PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES (PROMs) FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

Report to the Department of Health
2009

Patient-Reported

OUTCOME

Measures

A STRUCTURED REVIEW OF PATIENT--REPORTED OUTCOME MEASURES FOR COPD: AN UPDATE 2009.

Nicola Davies
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 CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD): An update 2009

CONTENTS

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES IN RELATION TO COPD, 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 Chronic Obstructive Pulmonary Disease (COPD) and to provide recommendations to the Department of Health of PROMs for COPD 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 COPD-specific instruments which were then presented to a multi-disciplinary 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 COPD with some recommendations[1].

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 COPD, followed by COPD-specific PROM results. A number of newly identified COPD-specific measures are also briefly discussed. The report concludes with discussion and recommendations.

Results and short-list of PROMs for people with COPD

The previous review reported evidence for the following PROMs: SF-36, SF-20, SF-12, EQ-5D, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), and the COOP Charts. This update identified further evidence of performance for the SF-36, EQ-5D and the COOP Charts. An additional generic PROM, the Health Utilities Index-3 (HUI-3), was identified with reporting of some evidence of performance.

Five condition-specific instruments were previously identified:

a) Breathing Problems Questionnaire (BPQ)
b) Chronic Respiratory Disease Questionnaire (CRQ)
c) Functional Performance Inventory (FPI)
d) St George's Respiratory Questionnaire (SGRQ)
e) Seattle Obstructive Lung Disease Questionnaire (SOLQ)

In carrying out the current update of evidence, two newer versions of previously recommended condition-specific PROMs were identified:

f) Short-Form Chronic Respiratory Disease Questionnaire (SF-CRQ)

[1] A Structured review of PROMs for COPD 2006, can be downloaded from http://phi.uhce.ox.ac.uk/

g) COPD-Specific St George's Respiratory Questionnaire (SGRQ-C)

Short-listed PROMs for discussion
Based on volume of evaluations and good measurement and operational characteristics, the following PROMs were presented to the multidisciplinary panel for discussion:

1. CRQ
2. SF-CRQ
3. SGRQ
4. SGRQ-C
5. SF-36
6. EQ-5D

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 COPD. 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 SF-CRQ is recommended as a COPD- specific instrument. These two measures used together will provide complementary evidence of health status of people with COPD in the context of potential population-level applications in the NHS.

Further evidence has emerged since this review in 2009. The findings indicate more favourable evidence for the Clinical COPD Questionnaire (CCQ). Although the research evidence is predominantly with non-English populations there is considerable support for its clinical utility in the UK and is highlighted as a potentially useful questionnaire in the Quality Outcomes Framework.

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. PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH COPD: 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' experience and views. Guidance has now been issued regarding the routine collection of PROMs for the selected elective procedures (DH, 2009).

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). The report included a review of evidence regarding PROMs for COPD with some recommendations.

The aim of this report is two-fold: to provide an update of more recently published evidence for PROMs in COPD 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 COPD 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.

For full details of measurement and operational performance for each PROM for COPD identified in the earlier review, the reader should refer to Section II. 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.

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.

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 COPD, followed by COPD-specific PROM results. A number of newly identified COPD-specific measures are also briefly discussed. The report concludes with discussion and recommendations for short-listed PROMs.

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., 2005). 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 'respiratory' and further searches conducted using title OR abstract using terms 'chronic obstructive pulmonary disease' OR 'COPD.'

A further 14,296 records covering the period January 2006-July 2007 had been downloaded using a revised search strategy (Appendix 2, version b) but not assessed or keyworded. The terms 'respiratory' OR 'chronic obstructive pulmonary disease' OR 'COPD' from title OR abstract were performed.

Supplementary searches included scanning the reference lists of key articles, checking instrument websites, where found, and drawing on other bibliographic resources. Hand searching of titles of key journal was conducted from October 2006 to October 2008. The following journals were selected:

- Chest
- Health and Quality of Life Outcomes
- Medical Care

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

- Quality of Life Research
- Respiratory Medicine
- Thorax
- Journal of COPD

In addition, PubMed records for the past two years (i.e. September 2006-2009) were searched using the terms 'chronic obstructive pulmonary disease' and 'COPD' 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, 34 articles were included in the updated review. The recommendations are based on an assessment of the evidence from both the previous review and the update review combined. The previous review identified 46 articles. Results are presented in Table 1.

Identification of patient-reported outcome measures

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: original search (up to December 2005) **Total number = 16,054**	899	192	7
Additional PROM database search (January 2006-July 2007) **Total number = 14,296**	284	126	15
Supplementary searching	-	-	12
TOTAL	-	-	**34**
2006 review			**46**
TOTAL			**80**

2. Results: Generic PROMs evaluated with people with COPD

The previous review reported evidence for the following PROMs:

a) SF-36
b) SF-20
c) SF-12
d) EQ-5D
e) Nottingham Health Profile (NHP)
f) COOP Charts
g) Sickness Impact Profile (SIP)

This update identified further evidence of performance for the SF-36, EQ-5D and the COOP Charts. An additional generic PROM, the Health Utilities Index-3 (HUI-3), was identified with reporting of some evidence of performance.

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

a) SF-36

Eleven studies describe the evaluation of the SF-36 in the previous review with all measurement and operational criteria reported. Ten further studies were identified in this update, offering further evidence on the construct validity, responsiveness, and interpretability of the SF-36.

Convergent validity has been supported with comparative domains of another generic instrument, the COOP (Eaton et al., 2005) and the Dyadic Adjustment Scale, as was expected (Ashmore et al., 2005). However, there was weak correlation between the SF-36 and respiratory function measured by fifteen-count breathlessness scores (Williams et al., 2006), as well as between the SF-36 and SGRQ (Coultas et al., 2005).

The SF-36 PCS and PF, RE and GHP domains could distinguish between patients with and without prior rehabilitation experience in patients with severe emphysema (Ries et al., 2005), but has failed to discriminate according to degree of breathlessness (Williams et al., 2006).

The RF, RE, SF, MH, GH, and V subscales of SF-36 have been found to be responsive to change in patients participating in a rehabilitation programme, consistent with changes in CES-D and physiological outcomes (Ashmore et al., 2005). Furthermore, the responsiveness is supported with score changes similar to the Fatigue, Emotions and Mastery domains of the condition-specific Chronic Respiratory Questionnaire as well as to the Activities and Impact domains of the St George's Respiratory Questionnaire in patients receiving an 8-week respiratory rehabilitation intervention. Responsiveness of the PF, GH, and V subscales has been demonstrated with chronically ill COPD patients receiving a palliative care intervention (Aiken et al., 2006). Further responsiveness has been demonstrated for RP, V, and PCS with patients involved in a dyspnoea self-management programme, consistent with changes in physiological outcomes (Carrieri et al., 2005).

In intervention studies, the PCS was found to be more responsive than condition-specific instruments (Puhan et al., 2007). Statistically significant improvements were evidenced for

all scales except P with severe emphysema patients undergoing pulmonary rehabilitation (Ries et al., 2005). The RP domain demonstrated responsiveness to dyspnoea self-management (Carrieri et al., 2005) and V to exercise (Brown et al., 2008). The only significant improvement to scores in patients involved in a nursing intervention was for RP (Coultas et al., 2005), although the authors note the insufficient power and potential biases in their study.

When comparing patient and primary care physicians' perceived clinically minimal important changes in SF-36 Total and subscale scores, divergence was so great that the authors rendered it difficult to interpret the trends in many of the SF-36 scale results (Wyrwich et al., 2005).

b) SF-20

One study provided evidence of the measurement properties of the SF-20 in the previous review. No further studies providing evidence of the measurement properties of this instrument were identified in this update.

c) SF-12

One study provided evidence of the measurement properties of the SF-12 in the previous review. No further studies providing evidence of the measurement properties of this instrument were identified in this update.

d) EQ-5D

Three UK studies describe the evaluation of the EQ-5D in the previous review providing some good evidence of validity, responsiveness and acceptability. One further study was identified in this update, offering further evidence on the construct validity and acceptability of the EQ-5D.

The EQ-5D demonstrated convergent validity with the MRC Dyspnoea Scale in both primary and specialist care settings within the UK and USA and across five EU countries (Punekar et al., 2007).

The Mobility, Self-Care, and Usual Activities domains demonstrated small but significant responsiveness to specialist care when compared to usual care, whilst scores on Pain/Discomfort and Anxiety/Depression did not significantly change. (Punekar et al., 2007).

Out of 2,933 patients, 92% returned the EQ-5D with no missing data, supporting the acceptability of this instrument (Punekar et al., 2007).

e) Nottingham Health Profile

One study providing weak evidence of the measurement properties of the NHP in the previous review. No further studies providing evidence of the measurement properties of this instrument were identified in this update.

f) COOP Charts

Evidence for the COOP Charts was evaluated in one study in the previous review, providing limited evidence of reliability, validity and responsiveness. One further study identified in this update, conducted by the same author of the previous evaluation (Eaton et al., 2005), offers limited evidence on the validity, responsiveness, and acceptability of the COOP Charts.

Reproducibility for the COOP domains following a two-month non-intervention period post 6-week rehabilitation were unacceptable, all having an ICC lower than 0.70 (Eaton et al., 2005).

Comparative domains of the COOP demonstrated highly significant convergent validity with the SF-36. The Feelings domain of the COOP was also highly significantly convergent with CRQ Emotional Function and the Anxiety and Depression domains of the Hospital Anxiety and Depression Scale. The COOP Physical Fitness and Daily Activities converged with the CRQ Dyspnoea and the COOP Social Support and Feelings with the CRQ Mastery (Eaton et al., 2005).

The COOP Charts demonstrated discriminate validity according to type of ambulatory oxygen patients were receiving (Eaton et al., 2005).

Social Activities, Feelings, and Health domains of the COOP were responsive to change in patients receiving both cylinder oxygen and air interventions (Eaton et al., 2005). Eaton et al. (2005) used effect sizes above 0.2 as an indicator of significant responsiveness.

Acceptability is supported in a study by Eaton et al. (2005) reporting that patients with significant COPD did not require any assistance completing the COOP questionnaires and that no problems were encountered with missing values.

g) Sickness Impact Profile

Two studies provided weak evidence of the measurement properties of the SIP in the previous review. No further studies providing evidence of the measurement properties of this instrument were identified in this update.

h) HUI-3

No evidence was identified in the previous review for the HUI-3.

Two evaluative studies were identified from this update, offering weak evidence for the responsiveness of the HUI-3.

The SRM for the HUI-3 was 0.20 compared to 0.51-0.84 for condition-specific instruments when administered to patients after an 8-week respiratory rehabilitation intervention (Puhan et al., 2007).

The HUI was responsive to changes in patients with advanced emphysema following Lung Volume Reduction Surgery (Miller et al., 2006).

3. Results: Condition-specific PROMs

Five condition-specific instruments were included in the previous review:

a) Breathing Problems Questionnaire (BPQ)
b) Chronic Respiratory Disease Questionnaire (CRQ)
c) Functional Performance Inventory (FPI)
d) St. George's Respiratory Questionnaire (SGRQ)
e) Seattle Obstructive Lung Disease Questionnaire (SOLQ)

This update identified further evidence of performance for the BPQ, CRQ, and the SGRQ. Two newer versions of the CRQ and SGRQ were identified this update: a short-form of the CRQ and a COPD-specific SGRQ.

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

a) Breathing Problems Questionnaire (BPQ)

Two studies were identified which evaluated the BPQ in the previous review (2006), both of which were conducted in the UK (Hyland et al., 1994; Yohannes et al., 1998) and reported evidence of good measurement performance but offered little support for operational characteristics. A further two studies have been identified in this update, one of which was conducted in the UK (Yohannes et al., 2005).

Significant correlations have been demonstrated between the BPQ and Manchester Respiratory Activities of Daily Living (MRADL) score (Yohannes et al., 2005), supporting construct validity.

Discriminant validity has not been established in terms of length of supervised pulmonary rehabilitation (Sewell et al., 2006). However, in a UK study, BPQ scores were predictive of mortality in older patients hospitalised for acute exacerbation COPD (Yohannes et al., 2005).

Scores on the BPQ changed significantly at both four and seven weeks into pulmonary rehabilitation, but this responsiveness was not evident at 6 months post-rehabilitation (Sewell et al., 2006).

b) Chronic Respiratory Disease Questionnaire (CRQ)

Eighteen studies evaluating the CRQ, five of which were conducted in the UK (Brightling et al., 2001; Harper et al., 1997; Singh et al., 2001; Williams et al., 2001, 2006), were identified in the previous review (2006). A further eleven studies have been identified in this update, offering further evidence on the reliability, validity, responsiveness, interpretability, precision, and operational characteristics of the CRQ.

One of the studies reviewed substantiated the established high internal consistency of the CRQ, with Cronbach's alpha being above 0.80 (Kapella et al., 2006).

The Emotional Function domain of the CRQ was highly significantly correlated with the Feelings domain of the generic COOP in patients receiving ambulatory oxygen (Eaton et al.,

2005). Furthermore, CRQ Dyspnoea was significantly correlated with both the COOP Physical Fitness and Daily Activities. CRQ Mastery was significantly correlated with both COOP Social Support and Feelings. In another study, the CRQ Dyspnoea correlated with the Profile of Mood States Fatigue domain (Kapella et al., 2006).

Convergent validity with physiological measures was demonstrated with FEV^1 for Mastery but not Dyspnoea, Fatigue or Emotion domains, with VO^2peak for Mastery and Fatigue, but not Dyspnoea and Emotion domains (Berry et al., 2006).

The CRQ has failed to demonstrate discriminant validity in terms of length of supervised pulmonary rehabilitation (Sewell et al., 2006), type of exercise programme (Sewell et al., 2005), severity of exacerbations (Carr et al., 2007), or type of post-rehabilitation care (Gupta et al., 2006). However, in the former study, the Mastery and Dyspnoea domains demonstrated discriminant validity for frequency of rollater use.

Scores on the Fatigue and Emotional subscales did not change significantly in patients undergoing Dyspnoea self-management (Carrieri et al., 2005), demonstrating weak responsiveness. All domains were responsive to change in patients receiving an 8-week rehabilitation programme (Eaton et al., 2006) and a 7-week exercise programme (Sewell et al., 2005). The Dyspnoea, Emotional Function, and Fatigue domains have demonstrated responsiveness in patients receiving a 4-week pulmonary rehabilitation programme from baseline to 4-weeks but no significant difference from baseline to 6 months post-rehabilitation (Sewell et al., 2006).

In a concurrent evaluation of the self-administered (CRQ-SA) and interview-administered (CRQ-IA) versions of the CRQ, SRMs were found to be significantly higher for the CRQ-SA than the CRQ-IA in patients receiving an 8-week respiratory rehabilitation intervention (Puhan et al., 2007). This may suggest the self-administered version was more responsive that the interview-administered version.

The majority of studies used an MCID of 0.5 per domain (Sewell et al., 2006; Gupta et al., 2006; Carr et al., 2007; Eaton et al., 2005), whilst Eaton et al. (2006) and Carrieri et al. (2005) used an MCID in total CRQ score of $\geq$10. Wyrwich et al. (2005) demonstrated the CRQ was able to detect small changes at levels reported by patients (1-2 points) and their primary care physicians (1-5 points), confirming minimal important difference standards developed by Jaeschke et al. (1989) anchored on patient-perceived changes in health-related QoL. In general, patients tended to report smaller changes as more clinically meaningful than do physicians, indicating potential interpretation problems.

Short-Form - Chronic Respiratory Disease Questionnaire (SF-CRQ)

The SF-CRQ is an 8-item version of the original 20-item CRQ. The SF-CRQ maintains the original four domains of Dyspnoea (2-items), Fatigue (2-items), Emotional Functioning (2-items), and Mastery (2-items) as well as the seven-point Likert scale reflecting level of impairment; higher scores reflect no impairment. Item selection was based on previous research (Moran et al., 2001) and consultation with the developer of the original CRQ, Dr Guyatt. The instrument was pilot tested with consecutive emergency department patients with COPD (n = 301).

No evidence was identified in the previous review for the SF-CRQ. One study providing evidence of the psychometric properties of the SF-CRQ was identified in this update.

High internal consistency was supported, Tsai et al. (2008) reporting a Cronbach's alpha above 0.80 (Tsai et al., 2008).

The SF-CRQ has demonstrated significant convergent validity with the dimension-specific instruments Symptom and Activity Scale (SAS) and VAS Dyspnoea (Tsai et al., 2008).

The SF-CRQ has demonstrated discriminant validity between patients who reported having 'improved' or 'worsened/unchanged' over a 2-week period post-emergency department visit for acute exacerbations (Tsai et al., 2008). Discriminant validity was further supported by a significant difference in the Mastery domain of the SF-CRQ between the relapse and non-relapse group (Tsai et al., 2008).

The SF-CRQ has demonstrated significant responsiveness with effect sizes being large in Dyspnoea and Mastery scores and moderate in Emotional function and Fatigue domains to post-emergency department visits for acute exacerbations at 2-weeks (Tsai et al., 2008).

There is some evidence of lower responsiveness of the SF-CRQ in relation to worsening COPD. The authors of the study speculate without direct evidence that this may be due to floor effects (Tsai et al., 2008). No data is supplied pertaining to possible floor effects.

The SF-CRQ demonstrated low respondent burden in a cohort of COPD patients aged 62-75 years, taking less than 5 minutes to complete (Tsai et al., 2008).

The SF-CRQ has been found to be practical, simple, and easy to administer (Tsai et al., 2008).

c) Functional Performance Inventory (FPI)

Three USA studies provided adequate evidence of the measurement properties of the FPI in the previous review. No further studies providing evidence of the measurement properties of this instrument were identified in this update.

d) St. George's Respiratory Questionnaire (SGRQ)

Ten studies evaluating the measurement and operational properties of the SGRQ were identified in the previous review, five of which were conducted in the UK (Jones et al., 1991; Okubadejo et al., 1996; Singh et al., 2001; Wilson et al., 1997). A further eighteen studies have been identified in this update, offering further evidence of the validity, responsiveness, interpretability, and operational characteristics of the SGRQ.

The SGRQ has demonstrated significant convergent validity with the Shortness of Breath Questionnaire (SOBQ) (Brown et al., 2008) and a number of physiological measures (Johnson et al., 2007). Furthermore, the SGRQ scores from the original development study (Jones et al., 1992) and scores from the SGRQ-C, demonstrated convergent validity (Meguro et al., 2007). The SGRQ was found to converge with other measures of both respiratory function, such as FEV, FVC, 6 minute walking test (6MWD), MRC Dyspnoea grade, and other patient-reported measures, the Hospital Anxiety and Depressions Scale (HADS),

Sickness Impact Profile (SIP), Global Health, Cough/phlegm, and Daily Wheeze. Convergence was not significant between Symptoms and FVC on the SGRQ but was significant for the Impacts component.

The SGRQ Total score demonstrated discriminant validity in terms of levels of Physical Activity, and the Total and sub-component scores could discriminate according to COPD severity as measured via spirometry (McGlone et al., 2006). Discriminant validity has also been demonstrated between population norms and COPD patients; symptomatic and asymptomatic patients with stage 0 and I COPD, as well as between different stages of COPD, for the Symptom domain scores, Impact domain scores (except for stage I), and Total SGRQ (except for stage I) (Maleki-Yazdi et al., 2007). Although the Activity domain scores could not discriminate stages 0 and I COPD, they could discriminate stages 0 and 1 from stages II and III. Yeo et al. (2006) found that the SGRQ could not discriminate COPD severity, but could discriminate secondary care attendance and those with co-morbidities. SGRQ scores have also demonstrated discriminant validity in terms of hospitalisation frequency and exacerbations (Fan et al., 2007), type of nursing care (Coultas et al., 2005), and treatment type (Gross et al., 2008; Roth et al., 2006; Ferguson et al., 2008, Tashkin et al., 2008), except for in one study examining different medications (Donohue et al., 2006). The Total score and Activity and Impacts scores could distinguish between patients with and without prior rehabilitation experience in patients with severe emphysema (Ries et al., 2005).

Discriminant validity was not established for the SGRQ in terms of palliative care preferences (Stapleton et al., 2005), use of distraction during rehabilitation (Bauldoff et al., 2005), or according to the fifteen-count breathlessness score (Williams et al., 2006).

No significant difference was found in SGRQ scores between exercise rehabilitation patients who participated in additional non-walking exercise and those who did not (McGlone et al., 2006), but this was consistent with physiological measures such as lung function. The responsiveness of the SGRQ in patients receiving an 8-week respiratory rehabilitation intervention was found to be significantly larger than similar domains on generic and preference-based instruments (Puhan et al., 2007). Further support for the responsiveness of the SGRQ has been evidenced for pulmonary rehabilitation in patients with severe emphysema (Ries et al., 2005). Activity, Impacts, and Total SGRQ responsiveness to usual care has been supported, as has the responsiveness of the Activity domain in patients receiving nurse-assisted medical management (Coultas et al., 2005). Significant changes in SGRQ scores have been evidenced in a number of medication efficacy studies, demonstrating strong responsiveness (Roth et al., 2006; Gross et al., 2008; Jones and Bosh, 1997; Ferguson et al., 2008; Tashkin et al., 2008).

SGRQ scores have been compared to population means to facilitate interpretability (McGlone et al., 2006). An MCID of 4 units is the most commonly utilised measure of interpretability (Ferguson et al., 2008; Gross et al., 2008; Coultas et al., 2005; Bauldoff et al., 2005; Ries et al., 2005), as recommended by the instrument authors (Jones et al., 1992).

A 9% refusal rate has been reported in a pilot study examining early detection of COPD (n = 244) (Maleki-Yazdi et al., 2007).

Time and cost restraints of administering the SGRQ in busy primary care practices have been reported in one study (Maleki-Yazdi et al., 2007).

St. George's Respiratory Questionnaire – COPD Version (SGRQ-C)

The SGRQ-C is 40-item version of the original 50-item SGRQ and was derived from the original version following detailed analysis of data from large studies in COPD. The intention was to remove the items with the weakest measurement properties in the original instrument, but at the same time ensure that its scores were directly comparable with the original SGRQ (Meguro et al., 2007). The instrument consists of two parts: Part I produces the Symptoms score, and Part 2 the Activity and Impacts scores. A Total score is also produced.

No evidence was identified in the previous review for the SGRQ-C. One UK study providing evidence of the psychometric properties of the SGRQ-C was identified in this update.

The content validity of the SGRQ-C is reported in a UK study by Meguro et al. (2007), where steps were taken to identify weaknesses in the SGRQ in order to revise the instrument to make it COPD-specific. Weak items within the SGRQ were identified via a five-stage study assessing the psychometric properties of the original SGRQ and the smaller COPD specific version. Patients were involved in the completion of both questionnaires whilst undergoing various treatment regimens.

Excellent reliability has been established, with ICCs ranging from 0.96 to 0.99 (Meguro et al., 2007).

The SGRQ-C was found to significantly converge with the original SGRQ and with other measures of both respiratory function, such as FEV, FVC, 6 minute walking test (6MWD), MRC Dyspnoea grade, and other patient-reported measures, such as the Hospital Anxiety and Depressions Scale (HADS), Sickness Impact Profile (SIP), Global Health, Cough/phlegm, and Daily Wheeze. Convergence was not significant between the Impacts component and FEV, but convergence was established for the Symptoms component and FVC.

The item referring to ability to work was found to have a large number of missing responses (19% compared with a mean of 1% for the other items) (Meguro et al., 2007), demonstrating low acceptability. Overall, the study reports that the SGRQ-C is shorter, contains the best of the original items and produces scores equivalent to the existing instrument.

e) Seattle Obstructive Lung Disease Questionnaire (SOLQ)

Three studies provided limited evidence of the measurement properties of the SOLQ in the previous review. No further studies evidencing the measurement properties of this instrument were identified in this update.

March 2010 Updated evidence

Further evaluations SGRQ-C. March 2010

The SGRQ-C scores have been used as indicators of quality of life for the evaluation of two other PROMs related to activities of daily living and symptoms in COPD patients Partridge et al., 2009 UK). No psychometric information can be drawn from this.

Responsiveness to change is reported in a multi-centred trial comparing bi-therapy with tri-therapy for COPD; significant improvement was reported in the tri-therapy group (Welte et al., 2009). The participants in this study were mainly from Europe but included a Canadian population.

Several studies report proposals to use the SGRQ-C in prospective studies. For example, Vestbo et al. (2008) outline a longitudinal international study which aims to identify predictive surrogate end-points. The SGRQ-C is included in a several health outcome measures in this study. The SGRQ-C will be administered annually. Some results have been published relating to this study although limited information is given. Significantly worse scores were reported at baseline for patients who were suffering from COPD but also were users of anti-depressants.

f) Clinical COPD Questionnaire (CCQ)

The PROM group were alerted to the CCQ by members of the multidisciplinary panel during the 2009 update of PROMs for COPD. At the time there was very limited UK evidence for the group to be confident in highlighting its potential value in the NHS.

The measure has specifically been developed to measure clinical disease control and not for the assessment of well-being or health-related quality of life. There are two versions to evaluate clinical COPD control: 24 recall and one week recall. There are 10 items divided into three domains: Symptoms, Functional state, and Mental state and responses are obtained on a seven-point Likert scale.

Registration is required to obtain the manual
http://www.ccq.nl

- The use of the CCQ in non funded academic research and individual clinical practice is free of charge.

Psychometric evidence

The development study reported the qualitative methodology used with patients to generate items. Items were generated and experts participated in an item reduction phase. Focus groups with patients included two from the NL and one from the UK (total participants 34). The Dutch interviews and focus groups results were analysed in Dutch and then translated in English for further development of the questionnaire.

The primary study of psychometric performance (van der Molen et al., 2003) included 119 patients from NL. High internal consistency was reported; the measure discriminated between healthy ex-smokers and COPD patients; moderate to strong correlation of scores were reported between CCQ and SF-36 domains and the SGRQ. Test-retest reliability was reported in a small sample (20) as high (ICC 0.94) and responsiveness in 36 smokers with or without COPD who significantly improved following smoking cessation. The minimally clinically important difference of CCQ total score of 0.4 has been reported in a Dutch population (Kocks et al., 2006).

Several evaluations of the CCQ with patients with COPD have been published since the previous review. Most of these are conducted in Europe using a non-English version of the CCQ.

Ställberg et al., (2009 Sweden) report good reproducibility; normal distribution of scores for Total and Symptoms with no floor or ceiling effects; and strong correlation of scores between CCQ and SGRQ.

Several respiratory indicators were associated with poorer scores in a large study conducted in Spain (Izquierdo et al., 2009). The CCQ has been used in several published trials and in ongoing studies in a Dutch population. These include effectiveness of home-based tele-monitoring (Trappenburg et al., 2008) and a research proposal of self management plans for patients with COPD (Trappenburg et al., 2009).Both of these studies refers to the use of the CCQ as a measure of HRQL and Trappenburg et al. (2009) will use the 24 hour recall version to assess recovery following an exacerbation (results unpublished at present). Several other evaluations of translations of the CCQ are available but not reported here.

Two studies report some very limited evidence from an English population (Jones et al., 2009a; Nouraei et al., 2009). The CCQ was used as a reference instrument for the evaluation of convergent validity of a newly developed brief assessment of COPD severity (Jones et al., 2009b). Nouraei et al. (2009) report high internal consistency and reproducibility; evidence of construct validity with significant correlations between CCQ and respiratory physiology and responsiveness to changes following surgery- this was a small study of patients with laryngotracheal stenosis.

The CCQ is referred to in the QOF for COPD:

COPD11 - The percentage of patients with COPD receiving inhaled treatment in whom there is a record that inhaler technique has been checked in the previous 15 months
COPD is increasingly recognised as a treatable disease with large improvements in symptoms, health status, exacerbation rates and even mortality if managed appropriately. Appropriate management should be based on NICE guideline CG12 and international GOLD guidelines in terms of both drug and non-drug therapy. *In making assessments of the patient's condition as part of an annual review and when considering management changes it is essential that health care professionals are aware of:* *• current lung function* *• exacerbation history* *• degree of breathlessness (MRC dyspnoea scale) and* ***A tool such as the Clinical COPD Questionnaire could be used to assess current health status.***

Two other instruments were identified

1. EXA**cerbations of** C**hronic Pulmonary Disease** T**ool (EXACT)**

The **EXA**cerbations of **C**hronic Pulmonary Disease **T**ool (EXACT) is a patient daily diary which is completed in a personal digital assistant (PDA). There are 14 items which were derived with the involvement of patients. It has predominantly been developed for use in clinical trials and for the assessment and evaluation of exacerbations of COPD. Several evaluations of good measurement criteria and patient acceptability have been reported at conferences. This is diary although appears attractive, will have cost implications in terms of purchasing the PDAs and it is too narrow in its focus of exacerbations.

Further details can be found at: **ttp://www.exactproinitiative.com/default.php**

2. COPD Assessment Test (CAT): Jones et al. (2009a, 2009b, UK)

The CAT has been specifically developed specifically for routine use in clinical practice. 8 items for this instrument were identified through qualitative research with COPD patients in three prospective international studies (Europe and the USA, n = 1,503) (Jones et al., 2009a). Good measurement criteria are reported with Rasch analysis identifying eight items fitting a unidimensional model. High internal consistency is reported (Cronbach's α= 0.88) and reproducibility is reported (ICC 0.80). Strong correlation of scores is reported between the CAT and the SGRQ-C (r=0.8). A five point difference of scores is reported as significant between stable and exacerbation. It can be downloaded free of charge from the developer website: http://www.catestonline.org/

4. DISCUSSION AND RECOMMENDATIONS

Generic PROMs

Table 2 summarises the psychometric criteria and operational characteristics of the generic PROMs included in the review using the Appraisal criteria found in Appendix B. The previous review reported evidence for the SF-36, SF-20, SF-12, EQ-5D, NHP, SIP, and the COOP Charts. This update provided further evidence for the SF-36, EQ-5D, and COOP Charts and also identified a further generic PROM used with COPD, this being the HUI. It is clear from the appraisal of the evidence, as detailed in Table 2, that the SF-36 remains the most frequently evaluated generic PROM overall with people with COPD with good measurement and operational performance reported.

Although the number of articles reviewed providing evidence for the EQ-5D is small (n = 4), it is the second most utilised generic measure in COPD patients. This update has provided further evidence of construct validity and acceptability, though further evidence is required.

Further evidence was provided for the construct validity and responsiveness of the COOP Charts with COPD patients, as well as further support for the operational characteristics of ease of administration and patient acceptability for people with chronic COPD. The COOP Charts show promise with COPD, but much more research is required before any recommendations could be made.

The HUI-3 was identified in this update, with reasonable evidence of responsiveness but no evidence for any other psychometric or operational criteria being assessed.

There is no further reported evidence for the SF-20, SF-12, SIP, or NHP.

COPD-specific PROMs

Table 3 summarises the psychometric criteria and operational characteristics of the condition-specific PROMs included in the review using the Appraisal criteria outlined in Appendix B. The previous review reported and recommended the CRQ (SAS and IA) and SGRQ based on the number of evaluations reporting good measurement and operational performance. With the evidence that has emerged from this review, the recommended instruments remain the CRQ and SGRQ, although newer and more pragmatic versions of these instruments are discussed and recommended for expert review.

The majority of evidence obtained for this update was for the previously recommended CRQ and SGRQ. This update reports evidence for the development of a CRQ short-form (SF-CRQ), which has demonstrated strong operational characteristics in terms of acceptability and feasibility. UK evidence outlining the development of a COPD-specific version of the SGRQ (SGRQ-C) is also reported. The SGRQ-C is purported to be shorter, to contain the best of the original items and to produce scores equivalent to the existing instrument. Both of these shorter and more condition-specific versions of these instruments are worthy of consideration for pragmatic reasons.

Limited support for the construct validity of the BPQ has been reported, and there is not enough data to recommend this instrument for use within the clinical environment.

No further evaluations were identified for the FPI or SOLQ.

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

PROM	Reproducibility	Internal consistency	Validity: Content	Validity: Construct	Responsiveness	Interpretability	Floor/ceiling/ precision	Acceptability	Feasibility
SF-36	+	++	0	++	++	+	_	+	0
SF-20	0	0	0	+	0	0	0	0	0
SF-12	0	0	0	+	0	+	0	0	0
EQ-5D	+	n/a	0	++	+	0	+	+	0
NHP	0	0	0	+	0	0	0	0	0
COOP Charts	_	0	0	+	+	+	0	+	+
SIP	0	0	0	+	0	0	0	0	0
HUI-3	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.*

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

PROM	Reproducibility	Internal consistency	Validity: Content	Validity: Construct	Responsiveness	Interpretability	Floor/ceiling/ precision	Acceptability	Feasibility
BPQ	+	+	+	++	+	0	0	0	0
CRQ-IA	++	+	0	++	++	++	0	++	0
CRQ-SAS	+	+	+	+	++		0	+	
SF-CRQ	0	+	+	+	+	0	_	+	+
FPI	+	+	0	+	0	0	+	+	0
SGRQ	++	+	0	++	++	++	+	+	_
SGRQ-C	+	0	+	+	0	0	0	_	+
SOLQ	+	+	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.*

RECOMMENDATIONS

For this update, tables 2 and 3 provide ratings of the evidence of measurement and operational performance applying the appraisal criteria for PROMs described in Appendix B to all of the available evidence. . Based on this appraisal, the following instruments were recommended for consideration by a multidisciplinary panel (see Appendix F).

The CRQ, SF-CRQ, SGRQ, and SGRQ-C were shortlisted as promising COPD-specific PROMs and the SF-36 and EQ-5D as generic measures.

Conclusions of scoring

Generic measures: Although comments suggested preference for the EQ-5D, total scores slightly favoured the SF-36. Based on this, though, there appears broadly similar support for both the SF-36 and EQ-5D as a generic health measure. However, the EQ-5D may be more favourable on the basis of its brevity of items, response rates and the provision of a preference score.

Condition-specific: Three condition-specific questionnaires were scored with approximate equal levels of support:

1. SGRQ-C
2. SGRQ
3. SF-CRQ

The comments indicated potentially more problems with the CRQ outside a research context.

There was considerable enthusiasm for the use of the Clinical COPD Questionnaire. Despite a lack of published evidence, many members have experience of using it in clinical practice and research contexts in the UK.

The research evidence, panel's ratings and views of the SGRQ, SGRQ-C and SF-CRQ were similar with moderate level of support for the SGRQ-C. There was a small amount of very positive evidence in the review for the short form of the CRQ; it also has a broad range of relevant items and its relative brevity suggesting more promising response rates. However, it appears that the CRQ-SF was validated principally for use in exacerbations of COPD. A pilot study of PROMs for COPD for general use in primary care may need an instrument with intended relevance beyond exacerbations of COPD. Amongst remaining COPD-specific instruments, the CRQ-SAS was found to have substantial supporting evidence in the review and its format appeared likely to be acceptable.

Having in mind an overall strategy of a generic and condition-specific measure being used in combination to assess complementary aspects of health status, and also having in mind the need for an approach that reduces the volume of questions and likely burden of responding, the current review recommends the combination of EQ-5D and CRQ-SAS for use in potentially large scale population studies. The simplicity and the brevity of the EQ-5D, make it likely that it will not adversely influence response rates. The fact that it yields UK-derived preference values, makes it an attractive generic measure providing complementary evidence on health status alongside the CRQ-SAS.

Evidence update: March 2010

SGRQ-C:
One study was included in the previous review for SGRQ-C. Further evidence of responsiveness is reported from one study (Welte et al., 2009) and proposal to use the instrument in another (Vestbo et al., 2009). No convincing evidence has emerged to favour this measure above the others.

Clinical COPD Questionnaire (CCQ).
Despite the CCQ clearly recommended in the QOF for COPD and anecdotal evidence of its widespread use clinically and in a research context in the UK, there remains a distinct lack of English language published evaluations. The two more recent UK evaluations add very little: one study used the CCQ as a reference measure; one with patients pre and post surgery for laryngotracheal stenosis. Several non-English language evaluations have been published reporting good measurement performance.

The developers state that the CCQ is for the measurement of COPD control and not for the assessment of HRQL and well-being. It has though been evaluated as a health status measure and for the impact of COPD on the patient in several studies

The UK developed **COPD Assessment Test** appears attractive and is developed specifically for clinical use. Further published evaluations are needed to strengthen its utility.

If a choice has to be made, the CCQ seems to have good face and content validity; has professional endorsement and is specifically sign-posted in QOF guidance and policy and its brevity an attractive feature.

.

APPENDIX Ai: 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 Aii: 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 hand searched 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. 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 xxx).

Table XXX

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 xxxxxxxxxxxxxxxxx

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-retest 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 hypothesised direction; assessed on the basis of correlations between the measure and other similar measures	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted 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 Ci: 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 summarised 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-20: Medical Outcomes Study 20-item Short Form Health Survey (Stewart et al., 1988; Ware, Sherbourne and Davies, 1992)

The Medical Outcomes Study (MOS) 20-item Short Form Health Survey (SF-20) is a 20-item abbreviation of the same Rand instrument from which the SF-36 originates (Stewart et al., 1988; Ware et al., 1992; McDowell and Newell, 1996). The abridged instrument was intended to reduce respondent burden, whilst comprehensively addressing important issues in health status measurement.

The SF-20 assesses health across six domains, namely bodily pain (BP: one item), general health perception (GH: five items), physical function (PF: six items), mental health (MH: five items), social function (SF: one item), and role function (RF: two items), as shown in Table1.

Items have categorical response options (range: 3-6 options); several items have reversed scoring. Domain item summation scores are transformed into a scale from 0 to 100, where higher values denote better health. The instrument may be self-, interview-, or telephone-administered. Instrument self-administration takes approximately four minutes (McDowell and Newell, 1996), but longer completion times have been reported for older people (Siu et al., 1993a, b).

c) 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). 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).

d) 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.

e) Nottingham Health Profile (Hunt et al., 1980)

The Nottingham Health Profile (NHP) was developed in the UK during the 1970s for use in the evaluation of medical or social interventions (Hunt et al., 1980). Instrument content was derived from over 2000 statements given by 768 patients with a variety of chronic ailments and other lay people.

Part I of the instrument has 38 items across six domains: bodily pain (BP), emotional reactions (ER), energy (E), physical mobility (PM), sleep (S), and social isolation (SI), as shown in Table 1. All items are statements that refer to departures from normal functioning, and relate to feelings and emotional state rather than change in behaviour. Respondents answer 'yes' or 'no' according to whether or not they feel the item applies to them in general. Positive responses are weighted and summed to give six domain scores between 0 and 100, where 100 denote maximum limitation.

Part II of the NHP is less widely used and provides a brief indicator of handicap. The instrument may be self- interview-, or telephone-administered.

f) COOP Charts for Primary Care Practice (Nelson et al., 1987)

The Dartmouth Primary Care Cooperative Information Project developed the COOP charts in the late 1980s to provide a screening tool for use by doctors in routine practice (Nelson et al., 1987). The charts support the assessment of patient health status and functioning.

The original instrument, developed in the USA, has nine charts, each containing a single question about health, functioning, or quality of life during the previous month (Table 1). Eight charts assess bodily pain (BP), daily activities (DA), emotional condition/feelings (EC), physical fitness (PF), quality of life (QoL), social activities (SA), social support (SS), and current overall health (OH) perceptions. An additional chart assesses change in overall health. Literature reviews, existing instruments, and discussion with practising physicians and experts in health status measurement informed item derivation (Nelson et al., 1990).

Following a multinational feasibility study, item content was revised to seven charts, omitting quality of life and social support, with a reduced recall period of two weeks (World Organisation of National Colleges, Academies and Academic Associations of General Practitioners and Family Physicians [WONCA]: WONCA/COOP Health Assessment Charts. Froom, 1988; Langraf and Nelson, 1992). Each chart within the WONCA/COOP includes a descriptive title, a question, and a pictorially illustrated five-point response scale, where five is the most severe limitation. Each represents a separate domain; an overall score is not calculated (McDowell and Newell, 1996). The charts can be self or interview-administered.

g) 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).

h) 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).

Appendix Cii: 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-20: MOS 20-item Short Form Health Survey (20)	Bodily pain (BP) (1), General health (GH) (5) Mental health (MH) (5), Physical functioning (PF) (6) Role functioning (RF), Social functioning (SF) (1)	Categorical: 3-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Summation Domain profile (0-100, 100 best health)	Self (5-7)
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

Instrument	*Domains (no. items)*	*Response options*	*Score*	*Completion (time in minutes)*
Nottingham Health Profile (NHP) (38)	Bodily pain (BP) (8), Emotional reactions (ER) (9), Energy (E) (3), Physical mobility (PM) (8), Sleep (S) (5), Social isolation (SI) (5)	Yes/no; positive responses weighted Recall 'general' health	Algorithm Domain profile 0-100, 100 is maximum limitation	Interview Self (10-15)
COOP Charts for Primary Care Practice (COOP) (8+1)	Bodily pain (BP) (1), Daily activities (ADL) (1), Emotional condition (EC) (1), Physical fitness (PF) (1), Quality of life (QL) (1), Social activities (SA) (1), Social support (SS) (1), Overall health perception (OH) (1), Change in health status (1)	Categorical: 1-5 (illustrated) 2-week recall	Chart profile (1-5, 5 no limitations)	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

Appendix Ciii: 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 (36)	x	x	x	x	x		x	
SF-20 (20)	x	x	x	x	x		x	
SF-12 (12)	x	x	x	x	x		x	
EQ-5D (5+1)	x	x	x	x	x		x	
NHP (38)	x	x		x	x			
COOP (9) Charts	x	x	x	x	x		x	
SIP (136)	x	x		x	x	x	x	
HUI (8)	x			x		x		

Appendix D: Condition-specific PROMS

a) Breathing Problems Questionnaire (BPQ)

The Breathing Problems Questionnaire (BPQ) items were derived from focus groups with fifteen patients and refined by 89 COPD patients (Hyland 1994). The instruments foundation was based on three constructs of quality of life: Problems, Negative evaluations and Positive evaluations. Following factor analysis two factors emerged with 27 items constituting the BPQ problems score and 6 items the BPQ negative evaluations score. Further modifications have been made with a shortened version of ten items and a single scale score (Hyland et al., 1998).

b) Chronic Respiratory Disease Questionnaire (CRQ)

The Chronic Respiratory Disease Questionnaire (CRQ) was developed following interviews with 100 patients with chronic airflow limitation to identify the impact on their quality of life and how important their symptoms were (Guyatt 1987). The most frequently reported and important items were selected and provided the conceptual framework for the instruments which were categorised into four domains: Dyspnoea, Fatigue, Emotional function and Mastery. The Dyspnoea domain is individualised and related to activities which patients report their degree of dyspnoea. A list is provided to aid selection where needed.

The CRQ has four domains with a total of 20-items and include Dyspnoea (5-items); Fatigue (4-items); Emotional functioning (7-items); Mastery (4-items). Scoring is by domains and uses a seven-point Likert scale with higher scores reflecting no impairment.

Short-Form Chronic Respiratory Disease Questionnaire (SF-CRQ)

The SF-CRQ is an 8-item version of the original 20-item CRQ. The SF-CRQ maintains the original four domains of Dyspnoea (2-items), Fatigue (2-items), Emotional Functioning (2-items), and Mastery (2-items) as well as the seven-point Likert scale reflecting level of impairment; higher scores reflect no impairment. Item selection was based on previous research (Moran et al., 2001) and consultation with the developer of the original CRQ, Dr Guyatt. The instrument was pilot tested with consecutive emergency department patients with COPD (n = 301).

c) Functional Performance Inventory (FPI)

The FPI was developed in the USA involving both patients and clinical experts. Focus groups with patients informed an activity profile and content evaluated by clinical experts. Pre-testing of the instruments face validity was evaluated with a group of patients. The FPI is based on a conceptual framework of functional status as a multidimensional concept involving activities carried out to meet basic need, fulfill roles and maintain health and well-being (Leidy and Knebel 1999). The FPI has six domains (65items): Body Care (9); Household Maintenance (21); Physical Exercise (7); Recreation (11); Spiritual Activities (5) and Social Activities (12). Response options range from 1 where the activity can be performed easily to 4 where the activity is no longer performed for health reasons. Higher scores reflect high functioning. Domain and Total scores are computable. Modifications have been made to scoring by Larson et al., (1998).

d) St. George's Respiratory Questionnaire (SGRQ)

The SGRQ was developed in the UK to measure the impact of asthma and chronic obstructive pulmonary disease (COPD) from a patient perspective. There are two parts of the instrument. Part 1 is concerned with symptoms focusing on the severity, frequency and effect

of respiratory symptoms over the last year and responses are obtained with a 5 point Likert scale. Part 2 includes two domains Activity limitations and social and psychological Impact and focuses on the patient's current state with True or False responses. Three components scores are calculated and a total score. All items have empirically derived weights and normative data are available. Scoring algorithms and calculators are available from the developers. Scores are expressed as the percentage of overall impairment with 100 equaling to worst possible health and zero the best.

Items were initially derived from studies with adult patients with asthma examining distress ratings relating to symptoms and the impacts of asthma (Quirk and Jones1990) and the influence of demographic and disease factors with the degree of distress (Quirk et al., 1991). Empirical weights were obtained from one hundred and forty patients with asthma (Quirk et al., 1991). Further analysis of previously derived weights was compared with patients with COPD with thirty-six patients (mean age 66) (Jones 1991) and no significant differences between the item weights from the asthma patients (Quirk et al., 1991) and COPD patients.

The final instrument has 50 items and 76 weighted responses divided into three components Symptoms, Activities and Impacts

St. George's Respiratory Questionnaire – COPD Version (SGRQ-C)
The SGRQ-C is 40-item version of the original 50-item SGRQ and was derived from the original version following detailed analysis of data from large studies in COPD. The intention was to remove the items with the weakest measurement properties in the original instrument, but at the same time ensure that its scores were directly comparable with the original SGRQ (Meguro et al., 2007).

The instrument consists of two parts: Part I produces the Symptoms score, and Part 2 the Activity and Impacts scores. A Total score is also produced. No specific recall period is used except for one item. Each response has a unique empirically derived 'weight' (Quirk, et al., 1990). The lowest possible weight is zero and the highest is 100. In cases where the two response options to an item in the original SGRQ were combined in the SGRQ-C, the weight for the new response option was calculated from the mean of the two that were combined

In a small number of items on the SGRQ-C there is a reduction in the number of response categories when compared to the original version. There has also been a change in the wording of Part 1.

e) <u>Seattle Obstructive Lung Disease Questionnaire</u>
The Seattle Obstructive Lung Disease Questionnaire (SOLQ) was developed using the CRQ model of functional status but with the intention of providing a self-reported questionnaire which can be computer scanned and therefore processing and scoring enabling feasibility. Dimensions were selected from patient interviews, literature, and clinical experience of the developers and the CRQ model of COPD specific health-related quality of life.

The instrument comprises four domains: Physical functioning, Emotional functioning, Coping skills and Treatment satisfaction with 29 items. Scoring is on a simple linear scale with lowest scores indicating poorer functioning. Domain scores are computed. Permission for use is required from the author.
.

APPENDIX Dii: CONDITION-SPECIFIC PROMs

Condition-specific patient-reported health instruments

Instrument	***Domains (no. items)***	***Response options***	***Score***	***Administration/ Completion (time)***
Breathing Problems Questionnaire (BPQ)	*Thirteen domains (33),* Two subscales: Problems and Emotional evaluations Walking (3) Bending or reaching (2) Washing and bathing (2) Household chores (3) Social interactions (3) Effects of weather and temperature (4) Effects of smells and fumes (2) Effects of colds (1) Sleeping (2) Medicine (2) Dysphoric states (5) Eating (2) Excretion urgency (2)	4 point Likert	Subscale scores Lower scores better quality of life	
Chronic Respiratory Disease Questionnaire (CRQ)	*Four domains (20 items)* Dyspnoea (5) Fatigue (4) Emotional functioning (7) Mastery (4)	7 point Likert	Domain Higher scores indicate no impairment	Maximum 30 minutes
Chronic Respiratory Disease Questionnaire Short-Form(SF-CRQ)	*Four domains (8 items)* Dyspnoea (2) Fatigue (2) Emotional functioning (2) Mastery (2)	7 point Likert	Domain Higher scores indicate no impairment	Less than 5 minutes
Functional Performance Inventory (FPI)	*Six domains (65 items*): Body Care (9); Household Maintenance (21); Physical Exercise (7); Recreation (11); Spiritual Activities (5) and Social Activities (12).	5 point Likert	Total and Domain Higher scores reflect high functioning	

Instrument	*Domains (no. items)*	*Response options*	*Score*	*Administration/ Completion (time)*
Seattle Obstructive Lung Disease Questionnaire (SOLQ)	*Four domains (29 items)* Physical functioning Emotional functioning Coping skills Treatment satisfaction	Linear scale	Domain score Lower score indicate lowest function	5-10 minutes completion Can be computer scanned
St. George's Respiratory Questionnaire (SGRQ)	*Two parts; Domains (50)* Part 1: Symptom scores Part 2: Activity and Impact	Part 1: Likert Scale Part 2: True or False	Weighted scoring Total and domain scores Percentage of overall impairment 0=best possible health and 100 worse	Self- report but recommended interview administered 8- 15 minutes to complete
St. George's Respiratory Questionnaire - COPD Version (SGRQ-C)	*Two parts; Domains (40)* Part 1: Symptom scores Part 2: Activity and Impact	Part 1: Likert Likert Part 2: True or False	Weighted scoring Total and domain scores Percentage of overall impairment 0 = best possible health and 100 = worse	Not reported

Summary of condition-specific 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 constructs**	**Treatment satisfaction**
BPQ	x	x		x	x		x		x
CRQ	x	x		x				x	
SF-CRQ	x	x		x				x	
FPI	x				x		x	x	
SGRQ	x	x	x	x	x		x	x	x
SGRQ-C	x	x	x	x	x		x	x	
SOLQ	x			x				x	x

APPENDIX E: Availability and Licensing Details for Short-Listed PROMs

PROMs	Country	Licensing Fee, etc.	Contact
St George's Respiratory Questionnaire (SGRQ) (Jones et al., 1991) St George's Respiratory Questionnaire – COPD version (SGRQ-C) (Meguro et al., 2007)	UK	Academics and medical professionals can obtain the questionnaire, scoring manual, and scoring calculator free of charge.	Manual and scoring instructions available from Dr PW Jones, Division of Physiological Medicine, St. Georges Hospital Medical School, Cranmer Terrace, London SW17 ORE. Email: pjones@sghns.ac.uk.
Chronic Respiratory Disease Questionnaire (CRQ) (Guyatt et al., 1987) Chronic Respiratory Disease Questionnaire – Short-Form (SF-CRQ) (Tsai et al., 2008)	Canada	License Agreement required from the Licensing Officer. $500 for the training manual, training cassette, and the CRQ.	Licensing Officer Contact: Mrs. Peggy Austin, Dr. Holger or Dr. Gordon Guyatt Dept. of Clinical Epidemiology & Biostatistics McMaster University 1200 Main Street West, HSC 3V43A Hamilton, Ontario CANADA L8N 3Z5 PHONE: 905-525-9140 ext. 22154 FAX: 905-540-1144
SF-36 (Ware and Sherbourne, 1992)	USA	QualityMetric, the Medical Outcomes Trust, and the Health Assessment Lab are the co-copyright and trademark holders of the SF-36. A commercial license is required. Permission for use is required for scholarly research. The shared licensing program allows individuals and organisations that benefit from commercial uses of the intellectual property to pay royalties or other user fees that will support the research community that made the original surveys possible.	

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 COPD and special interest in Patient-reported Outcome Measures.

The panel were sent the following documents:

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

The panel were sent by email rating scales to judge the suitability of the questionnaire for use in the NHS for the evaluation of services. There was a section for comments.
The rating scale used the following responses:

'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).

Scores for each questionnaire were ranked in order of preference. The Total maximum score=36.

The results and comments were then distributed by email to the panel for further rating should they wish to change their vote.

Ratings and comments

Generic measures

Overall there was agreement that generic health status measures are useful but that using a disease-specific measure would be the preferred option, specifically in a clinical setting. It was thought to be too burdensome to include both a generic particularly the SF-36 and disease-specific measure. However, using a generic measure to compare groups of patients with other chronic conditions as well as capturing co-morbidities associated with the condition is an attraction.

The activity item in the SF-36 was deemed not to capture the extreme limitations people with COPD experience. Although the EQ-5D was favoured by most people due to its brevity and the ability to generate a utility value, there were items not included which were relevant to people with COPD such as symptoms. It was acknowledged that a utility can be generated from the SF-36 but that this was more difficult. The lack of responsiveness in patients with COPD was noted as well as the possibility of it not capturing well experiences of those patients with less severe disease.

Generic Total maximum score= 28

FIRST RATING	'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	0		2	9	8	**19**
EQ-5D	0	1	2	9	4	**16**

Conclusions of scoring

The panel rated each questionnaire using the response categories listed above. Although comments suggested preference for the EQ-5D, total scores slightly favoured the SF-36. Based on this, though, there appears broadly similar support for both the SF-36 and EQ-5D as a generic health measure. However, the EQ-5D may be more favourable on the basis of its brevity of items, response rates and the provision of a preference score.

COPD-Specific PROMs

Chronic Respiratory Questionnaire (CRQ)
Several members commented on the length and complexity of completion of the CRQ. It was considered to be impractical for use in clinical practice especially if other questionnaires are being administered alongside.

The Domain scoring was considered to be a less attractive feature. The Emotional domain was judged to be important because of the incidence of depression in people with COPD.

Reference was made to the self-administered version of the CRQ. Evidence has been presented in the 2006 review (Fitzpatrick et al., 2006) with equivalent evidence of measurement performance.

It its original value it considered not to be a practical tool for measuring quality of NHS services.

Short Form-CRQ
Several members favoured this shorter version than the original instrument based on acceptability to patients and feasibility in terms of a short completion time. It was deemed to be a promising instrument but that further evaluations are needed to be confident to recommend it.

One member pointed out that the questionnaire has been significantly changed in this version. The Mastery domain reflecting the patient's perceived degree of control of the condition has been reduced to questions about panic and fear which is a different construct.

St. George's Respiratory Questionnaire (SGRQ)
A number of members indicated that the SGRQ is widely known and used specifically in research trials and only in clinical practice as part of pulmonary rehabilitation. It was reported by some to be acceptable to clinicians. However, one member had used it in a workshop and clinician's feedback was negative.

It was reported to be difficult for patients to complete, lacks clarity and its length limited its acceptability to patients. Some patients have needed assistance to complete the questionnaire. It was noted that patients were not included during the development of the instrument.

Difficulty in the interpretation of results hampered its utility particularly for patients with less severe symptoms. The absence of an emotional domain was considered problematic particularly as patients with COPD are prone to depression. It requires the use of an additional measure for psychological well-being.

Despite these issues, it was considered to have some utility in research and clinically in pulmonary rehabilitation programmes and settings.

SGRQ-C
Overall consensus was that this was more appropriate than the parent version due to it's relevance to COPD patients and ease of administration and scoring methods. Despite this, there are still 40 items which may be burdensome for some patients.

The group overall, considered it to be promising.

Other instruments

The Clinical COPD Questionnaire (CCQ) was identified as a useful questionnaire by several members of the panel (www.ccq.nl). This was developed with English and Dutch speaking populations (van de Molen et al., 2003). There are ten items and three domains: Symptoms, Functional state and Mental. Responses are obtained on a 7 point scale where 0= asymptomatic/no limitations to 6= extremely symptomatic/total limitation. Domain and Total scoring is suggested. Reliability, discriminative validity and responsiveness have been reported during development. An evaluation in a Dutch population established minimal clinically important difference for the scale (Kocks et al., 2006). Further UK evaluations are required to be confident in recommending this questionnaire above those already short-listed.

Specific Total in ranked order Maximum score = 28

FIRST RATING	**'not at all suitable' (score 0)**	**'to some extent unsuitable' (score 1)** *Max score 7*	**'uncertain' (score 2)** *Max score 14*	**'to some extent suitable' (score 3)** *Max score 21*	**'very suitable' (score 4)** *Max score 28*	**TOTAL** *Max score 28*
SGRQ-C	0			12	8	**20**
SGRQ		1	2	12	4	**19**
SF-CRQ			4	15		**19**
CRQ		2		9	4	**15**

Conclusions of scoring

The panel rated each questionnaire using the response categories listed above. Three questionnaires were scored with approximately equal levels of support:

-SGRQ-C
-SGRQ
-SF-CRQ

The comments indicated potentially more problems with the CRQ outside a research context.

There was considerable enthusiasm for the use of the Clinical COPD Questionnaire. Despite a lack of published evidence, many members have experience of using it in clinical practice and research contexts in the UK.

The research evidence, panel's ratings and views of the SGRQ, SGRQ-C and SF-CRQ were similar with moderate level of support for the SGRQ-C. If nevertheless a choice has to be made, we would favour the shorter form of the CRQ based on having a broad range of relevant items and its relative brevity suggesting more promising response rates.

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 COPD? (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

COPD OUTCOMES CONSENSUS GROUP MEMBERS.

Dr Jennifer Cleland
Senior Clinical Lecturer in Medical Education and Primary Care, University of Aberdeen

Dr Kevin Gruffyd-Jones
General Practitioner, Box Surgery, Wilts.

Dr Maxine Hardinge
Consultant Physician in Respiratory and General Medicine , Respiratory Medicine Churchill Hospital , Oxford

Dr Rupert Jones
General Practitioner, Respiratory Research Unit, Plymouth

Sharon Lee
Specialist Community Matron, Ashford, Kent

Dr Hilary Pinnock
Senior Clinical Research Fellow, University of Edinburgh

Dr Mike Thomas
Asthma UK Senior Research Fellow , University of Aberdeen

PROM Group

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

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

REFERENCES

Aiken LS, Butner J, Lockhart CA, Volk-Craft BE, Hamilton G and Williams FG. Outcome evaluation of a randomized trial of the PhoenixCare intervention: Program of case management and coordinated care for the seriously chronically ill. *Journal of Palliative Medicine* 2006; 9: 111-26.

Ashmore JA, Emery CF, Hauck ER and MacIntyre NR. Marital adjustment among patients with chronic obstructive pulmonary disease who are participating in pulmonary rehabilitation. *Heart and Lung* 2005; 34: 270-8.

Bauldoff GS, Rittinger M, Nelson T, Doehrel J and Diaz PT. Feasibility of distractive auditory stimuli on upper extremity training in persons with chronic obstructive pulmonary disease. *Journal of Cardiopulmonary Rehabilitation* 2005; 25: 50-5.

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

Bergner M, Bobbitt RA, Carter WB et al. The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care;* 1981;19:787-805.

Berry MJ, Adair NE and Rejeski WJ. Use of peak oxygen consumption in predicting physical function and quality of life in COPD patients. *Chest* 2006; 129: 1516-1522.

Brazier JE, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF-36 Health Survey questionnaire. *Quality of Life Research* 1993; 2(3):169-180.

Brightling CE, Monterio W, Green RH, Parker D, Morgan MDL, Wardlaw AJ *et al*. Induced sputum and other outcome measures in chronic obstructive pulmonary disease: safety and repeatability. *Respiratory Medicine* 2001; 95:999-1002.

Brown CD, Benditt JO, Sciurba FC, Lee SM, Criner GJ, Mosenifar Z, Shade DM, Slivka WA and Wise RA. Exercise testing in severe emphysema: Association with quality of life and lung function. *COPD: Journal of Chronic Obstructive Pulmonary Disease 2008;* 5(2): 117-124.

Carr SJ, Goldstein RS and Brooks D. Acute exacerbations of COPD in subjects completing pulmonary rehabilitation. *Chest* 2007; 132(1):127-34.

Carrieri KV, Nguyen HQ, Donesky CD, Demir DS, Neuhaus J and Stulbarg MS. Impact of brief or extended exercise training on the benefit of a dyspnoea self-management program in COPD. *Journal of Cardiopulmonary Rehabilitation* 2005; 25: 275-84.

Coultas D, Frederick J, Barnett B, Singh G and Wludyka P. A randomized trial of two types of nurse-assisted home care for patients with COPD. *Chest* 2005; 128: 2017-24.

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

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

Department of Health (2009) Guidance on the routine collection of Patient Reported Outcome Measures (PROMs)

Donohue JF, Parsey MV, Andrews C, D'Urzo T, Sharma S, Schaefer K, Claus R, Baumgartner RA and Levalbuterol COPD Study Group. Evaluation of the efficacy and safety of levalbuterol in subjects with COPD. *COPD* 2006; 3(3):125-32.

Eaton T, Young P, Fergusson W, Garrett JE, and Kolbe J. The Dartmouth COOP Charts: A simple, reliable, valid and responsive quality of life tool for chronic obstructive pulmonary disease. *Quality-of-Life-Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation* 2005;14:577-85.

EuroQol Group, The. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16(3):199-208.

Fan VS, Ramsey SD, Make BJ, and Martinez FJ. Physiologic variables and functional status independently predict COPD hospitalizations and emergency department visits in patients with severe COPD. *COPD* 2007; 4(1):29-39.

Ferguson GT, Anzueto A, Fei R, Emmett A, Knobil K, and Kalberg C. Effect of fluticasone propionate/salmeterol (250/50 microg) or salmeterol (50 microg) on COPD exacerbations. *Respir Med.* 2008;102(8):1099-108.

Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating Patient-Based Outcome Measures For Use In Clinical Trials. *Health Technology Assessment* 1998, 2, I-Iv, 1-74.

Fitzpatrick R, Bowling A, Gibbons E, Haywood K, Jenkinson C, Mackintosh A, Peters M. A structured review of patient-reported measures in relation to selected chronic conditions, perceptions of quality of life and carer impact. *University of Oxford and the Department of Health* 2006.

Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. (2002a) Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal;* 324 (7351): 1417-1421.

Gross NJ, Nelson HS, Lapidus RJ, Dunn L, Lynn L, Rinehart M, Denis-Mize K, and Formoterol Study Group. Efficacy and safety of formoterol fumarate delivered by nebulization to COPD patients. *Respir Med.* 2008 102(2):189-97.

Gupta RB, Brooks D, Lacasse Y, and Goldstein RS. Effect of rollator use on health-related quality of life in individuals with COPD. *Chest* 2006 130: 1089-1095.

Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42:773-8.

Harper R, Brazier JE, Waterhouse JC, Walters SJ, Jones NMB, Howard P. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax* 1997; 52:879-87.

Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *Journal of Epidemiology and Community Health;* 1980; 34: 281-286.

Hyland ME, Bott J, Singh SJ, Kenyon CAP. Domains, constructs and the development of the breathing problems questionnaire. *Quality of Life Research* 1994; 3:245-56.

Hyland ME, Singh SJ, Sodergren SC, Morgan MP. Development of a Shortened Version of the Breathing Problems Questionnaire Suitable for Use in a Pulmonary Rehabilitation Clinic: A Purpose-Specific, Disease-Specific Questionnaire *Quality of Life Research,* Volume 7, Number 3, 1998 , pp. 227-233(7)

Izquierdo JL, Barcina C, Jimenez J, Munoz M, Leal M. Study of the burden on patients with chronic obstructive pulmonary disease. *International Journal of Clinical Practice* 2009; 63(1):87-97.

Jaeschke R, Singer J, and Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10(4):407-15.

Jenkinson C, McGee H. Health Status Measurement, a Brief but Critical Introduction. Oxford: Radcliffe Medical Press Ltd, 1998.

Johnson MK, Birch M, Carter R, Kinsella J, and Stevenson RD. Measurement of physiological recovery from exacerbation of chronic obstructive pulmonary disease using within-breath forced oscillometry. Thorax 2007; 62: 299-306.

Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85:25-31.

Jones PW and Bosh TK. Quality of life changes in COPD patients treated with salmeterol. *Am. J. Respir. Crit. Care Med.* 1997; 155(4) 1283-1289.

Jones PW, Quirk FH, Baveystock CM, AND Littlejohns P. A self-complete measure of health status for chronic airflow limitation. *Am Rev Respir Dis* 1992;145;1321-1327.

Jones P, Harding G, Wiklund I, Berry P, Leidy N. Improving the process and outcome of care in COPD: development of a standardised assessment tool. *Prim Care Respir J.* 2009a Sep;18(3):208-15.

Jones P, Harding G, Berry P, Wiklund I, Chen W-H, Kline Leidy N Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009b; 34:648-654.

Kapella MC, Larson JL, Patel MK, Covey MK, AND Berry JK. Subjective fatigue, influencing variables, and consequences in chronic obstructive pulmonary disease. *Nursing Research* 2006; 55:10-7.

Kocks J, Tuinenga M, Uil S, van den Berg J, Ståhl E, van der Molen T. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respir Res.* 2006; 7(1): 62.

Larson JL, Kapella MC, Wirtz S, Covey MK, Berry J. Reliability and validity of the functional performance inventory in patients with moderate to severe chronic obstructive pulmonary disease. *Journal of Nursing Measurement* 1998; 6:55-73.

Leidy NK, Knebel AR. Clinical validation of the Functional Performance Inventory in patients with chronic obstructive pulmonary disease. *Respiratory Care* 1999; 44:932-9.

Maleki-Yazdi MR, Lewczuk CK, Haddon JM, Choudry N, and Ryan N. Early detection and impaired quality of life in COPD GOLD stage 0: a pilot study. *COPD* 2007; 4(4):313-20.

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

McGlone S, Venn A, Walters EH, and Wood-Baker R. Physical activity, spirometry and quality-of-life in chronic obstructive pulmonary disease. *COPD* 2006; 3(2):83-8.

Meguro M, Barley EA, Spencer S, and Jones PW.
Development and validation of an improved, COPD-specific version of the St. George Respiratory Questionnaire. *Chest* 2007: 456–463.

Miller JD, Malthaner RA, Goldsmith CH, Goeree R, Higgins D, Cox PG, Tan L, Road JD; Canadian Lung Volume Reduction Surgery Study. A randomized clinical trial of lung volume reduction surgery versus best medical care for patients with advanced emphysema: a two-year study from Canada. *Ann Thorac Surg.* 2006; 81(1):314-20; discussion 320-1.

Moran LA, Guyatt GH, et al. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol* 2001;54(6): 571-9.

Nouraei SAR, Randhawa PS, Koury EF, Abdelrahim A, Butler CR, Venkataraman A, Howard DJ, Sandhu GS. Validation of the Clinical COPD Questionnaire as a psychophysical outcome measure in adult laryngotracheal stenosis *Clinical Otolaryngology* 2009;34(4):343-348.

Okubadejo AA, Jones PW, Wedzicha JA. Quality of life in patients with chronic obstructive pulmonary disease and severe hypoxaemia. Thorax 1996; 51:44-7.

Partridge MR, Miravitlles M, Ståhl E, Karlsson N, Svensson K, Welte T. Development and validation of the Capacity of Daily Living during the Morning and Global Chest Symptoms Questionnaires in COPD. *Eur Respir J.* 2009 Nov 6

Puhan MA, Guyatt GH, Goldstein R, Mador J, McKim D, Stahl E, Griffith L, and Schünemann HJ. Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respiratory Medicine* 2007; 101(2): 308-316.

Punekar YS, Rodriguez-Roisin R, Sculpher M, Jones P, and Spencer M. Implications of chronic obstructive pulmonary disease (COPD) on patients' health status: a western view. *Respir Med.* 2007; 101(3):661-9.

Resnick B, Nahm ES. Reliability and validity testing of the revised 12-item Short-Form Health Survey in older adults. *Journal of Nursing Measurement* 2001; Fall. 9(2):151-61.

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

Ries AL, Make BJ, Lee SM, Krasna MJ, Bartels M, Crouch R, and Fishman AP for the National Emphysema Treatment Trial Research Group. The effects of pulmonary rehabilitation in the National Emphysema Treatment Trial. *Chest* 2005; 128: 3799-3809.

Roth MD, Connett JE, D'Armiento JM, Foronjy RF, Friedman PJ, Goldin JG *et al*. Feasibility of retinoids for the treatment of emphysema study. *Chest* 2006;130:1334-45.
Schofield M, Mishra G. Validity of the SF-12 compared with the SF-36 health survey in pilot studies of the Australian longitudinal study on women's health. *Journal of Health Psychology* 1998; 3(2):259-271

Sewell L, Singh SJ, Williams JEA, Collier R, and Morgan M. Can individualized rehabilitation improve functional independence in elderly patients with COPD? *Chest* 2005; 128: 1194-1200.

Sewell L, Singh SJ, Williams JEA, Collier R, and Morgan M. How long should outpatient pulmonary rehabilitation be? A randomised controlled trial of four-weeks versus seven-weeks. *Thorax* 2006; doi:10.1136/thx.2005.048173.

Singh SJ, Sodergren SC, Hyland ME, Williams JEA, Morgan MDL. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. *Respiratory Medicine* 2001; 95:71-7.

Smith SC, Cano S, Lamping DL, Staniszewska S, Browne J, Lewsey J, van der Meulen J, Cairns J, Black N. Patient-reported outcome measures (PROMs) for routine use in treatment centres: recommendations based on a review of the scientific evidence. *London School of Hygiene and Tropic Medicine, London, UK, and the Department of Health*, 2004.

Stallberg B, Nokela M, Ehrs PO, Hjemdal P, Jonsson EW: Validation of the clinical COPD Questionnaire (CCQ) in primary care. *Health Qual Life Outcomes* 2009, 7: 26.

Stapleton RD, Nielsen EL, Engelberg RA, Patrick DL, and Curtis JR. Association of depression and life-sustaining treatment preferences in patients with COPD. *Chest* 2005;127:328-34.

Tashkin DP, Littner M, Andrews CP, Tomlinson L, Rinehart M, and Denis-Mize K. Concomitant treatment with nebulized formoterol and tiotropium in subjects with COPD: A placebo-controlled trial. *Respiratory Medicine* 2008; 102(4): 479 – 487.

Trappenburg JC, Koevoets L, de Weert-van Oene GH, Monninkhof EM, Bourbeau J, Troosters T, Verheij TJ, Lammers JW, Schrijvers AJ. Effects of Telemonitoring in Patients with Chronic Obstructive Pulmonary Disease. *Telemedicine and e-Health*. March 2008, 14(2):138-146.

Trappenburg JC, Koevoets L, de Weert-van Oene GH, Monninkhof EM, Bourbeau J, Troosters T, Verheij TJ, Lammers JW, Schrijvers AJ. Action Plan to enhance self-management and early detection of exacerbations in COPD patients; a multicenter RCT. *BMC Pulm Med*. 2009 Dec 29;9:52.

Tsai CL, Hodder RV, Page JH, Cydulka RK, Rowe BH, and Camargo CA Jr. The short-form chronic respiratory disease questionnaire was a valid, reliable, and responsive quality-of-life instrument in acute exacerbations of chronic obstructive pulmonary disease. *J Clin Epidemiol.* 2008; 61(5):489-97.

Van der Molen T, Willemse BW, Schokker S, Ten Hacken NH, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual Life Outcomes*, 2003 Apr 28;1(1):13

Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, Hagan G, Knobil K, Lomas DA, MacNee W, Silverman EK, Tal-Singer R, ECLIPSE investigators Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J.* 2008 Apr;31(4):869-73. Epub 2008 Jan 23

Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992; 30(6):473-483.

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

Ware JE. SF-36 Health Survey. Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Centre, 1997. 2nd 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.

Welte T, Miravitlles M, Hernandez P, Eriksson G, Peterson S, Polanowski T, Kessler R. Efficacy and tolerability of budesonide/formoterol added to tiotropium in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2009 Oct 15;180(8):741-50. Epub 2009 Jul 30.

Williams JEA, Singh SJ, Sewell L, Guyatt GH, Morgan MDL. Development of a self-reported Chronic Respiratory Questionnaire/CRQ-SR. *Thorax* 2001;56:954-9.

Williams M, De Palma L, Cafarella P, and Petkov J. Fifteen-count breathlessness score in adults with COPD. *Respirology* 2006; 11:627-32.

Wilson CB, Jones PW, O'Leary CJ, Cole PJ, Wilson R. Validation of the St. George's Respiratory Questionnaire in bronchiectasis. *American Journal of Respiratory and Critical Care Medicine* 1997;156:536-41.

Wyrwich KW, Tierney WM, Babu AN, Kroenke K, and Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. *Health Services Research* 2005;4 0:577-91.

Yeo J, Karimova G, and Bansal S. Co-morbidity in older patients with COPD -- its impact on health service utilisation and quality of life, a community study. *Age and Ageing* 2006; 35:33-7.

Yohannes AM, Roomi J, Waters K, Connolly MJ. Quality of life in elderly patients with COPD: measurement and predictive factors. *Respiratory Medicine* 1998;92:1231-6

Yohannes AM, Baldwin RC, and Connolly MJ. Predictors of 1-year mortality in patients discharged from hospital following acute exacerbation of chronic obstructive pulmonary disease. *Age and Ageing* 2005; 34(5):491-496.