered by the panel, the preference of the review team is in favour of the **Audit of Diabetes Dependant Quality of Life (ADDQoL)** as a diabetes-specific instrument, with two other options being the **Diabetes Quality of life (DQOL)** and **Diabetes Health Profile (DHP)** instruments. The combination of two measures, EQ-5D and a diabetes-specific measure together will provide complementary evidence of health status of people with diabetes in the context of potential population-level applications in the NHS. In making the final selection of PROMs considered suitable for piloting in the NHS, the DH will consider salient factors in addition to the evidence and multidisciplinary panel comments.

1: PATENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH DIABETES: AN UPDATE OF EVIDENCE

Background

Patient-reported outcome measures (PROMs) offer enormous potential to improve the quality and results of health services. They provide validated evidence of health from the point of view of the user or patient. They may be used to assess levels of health and need in populations, and in users of services, and over time they can provide evidence of the outcomes of services for the purposes of audit, quality assurance and comparative performance evaluation. They may also improve the quality of interactions between health professionals and individual service users.

Lord Darzi's Interim Report on the future of the NHS recommends that patient-reported outcome measures (PROMs) should have a greater role in the NHS (Darzi 2007). The new Standard NHS Contract for Acute Services, introduced in April 2008, includes a requirement to report from April 2009 on patient-reported outcome measures (PROMs) for patients undergoing Primary Unilateral Hip or Knee replacements, Groin Hernia surgery or Varicose Vein. Furthermore, Lord Darzi's report 'High Quality Care for All' (2008) outlines policy regarding payments to hospitals based on quality measures as well as volume. These measures include PROMs as a reflection of patients' experiences and views. Guidance has now been issued regarding the routine collection of PROMs for the selected elective procedures (DH, 2008).

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions, carer impact, and patient perceptions of quality (Fitzpatrick et al., 2006)[1]. The report included a review of evidence regarding PROMs for diabetes with some recommendations.

The aim of this report is two-fold: to provide an update of more recently published evidence for PROMs in diabetes and to provide as clear recommendations as possible to Department of Health of PROMs that could be used on a potentially large scale population basis to assess health status of people with diabetes to provide evidence relevant to determining the quality of services in the NHS. Recommended instruments would need to combine good measurement properties with the likelihood of low burden to respondents in order not to jeopardise responses rates. An additional consideration would be the availability of a PROM which yielded preference values derived from a UK source. As widely recommended, a strategy of combining a generic measure with a condition-specific measure was considered the most appropriate way of assessing complementary aspects of health status.

This current update review and recommendations draws on the existing evidence for each PROM up to 2006 but only provides fuller descriptive details of measurement and operational evidence which has emerged since 2006.

[1] A structured review of PROMs in diabetes, 2006 can be downloaded at http://phi.uhce.ox.ac.uk/

The full body of evidence was presented to a multidisciplinary panel for discussion. Details of their discussion and views are reported in Appendix F. The PROMs review group considered the combination of the full review of evidence and the multidisciplinary panel's views before reaching its own conclusions and recommendations. In making the final selection of PROMs considered suitable for piloting in the NHS, the DH will consider salient factors in addition to the evidence and multidisciplinary panel comments.

Structure of the report

The methods of the update review are described and the results of the search including sources and search terms used to identify specific published research. Details of this updated evidence are presented firstly for generic PROMs evaluated with people with diabetes, followed by diabetes-specific PROM results. A number of newly identified diabetes-specific measures are also briefly discussed. The report concludes with discussion and recommendations.

Methods for the update review

Methods adopted were as described in previous reviews performed by the PROM group, Oxford. Comprehensive searches were conducted; articles retrieved were assessed for relevance and checked by another reviewer; and evidence of measurement performance and operational characteristics abstracted for each PROM identified. Assessment and evaluation of the PROMs was performed independently by two reviewers adapting the London School of Hygiene appraisal criteria (Appendix B) outlined in their review (Smith et al., 2004). These criteria were modified for our reviews. The final short listing of promising PROMs to formulate recommendations was based on these assessments and discussion between reviewers. The most promising PROMs were then presented to a multidisciplinary panel for final agreement.

Search terms and results: identification of articles

The methods for searching were conducted using three main sources:

The primary source of evidence was the bibliography compiled by the PROM group[2]. At the time of the review update, the PROM database comprised 16,054 records (up to December 2005) downloaded from several electronic databases using a complex search strategy (Appendix A). These records had been assessed as eligible for inclusion in the bibliography and assigned keywords. These records were searched using keywords 'diabetes' and further searches conducted using title OR abstract using the term diabet*.

A further 14,296 records covering the period January 2006-July 2007 had been downloaded using a revised search strategy (Appendix B) but not assessed or keyworded. The term 'diabet*' was used to identify records from title OR abstract.

Supplementary searches included scanning the reference lists of key articles, checking instrument websites, where found, and drawing on other bibliographic resources. Hand

[2] Available online at http://phi.uhce.ox.ac.uk

searching of titles of key journal was conducted from 2006 to 2008. The following journals were selected:

- Diabetes Care
- Health and Quality of Life Outcomes
- Medical Care
- Quality of Life Research

In addition, PubMed records for the past two years (i.e. September 2006-2008) were searched using the term 'diabet*' and the names of the instruments identified in the previous review and this update.

All abstracts were reviewed. When assessed against the review inclusion criteria, articles were retrieved and reviewed in full. Of these, 25 articles were included in the review.
Results are presented in Table 1 combined with the results from the 2006 review (90).

Table 1. Number of articles identified by the literature review-

Source	***Results of search***	***No. of articles considered eligible***	***Number of articles included in review***
PROM database: 30,350	576	30	5
Supplementary search	71	41	20
TOTAL	-	-	25
2006 review			90
TOTAL			**115**

The recommendations are based on an assessment of all the evidence from both the previous review and the update review combined.

2. Result: Generic PROMs evaluated with people with diabetes

The previous review reported evidence for the following PROMs in relation to diabetes:

a) SF-36
b) SF-12
c) Sickness Impact Scale
d) Health Utilities Index
e) Quality of Well-Being Scale
f) EuroQol- EQ-5D

This update identified further evidence of performance for the SF-36, SF-12, HUI, and EQ-5D. No further evidence was found for the SIP or QWBS. An additional generic PROM, the WHOQOL-BREF was identified with reporting of some evidence of performance.

Full details of the development, domains and scoring methods for those instruments are detailed in Appendices C and D.

a). SF-36

Discriminative validity is supported for specific sub-scales from several studies. Older people with diabetes had significantly lower scores than controls on the PF domain (Sinclair et al., 2008). The GHP differentiated between patients with and without diabetic complications (Currie et al., 2006 UK). RP discriminated diabetic patients compared to population norms (Tahbaz et al., 2006; Sundaram et al., 2007). However, the SF-36 was not sensitive enough to discriminate diabetic patients with different levels of glycaemic control (Sundaram et al., 2007). Good evidence of validity is reported (Tapp et al., 2006) with adjusted odds ratios for scores in the lowest quartile of domain scores significant for BP, GHP, PF, RE, RP, SF—compared to those without diabetes in large population study (over 10,000). In addition, those with diabetes related complications were at an increased risk of being in the lowest quartile of the PF dimension.

Responsiveness is supported for the Vitality domain in a trial of different medications for patients with Type 2 diabetes, with statistically significant scores between the two groups as well as other PROMs in the trial (Secnik Boye et al., 2006).

b). SF-12

Both summary scores discriminated a group of diabetic patients with and without depression, increased age and those with higher income and co morbidities and those patients who were high risk diabetics and those with Type 2 diabetes (Grandy et al., 2008).

c) Sickness Impact Scale

No further evidence is found in this update

d) Health Utilities Index
Discriminative validity has been reported with lower scores for individuals with more long-standing diabetes, those receiving insulin treatment and patients experiencing more pain. Furthermore, the HUI differentiated between patients with co-morbidities associated with diabetes (depression, stroke, heart disease and cataracts); and lower scores were predictive of increased health care resource utilisation (Maddigan et al., 2006).

e) Quality of Well-Being Scale
No further evidence is found in this update

f). EQ-5D
The EQ-5D and VAS have some discriminative validity with lower utility scores for patients with diabetes compared to those with high-risk factors. Socio-demographic characteristics were significantly associated with EQ-5D scores for those with no established disease, with adjustment for diabetes and risk status, including age, income, obesity, gender, race, geographical location and household size (Grandy and Fox 2008). Further evidence of validity is reported with the EQ-5D scores discriminating between patients with those with more symptoms and those with daytime hypoglycaemic episodes. However, no differences in scores were reported between those having insulin or oral medications to maintain glycaemic control (Matza et al., 2007 UK) nor patients with or without diabetic complications such as end-stage renal disease or retinopathy (Currie et al., UK 2006).

Convergent validity is supported to some extent with moderate correlation between ED-5D and diabetes-specific measures (DSC-R and ADS). Weak correlation is reported for EQ-5D and BMI (Matza et al., 2007 UK).

Responsiveness is supported for the EQ-5D in a trial of different medications for patients with Type 2 diabetes, with statistically significant scores between the two groups as well as other PROMs in the trial (Secnik Boye et al., 2006) and following a pharmacy based intervention for diabetes, but for the VAS only (Krass et al., 2007). The EQ-5D detected change in patients who were transferring from oral medication control to insulin therapy with improvement in health status and also treatment satisfaction measured by the DTSQc (Kelley and Dempsey. 2006 UK).

Complete responses were reported for a large population (n=14,858) of patients with high risk of diabetes and those with diabetes (Greater than 75%) (Grandy and Fox 2008)
No missing data was reported (Matza et al., 2007 UK). Ceiling effects were observed with 40% of patients reporting the maximum value of 1. No ceiling effects were reported for the VAS.

g) WHOQOL-BREF
Some limited evidence supports construct validity with statistically significant improvement in Physical activity sub-score for patients following education sessions to facilitate self-management. (Skinner et al., 2006).

3. Results: Diabetes-specific PROMs

This section firstly provides a summary of the more recent evidence found for the six diabetes-specific instruments previously identified

a) Appraisal of Diabetes Scale/ADS
b) Audit of Diabetes-Dependent Quality of Life/ADDQoL
c) Diabetes 39/D-39
d) Diabetes Health Profile/DHP
e) Diabetes Quality of Life Measure/DQOL
f) Diabetes Quality of Life Clinical Trial Questionnaire/DQLCTQ

In carrying out the current update of evidence, a further ten diabetes-specific instruments were identified. These newly identified instruments and their evidence-base are each briefly described. The ten PROMs briefly reviewed here are:

g) Barriers to Physical Activity in Diabetes (Type 1) BAPADI
h) Diabetes Obstacles Questionnaire (DOQ)
i) Diabetes Treatment Satisfaction Questionnaire (DTSQs, DTSQc).
j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)
k) Diabetes Symptom Checklist-revised (DSC-R)
l) Diabetes-CAT (Computerised Adaptive Testing)
m) Diabetes Impact Survey (DIS)
n) Insulin Treatment Satisfaction Questionnaire (ITSQ)
0) Diabetes Empowerment Scale (DES)
p) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)

Full details of the development, domains and scoring methods of previously identified instruments and those identified in this update are detailed in Appendices C and D.

Update of evidence for previously identified diabetes-specific PROMs

a) Appraisal of Diabetes Scale/ADS

Some evidence of reliability, validity and acceptability were reported in the previous review from four evaluations.
No further evidence was found in this update.

b) Audit of Diabetes-Dependent Quality of Life/ADDQoL

Good evidence of reliability, validity, responsiveness, precision and acceptability evidence from five studies was reported in the previous review.

Three studies were identified in this update providing further evidence of internal consistency, validity and responsiveness.

High internal consistency has been found (Cronbach's alpha-0.92) (Sundaram et al., 2007).

Validity is supported with the ADDQoL discriminating between Type 2 diabetes patients receiving insulin for glycaemic control and those maintaining control with tablets and diet, with lower scores in the insulin group. Furthermore, patients with poor

glycaemic control had significantly lower scores than those with good control. Lower scores were reported for patients with depressive symptoms (CES-D score≥16) (Sundaram et al., 2007). Weak correlation has been reported between ADDQoL and SF 12 (Sundaram et al., 2007).

The ADDQoL was sensitive to change in patients receiving an intervention focused on education of insulin management and diabetes self-care with the scores increasing by 1.9 points (Lowe et al., 2008) and in a trial of alternative insulin regimens (Ashwell et al., 2008).

c) Diabetes 39/D-39
Some evidence of reliability, validity, acceptability from three studies was identified in the previous review.

No further evidence was found in this update

d) Diabetes Health Profile/DHP
Good evidence of reliability, validity, responsiveness, precision, acceptability from three studies was identified in the previous review.

No further evidence was identified in this update

e) Diabetes Quality of Life Measure/DQOL
Very good evidence of reliability, validity, responsiveness, precision, acceptability from ten studies was reported in the previous review.

One further study provides evidence of responsiveness with statistically significant improvement in scores in small sample of patients with diabetes undergoing islet transplantation (Poggioli et al., 2006).

f) Diabetes Quality of Life Clinical Trial Questionnaire/DQLCTQ
One study provided limited evidence in the previous review. No further evidence was found in this update.

New evidence identified
A further eleven diabetes-specific instruments were identified in the update.

g) Barriers to Physical Activity in Diabetes (Type 1) (BAPADI)
Developed with input from patients, this is a 11 -item questionnaire measures perceived barriers towards the practice of regular physical activity. Responses are obtained with patients indicating the likelihood (1=extremely unlikely to 7= extremely likely) that each of the items would keep them from practicing regular physical activity.

Good internal consistency and reproducibility are reported (Dube et al., 2006).

h) Diabetes Obstacles Questionnaire (DOQ)
This instrument was developed from a review of the literature to identify obstacles to self-management in Type 2 diabetes patients. There are 77 items in 8 domains: Medications (10), Self-monitoring (5), Knowledge and beliefs (9), Diagnosis (6),

Relationships with HCPs (18), Lifestyle changes (13), Coping (8), Advice and support (8). Responses are obtained on a five-point Likert scale.

Good internally consistency and some evidence of construct validity is reported; DOQ scores correlating with patient-reported problems, and glycaemic control (Hearnshaw et al., 2007 UK).

i) Diabetes Treatment Satisfaction Questionnaire (DTSQs, DTSQc).
This has been developed to measure satisfaction with diabetes treatment regimens with patients with Type 1 and 2 diabetes. There are eight items each rated with a 7-point Likert scale. Six items relate to treatment satisfaction with scores ranging from 0 to 36, higher scores indicating greater satisfaction. Two other items assess perceived frequency of hyperglycaemia and hypoglycaemia with higher scores reflecting greater problems (Bradley & Lewis 1990). The original 'status' form (DTSQs) focuses on current satisfaction which is reflected in the response options: 6=very satisfied, 0= very dissatisfied. Ceiling effects have been reported and the developers modified the response options to improve precision. This 'change' version (DTSQc) items are equivalent to the DTSQs but response options changed to 3= more satisfied now, -3= less satisfied now with a midpoint of 0 reflecting no change.

Responsiveness is reported for both versions but the DTSQc is more responsive than DTSQs version (Moon et al., 2007 UK; Ashwell et al., 2008). The DTSQc detected change in patients who were transferring from oral medication control to insulin therapy with improvement in satisfaction and also health status (EQ-5D) (Kelley and Dempsey. 2006 UK).

j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)
This has been developed with involvement from in-patients and healthcare professionals with17 items measuring treatment satisfaction and items aggregated to provide a summary score (Bradley et al., 2008).

A one factor structure has been established from principal components analysis. High internal consistency and responsiveness are reported. Discriminative validity is reported with differences in satisfaction scores based on length of stay, gender, and those who had longer duration of insulin use (Bradley et al., 2008; Moon et al., 2007).

k) Diabetes Symptom Checklist-revised (DSC-R)
This 32 item PROM was originally developed in the Netherlands (1994) but has since been translated for English speaking populations. The original instrument (DSC-2) formed the basis of the revised version DSC-R). This revised version measures both the frequency and perceived discomfort of physical and psychological symptoms associated with Type 2 diabetes and its potential complications. Responses to items are dichotomised yes/no answers. If 'yes' is selected, participants rate the perceived discomfort on a five-point scale ranging from 1 (not al all) to 5 (extremely). A Total score and domain scores are obtained. Higher scores indicate greater symptom burden.

Responsiveness has been reported in several studies (Secnik and Boye. 2006; McGill et al., 2007; Vinik and Zhang 2007).

l) Diabetes-CAT (Computerised Adaptive Testing)
The Diabetes-CAT system utilising Item Response Theory (ITR) methods was developed and allows a maximum of seven items to be administered but with stepping rules depending on the persons IRT score. Items were selected from the DIS. In comparison to a paper administration of the DIS, respondent burden was less for the Diabetes-CAT which was determined by time to complete, respondent impression of relevant items, and preference.

Acceptable internal consistency, reproducibility, construct and discriminative validity has been reported (Schwartz et al., 2006).

m) Diabetes Impact Survey (DIS)
This is a new 37-item measure that assesses the degree to which diabetes influences role function, social function, vitality and mental health.. The time frame for all items is the past four weeks with responses obtained from a 5-point Likert scale assessing frequency. The items are taken from the Global Disease Impact Survey (Quality-Metric Incorporated 2000) (Schwartz et al., 2006). English and Spanish language versions are available. Confirmatory factor analysis supports the uni-dimensionality of the instrument. No other evidence has been identified of other measurement properties.

n) Insulin Treatment Satisfaction Questionnaire (ITSQ)
This was developed involving patients and literature reviews. Confirmatory factor analysis produced a 5-factor, 22-item instrument that assess Inconvenience of Regimen (IR 5 items), Lifestyle Flexibility (LF 3 items), Glycaemic Control (GC 3 items), Hypoglycaemia Control (HC 5 items) and Insulin Delivery Device Satisfaction (DS 6 items) (Anderson et al., 2004). Items are scored on a 7-point Likert scale of 0 to 100 with the higher score indicating better treatment satisfaction.

Internal validity has been reported with item-to-item correlations within the 0.50-0.60 range. Confirmatory factor analysis supports the five factor structure (Brod et al., 2007).

Internally consistency is reported with alphas $\geq$0.85 and reproducibility Spearman rank correlation coefficients ranging from 0.63 to 0.94 (Anderson et al., 2004; Brod et al., 2007).

Construct validity is supported with ITSQ scores highly correlated with related measures of treatment burden. ITSQ scores differentiated between insulin delivery methods, glycosylated hemoglobin values, the number of times the patient required assistance administering insulin, and insulin adherence (Anderson et al., 2004). Moderate correlations between ITSQ and other diabetes satisfaction questionnaires, self-reported medical compliance, glycaemic control, are also reported (Brod et al., 2007).

Baseline satisfaction scores were also shown to be predictive of future satisfaction (Brod et al., 2007).

Ceiling effects are reported but were less than 30% (Brod et al., 2007).

Some missing data has been reported but this was less than 2 % (Brod et al., 2007).

o) Diabetes Empowerment Scale (DES)
The DES assesses the patients' self-efficacy in dealing with their condition. It divides self-efficacy into three domains: Psychological aspects, assessing satisfaction and readiness to change, and setting and achieving diabetic goals. Questions are rated from 5 (agree) to 1 (strongly agree).

Responsiveness is reported with statistically significant scores reported in patients receiving an intervention focused on education of insulin management and diabetes self-care. The DES change scores were statistically significant (Lowe et al., 2008).

p) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)
Interviews with patients and cognitive de-briefing informed instrument development. Six items focus on concepts relevant to patient satisfaction with oral anti-diabetic medications. Each item is scored on a 5 point scale 1 = extremely dissatisfied, 5 extremely satisfied. Aggregating scores generates a total score ranging for 6 to 30 with higher scores indicating greater degree of satisfaction.

Factor analysis supports one underlying factor; satisfactory Test re-test reliability has been reported and construct validity with hypothesised correlations between SOADAS and the SF 12 with greater correlations for PCS. The SOADAS discriminates between those patients with or without diabetic symptoms (Donatti et al., 2008).

4. DISCUSSION AND RECOMMENDATIONS

The evidence identified in the previous review (Fitzpatrick et al., 2006) together with new evidence reported in the current update are assessed for generic and disease-specific measures in order to reach recommendations of a short-list of more promising PROMs for diabetes.

Generic measures

Seven generic measures were considered in detail in the previous review: SF-36 (and its variant SF-12), SIP, HUI, QWB, and the EQ-5D. One further generic PROM was identified in this update: WHOQOL-BREF.

Table 2 summarises the psychometric criteria and operational characteristics of the generic PROMs included in the review using the Appraisal criteria outlined in Appendix B. The previous review reported and recommended the SF-36 for assessment of broader aspects of health status in diabetes and where, specifically, utility values are required, the EQ-5D and HUI.

The most substantial evidence remains in support of the SF-36 to capture the broader aspects of health for people with diabetes. High internal consistency has been reported in the previous review. Strong evidence of construct validity is available from the previous review, particularly the discriminative properties. The update provides further evidence of the SF-36 differentiating between people with diabetes and population norms and those with different levels of glycaemic control. Evidence of responsiveness has been provided in longitudinal measurement and in trials of effective interventions. Some evidence from one study suggests that people with diabetes have scores closer to population norms with skewed distribution of scores. Acceptability has been indicated with high response rates to postal administration.

Strong evidence has been reported for the construct validity of the EQ-5D with scores reflecting severity of diabetic symptoms; co morbidities and sociodemographic variables. There is also some convergence with diabetic-specific measures. The EQ-5D has some evidence of responsiveness. Ceiling effects have been observed for the utility measure in this diabetic population.

Some evidence is reported of discriminative validity with the EQ-5D distinguishing patients with or without diabetes; patients receiving different treatment regimens and those with diabetic co morbidities. No other measurement criteria are addressed by the evidence.

Less convincing evidence is available either to recommend the SF-12 from the previous review or this update. Limited evidence of validity is available and missing values have been reported.

The SIP, QWB and WHOQoL-BREF have very limited evidence to support application in the NHS and further evaluations are needed.

Table 2 Appraisal of psychometric and operational performance of generic PROMs for people with diabetes

PROM	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Floor/ceiling/ precision	Acceptability	Feasibility
SF-36	0	+++	+++	+++	+++	0	—	+	0
SF-12	0	0	+	++	0	0	0	—	0
EQ-5D	0	n/a	++	+++	+	0	—	0	0
SIP	+	0	0	+	0	0	0	0	0
HUI	0	0	0	+	0	0	0	0	0
QWB	+	0	0	+	0	0	0	0	0
WHOQoL-BREF	0	0	0	+	0	0	0	0	0

Psychometric and operational criteria

0 *not reported*

— *no evidence in favour*

\+ *some limited evidence in favour*

\++ *some good evidence in favour*

\++ + *good evidence in favour.*

Diabetes-specific measures

Table 3 summarises the psychometric criteria and operational characteristics of the diabetes-specific PROMs included in the review using the Appraisal criteria outlined in Appendix B. The previous review reported and recommended the ADDQOL, DHP and DQOL based on the number of evaluations reporting good measurement and operational performance.

Based on the updated evidence, the ADDQoL, DHP and DQOL remain the most frequently evaluated diabetes-specific measures. The ADDQoL and DHP have been developed in the UK and are available from the developers.

The ADDQoL has good evidence of internal consistency from several studies both in the previous review and this update. Discriminative and construct validity is supported with scores reflecting different glycaemic control and symptoms and diabetic patients with complications. Correlation with other related constructs is also suggested. The ADDQoL is responsive and there is some evidence of precision of the scale. However, some missing data has been reported and one of the questions seems potentially problematic. It refers to eliciting information regarding not having diabetes and it is suggested that this is a complex cognitive task.

The DQOL has mostly good evidence of internal consistency although some items have low alphas. Discriminative validity supports its sensitivity at identifying patients with different levels of symptoms and co morbidities. Some evidence of responsiveness is reported and it appears to be acceptable to patients.

The DHP has some good evidence of measurement performance, although has not been evaluated as frequently as the ADDQoL and DQOL.

No further evidence was found for the ADS, Diabetes 39 and DQLCTQ.

During this update, several other newly developed diabetes-specific measures were identified. There is limited volume of evidence to recommend any of them for short listing and discussion at present. Despite this, there seems to be a trend for the development of more narrowly focused patient-reported measures for people with diabetes which focus on for example, treatment satisfaction (DTSC, DTSC-IP, ITSQ, SOADAS). Others relate to symptoms (DSC-R) and barriers (BAPADI, DOQ). Further evaluations are needed of these instruments.

It may be that, in addition to the broadly based and general measures discussed below in the recommendations, the growing emphasis on issues such as self management in diabetes will require use of instruments addressing more specific experiences such as control.

Discussion

For assessment of broader aspects of health status in diabetes, the SF-36 and EQ-5D offer options in that there is supportive evidence for both from the literature. Whilst there is somewhat more supportive evidence for SF-36 on some criteria, as acknowledged by the multi-disciplinary group there are distinct advantages for the EQ-5D in terms of brevity and simplicity. An additional advantage for the EQ-5D is that it provides UK-derived preference values. For diabetes-specific impact the following were presented to a multidisciplinary panel: ADDQOL, DHP and the DQOL. The review group agree with the views of the multi-disciplinary panel that no existing diabetes-specific measure offers an optimal assessment of the full spectrum of experiences of diabetes.

Recommendations

On the basis of appraisal of evidence by the PROM Group, and taking account of ratings and comments from the panel, the SF-36 and EQ-5D are considered suitable as generic measures in diabetes. Given its brevity and the fact that it yields UK-derived preferences, the **EQ-5D** is recommended for use in combination with a condition-specific PROM. The review group agreed with the views of the multi-disciplinary panel that it is difficult to see any of the reviewed instruments as ideal to address the spectrum of experiences of diabetes. Although it is difficult to make a case in favour of one of the three diabetes-specific instruments considered by the panel, the preference of the review team is in favour of the Audit of Diabetes Dependant Quality of Life (ADDQoL) as a diabetes-specific instrument, with two other options being the Diabetes Quality of life (DQOL) and Diabetes Health Profile (DHP) instruments. .

In making the final selection of PROMs considered suitable for piloting in the NHS, the DH will consider salient factors in addition to the evidence and multidisciplinary panel comments.

Table 3: Appraisal of psychometric and operational performance of diabetes-specific PROMs

PROM	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Floor/ceiling/ precision	Acceptability	Feasibility
ADS	+	0	0	+	0	0	0	+	0
ADDQoL	0	++	++	+++	++	0	0	+	0
D-39	+	0	+	+	0	0	0	+	0
DHP	+	+	++	++	+	0	+	+	0
DQOL	+	+	++	+++	+	0	0	++	0
DQLCTQ	+	+	+	+	0	0	0	+	+
BAPADI	+	+	0	0	0	0	0	0	0
DOQ	+	0	0	+	0	0	0	0	0
DTSQs, DTSQc	0	0	0	0	+	0	0	0	0
DTSQ-IP	0	+	+	0	+	0	0	0	0
DSC-R	0	0	0	0	++	0	0	0	0
Diabetes-CAT	+	+	+	+	0	0	0	0	0
DIS	+	+	0	+	0	0	0	0	0
ITSQ	0	0	0	+	++	0	0	+	0
DES	0	0	0	0	+	0	0	0	0
SOADAS	+	0	+	+	0	0	0	0	0

0 not reported — no evidence in favour + some limited evidence in favour ++ some good evidence in favour +++ good evidence in favour.

Paediatric diabetic-specific PROMs

There are unique challenges for children with type 1 diabetes (T1DM) with management of glycaemic control often carried out by parents and carers. For adolescents, during puberty, there are specific physiological changes due to the reduction of insulin production which subsequently affects metabolic control. Therefore, potentially, during puberty the teenager can experience fluctuations in glycaemic control which are attributable to endocrine changes, greater insulin resistance, as well as erratic meal and exercise patterns (Court et al., 2008). Combined with the challenges associated with the transition from childhood to adulthood, the life-style changes which diabetes can impose may impact on psychosocial functioning and more specifically treatment compliance (Insabella et al., 2007).

Involving children and adolescents in assessing the impact of diabetes is important and proxy judgment by parents or carers may have limitations. There is very little evidence of the measures developed for adults with diabetes being tested for use in children. Two instruments the ADDQoL and DQOL have specific versions for younger populations. A number of other instruments have been developed specifically for younger respondents.

Formal appraisal of the psychometric evidence of these instruments with children and adolescents is beyond the scope of this review.

A recent review of measures identified four generic and five diabetic-specific PROMs which have been evaluated with adolescents (de Wit et al., 2006).

Generic:
- Child Health Questionnaire-Child form 87 (CHQ-CF87)
- Pediatric Quality of Life Inventory 4.0 (PedsQL 4.0)
- KINDL-R
- DISABAKIDS chronic generic module

Diabetes-specific:
- Diabetes Quality of Life-Youth questionnaire (DQOL-Youth)
- PedsQL-Diabetes module (PedsQL-DM)
- KINDL-R-Diabetes Module (KINDL-R-DM)
- ADDQoL-Teen
- DISABAKIDS-Diabetes module (DISABAKIDS-DM)

Generally, the psychometric properties of these PROMs are reported to be satisfactory (de Wit et al., 2006),

The remit of this narrative was to examine whether adult diabetes-specific PROMs could be applied with children in the NHS. There is very little evidence to support recommending any PROM identified in this review for use with children. Moreover, the conceptual and methodological challenges of adopting adult PROMs with children strengthen such judgement.

APPENDIX A: sources for PROM bibliography

1. AMED: Allied and Complementary Medicine Database
2. Biological Abstracts (BioAbs)
3. BNI: British Nursing Index Database, incorporating the RCN (Royal College of Nursing) Journals Database
4. CINAHL: Cumulative Index to Nursing and Allied Health Literature
5. Econlit - produced by the American Economic Association
6. EMBASE - produced by the scientific publishers Elsevier
7. MEDLINE - produced by the US National Library of Medicine
8. PAIS: Public Affairs Information Service
9. PsycINFO (formerly PsychLit) - produced by the American Psychological Association
10. SIGLE: System for Information on Grey Literature in Europe
11. Sociofile: Cambridge Scientific Abstracts Sociological Abstracts Database
12. In addition, all records from the journal 'Quality of Life Research' are downloaded via Medline.

APPENDIX A: PROM Bibliography search strategy[3]

a. records to December 2005 (downloads 1-12)

((acceptability or appropriateness or (component$ analysis) or comprehensibility or (effect size$) or (factor analys$) or (factor loading$) or (focus group$) or (item selection) or interpretability or (item response theory) or (latent trait theory) or (measurement propert$) or methodol$ or (multi attribute) or multiattribute or precision or preference$ or proxy or psychometric$ or qualitative or (rasch analysis) or reliabilit$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or (standard gamble) or (summary score$) or (time trade off) or usefulness$ or (utility estimate) or valid$ or valuation or weighting$)

and

((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY$ or QL or QoL or (qualit$ of life) or (quality adjusted life year$) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))

or

((bibliograph$ or interview$ or overview or review) adj5 ((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY$ or QL or QoL or (qualit$ of life) or (quality adjusted life year$) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))

b. records from January 2006 (download 13)

(((acceptability or appropriateness or component$ analysis or comprehensibility or effect size$ or factor analys$ or factor loading$ or feasibility or focus group$ or item selection or interpretability or item response theory or latent trait theory or measurement propert$ or methodol$ or multi attribute or multiattribute or precision or preference$ or proxy or psychometric$ or qualitative or rasch analysis or reliabilit$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or valid$ or valuation or weighting$)

and

(HRQL or HRQoL or QL or QoL or qualit$ of life or quality adjusted life year$ or QALY$ or disability adjusted life year$ or DALY$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score$ or time trade off or health index or health profile or health status or ((patient or self$) adj (rated or reported or based or assessed)) or ((disability or function$ or subjective or utilit$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))

or

((bibliograph$ or interview$ or overview or review) adj5 (HRQL or HRQoL or QL or QoL or qualit$ of life or quality adjusted life year$ or QALY$ or disability adjusted life year$ or DALY$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score$ or time trade off or health index or health profile or health status or ((patient or self$) adj (rated or reported or based or assessed)) or ((disability or function$ or subjective or utilit$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$)))))

[3] Note: the bibliography includes approximately 1,650 handsearched additions.

APPENDIX B: Psychometric criteria

Appraisal of PROMs

The methods that will be used for assessing the performance of PROMs were developed and tested against multi-disciplinary consensus and peer review (8). They focus on explicit criteria to assess reliability, validity, responsiveness, precision, acceptability and feasibility. A pragmatic combination of the criteria developed and used in previous reports to DH by the Oxford and LSHTM groups will be used.

The appraisal framework focuses on psychometric criteria and PROMs must fulfil some or all to be considered as a short-listed instrument. Practical or operational characteristics are also assessed (acceptability and feasibility) (Appendix B: Appraisal framework).

Once evidence has been assessed for eligibility, records considered as inclusions will be assembled for each PROM identified. Measurement performance and operational characteristics will be appraised using the following rating scale independently by two reviewers and inter-rater reliability calculated (Table 1).

Table 1

Psychometric and operational criteria

0	*not reported*
—	*no evidence in favour*
+	*some limited evidence in favour*
++	*some good evidence in favour*
++ +	*good evidence in favour.*

Appraisal of psychometric and operational performance of PROMs for diabetes

PROM	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Floor/ceiling/ precision	Acceptability	Feasibility

Appraisal criteria (adapted from Smith et al., 2005 and Fitzpatrick et al., 1998; 2006)

Appraisal component	Definition/test	Criteria for acceptability
Reliability		
Test-retest reliability	The stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and re-test scores	Test re-test reliability correlations for summary scores 0.70 for group comparisons
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of items in a scale); assessed by Cronbach's alpha's and item-total correlations	Cronbach's alphas for summary scores ≥0.70 for group comparisons Item-total correlations ≥ 0.20
Validity		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to cover; assessed qualitatively during the questionnaire development phase through pre-testing with patients. Expert opinion and literature review	Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale represent the construct being measured Patients involved in the development stage and item generation
Construct validity	Evidence that the scale is correlated with other measures of the same or similar constructs in the	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted

	hypothesised direction; assessed on the basis of correlations between the measure and other similar measures	strength of correlation
	The ability of the scale to differentiate known-groups; assessed by comparing scores for sub-groups who are expected to differ on the construct being measured (e.g a clinical group and control group)	Statistically significant differences between known groups and/or a difference of expected magnitude
Responsiveness	The ability of a scale to detect significant change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including t-tests, effect sizes (ES), standardised response means (SRM) or responsiveness statistics	Statistically significant changes on scores from pre to post-treatment and/or difference of expected magnitude
Floor/ceiling effects	The ability of an instrument to measure accurately across full spectrum of a construct	Floor/ceiling effects for summary scores <15%
Practical properties		
Acceptability	Acceptability of an instrument reflects' respondents' willingness to complete it and impacts on quality of data	Low levels of incomplete data or non-response
Feasibility/burden	The time, energy, financial resources, personnel or other resources required of respondents or those administering the instrument	Reasonable time and resources to collect, process and analyse the data.

APPENDIX C:
GENERIC INSTRUMENTS

This Appendix provides a brief description of the generic PROMs included in this review. Their origins, development and content are briefly summarized (Section i). Content and format are further summarized in Sections i and ii.

a) SF-36: Medical Outcomes Study 36-item Short Form Health Survey (Ware and Sherbourne, 1992; Ware et al., 1994; Ware, 1997)

The Medical Outcomes Study (MOS) Short Form 36-item Health Survey (SF-36) is derived from the work of the Rand Corporation during the 1970s (Ware and Sherbourne, 1992; Ware et al., 1994; Ware, 1997). It was published in 1990 after criticism that the SF-20 was too brief and insensitive. The SF-36 is intended for application in a wide range of conditions and with the general population. Ware et al., (1994; 1997) proposed that the instrument should capture both mental and physical aspects of health. International interest in this instrument is increasing, and it is by far the most widely evaluated measure of health status (Garratt et al., 2002a).

Items were derived from several sources, including extensive literature reviews and existing instruments (Ware and Sherbourne, 1992; Ware and Gandek, 1998; Jenkinson and McGee 1998). The original Rand MOS Questionnaire (245 items) was the primary source, and several items were retained from the SF-20. The 36 items assess health across eight domains (Ware, 1997), namely bodily pain (BP: two items), general health perceptions (GH: five items), mental health (MH: five items), physical functioning (PF: ten items), role limitations due to emotional health problems (RE : three items), role limitations due to physical health problems (RP: four items), social functioning (SF: two items), and vitality (V: four items), as shown in Table 3.1. An additional health transition item, not included in the final score, assesses change in health. All items use categorical response options (range: 2-6 options). Scoring uses a weighted scoring algorithm and a computer-based programme is recommended. Eight domain scores give a health profile; scores are transformed into a scale from 0 to 100 scale, where 100 denotes the best health. Scores can be calculated when up to half of the items are omitted. Two component summary scores for physical and mental health (MPS and MCS, respectively) can also be calculated. A version of the SF-36 plus three depression questions has been developed and is variously called the Health Status Questionnaire (HSQ) or SF-36-D.

The SF-36 can be self-, interview-, or telephone-administered.

b) SF-12: Medical Outcomes Study 12-item Short Form Health Survey (Ware et al., 1995)

In response to the need to produce a shorter instrument that could be completed more rapidly, the developers of the Medical Outcomes Study (MOS) 36-item Short Form Health Survey (SF-36) produced the 12-item Short Form Health Survey (SF-12) (Ware et al., 1995).

Using regression analysis, 12 items were selected that reproduced 90% of the variance in the overall Physical and Mental Health components of the SF-36 (Table 3.1). The same eight domains as the SF-36 are assessed and categorical response scales are

used. A computer-based scoring algorithm is used to calculate scores: Physical Component Summary (PCS) and Mental (MCS) Component Summary scales are generated using norm-based methods. Scores are transformed to have a mean value of 50, standard deviation (SD) 10, where scores above or below 50 are above or below average physical or mental well-being, respectively. Completion by UK city-dwellers reporting the absence of health problems yielded a mean PCS score of 50.0 (SD 7.6) and MCS of 55.5 (SD 6.1) (Pettit et al., 2001). Although not recommended by the developers, Schofield and Mishra (1998) report eight domain scores and two summary scores. The SF-12 may be self-, interview-, or telephone-administered.

Several authors have proposed simplification of the scoring process and revision of the SF-12 summary score structure, where norm-based weighting is replaced by item summation to facilitate score interpretation (Resnick and Nahm, 2001; Resnick and Parker, 2001).

c) EuroQol-EQ-5D (The EuroQol Group, 1990; revised 1993)

The European Quality of Life instrument (EuroQol) was developed by researchers in five European countries to provide an instrument with a core set of generic health status items (The EuroQol Group, 1990; Brazier et al., 1993). Although providing a limited and standardized reflection of HRQL, it was intended that use of the EuroQol would be supplemented by disease-specific instruments. The developers recommend the EuroQol for use in evaluative studies and policy research; given that health states incorporate preferences, it can also be used for economic evaluation. It can be self or interview-administered.

Existing instruments, including the Nottingham Health Profile, Quality of Well-Being Scale, Rosser Index, and Sickness Impact Profile were reviewed to inform item content (The EuroQol Group, 1990). There are two sections to the EuroQol: the EQ-5D and the EQ thermometer. The EQ-5D assesses health across five domains: anxiety/depression (AD), mobility (M), pain/discomfort (PD), self-care (SC), and usual activities (UA), as shown in Table 3.1. Each domain has one item and a three-point categorical response scale; health 'today' is assessed. Weights based upon societal valuations of health states are used to calculate an index score of –0.59 to 1.00, where –0.59 is a state worse than death and 1.00 is maximum well-being. A score profile can be reported. The EQ thermometer is a single 20 cm vertical visual analogue scale with a range of 0 to 100, where 0 is the worst and 100 the best imaginable health.

d) Sickness Impact Profile (Bergner et al., 1976; revised: Bergner et al., 1981)

The Sickness Impact Profile (SIP) was developed in the USA to provide a broad measure of self-assessed health-related behaviour (Bergner et al., 1976; Bergner et al., 1981). It was intended for a variety of applications, including programme-planning and assessment of patients, and to inform policy decision-making (Bergner et al., 1976; Bergner et al., 1981; McDowell and Newell, 1996).

Instrument content was informed by the concept of 'sickness', which was defined as reflecting the change in an individual's activities of daily life, emotional status, and attitude as a result of ill-health (McDowell and Newell, 1996). Item derivation was based on literature reviews and statements from health professionals, carers, patient groups, and healthy subjects describing change in behaviour as a result of illness. The

SIP has 136 items across 12 domains: alertness behaviour (AB: ten items), ambulation (A: 12 items), body care and movement (BCM: 23 items), communication (C: nine items), eating (E: nine items), emotional behaviour (EB: nine items), home management (HM: ten items), mobility (M: ten items), recreation and pastimes (RP: eight items), sleep and rest (SR: seven items), social interaction (SI: 20 items) and work (W: nine items).

Each item is a statement. Statements that best describe a respondent's perceived health state on the day the instrument is completed are ticked. Items are weighted, with higher weights representing increased impairment. The SIP percentage score can be calculated for the total SIP (index) or for each domain, where 0 is better health and 100 is worse health. Two summary scores are calculated: Physical function (SIP-PhysF), a summation of A, BCM, and M, and psychosocial function (SIP-PsychF), a summation of AB, C, EB, and SI. The five remaining categories are scored independently. The instrument may be self or interview-administered.

The Functional Limitation Profile (FLP) is an Anglicized version of the SIP (Patrick and Peach, 1989; McDowell and Newell, 1996). Wording and some weightings have been altered, and summary scores are calculated using different dimensions to those used in the SIP (i.e. FLP Physical summary calculated by summing A, BCM, M and HM; FLP Psychosocial summary calculated by summing RP, EB, AB, SI and SR. Several abbreviated versions of the SIP have been developed, including a 68-item version (De Bruin et al., 1992; Post et al, 1996).

e) Health Utilities Index

The Health Utilities Index (HUI) was designed as a comprehensive measure of health status and health related quality of life. The Health Utilities Index (Mark 3) is a system composed of a health status classification which defines 972,000 discrete health states, and a preference, or utility, function which can be used to calculate the desirability for each health state. The HUI3 health status classification was developed by Feeny et al., (1995) to assess capacity on eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain/discomfort. The utility function reflects community preferences and scores each unique health state on a scale ranging from 0 (death) to 1 (perfect health). An excellent summary of the development of the HUI measures can be found in Feeny et al., (1996). The HUI3 is a development of the Health Utilities Index containing a sub-set of items which constituted the HUI2. This report summarises data for the most recent version of the HUI (i.e. the HUI3).

f) Quality of Well-Being Scale (formerly the Index of Well-Being) (Kaplan et al., 1976; Kaplan et al., 1984; Kaplan et al., 1993)

The Index of Well-Being was modified and renamed the Quality of Well-Being scale (QWB) to emphasize the focus on quality of life evaluation (Kaplan et al., 1993; McDowell and Newell, 1996, 2006).

The QWB uses a three-component model of health (Kaplan and Anderson, 1988, cited by McDowell and Newell, 1996) comprising: 1) functional assessment, 2) a value reflecting the utility or desirability of each functional level, and 3) an assessment of

illness prognosis to anticipate future health-care need, which may describe positive health. The QWB is interview-administered.

Completion corresponds to the three-component model. First, three domains of self-reported function are assessed, namely mobility and confinement (MOB: three categories), physical activity (PAC: three categories), and social activity (SAC: five categories). Respondents select the most appropriate category to describe their perceived functional level. Domain categories give 45 possible combinations (3 x 3 x 5); with the inclusion of death, 46 function levels are defined for the second stage of completion (McDowell and Newell, 1996). In addition, respondents select from a list of 27 items symptoms or medical problems experienced over the previous eight days.

Social preference weights for each possible health state have been derived from empirical studies. At the second stage, the assignment of an appropriate weight, or utility, to a health state or functional level gives the QWB index score from 0 to 1, where 0 equates to death and 1 to complete well-being. A negative score may be generated, representing a state ‘worse than death’. QWB index scores can be converted into Quality-Adjusted Life-Years (QALYs), supporting their application in economic and policy analysis.

Stage three of the QWB addresses issues of prognosis to produce well-life expectancy score (McDowell and Newell, 1996). This stage is not necessary for calculating the QWB index.

A self-administered version has been developed: the QWB-SA (Andersen et al., 1995). Following a review of QWB items, five items were added to a mental health section and three self-rated health items were included. The QWB-SA has five domains: symptoms and problem complexes (58 acute and chronic items), self-care (two items), mobility, physical functioning (11 items for these two), and performance of usual activity (three items). For the first domain, respondents indicate the presence or absence (‘yes’ or ‘no’) of chronic (18), acute physical (25), and mental health symptoms (11) over the previous three days. The remaining four domains all use a three-day recall response option. The total number of items is inconsistent, ranging from 71 to 74. Symptom/problem weights for the QWB-SA are based on the original QWB weighting system. The focus of the original QWB is utility measurement and quality of life; the focus of the QWB-SA is symptoms and assessment of function. The QWB-SA has been recommended for self-completion by older adults (Andersen et al., 1995).

APPENDIX C

Generic patient-reported outcome measures

:

Instrument	*Domains (no. items)*	*Response options*	*Score*	*Completion (time in minutes)*
SF-36: MOS 36-item Short Form Health Survey (36)	Bodily pain (BP) (2), General health (GH) (5) Mental health (MH) (5), Physical functioning (PF) (10) Role limitation-emotional (RE) (3), Role limitation-physical (RP) (4), Social functioning (SF) (2), Vitality (V) (4)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean values 14-15) Self (mean 12.6)
SF-12: MOS 12-item Short Form Health Survey (12)	Bodily pain (BP) (1), Energy/Vitality (V) (1), General health (GH) (1), Mental health (MH) (2), Physical functioning (PF) (2), Role limitation-emotional (RE) (2), Role limitation-physical (RP) (2), Social functioning (SF) (1)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview or self
European Quality of Life Questionnaire (EuroQol-EQ5D) (5+1)	EQ-5D Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) EQ-thermometer Global health (1)	EQ-5D Categorical: 3 options *EQ-thermometer* VAS Current health	EQ-5D Summation: domain profile Utility index (–0.59 to 1.00) *Thermometer* VAS (0-100)	Interview or self
Sickness Impact Profile (136)	Alertness behaviour (AB) (10), Ambulation (A) (12) Body care and movement (BCM) (23), Communication (C) (9) Eating (E) (9), Emotional behaviour (EB) (9) Home management (HM) (10), Mobility (M) (10) Recreation and pastimes (RP) (8), Sleep and rest (SR) (7) Social interaction (SI) (20), Work (W) (9)	Check applicable statements. Items weighted: higher weights indicate increased impairment Recall current health	Algorithm Domain profile (0-100%, 100 worst health); Index (0-100%) Summary: Physical (A, BCM, M), Psychosocial function (AB, C, EB, SI)	Interview (range: 21-33) Telephone: PF only (11.5) Self (19.7)
Health Utility Index 3 (Feeny et al, 1995) (8)	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, Pain	Four domains have five response options and five have six response options	Global Utility index and single attribute utility scores for the eight separate dimensions	Self report, face to face and telephone interview
Quality of Well-being Scale (QWB) (30)	Mobility and confinement (MOB) (3 categories) Physical activity (PAC) (3 categories) Social activity (SAC) (5 categories) Symptoms and medical problems (27)	Categorical: yes/no Recall 6 days Symptoms 8 days	Algorithm Index 0-1, 1 complete well-being	Interview Telephone (mean 17.4, range 6-30)

Summary of generic instruments: health status domains ***(after Fitzpatrick et al., 1998)***

Instrument	*Instrument domains*							
	Physical function	**Symptoms**	**Global judgement**	**Psychol. well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal construct**
SF-36	x	x	x	x	x		x	
SF-12	x	x	x	x	x		x	
EQ-5D	x	x	x	x	x		x	
SIP	x	x		x	x	x	x	
HUI	x			x		x		
QWB	x				x			

APPENDIX D: DIABETES-SPECIFIC PROMs

a) Appraisal of Diabetes Scale/ADS

The ADS is a brief self-report questionnaire which assesses an individual's thoughts about coping with diabetes (Carey et al. 1991). It was developed in light of 'the transactional relationship between stress and diabetes' – the fact that whilst external stressors can disturb glucose metabolism, hence the course of the disease, adherence to a strict diabetic regime can of itself be stressful. The authors suggest that the ADS may be useful as a screening instrument for adjustment to diabetes, specifically to identify those patients experiencing, or at risk for, dysphoric reactions and problems of adherence to their diabetic regime. The content of the scale is based on theory and research regarding appraisal processes; some items were adapted from a generic Attribution Questionnaire (Hammen and Mayol, 1982).

The instrument consists of seven items covering distress caused by diabetes, control over diabetes (two items), uncertainty due to diabetes, anticipated future deterioration, coping, and effect of diabetes on life goals. The items use a five-point adjectival scale scored from 1 (e.g. control – none at all) to 5 (control – total amount). ADS items are summed to produce a score from 0-35, 0 representing the least and 35 the greatest impact of diabetes.

b) Audit of Diabetes Dependent Quality of Life/ADDQoL

The ADDQoL is an individualized instrument designed to measure an individual's perceptions of the impact of diabetes on their quality of life (Bradley et al., 1999; Speight & Bradley, 2000; Bradley & Speight, 2002). The intention was to create a detailed version for research and in-depth clinical work, and a short form for audit purposes. No further information has been found regarding the latter.

The instrument comprises 18 (originally 13) items where the respondent is invited to indicate, firstly, the effect of diabetes on a particular aspect of life (for example, enjoyment of food, ease of travelling) and, secondly, how important this aspect of life is to overall quality of life. Three (originally ten) of the items – namely, family life, working life, and sex life - have a 'not applicable' response option, allowing patients to exclude items which are not relevant to them. Patients respond by circling a number on a seven-point scale which asks how a particular aspect of their life would be if they did not have diabetes (from -3: 'very much better' to +3: 'very much worse'). They then rate the importance of this aspect of their life on a four-point scale (from 3: very important, to 0: not at all important). Impact ratings are multiplied by importance ratings to produce a –9 to +9 score, then summed and divided by the number of applicable domains to produce a final score from –9 to +9.

In the original version of the ADDQoL, two additional summary items asked respondents to rate their general QoL, and what their QoL would be if they did not have diabetes, each on a seven-point verbal rating scale. The revised version has a single summary item measuring 'present quality of life' on a seven-point scale from -3 (extremely bad) to +3 (excellent). The wording has also been simplified and amended in order to reduce the number of 'non-applicable' items.

c) Diabetes 39/D-39

The authors of the D-39 intended it to have 'range and reliability', in other words, to be highly relevant to a wide range of diabetes patients over time, easy to use and understand, and to possess good psychometric properties (Boyer and Earp, 1997). A slightly modified version has been developed for use in clinical trials.

The D-39 comprises 39 items in five domains, namely energy and mobility (15 items), diabetes control (12 items), anxiety and worry (four items), social and peer burden (five items), and sexual functioning (three items). Scores are marked on seven-point visual analogue scales ranging from 'not affected at all' to 'extremely affected', then transformed linearly to 0 to 100 scales.

d) Diabetes Health Profile/DHP

The DHP-1 is a multidimensional self-completion instrument originally designed to identify psychosocial dysfunction among adult insulin-dependent and insulin-requiring patients in an ambulatory care setting (Meadows et al., 1996). The instrument has also been adapted for use in non-insulin dependent patients (Meadows et al., 2000). Content was derived from a literature review, a review of available instruments, interviews with IDDM and insulin-requiring patients, and discussions with diabetes health-care professionals (Meadows et al., 1996).

The DHP-1 comprises 32 items covering three dimensions: psychological distress (14 items), barriers to activity (13 items), and disinhibited eating (5 items); it is suggested this last may be appropriate as a screening tool for eating problems. Each item has a four-point adjectival scale; items are summed within the three dimensions and transformed to produce a score from 0-100 where 0 represents no dysfunction.

e) Diabetes Quality of Life Measure/DQOL

The DQoL was originally developed for use in a clinical trial comparing the efficacy of two different treatment regimens on the appearance and progression of chronic complications of patients with IDDM (DCCT Research Group 1988). However its structure allows for application to other patients with IDDM and NIDDM. The developers state that the DQoL could be used in clinical settings as a screening measure to identify patients with concerns about diabetes.

The instrument has 46 core items forming four scales: satisfaction with treatment (15 items), impact of treatment (20 items), worries about future effects of diabetes (four items), and worries about social and vocational issues (seven items). The instrument also includes a generic health item that does not contribute to the scales. Adolescent and youth versions of the DQoL have been developed (Ingersoll and Marrero, 1991). The dimensions and DQoL total scores (average score across the four dimensions) are scored 0-100 where 0 represents the lowest possible quality of life and 100 the highest.

f) Diabetes Quality of Life Clinical Trials Questionnaire-Revised/DQLCTQ-R

The DQLCTQ was developed for use in multinational clinical trials of patients with IDDM and NIDDM (Shen et al., 1999). It was developed and published alongside a revised version of the instrument referred to as the DQLCTQ-R. The DQLCTQ comprises 142 items across 20 domains, three self-efficacy questions and four demographic questions. For the most part, items use five-point adjectival scales. The

DQLCTQ-R, comprises 57 items across eight domains, with between three and ten response options. Mean scores for each domain are transformed into a 100-point scale where higher scores represent better quality of life.

g) Barriers to Physical Activity in Diabetes (Type 1) (BAPADI)

Developed with input from patients, this is an 11 -item questionnaire measures perceived barriers towards the practice of regular physical activity. Responses are obtained with patients indicating the likelihood (1=extremely unlikely to 7= extremely likely) that each of the items would keep them from practicing regular physical activity.

h) Diabetes Obstacles Questionnaire (DOQ)

This instrument was developed from a review of the literature to identify obstacles to self-management in Type 2 diabetes patients. There are 77 items in 8 domains: Medications (10), Self-monitoring (5), Knowledge and beliefs (9), Diagnosis (6), Relationships with HCPs (18), Lifestyle changes (13), Coping (8), Advice and support (8). Responses are obtained on a five point Likert scale.

i) Diabetes Treatment Satisfaction Questionnaire (DTSQs, DTSQc).

This has been developed to measure satisfaction with diabetes treatment regimens with patients with Type 1 and 2 diabetes. It was referred to in the previous review. There are eight items each rated with a 7-point Likert scale. Six items relate to treatment satisfaction with scores ranging from 0 to 36, higher scores indicating greater satisfaction. Two other items assess perceived frequency of hyperglycaemia and hypoglycaemia with higher scores reflecting greater problems (Bradley & Lewis 1990). The original 'status' form (DTSQs) focuses on current satisfaction which is reflected in the response options: 6=very satisfied, 0= very dissatisfied. Ceiling effects have been reported and the developers modified the response options to improve precision. This 'change' version (DTSQc) items are equivalent to the DTSQs but response options changed to =3= more satisfied now, -3= less satisfied now with a midpoint of 0 reflecting no change.

j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)

This has been developed with involvement from in-patients and healthcare professionals with17 items measuring treatment satisfaction with an aggregated Total score (Bradley et al., 2008).

k) Diabetes Symptom Checklist-revised (DSC-R)

This 32 item PROM was originally developed in the Netherlands (1994) but has since been translated for English speaking populations. The original instrument (DSC-2) formed the basis of the revised version DSC-R). This revised version measures both the frequency and perceived discomfort of physical and psychological symptoms associated with Type 2 diabetes and it's potential complications. Responses to items are dichotomised yes/no answers. If 'yes' is selected, participants rate the perceived discomfort on a five-point scale ranging from 1 (not al all) to 5 (extremely). A Total score and domain scores are obtained. Higher scores indicate greater symptom burden.

l) Diabetes-CAT (Computerised Adaptive Testing)

The Diabetes-CAT system utilising Item Response Theory (ITR) methods was developed and allows a maximum of seven items to be administered but with stepping rules depending on the persons IRT score. Items were selected from the DIS. In comparison to a paper administration of the DIS, respondent burden was less for the Diabetes-CAT which was determined by time to complete, respondent impression of relevant items, and preference.

m) Diabetes Impact Survey (DIS)

This is a new 37-item measure that assesses the degree to which diabetes influences role function, social function, vitality and mental health.. The time frame for all items is the past four weeks with responses obtained from a 5-point Likert scale assessing frequency. The items are taken form the Global Disease Impact Survey (Quality-Metric Incorporated 2000) (Schwartz et al., 2006). English and Spanish language versions are available.

n) Insulin Treatment Satisfaction Questionnaire (ITSQ)

This was developed involving patients and literature reviews. Confirmatory factor analysis produced a 5-factor, 22-item instrument that assess Inconvenience of Regimen (IR 5 items), Lifestyle Flexibility (LF 3 items), Glycaemic Control (GC 3 items), Hypoglycaemia Control (HC 5 items) and Insulin Delivery Device Satisfaction (DS 6 items) (Anderson et al., 2004). Items are scored on a 7-point Likert scale of 0 to 100 with the higher score indicating better treatment satisfaction.

o) Diabetes Empowerment Scale (DES)

The DES assesses the patients' self-efficacy in dealing with their condition. It divides self-efficacy into three domains: Managing the psychological aspect, assessing satisfaction and readiness to change and setting and achieving diabetic goals. Questions are rated from 5 (agree) to 1 (strongly agree).

p) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)

Interviews with patients and cognitive de-briefing informed instrument development. Six items focus on concepts relevant to patient satisfaction with oral anti-diabetic medications. Each item is scored on a 5 point scale 1 = extremely dissatisfied, 5 extremely satisfied. Aggregating scores generates a total score ranging for 6 to 30 with higher scores indicating greater degree of satisfaction.

DIABETES-SPECIFIC INSTRUMENTS: Domains, items and scoring methods

Instrument (no. items)	***Domains (no. items)***	***Response options***	***Score***	***Administration/ Completion (time)***
Appraisal of Diabetes Scale/ADS	*Single index (7)* Distress Control (2 items) Uncertainty Future condition Coping Impact on life goals	5-point adjectival scales: 1 (not at all) to 5 (extremely/totally)	Scale scores summed to give an overall total 0-35	5 mins
Audit of Diabetes-Dependent Quality of Life/ADDQoL	*18 items:* Freedom to eat as I wish Enjoyment of food Family life* Working life* Sex life* Physical activity Worries about the future Holidays/leisure activities Freedom to drink as I wish Self-confidence Friendships, social life Motivation to achieve things Ease of travelling Physical appearance Finances Living conditions Unwanted dependence on others Reaction of society *1 summary item:* Present QoL	Impact: -3 (very much better without diabetes) to +3 (very much worse) Importance: 0 (not at all important) to 3 (very important) 3 items with N/A option (*)	Impact x importance = weighted score (range -9 to +9). Scores for each item summed, then divided by no. applicable items to give average weighted impact (AWI) score (i.e. N/A items do not contribute to score).	<10 mins

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration/ Completion (time)*
Diabetes 39/D-39	*39 items:* Anxiety and worry (4) Social and peer burden (5) Sexual functioning (3) Energy and mobility (15) Diabetes control (12)	7-point visual analogue scales; 1 = not affected at all, 7 = extremely affected	Scores transformed into 0-100 scores; 0 – lowest, 100 – highest possible score	*Not reported*
Diabetes Health Profile/DHP 1/18	Psychological distress (14/6) Barriers to activity (13/7) Disinhibited eating (5/5)	Four-point adjectival scales	Item scores 0-3 in each dimension summed & transformed to produce score between 0 (no dysfunction) and 100	*Not reported – probably 5-10*
Diabetes Quality of Life Measure/ DQOL	Worries - future effects of diabetes (4) Worries - social/vocational issues (7) Impact of treatment (20) Satisfaction with treatment (15)	5-point Likert scale	*No details*	15- 20 minutes
Diabetes Quality of Life Clinical Trials Questionnaire-Revised/ DQLCTQ-R	*57 items in 8 domains:* Physical function Energy/fatigue Health distress Mental health Satisfaction (DQOL) Treatment satisfaction Treatment flexibility Frequency of symptoms *1 global health question* *1 transition question*	Variety of ordinal scales, with 3 to 10 response options.	Mean scores for each domain converted to a 100-point scale	'10 mins' – probably 15-20

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration/ Completion (time)*
g) Barriers to Physical Activity in Diabetes (Type 1) BAPADI	11 -item questionnaire measures perceived barriers towards the practice of regular physical activity.	7 point Likert scale	Responses are obtained with patients indicating the likelihood (1=extremely unlikely to 7= extremely likely) that each of the items would keep them from practicing regular physical activity.	
h) Diabetes Obstacles Questionnaire (DOQ)	Medications (10), Self-monitoring (5), Knowledge and beliefs (9), Diagnosis (6), Relationships with HCPs (18), Lifestyle changes (13), Coping (8), Advice and support (8).			
i) Diabetes Treatment Satisfaction Questionnaire (DTSQs, DTSQc).	*8 items*- Treatment satisfaction(6) Perceived frequency of hyper and hypo-glycaemia (2) scores ranging from 0 to 36, higher scores indicating greater satisfaction. Two other items assess perceived frequency of hyperglycaemia and hypoglycaemia with higher scores reflecting greater problems	7-point Likert scale	DTSQs: Treatment satisfaction scores ranging from 0 to 36, higher scores indicating greater satisfaction Perceived frequency of hyperglycaemia and hypoglycaemia with higher scores reflecting greater problems DTSQc: 'change' version items are equivalent to the DTSQs but response options changed to =3= more satisfied now, -3= less satisfied now with a midpoint of 0 reflecting no change	

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration/ Completion (time)*
j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)	Treatment satisfaction (17 items)		Aggregated Total score	
k) Diabetes Symptom Checklist-revised (DSC-R)	Symptoms (32)	5 point Likert scale	Responses to items are dichotomised yes/no answers. If 'yes' is selected, participants rate the perceived discomfort on a five-point scale ranging from 1 (not al all) to 5 (extremely). A Total score and domain scores are obtained. Higher scores indicate greater symptom burden.	
l) Diabetes-CAT (Computerised Adaptive Testing)	The Diabetes-CAT system utilising Item Response Theory (ITR) methods was developed and allows a maximum of seven items to be administered but with stepping rules depending on the persons IRT score. Items were selected from the DIS. In comparison to a paper administration of the DIS, respondent burden was less for the Diabetes-CAT which was determined by time to complete, respondent impression of relevant items, and preference.			

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration/ Completion (time)*
m) Diabetes Impact Survey (DIS)	37 items Role function Social function Vitality Mental health .	5 point Likert sclae	The time frame for all items is the past four weeks with responses obtained from a 5-point Likert scale assessing frequency	
n) Insulin Treatment Satisfaction Questionnaire (ITSQ)	*22-item* / 5 domains i Inconvenience of Regimen (IR 5 items) Lifestyle Flexibility (LF 3 items) Glycaemic Control (GC 3 items) Hypoglycaemia Control (HC 5 items) Insulin Delivery Device Satisfaction (DS 6 items)	7-point Likert scale	0 to 100 with the higher score indicating better treatment satisfaction.	
0) Diabetes Empowerment Scale (DES)	It divides self-efficacy into three domains: Managing the psychological aspect, assessing satisfaction and readiness to change and setting and achieving diabetic goals.	5 point Likert scale	Questions are rated from 5 (agree) to 1 (strongly agree).	
q) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)	*Six items* Patient satisfaction with oral anti-diabetic medications.	5 point Likert scale	Each item is scored on a 5 point scale 1 = extremely dissatisfied, 5 extremely satisfied. Aggregating scores generates a total score ranging for 6 to 30 with higher scores indicating greater degree of satisfaction	

Summary of diabetes-specific instruments: health status domains

Instrument	*Instrument domains (after Fitzpatrick et al., 1998)*								
	Physical function	**Symptoms**	**Global judgment**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**	**Treatment satisfaction**
Appraisal of Diabetes Scale/ADS				X	X		X	X	
Audit of Diabetes-Dependent Quality of Life/ADDQoL	X			X	X		X	X	
Diabetes 39/ D-39	X			X	X			X	
Diabetes Health Profile/DHP 1/18				X	X		X		
Diabetes Quality of Life Measure/ DQOL		X		X	X		X		X
Diabetes Quality of Life Clinical Trials Questionnaire-Revised/DQLCTQ-R	X	X	X	X	X		X	X	X
g) Barriers to Physical Activity in Diabetes (Type 1) BAPADI	X								
h) Diabetes Obstacles Questionnaire (DOQ)				X			X	x	X

Instrument domains (after Fitzpatrick et al., 1998)									
Instrument	**Physical function**	**Symptoms**	**Global judgment**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**	**Treatment satisfaction**
j) Diabetes Treatment Satisfaction Questionnaire for Inpatients (DTSQ-IP)									X
k) Diabetes Symptom Checklist-revised (DSC-R)		X		X					
l) Diabetes-CAT (Computerised Adaptive Testing)									
m) Diabetes Impact Survey (DIS)		X		X	X		X		
n) Insulin Treatment Satisfaction Questionnaire (ITSQ)									X
0) Diabetes Empowerment Scale (DES)								X	
p) Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS)									X

APPENDIX E: Contact details

Dr. Clare Bradley (c.bradley@rhul.ac.uk)
Health Psychology Research,
Department of Psychology,
Royal Holloway, University of London,
Egham, Surrey, TW20 0EX UK

Diabetes Health Profile

Dr Keith Meadows
DHP Research & Consultancy Ltd
113 Lower Camden
Chislehurst
Kent BR75JD
Email:kmeadows@dhpresearch.com
Tel: +44 (0)2084673739

DQOL

It seems this is free and attached as an appendix to Jacobson et al. (1988).

APPENDIX F: Methods of working, membership and conclusions of multi-disciplinary panel

Members of the multi-disciplinary panel were invited to participate based on their clinical or research experience of diabetes and special interest in Patient-reported Outcome Measures.

The panel were sent the following documents:

- A structured review of patient-reported outcome measures for diabetes: An update 2009
- A structured review of patient-reported health instruments for people with diabetes (2006)
- Copies of the PROMs short-listed for discussion.

Minutes: Meeting 27th March 2009 at Nuffield College, Oxford

Present
Multidisciplinary panel: Florence Brown (FB), Andrew Farmer (AF), Carla Gianfrancesco (CG), Hilary Hearnshaw (HH), Richard Morgan (RM), Bridget Turner (BrT)

PROMs Group: Ray Fitzpatrick (RF), Elizabeth Gibbons (EG), Anne Mackintosh (AM)

E-mail contributors
John Brazier (JB), Simon Heller (SH), Mike Robling (MB), Ben Turner (BeT)

Summary of discussions

1. **Context:** RF outlined the purpose of the meeting: to formulate recommendations for the most suitable Patient-reported Outcome Measure (PROM) for routine implementation in the NHS to measure the quality and outcomes of services for people with diabetes. The structured review of PROMs for people with diabetes had previous been sent to the members and a summary was provided of the criteria used to select the instruments short-listed for discussion. It was acknowledged that there were several instruments which failed to meet the criteria for selection. Mainly, this was based either on a lack of evidence of measurement properties or because of the narrow focus of the questionnaire. Doubts were expressed as to the use that would be made of the evidence gathered. RF explained that PROMs would be used alongside other measures of quality of care and patient satisfaction.

2. **General discussion:** FB referred to the Problem Areas in Diabetes instrument had not been considered. AM referred the group to the 2006 review (Fitzpatrick et al.). This measure had been identified and discussed briefly however; it was not shortlisted given its unidimensionality (psychological well-being). The Diabetes Treatment Satisfaction Questionnaire was also referred to as being very specific to people with diabetes. This had been identified by the PROM group in the review but although, more recent evidence was emerging, there was still insufficient to qualify as a shortlisted instrument.

   BrT questioned whether a single diabetes-specific measure could capture all aspects of patient experience; diabetes is a life-long condition with different needs at different stages. It was considered that a narrowly focused instrument may be preferable. However, for DH purposes, more broad-ranging measures are required in order to measure the overall impact of the disease and interventions. There was concern that a generic health status measure was not sensitive enough to detect change and that a disease-specific measure would also be needed. RF affirmed that the intention would be do include both as well as questionnaires to measure satisfaction with care and was at least a way of having some patient input where at present there is none.

3. **Discussion of generic instruments:**

   SF-36 –
   Positives:
   - simple, easy to follow, readable
   - good responsiveness evidence
   - useful for comparison across different conditions
   - best of the generic measures, based on the review
   - utility value can be derived (SF-6D) or by predicting EQ-5D values
   - useful for identifying psychological co-morbidities

   Negatives:
   - lengthy.
   - some questions - e.g. 5, 6 re. impact of emotional problem on other activities are over-complex and may be hard for some respondents to interpret
   - diabetes-specific aspects not covered: hypoglycaemia, weight gain, self-management, confidence re. glycaemic control, self-efficacy
   - may reflect health states not related to diabetes

   General point:
   - possible overlap of items with a specific questionnaire; if combined with a lengthy specific measure, the SF-36 may be too burdensome.

   EQ-5D –
   Positives:
   - shorter – patients are more likely to complete it

   Negatives:
   - too general

- categories for responses v. broad – patients scoring in lowest range would be too ill to respond to questions 1-3; too little differentiation at upper end – diabetes patients with problems could appear very healthy
- Feeling Thermometer does not add anything useful
- problems with wording of 2nd question – 'self-care' in diabetes is more complex than the ability to wash and dress. Self-management of the condition was considered to be more important

4. **Discussion of specific instruments**

ADDQoL – developed in the UK
Positives:
- individual weighting is good
- good layout. Some very relevant questions
- It has been used in DAFNE trial and yielded very rich data but CF questioned its usefulness outside a trial context
- can be used with Type 1 and Type 2 diabetes
- performs well in psychometric terms
- appears sensitive –can be used to assess impact of interventions

Negatives:
- too complex for some
- lacks questions on experience of insulin and side effects; vision
- problems with face validity: questions ask respondents to imagine not having diabetes - patients may find this a complex task particularly if they have had the condition for many years
- lengthy
- the free text section was considered to have little value

The group was alerted to the proposed development of a short-form version referred to in the 2006 review.

DQOL – developed in the US
Positives:
- some questions were deemed to be useful such as the 'knowledge' question which has not been highlighted in other questionnaires presented.
- the domain structure was an attractive feature

Negatives:
- lengthy – but CG has encountered reference to a short form, discussed at a recent forum on diabetes outcomes and indicators in Australia
- whilst some questions are relevant they do appear conceptually difficult
- there may be cultural differences US/UK
- little evidence for responsiveness – surprising, given wide use of the measure
- wording of questions and answers discordant in some instances (e.g. Impact 6)

DHP – developed in the UK; items derived from people with diabetes
Positives:
- greater face validity than ADDQoL
- straightforward language, simply phrased

- some very relevant items e.g. disinhibited eating

Negatives:

- seems out-of-date, and more suitable to assess distress
- DHP-32 starts with question about self-harm which was considered to be far too difficult as an opening question and could impact enormously on mental health
- DHP-18 (developed for NIDDM) seems much more accessible/acceptable, although there are no items regarding insulin.
- there seems to be a focus on problem identification
- limited in its scope although the developers suggest it be used in conjunction with another diabetes-specific measure, such as PAID, DTSQ.

RF summarised a preliminary report of work carried out by JB and colleagues with patients with Type 1 diabetes. Some of the group amended their scores in light of preliminary findings from JB's patient focus group regarding the acceptability and relevance of particular measures. In particular, the shorter form DHP received favourable comment; however, no recent evidence to support the use of the DHP-18 was found in the 2009 review. RF contacted the developers following the meeting and one further publication of work in progress was identified which was published in the PRO-Newsletter 2008 (Mapi Research). Construct, discriminative and convergent validity is presented for the DHP-18 from a UK and French population.

Electronic comments

SF-36

Comments were made about the review combining the evidence of the measures in relation to both Type 1 and 2 diabetes. The review was scoped broadly to be inclusive of instruments which could be used for both types of diabetes. Concern was expressed about the item 'downhearted and low' which may be miss-interpreted to relate to hypos. The length was considered a barrier to completion.

EQ-5D

Ceiling effects were noted but in comparison to the SF-36, it was considered to be easier to complete for patients.

ADDQOL

The derivation of items from a patient's perspective was considered an attractive feature. All commented on the complexity of answering the question relation to if the patient did not have diabetes. It was though considered a useful measure for assessing the impact of interventions. Overall it was considered to perform well in psychometric terms.

DHP

Concern was expressed about the content validity of the two versions of this scale and its bias towards distress. The evidence presented also suggests lack of responsiveness.

DQOL
The content validity of the measure was considered to be good. The satisfaction with treatment scale covers quite a diverse range of items which may not necessarily directly address clinical management. There may not be sufficient delineation between life with diabetes and life in general.

The response descriptors for 'Impact' seem inappropriate given the item stem (e.g. how often do you have a bad night's sleep?). From a clinical perspective the items are all negatively framed (apart from 'feel good about yourself' which may cause response error – which may be undesirable. Many of the items for Worry: Social/vocational will be not applicable for some respondents. It was noted that most of the validation has been in clinical trials and in the US. The length was considered a potential barrier.

General discussion

It was agreed that none of the measures is ideal. All are trial-oriented, rather than appropriate for life-long monitoring, and were not designed for the purpose intended by the DH. It was also noted that mode of administration of questionnaires is particularly important for this clinical group, given problems of poor vision which affect many diabetics.

CG referred to recent publications from the Australian diabetes forum relating to outcomes for diabetes. The PROM group assessed these reports and although useful information is presented, the instruments identified were specifically focused on the evaluation of diabetes education programmes.

Overall comments suggest that there is a need for different outcome measures for different purposes. Specifically this relates to measures for Type 1 and 2 diabetes but generally different measures may be needed for evaluation of education interventions compared to assessment of the impact of new technologies.

8. Voting and discussion of results: including scores of e-mail contributors, the results were as follows:

SF-36 – 33 points
EQ-5D – 25 points
ADDQoL – 32 points
DQOL – 21 points
DHP – 20 points

There was expressed concern at the 'combination effect' of the two highest rated measures (SF-36 and ADDQoL) in that both are relatively long, with some complexity of language - although it was pointed out that repeated administrations may help to overcome this problem.

Distribution of scores
Generic total

PROM	**'not at all suitable' (score 0)**	**'to some extent unsuitable' (score 1)**	**'uncertain' (score 2)**	**'to some extent suitable' (score 3)**	**'very suitable' (score 4)**	**TOTAL**
SF-36				7 (21)	3 (12)	**33**
EQ-5D		2 (2)	2 (4)	5 (15)	1 (4)	**25**

Specific instruments

PROM	**'not at all suitable' (*score 0*)**	**'to some extent unsuitable' *(score 1)***	**'uncertain' *(score 2)***	**'to some extent suitable' *(score 3)***	**'very suitable' *(score 4)***	**TOTAL**
ADDQoL			1 (2)	6 (18)	3 (12)	**32**
DQOL		3 (3)	4 (8)	2 (6)	1 (4)	**21**
DHP	1 (0)	3 (3)	2 (4)	3 (9)	1 (4)	**20**

Patient-reported Outcome Measure Rating Scale

1. On the basis of the review of evidence and your personal experience, is this questionnaire suitable for the measurement of the quality and outcomes of services for people with diabetes? (please tick one box)

☐ Not at all suitable ☐ To some extent unsuitable ☐ Uncertain ☐ To some extent suitable ☐Very suitable

Do you have another questionnaire you could suggest?

Any additional comments

DIABETES OUTCOMES CONSENSUS GROUP MEMBERS.

Florence Brown. National Care Advisor Diabetes UK Northern Ireland

Dr Andrew Farmer. Lecturer in General Practice. Fellow of Exeter College at the University of Oxford

Carla Gianfrancesco. Specialist Diabetes Dietitian. Dietetic Dept, Northern General Hospital

Hilary Hearnshaw. Associate Professor of Primary Care, University of Warwick

Richard Morgan. Patient representative

Bridget Turner. Head of Healthcare Policy. Diabetes UK

Members unable to attend the meeting.

These people contributed by email

John Brazier. Professor of Health Economics. School of Health and Related Research, University of Sheffield

Simon Heller. Professor of Clinical Diabetes Academic Unit of Diabetes, Endocrinology and Metabolism. School of Medicine and Biomedical Sciences, Sheffield

Dr Ben Turner. Diabetic specialist. Dept. of Diabetes & Endocrinology. Basingstoke & North Hampshire NHS

Dr Mike Robling. Associate Director, South East Wales Trials Unit (SEWTU)
School of Medicine Cardiff University

PROM group

Ray Fitzpatrick
Professor of Public Health and Primary Care, Deputy Head, Department of Public Health, University of Oxford.

Elizabeth Gibbons
Senior Research Officer, Department of Public Health, University of Oxford.

REFERENCES

Anderson JP, Kaplan RM, Berry CC, Bush JW, Rumbaut RG. Interday reliability of function assessment for a health status measure: the Quality of Well-Being scale. *Medical Care* 1989;**27**:1076-84.

Anderson RM, Fitzgerald JT, Wisdom K, Davis WK, Hiss RG. A comparison of global versus disease-specific quality-of-life measures in patients with NIDDM. *Diabetes Care* 1997;**20**:299-305.

Ashwell SG, Bradley C, Stephens JW, Witthaus E, Home PD. Treatment satisfaction and quality of life with insulin glargine plus insulin lispro compared with NPH insulin plus unmodified human insulin in individuals with type 1 diabetes. *Diabetes Care* 2008;31:1112-7.

Bergner M, Bobbitt RA, Kressel S, et al. (1976) The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Service;* 6: 393-415.

Boyer JG, Earp JA. The development of an instrument for assessing the quality of life of people with diabetes: Diabetes-39. *Medical Care* 1997;**35**:440-53.

Bradley C. Design of a renal-dependent individualised quality of life questionnaire. *Advances in Peritoneal Dialysis* 1997;**13**:116-20.

Bradley C, Lewis KS. Measures of psychological well-being and treatment satisfaction developed from the responses of people with tablet-treated diabetes. *Diabetic Medicine* 1990;**7** :445-51.

Bradley C and Lewis KS (1990) Measures of psychological well-being and treatment satisfaction developed from the responses of people with tablet-treated diabetes. *Diabetic Medicine* **7**, 445-451.

Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. *Quality of Life Research* 1999;**8**:79-91.

Bradley C,.Speight J. Patient perceptions of diabetes and diabetes therapy: assessing quality of life. *Diabetes and Metabolism Research and Reviews* 2002;**18**:S64-S69.

Brod MI, Skovlund SE, Wittrup-Jensen KU. Measuring the impact of diabetes through patient report of treatment satisfaction, productivity and symptom experience. *Quality of Life Research* 2006;15 :481-91.

Court JM, Cameron FJ, Berg-Kelly K, Swift PGF. Diabetes in adolescence. *Pediatric Diabetes* 2008: 9 (Part I): 255–262.

Currie CJ, Poole CD, Woehl A, Morgan CL, Cawley S, Rousculp MD *et al*. The financial costs of healthcare treatment for people with Type 1 or Type 2 diabetes in the UK with particular reference to differing severity of peripheral neuropathy. *Diabet.Med.* 2007;24:187-94.

Darzi A. Our NHS Our Future: NHS Next Stage Review Interim Report. Department of Health, London, October 2007.

Darzi A. Our NHS Our Future: High Quality Care for All. NHS Next Stage Review Final Report. Department of Health, London, June 2008.

De Bruin AF, De Witte LP, Stvene F, Diederiks JPM. (1992) Sickness Impact Profile: the state of art of a generic functional status measure. *Social Science and Medicine;* 35: 1003-14.

Delamater AM. Psychological care of children and adolescents with diabetes. *Pediatric Diabetes* 2007: 8: 340–348.

Department of Health (2008) Guidance on the routine collection of Patient Reported Outcome Measures (PROMs)

Donatti C, Wild D, Horblyuk R, Beusterien KM, Anderson RT, Arondekar B *et al*. Psychometric evaluation of the Satisfaction with Oral Anti-Diabetic Agent Scale (SOADAS). *Diabetes Res.Clin.Pract.* 2008;80:108-13.

Dube MC, Valois P, Prud'homme D, Weisnagel SJ, Lavoie C. Physical activity barriers in diabetes: development and validation of a new scale. *Diabetes Research & Clinical Practice* 2006;72:20-7.

Feeny, David, William Furlong, Michael Boyle, and George W. Torrance, "Multi-Attribute Health Status Classification Systems: Health Utilities Index." *PharmacoEconomics*, Vol 7, No 6, June, 1995, pp 490-502

Gardner Stephen G, Bingley Polly J, Sawtell Pamela A, Weeks Suzanne, Gale Edwin AM. Rising incidence of insulin dependent diabetes in children aged under 5 years in the Oxford region: time trend analysis. *BMJ* 1997;315:713-717

Garratt AM, Schmidt LJ, Fitzpatrick R. Patient-assessed health outcome measures for diabetes: a structured review. *Diabetic Medicine* 2002;**19**:1-11.

Grandy S, Chapman RH, Fox KM. Quality of life and depression of people living with type 2 diabetes mellitus and those at low and high risk for type 2 diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). *Int.J.Clin.Pract.* 2008;62:562-8.

Grandy S,.Fox KM. EQ-5D visual analog scale and utility index values in individuals with diabetes and at risk for diabetes: Findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). *Health Qual.Life Outcomes.* 2008;6:18.

Hammen C, Mayol A. Depression and cognitive characteristics of stressful life-event types. *Journal of Abnormal Psychology* 1982;**91**:165-74.

Hearnshaw H, Wright K, Dale J, Sturt J, Vermeire E, van Royen P. Development and validation of the Diabetes Obstacles Questionnaire (DOQ) to assess obstacles in living with Type 2 diabetes. *Diabet.Med.* 2007;24:878-82.

Ingersoll GM, Marrero DG. A modified quality-of-life measure for youths: psychometric properties. *The Diabetes Educator* 1991;**17**:114-8.

Kaplan RM, Bush JW, Berry CC. (1976) Health status: Types of validity and the index of well-being. *Health Service Research;* 11:478.

Kaplan RM, Anderson JP. (1988) A General Health Policy Model: update and applications. *Health Service Research;* 23: 203-235.

Kaplan RM, Anderson JP, Ganiats TJ. (1993) The Quality of Well-Being Scale: rationale for a single quality of life index. In: Walker SR, Rosser RM (Eds). Quality of Life Assessment: key issues in the 1990s. Dordrecht, Netherlands: Kluwer Academic Publishers: 65-94.

Kelley K,.Dempsey C. An evaluation of an insulin transfer programme delivered in a group setting. *J.Clin.Nurs.* 2007;16:152-8.

Krass I, Armour CL, Mitchell B, Brillant M, Dienaar R, Hughes J *et al*. The Pharmacy Diabetes Care Program: assessment of a community pharmacy diabetes service model in Australia. *Diabet.Med.* 2007;24:677-83.

Lange LJ,.Piette JD. Personal models for diabetes in context and patients' health status. *Journal of Behavioral Medicine* 2006;29:239-53.

Lowe J, Linjawi S, Mensch M, James K, Attia J. Flexible eating and flexible insulin dosing in patients with diabetes: Results of an intensive self-management course. *Diabetes Res.Clin.Pract.* 2008;80:439-43.

Maddigan SL, Feeny DH, Majumdar SR, Farris KB, Johnson JA. Health Utilities Index mark 3 demonstrated construct validity in a population-based sample with type 2 diabetes. *Journal of Clinical Epidemiology* 2006;59:472-7.

Matza LS, Boye KS, Yurgin N. Validation of two generic patient-reported outcome measures in patients with type 2 diabetes. *Health Qual.Life Outcomes.* 2007;5:47.

Matza LS, Yurgin N, Boye KS, Malley K, Shorr JM. Obese versus non-obese patients with type 2 diabetes: patient-reported outcomes and utility of weight change. *Curr.Med.Res.Opin.* 2007;23:2051-62.

Matza LS, Boye KS, Yurgin N, Brewster-Jordan J, Mannix S, Shorr JM *et al*. Utilities and disutilities for type 2 diabetes treatment-related attributes. *Qual.Life Res.* 2007;16:1251-65.

McCollum M, Ellis SL, Regensteiner JG, Zhang W, Sullivan PW. Minor depression and health status among US adults with diabetes mellitus. *American Journal of Managed Care* 2007;13:65-72.

McDowell I, Jenkinson C. (1996) Development standards for health measures. *Journal of Health Service Research Policy;* October. 1(4): 238-246.

McDowell I, Newell C. (1996) Measuring Health: a guide to rating scales and questionnaires. Oxford University Press, New York.

McDowell I, Jenkinson C. (1996) Development standards for health measures. *Journal of Health Service Research Policy;* October. 1(4): 238-246.

McDowell I, Newell C. (2006) Measuring Health, 3rd ed. Oxford: Oxford University Press.

McDowell Nelson EC, Wasson J, Kirk J, et al. (1987) Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. *Journal of Chronic Disease;* 40(suppl 1): 55S-63S.

McGill JB, Bakris GL, Fonseca V, Raskin P, Messerli FH, Phillips RA *et al*. beta-blocker use and diabetes symptom score: results from the GEMINI study. *Diabetes Obes.Metab* 2007;9:408-17.

Meadows KA, Steen N, McColl E, Eccles MP, Shiels C, Hewison J *et al*. The Diabetes Health Profile (DHP), a new instrument for assessing the psychosocial profile of insulin requiring patients: development and psychometric evaluation. *Quality of Life Research* 1996;**5**:242-54.

Meadows KA, Abrams C, Sandbaek A. 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* 2000;**17**:572-80.

Moon RJ, Bascombe LA, Holt RI. The addition of metformin in type 1 diabetes improves insulin sensitivity, diabetic control, body composition and patient well-being. *Diabetes, Obesity & Metabolism* 2007;9:143-5.

Poggioli R, Faradji RN, Ponte G, Betancourt A, Messinger S, Baidal DA *et al*. Quality of life after islet transplantation. *American Journal of Transplantation* 2006;6:371-8.

Resnick B, Parker R. (2001) Simplified scoring and psychometrics of the revised 12-item Short-Form Health Survey. *Outcomes Management for Nursing Practice;* Oct-Dec. 5(4): 161-6.

Rosenbloom AL, Silverstein JH, Amemiya S, Zeitler P, Klingensmith GJ. Type 2 diabetes in the child and adolescent. *Pediatric Diabetes* 2008: 9: 512–526.

Schwartz C, Welch G, Santiago-Kelley P, Bode R, Sun X. Computerized adaptive testing of diabetes impact: a feasibility study of Hispanics and non-Hispanics in an active clinic population. *Quality of Life Research* 2006;15:1503-18.

Secnik BK, Matza LS, Oglesby A, Malley K, Kim S, Hayes RP *et al*. Patient-reported outcomes in a trial of exenatide and insulin glargine for the treatment of type 2 diabetes. *Health & Quality of Life Outcomes* 2006;4:80, 2006.:80, 2006.

Schwartz CE, Genderson MW, Kaplan RM, Anderson JP, Holbrook T. Covariation of physical and mental symptoms across illnesses: results of a factor analytic study. *Annals of Behavioral Medicine* 1999;**21**:122-7.

Shen W, Kotsanos JG, Huster WJ, Mathias SD, Andrejasich CM, Patrick DL. Development and validation of the Diabetes Quality of Life Clinical Trial Questionnaire. *Medical Care* 1999;**37**:AS45-AS66.

Sinclair AJ. Towards a minimum data set for intervention studies in type 2 diabetes in older people. *J.Nutr.Health Aging* 2007;11:289-93.

Sinclair AJ, Conroy SP, Bayer AJ. Impact of diabetes on physical function in older people. *Diabetes Care* 2008;31:233-5.

Skinner TC, Hoey H, McGee HM, Skovlund SE, Hvidore Study Group on Childhood Diabetes. A short form of the Diabetes Quality of Life for Youth questionnaire: exploratory and confirmatory analysis in a sample of 2,077 young people with type 1 diabetes mellitus. *Diabetologia* 2006;49:621-8.

Speight J, Bradley C. ADDQol indicates negative impact of diabetes on quality of life despite high levels of satisfaction with treatment. *Diabetologia* 2000;**43**:A225 (Abstract 864).

Sundaram M, Kavookjian J, Patrick JH, Miller LA, Madhavan SS, Scott VG. Quality of life, health status and clinical outcomes in Type 2 diabetes patients. *Quality of Life Research* 2007;16:165-77.

Tahbaz F, Kreis I, Calvert D. An audit of diabetes control, dietary management and quality of life in adults with type 1 diabetes mellitus, and a comparison with nondiabetic subjects. *Journal of Human Nutrition and Dietetics* 2006;19:3-11.

Tapp RJ, Dunstan DW, Phillips P, Tonkin A, Zimmet PZ, Shaw JE. Association between impaired glucose metabolism and quality of life: results from the Australian diabetes obesity and lifestyle study. *Diabetes Res.Clin.Pract.* 2006;74:154-61.

Vinik AI,.Zhang Q. Adding insulin glargine versus rosiglitazone: health-related quality-of-life impact in type 2 diabetes.[erratum appears in Diabetes Care. 2007 Jun;30(6):1684]. *Diabetes Care* 2007;30:795-800.

Ware JE, Sherbourne CD, Davies AR. (1992) Developing and testing the MOS 20-Item Short-Form Health Survey: A general population application. In: Stewart AL,

Ware JE. (editors). Measuring functioning and well-being: the Medical Outcomes Study approach (pp.277-290). Durham, NC: Duke University Press.

Ware JE, Sherbourne CD. (1992) The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care;* 30: 473-483.

Ware J, Kosinski M, Keller SD. (1994) SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: The Health Institute, New England Medical Centre.

Ware JE, Kosinski M, Keller SD. (1995) SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. The Health Institute, New England Medical Center. Boston, MA. Second Edition.

Ware JE. (1997) SF-36 Health Survey. Manual and Interpretation Guide. The Health Institute, New England Medical Centre. Boston, MA. Nimrod Press. Second Edition.

Ware JE, Gandek B. (1998) Methods for testing data quality, scaling assumptions and reliability: the IQOLA project approach. *Journal of Clinical Epidemiology;* 51(11): 945-952.