PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

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

Report to the Department of Health, 2009

Patient-Reported

OUTCOME

Measures

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

Anne Mackintosh
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/

CONTENTS

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

Glossary of abbreviations

6MWT	6-minute walk test
ACE	Angiotensin-converting enzyme
AMI	Acute myocardial infarction
BDI	Beck Depression Inventory
BEST	Beta-Blocker of Survival Trial
BPAP	Bi-level positive airway pressure
BNP	B-type natriuretic peptide
CHARM	Candesartan in Heart Failure: Assessment of Reduction in Mortality and Morbidity
CHAT	Chronic Heart failure Assessment Tool
CHF	Congestive heart failure
CHQ	Chronic Heart Failure Questionnaire
CORC	Cardiovascular Outcomes Research Consortium
CPAP	Continuous positive airway pressure
CR	Cardiac rehabilitation
DASI	Duke Activity Status Index
DHF	Diastolic heart failure
DIG	Digitalis Investigation Group
EECP	Enhanced External Counterpulsation
ELITE	Evaluation of losartan in the elderly
EPHESUS	Eplerenone postacute myocardial infarction heart failure efficacy and survival study
EQ-5D	EuroQol 5 Dimensions Index
EQ-VAS	EuroQol Visual Analogue Scale *or* EQ Thermometer
ESCAPE	Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterisation Effectiveness
EVEREST	Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study With Tolvaptan
HADS	Hospital Anxiety and Depression Scale
HF	Heart failure
HRQoL	Health-related Quality of Life
ICD	Implantable cardioverter defibrillator
ICC	Intraclass correlation coefficient
KCCQ[-os]	Kansas City Cardiomyopathy Questionnaire[-overall summary score]
LVAD	Left ventricular assist device
[LV]EF	[Left ventricular] ejection fraction
LV[S]D	Left ventricular [systolic] dysfunction
MADIT	Multicenter Automatic Defibrillator Implantation Trial
[M]CID	[Minimum] Clinically important difference
MCS	Mental Component Summary *(of the SF-36 and related measures)*
MLHFQ	Minnesota Living with Heart Failure Questionnaire
MOMENTUM	Multicenter Trial of the Orqis Medical Cancion System for the Enhanced Treatment of Heart Failure Unresponsive to Medical Therapy
MOS	Medical Outcomes Study
NYHA	New York Heart Association
PAC	Pulmonary artery catheterisation

PCS	Physical Component Summary *(of the SF-36 and related measures)*
POMS-F	Profile of Mood States-Fatigue
QoL	Quality of Life
RCT	Randomised controlled trial
SDHFQ	San Diego Heart Failure Questionnaire
SF-36	Medical Outcomes Study 36-Item Short-Form Health Survey *domains:* BP Bodily pain, GH General health, MH Mental health, PF Physical functioning, RE Role limitation-emotional, RP Role limitation-physical, SF Social functioning, V Vitality
SHF	Systolic heart failure
SIP	Sickness Impact Profile
SOLVD	Studies of Left Ventricular Dysfunction

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

Executive Summary

Aims of the report

The aims of this report are to review the evidence of patient-reported outcome measures (PROMs) for people with heart failure and to provide recommendations to the Department of Health of PROMs for heart failure 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 heart failure-specific instruments which were then presented to a multidisciplinary panel for discussion. The literature review of the evidence-base and the discussions of the multi-disciplinary panel underpin final recommendations to the Department of Health.

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions (Fitzpatrick et al., 2006). The report included a review of evidence regarding PROMs for heart failure 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 heart failure, followed by heart failure-specific PROM results. The report concludes with discussion and recommendations.

Results and short-list of PROMs for people with heart failure

The previous review reported evidence for the following generic PROMs:

a) EQ-5D
b) SF-36
c) SF-12
d) Sickness Impact Profile (SIP)

This update identified further evidence of performance for all of these instruments. An individualised measure, the Patient Generated Index (PGI), and an additional generic PROM, the SF-8, were identified with reporting of some evidence of performance.

Four heart failure-specific instruments were previously identified:

a) Chronic Heart Failure Questionnaire (CHQ)
b) Kansas City Cardiomyopathy Questionnaire (KCCQ)
c) MacNew (ex-QLMI: Quality of Life after Myocardial Infarction Questionnaire)
d) Minnesota Living with Heart Failure Questionnaire (MLHFQ

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

This update identified further evidence of performance for the CHQ, KCCQ, and MLHFQ, but not for the MacNew.

In carrying out the current update of evidence, three further heart failure-specific instruments were identified, and are reported briefly:

e) Chronic Heart Failure Assessment Tool (CHAT)
f) HeartQoL
g) San Diego Heart Failure Questionnaire

Some additional evidence was found for several heart failure-specific instruments which were identified in the previous review but had limited evidence of performance:

h) Cardiac Depression Scale
i) Duke Activity Status Index
j) Heart Failure Functional Status Inventory/HFFSI
k) LVD-36
l) Memorial Symptom Assessment Scale-Heart Failure
m) Multidimensional Index of Life Quality
n) Quality of Life in Heart Failure Questionnaire – QLQ-CHF
o) Quality of Life Index – Cardiac Version

Short-listed PROMs for discussion

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

1. SF-36
2. SF-12
3. EQ-5D
4. Minnesota Living with Heart Failure Questionnaire
5. Kansas City Cardiomyopathy Questionnaire

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 heart failure. 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 MLHFQ is recommended as heart failure-specific instrument. These two measures used together will provide complementary evidence of health status of people with heart failure in the context of potential population-level applications in the NHS. In making the final selection of PROMs considered suitable for piloting in the NHS, the DH will consider salient factors in addition to the evidence and multidisciplinary panel comments.

1: INTRODUCTION & METHODS

Background

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

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

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions, carer impact, and patient perceptions of quality (Fitzpatrick et al., 2006). The report included a review of evidence regarding PROMs for heart failure with some recommendations[2].

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

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

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

Structure of the report

The methods of the review update 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 heart failure, followed by heart failure-specific PROM results. The report concludes with discussion and recommendations for short-listed PROMs.

Methods for the review update

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 were 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 primary source of evidence was the bibliography compiled by the PROM group[3]. At the time of the review update, the PROM database comprised 16,054 records (up to December 2005) downloaded from several electronic databases (see Appendix A i.) using a complex search strategy (see Appendix A ii a.). These records had been assessed as eligible for inclusion in the bibliography and assigned keywords, including disease-group (e.g. cardiovascular). 872 records keyworded 'cardiovascular' were identified; a further 177 records with 'heart' *or* 'cardiac' *or* 'cardiovascular' in the title or abstract, but not keyworded 'cardiovascular', were checked. Of all these, 895 were eliminated after scanning abstracts and, in some cases, full text, leaving 154 to be considered for potential inclusion. 12 were included in the final update.

A further 14,296 records covering the period January 2006-July 2007 had been downloaded using a revised search strategy (Appendix A ii b.) but not assessed or keyworded. Abstracts and titles were searched using the terms 'heart' *or* 'cardiac' *or* 'cardiovascular'. This search generated 966 records, of which 887 were eliminated after scanning abstracts or full text. The remaining 79 records were considered for potential inclusion; 11 articles were included in the final update.

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

Supplementary searches included scanning the reference lists of key articles and reviews. This yielded 105 articles, of which 50 were included in the final update.

All 1142 titles of issues of the following key journals published between January 2007 and August 2008 were scanned:
- Heart and Lung
- Journal of Cardiopulmonary Rehabilitation
- Journal of Cardiac Failure
- Health and Quality of Life Outcomes
- Medical Care
- Quality of Life Research

14 records were considered for potential inclusion; 9 were included in the final update.

In addition, PubMed records for the past two years (i.e. September 2006-August 2008) were searched using the term 'heart failure' and the names of the instruments identified in the previous review[4]. These additional searches yielded 393 records, of which 56 were considered for inclusion; 42 were included in the final update. The recommendations are based on an assessment of the evidence from both the previous review and the update review combined. The previous review identified 89 articles. Results are presented in Table 1.

Exclusions

Studies using non-English versions of instruments were excluded. Instruments focussing on a single dimension, for example, the Specific Activity Scale (physical function) and the Yale Scale or Dyspnea-Fatigue Index, were not considered for shortlisting.

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	952	154	12
Additional PROM database search (January 2006-July 2007) Total number = 14,296	966	79	11
Supplementary searching		175	101
TOTAL		**408**	**124**
2006 review			**89**
TOTAL			**213**

[4] Because of variants in instrument names, the terms 'Minnesota' and 'heart failure' and 'Kansas' and 'heart failure' were used instead of MLHFQ and KCCQ, respectively. For the MOS instruments, the terms 'SF-36' *or* 'SF36', 'SF-12' *or* 'SF12' were used.

2: Generic PROMs evaluated with people with heart failure

The previous review reported evidence for the following PROMs:

a) EQ-5D/EuroQol
b) SF-36
c) SF-12
d) Sickness Impact Profile (SIP)

This update identified further evidence of performance for the EQ-5D, SF-12, SF-36, and SIP. Full details of the development, domains, and scoring methods of these generic measures are detailed in Appendix C.

a) EQ-5D

The previous review identified five studies evaluating the performance of EQ-5D in HF, with good evidence for validity but mixed results for responsiveness.

The present update identified a further seven studies with, again, some good evidence for validity and variable findings regarding responsiveness.

Construct validity of the EQ-VAS was supported in a study comparing the prognostic significance of four summary measures in advanced HF (Sullivan et al., 2007) where the EQ-VAS correlated significantly with patient-reported symptoms, functional and psychological status, and with the other summary measures. However, unlike the other three measures (KCCQ overall summary, a single GH item from the SF-36, Standard Gamble) the EQ-VAS failed to predict the study endpoint.

Responsiveness of the EQ-5D was supported in a study reporting the results of a RCT of cardiac rehabilitation (Austin J. et al., 2005); statistically significant score changes were found in the experimental group, but not the control group, at 24 weeks. At five-year follow-up (Austin J. et al., 2008a, 2008b), the EuroQol score showed both groups to have deteriorated, whilst the EQ-VAS showed no change, and MLHFQ scores suggested a small improvement had been sustained in the CR group. The authors suggest this discrepancy between the MLHFQ and the EuroQol may be due to the latter measuring symptoms and impairments of more advanced age not particularly related to HF, such as pain, which is not detectable by the MLHFQ.

Conflicting results for responsiveness of the EQ-5D and the MLHFQ were found in the HeartMed trial to evaluate a community pharmacist intervention (Holland et al., 2007): EQ-5D scores favoured the intervention group whilst MLHFQ scores favoured controls, but neither difference was statistically significant.

Responsiveness of the EQ-VAS was supported in a RCT comparing placebo and spironolactone (Berry et al., 2007), which showed a statistically significant reduction in HRQoL in the spironolactone-treated group (a finding consistent with other studies), although there were no differences according to clinical and functional measures.

A concurrent evaluation study comparing the relative responsiveness of the KCCQ, SF-12 and EQ-5D (Eurich et al., 2006) in a sample of patients with moderate to severe

HF found the EQ-5D to be less responsive than the SF-12, and considerably less responsive than the condition-specific measure.

b) SF-36

The 2006 review identified 21 studies of heart failure using the SF-36, most providing evidence of construct validity. Some evidence was found for responsiveness, and reliability when used with the HF population. There was, however, evidence of floor and ceiling effects, indicating a lack of precision, and some doubt regarding acceptability.

The present update identified a further 23 studies, most illustrating construct validity or responsiveness of the measure.

Internal consistency reliability of the SF-36 was confirmed in a study examining coping, meaning in life, and HRQoL in HF (Park et al., 2008), with Cronbach's alphas of 0.74 and 0.77 for the PCS and MCS, respectively. Cronbach's alphas of over 0.50, considered acceptable by the authors, are reported in a pilot study which tested a case management approach with elderly HF patients (Pugh et al., 2001); however, these values fall well below the standard of 0.70, generally considered the minimum for group comparisons (Fitzpatrick et al., 1998).

Content validity of the SF-36 was supported in a descriptive study of patient perceptions of QoL using open-ended questions (Paul & Sneed, 2002). Analysis of the interviews was compared with the content of the SF-36 and the MLHFQ; it was concluded that the two measures fully addressed the QoL issues identified by these patients, with the possible exception of the spiritual dimension.

Construct validity of the SF-36 was demonstrated in several studies by discrimination between known groups. SF-36 PH items and the PF subscale discriminated patients in NYHA functional class III or IV from those in classes I or II in the Digitalis Intervention Group (DIG) QoL substudy (Lader et al., 2003). Two studies examining gender differences in the QoL of patients with HF showed the SF-36 discriminated men and women. Chin & Goldman (1998) found that women had significantly lower PCS scores than men at one year following an acute admission for HF; there were similar findings for the PF subscale in a study of older adults with HF (Friedman, 2003).

A study of the characteristics of diastolic and systolic HF (Kitzman et al., 2002) showed that both groups had substantially lower SF-36 scores than general population norms; however, except for the GH subscale, the SF-36 - unlike the MLHFQ - did not discriminate patients with DHF and those with SHF. In a small trial of exercise training (Smart et al., 2007), SF-36 GH and RP subscales did discriminate patients with systolic dysfunction from those with diastolic dysfunction. A study of the prevalence and impact of diastolic HF (O'Mahony et al., 2003) showed that the SF-36 PCS but not the MCS discriminated DHF subjects from those with no DHF or LVSD.

Construct validity of the SF-36 is also supported by correlation with other measures. In the study by Park et al. (2008) there were strong relationships between MCS and PCS scores and the Perceived Personal Meaning Scale, as well as with NYHA class.

Significant correlations were found between each of the SF-36 domains and the CHAT (Dunderdale et al., 2008), except for MH and V.

Responsiveness of the SF-36 has been demonstrated in a wide range of interventions.

Two studies illustrate sensitivity of the SF-36 to depression in HF patients. In a RCT studying the impact of depression on an intervention to enhance self-efficacy in a mixed group of patients with chronic illness, including HF (Jerant et al., 2008), there were significant MCS score changes in the intervention group at follow-up, consistent with a self-efficacy measure. A study examining the effect of paroxetine on depression and QoL in HF (Gottlieb et al., 2007) found significant changes in the treatment group for GH, social functioning (SF), and MH components of the SF-36.

The SF-36 has been used to evaluate the impact of different case management approaches. In a study of telehome monitoring in patients with cardiac disease (HF or angina) (Woodend et al., 2008) both intervention group and controls had significantly improved SF-36 scores over time, with significant between-group differences at three months, consistent with findings on the MLHFQ. Responsiveness of the SF-36 was also indicated in the study by Pugh et al. (2001) with positive, though statistically non-significant, trends in SF-36 scores consistent with results on the 6MWT, NYHA classification, and hospital readmission rates. In a RCT examining the impact of a home-based intervention with a subset of chronic CHF patients (Stewart et al., 1999) SF-36 and MLHFQ scores showed general improvement in QoL over time among survivors, with significant between-group differences at three months. Responsiveness of the PF and GH scales was supported in a study of intensive home-based case management for seriously chronically ill patients (PhoenixCare) by Aiken et al. (2006). Responsiveness of four subscales was confirmed in another home-based disease management study (Todero et al., 2002) with significant changes on RP, BP, MH, and V subscales, and positive trends for PF and GH.

Responsiveness of the SF-36 was found in a study evaluating the impact of an intervention to promote physical activity (Brodie et al., 2003), with significant changes in PF and V over time, and significant differences between groups on RP and SF subscales, and the health transition question. A small trial of exercise training comparing impact on systolic dysfunction (SD) and diastolic dysfunction (DD) (Smart et al., 2007) also illustrated responsiveness of the SF-36, with statistically significant improvement in scores post-training for patients with SD, but not for those with DD.

The SF-36 has also been used in trials of invasive procedures for HF. Responsiveness of the measure was demonstrated in a small pilot study of two modes of cardiac resynchronisation (Sirker et al., 2007), with significant improvements in both treatment groups and non-significant differences between them, consistent with a previous study. A small trial examining the impact of nocturnal CPAP on systolic heart function in CHF patients (Mansfield et al., 2004) found medium to large treatment effect sizes on all eight domains of the SF-36. Responsiveness of the PF domain was demonstrated in the Acorn trial of the CorCap cardiac support device (Mann et al., 2007) with significantly greater improvement in the treatment group. Both RP and RE subscales showed significant change in the intervention group in a RCT of long-term use of a LVAD in end-stage HF (Rose et al., 2001); this was

consistent with the Beck Depression Inventory (BDI) and NHYA functional class but not with the MLHFQ, which showed non-significant changes.

Sensitivity to change of selected SF-36 PH and PF items was not supported in the DIG QoL substudy (Lader et al., 2003), although given the absence of significant change in the other measures used, this may have been due to modest effects of the drug intervention.

Responsiveness of the SF-36 PCS was not supported in a RCT of behavioural management versus usual care (Shively et al, 2005), where the PCS failed to show any significant change but the MLHFQ physical functioning score did show change. Similar results were found in a RCT comparing usual (UC) care with transitional care (TC) in CHF patients discharged from hospital (Harrison et al., 2002). Despite statistically significant improvements on all components of the MLFHQ in the TC group relative to the UC group, there were no significant differences in SF-36 scores (PCS, MCS and GH).

Wyrwich et al. (2007) attempted to obtain clinically important differences (CIDs) for the SF-36 and the CHQ by comparing the perspectives of an expert physician panel, outpatients with heart disease, and primary care physicians. Small CIDs for the SF-36 varied greatly – suggesting that the measure may lack the sensitivity or precision needed to capture changes in HRQoL at the individual level.

Acceptability of the SF-36 was indicated in the QoL substudy of the DIG trial (Lader et al., 2003) where 90% and 78% of responses were complete at 4 and 12 months, respectively. However, this study used only part of the measure (two GH items and the PF subscale).

c) SF-12

The 2006 review identified eight studies which reported good evidence of construct validity and some evidence for reliability. There were mixed results regarding responsiveness, though no reported floor or ceiling effects.

The present update identified a further ten studies, mainly supporting construct validity and responsiveness of the SF-12.

Test-retest reliability of the SF-12 was moderate (ICC 0.59, where the standard is 0.70) in a study comparing several QoL measures used with CHF patients (Witham et al., 2007).

Construct validity of the SF-12 was supported in a study of patients with advanced heart disease referred for heart transplant (Evangelista et al., 2005), where the MCS discriminated between medically stable patients and other patient groups, and both PCS and MCS discriminated men and women. A strong relationship was also found between SF-12 PCS and MCS and depression (measured by the Beck Depression Inventory, BDI) although it was felt the SF-12 did not fully capture the psychological domain. The SF-12 PCS discriminated patients with HF and their spouses in a study of depression and QoL in HF (Mårtensson et al., 2003). In the patient group, there were also significant correlations between the SF-12 MCS and NYHA class, and between SF-12 PCS and depression (BDI).

Construct validity of the SF-12 was also demonstrated by a strong correlation with NYHA class in the study by Witham et al. (2007), and in a study of spirituality and HF (Quinn-Grifffin et al., 2007) where the PCS, but not the MCS, discriminated between HF and non-HF subjects.

A concurrent evaluation study comparing the relative responsiveness of the KCCQ, SF-12 and EQ-5D (Eurich et al., 2006) found the SF-12 to be much less responsive than the condition-specific measure, though somewhat more responsive than the EQ-5D. There were poorer results for the PCS than for the MCS.

Responsiveness of the SF-12 was partially supported in the MADIT II trial (Piotrowicz et al., 2007) where PCS (but not MCS) scores improved significantly. There were similar findings (significant change in PCS but not MCS) in a pre-post study examining the impact of a self-management programme (Harrison et al., 2007); the authors comment that the SF-12 may not have been sensitive to CHF-specific aspects of HRQoL.

Statistically and clinically significant changes in both PCS and MCS were found in a large, non-experimental study of the impact of disease management programmes on the QoL of patients with chronic health conditions, including heart failure (Walker et al., 2003). Responsiveness of the SF-12 MCS was also demonstrated in the evaluation of an education programme for older HF patients (Westlake et al., 2007), in which MCS scores over the three-month study period improved in the intervention group, and declined in the control group.

Responsiveness of the SF-12 was further indicated by a statistically non-significant trend towards improvement in the scores of the intervention group in a RCT of an electronic home monitoring device for patients with advanced HF (Goldberg et al., 2003), consistent with findings on the MLHFQ.

Acceptability of the SF-12 was found to be high, with a 99% completion rate in the study by Witham et al. (2007).

d) Sickness Impact Profile (SIP)

The 2006 review the SIP found four studies, which showed generally good results for validity and responsiveness.

The present update identified four further studies, which provide evidence of construct validity and responsiveness for the SIP.

Construct validity of the 68-item version of the SIP was supported in a prospective observational study of patients with advanced CHF (Blinderman et al., 2008) which found strong correlations between all components of the SIP, with the exception of the somatic autonomy subscale, and the Multidimensional Index of Life Quality composite score (the primary outcome measure in this study).

Construct validity of the 136-item SIP was demonstrated in a cross-sectional study examining the quality of life and psychological status of HF patients following ICD implantation (Friedmann et al., 2006). Overall SIP scores were highly correlated with

depression and anxiety (BDI, State-Trait Anxiety Inventory); there were also significant correlations between age and emotional behaviour, and sleep and rest subscales of the SIP (poorer scores in younger patients). Overall SIP and SIP physical dimension scores also discriminated patients who had had an ICD for over a year from those with ICDs for less than a year.

Responsiveness of the SIP was demonstrated in a substudy of the ELITE trial of losartan and captopril in patients with symptomatic HF (Cowley et al., 2000), which found significant score changes in both treatment groups. Better HRQoL scores were reflected in improvements in NYHA functional class and HF hospitalisation rates.

Responsiveness of the SIP was partially supported in a RCT of ACE inhibitors (Bulpitt et al., 1998) where the SIP showed non-significant changes, consistent with other measures (Profile of Mood States, Mahler dyspnoea index, and a disability index). However, the authors of this study suggest the SIP may not be sufficiently sensitive for use in mild to moderate CHF; they nevertheless recommend its use in future HF trials with large enough samples to detect small effect sizes.

Other measures

Patient Generated Index (PGI)

In a study which the authors believe to be the first to test an individualised measure in HF patients, Witham et al. (2007) administered the PGI, along with the CHQ, MLHFQ and SF-12. Acceptability of all four measures was high, with completion rates of 95% and above; however, the PGI is a complex tool requiring administration by a trained interviewer, limiting its feasibility. The authors conclude that, despite strong content validity, moderate reproducibility (ICC 0.65) and construct validity (correlation with NYHA class), the ability of the PGI to capture change, i.e. its responsiveness, is low.

SF-8

No studies using the SF-8, an eight-item version of the more familiar SF-36, were included in the original review; this update identified two studies.

Construct validity of the SF-8 was supported in a study of a disease management programme for HF patients (Martin et al., 2007), where baseline PCS and MCS of programme participants were significantly below the US population norm.

In the same study (Martin et al., 2007), responsiveness was shown by significant improvement in PF and MH at three months' follow-up, with sustained but less marked improvement through one year, in contrast to a MOS reference group which had undergone usual care.

Responsiveness of the SF-8 was further illustrated in a large, non-experimental study of the impact of four disease management programmes on the QoL of patients with chronic health conditions, including heart failure (Walker et al., 2003). Using either the SF-12 or the SF-8 (which the authors describe as having replaced the SF-12 from October 2001), statistically and clinically significant changes in PCS and MCS were found for HF patients.

3: Heart failure-specific PROMs

This chapter firstly provides a summary of the more recent evidence found for three of the four heart failure-specific instruments previously identified:

a) Chronic Heart Failure Questionnaire
b) Kansas City Cardiomyopathy Questionnaire
c) MacNew (ex-QLMI: Quality of Life after Myocardial Infarction Questionnaire
d) Minnesota Living with Heart Failure Questionnaire

In carrying out the current update of evidence, two further heart failure-specific instruments – the Chronic Heart Failure Assessment Tool, CHAT, and the HeartQoL were identified. These newly identified instruments and their evidence-base (where available) are each briefly described.

Findings in respect of eight previously identified heart failure-specific instruments with limited evidence are also presented.

Full details of the development, domains and scoring methods of the principal previously identified instruments are detailed in Appendix D.

a) Chronic Heart Failure Questionnaire (CHQ)
The 2006 review identified ten studies using the CHQ, supporting the reliability, validity, responsiveness, and precision of the instrument.

The present update identified eight further studies, confirming these properties.

Internal consistency reliability of the CHQ was confirmed as satisfactory in a study examining social support and HRQoL in patients with CHF (Bennett et al., 2001); Cronbach's alphas ranged 0.78-0.88 at baseline and 0.79-0.92 at 12-months follow-up. A study examining the relationship of age and sex to HRQoL in HF using the CHQ and the MLHFQ (Hou et al., 2004) found internal consistency reliability to be even higher, with Cronbach's alpha for the CHQ ranging 0.89-0.95. Test-retest reliability of the CHQ was high (ICC 0.93) in a study by Witham et al. (2007).

Construct validity of the CHQ was supported in the study by Bennett et al. (2001), with strong correlation between CHQ scores and changes in social support. Close correlation between instrument scores and severity of depression in the majority of patients was shown in a study using the combined CHQ-CRQ (Chronic Heart Failure Questionnaire-Chronic Respiratory Disease Questionnaire) to examine major depression and physical illness trajectories in a mixed group of patients with HF and pulmonary disease (Koenig et al., 2006). The CHQ discriminated between older and younger patients, and between men and women (known groups validity) in the study by Hou et al. (2004), and there was strong correlation between CHQ and NYHA class in the study by Witham et al. (2007).

Responsiveness of the CHQ was supported in a study evaluating an intervention to prevent readmission of elderly HF patients (Rich et al., 1995), with significant improvements on each of the CHQ subscales for the treatment group. A RCT of nocturnal CPAP on systolic heart function in CHF patients also supported responsiveness, with medium to large treatment effect sizes in three of four domains

of the CHQ (Mansfield et al., 2004). A study examining the relationship of age and sex to HRQoL in HF (Hou et al., 2004) also supports responsiveness, with significant changes over time on the total scale, and fatigue and emotional subscales. This was not reflected in the MLHFQ scores; the authors suggest that this may be because the CHQ is more sensitive in detecting individual changes in condition.

Wyrwich et al. (2007) attempted to obtain clinically important differences (CIDs) for the CHQ and the SF-36 by comparing the perspectives of an expert physician panel, outpatients with heart disease, and primary care physicians. The small CID for the CHQ domains was consistently one to two points – indicating precision.

Acceptability of the CHQ was high (100% completion rate) in the study by Witham et al, (2007).

Feasibility of the CHQ has been questioned, given the complexity of the instrument and the need for administration by a trained interviewer (Nanda & Andresen, 1998; Asadi-Lari et al., 2005).

b) Kansas City Cardiomyopathy Questionnaire (KCCQ)
13 studies were identified in the previous review, providing evidence of reliability, validity, and responsiveness.

The present update identified 19 further studies, mainly supporting construct validity of the KCCQ.

Test-retest reliability for the overall summary score of the KCCQ (KCCQ-os) was supported with an ICC of 0.88 in a study examining the psychometric properties of the KCCQ when used in HF patients with anaemia (Spertus et al., 2008). Internal consistency reliability was also confirmed in this study, with Cronbach's alphas of 0.92 and 0.93 for anaemic and non-anaemic patients, respectively.

Construct validity of the KCCQ was supported by correlation with known groups (NYHA classification) in the study by Spertus et al. (2008), and in an examination of the prevalence and burden of symptoms in a community-based sample of HF patients (Barnes et al., 2006). The latter found significant associations between lower KCCQ Total Symptom scores and being female, NYHA class, number of co-morbidities, and symptoms of depression (Geriatric Depression Scale). The KCCQ discriminated older and younger patients, and those whose NYHA functional status changed over time (Masoudi et al., 2004). Construct validity was also indicated in the SPPARO[5] trial (Earnest et al., 2004) where KCCQ scores were in line with those of a reference group.

The KCCQ physical impairment scale discriminated depressed and non-depressed patients in a study examining attitudes to impairment and depression in older HF patients (Turvey et al., 2006). A study based on the same sample (Klein et al., 2007) found that all KCCQ domains discriminated NYHA class, some coping styles, and (except for the QoL domain) comorbid illness. KCCQ score was an independent predictor for developing depressive symptoms in a study by Havranek et al. (2004) on

[5] System Providing Patients Access to Records Online

patients enrolled in the KCCQ Interpretability Study (Spertus et al., 2002). Construct validity of the KCCQ was further supported in a study of depression and symptoms in HF (Bekelman et al., 2007) where the number of symptoms was strongly inversely associated with KCCQ-os.

Significant relationships between KCCQ scores and anaemia were found in studies examining data from the QoL substudy of the EPHESUS trial, a large international RCT of aldosterone blockade in patients with HF post-AMI (Kosiborod et al., 2008). Construct validity of the KCCQ was also supported in a study comparing the prognostic significance of four summary measures in advanced HF (Sullivan et al., 2007) where the KCCQ summary score correlated significantly with clinical variables, and with the other measures.

Construct validity of the KCCQ was supported in a related study (Heidenreich et al., 2006) examining the prognostic value of the KCCQ-os in a large sample of outpatients, which stratified scores in 25-point ranges. There were significant associations between KCCQ scores and a range of clinical variables; one-year mortality was four-fold greater, and hospitalisation was five-fold greater, in those scoring less than 25 compared with those scoring 75 or above. This sample included patients with a broad range of HF aetiologies, increasing the generalisability of these findings.

Construct validity of the KCCQ was further demonstrated in a study by Subramian et al. (2008) which found the KCCQ-os was negatively associated with adherence to exercise, and in a small pilot study (Sayers et al., 2008) to develop an index of clinical exacerbations as a surrogate end point in HF research, where there was a significant correlation between the KCCQ and the proposed index.

Responsiveness of the KCCQ was supported by the preliminary findings of a CORC study to establish a MCID for the KCCQ (Spertus et al., 2002), which showed the instrument was at least as sensitive to change as traditional measures (6MWT and NYHA class). Significant changes in KCCQ-os and some KCCQ domains, with positive trends in the remainder, were found at study end in the EVEREST trial (Konstam et al., 2007).

Further support for KCCQ responsiveness was found in studies drawing on EPHESUS data. Change in KCCQ score was linearly associated with all-cause mortality, and with cardiovascular mortality or hospitalisation. This was true for each five-point decrease in KCCQ score, supporting interpretability of the KCCQ (Kosiborod et al., 2007).

In a similar study examining the prognostic value of the KCCQ (Soto et al., 2004) patients were (as in the Heidenreich study) stratified by range of KCCQ-os: those with scores of 75 or above were classified as ‘low-risk’, while those with scores below 25 were defined as ‘high-risk’. One-year event-free survival was significantly associated with KCCQ score-ranges. Soto and colleagues suggest that the KCCQ-os could usefully replace the NYHA classification - widely agreed to be crude and poorly standardised, although it is almost invariably used as a reference - as the instrument of choice for rapid clinical review.

Responsiveness of the KCCQ was also supported in the MOMENTUM trial of continuous aortic flow augmentation (CAFA) (Greenberg et al., 2008), where KCCQ scores showed non-significant increases in the treatment group, in line with clinical variables. Mean change in KCCQ score, though not clinically significant (less than five points), was linearly related to reported health change in the study by Spertus et al. (2008).

The relative responsiveness of the KCCQ, SF-12 and EQ-5D was examined in a concurrent evaluation study (Eurich et al., 2006) using the 6MWT, NYHA classification and clinician global rating of change as external criteria. Four responsiveness statistics were used: t-test, effect size (ES), Guyatt's responsiveness statistic, and standardised response mean (SRM). The KCCQ was found to be the most responsive, though for all three measures, responsiveness indices were larger for subjects who improved than for subjects who deteriorated, indicating possible floor effects; however, this may have been due to the (severely affected) clinical status of the sample.

Good acceptability of the KCCQ was indicated by an 80% response to a mailed survey (Subramian et al., 2008).

c) MacNew

Three studies were found in the previous review, providing some evidence of validity and responsiveness.

The present update identified no further studies evaluating the use of the MacNew with HF patients, though several with coronary heart disease patients.

d) Minnesota Living with Heart Failure Questionnaire (MLHFQ)

The previous review found 28 studies examining the MLHFQ. Internal reliability, construct validity, and responsiveness to change were all found to be good. However, some items appeared to need revision.

The present update identified a further 63 studies, the vast majority illustrating responsiveness of the MLHFQ, with considerable evidence also for construct validity.

Test-retest reliability of the MLHFQ was high (ICC 0.89), according to Witham et al. (2007).

High internal consistency reliability for the MLHFQ was found in a number of studies, with Cronbach's alpha ranging 0.86-0.95 (De Jong et al., 2005; Heo et al., 2008; Hou et al., 2004; Shuldham et al., 2007; Stephen, 2008).

Content validity of the MLHFQ was supported in a descriptive study of patient perceptions of QoL using open-ended questions (Paul & Sneed, 2002). Content analysis of the interviews was compared with the content of the MLHFQ and the SF-36; the authors conclude that the two measures fully addressed the QoL issues identified by this sample of patients, with the possible exception of the spiritual dimension.

Construct validity of the MLHFQ was supported in a number of studies by significant correlations with other measures.

Statistically significant correlation with NYHA class was found in studies by Witham et al. (2007) and De Jong et al. (2005); the latter also found a strong association with depression and anxiety, as measured by the Brief Symptom Inventory. The instruments used in the BEST study (Tate et al., 2007), namely the MLHFQ, SDHFQ (San Diego Heart Failure Questionnaire), Patient Global Assessment and Physician Global Assessment, were strongly correlated with each other; all four instruments predicted prognosis. In the study to develop the CHAT (Dunderdale et al., 2008), significant correlations were found between the new instrument and all aspects of the MLHFQ.

Correlation between the MLHFQ and measures of fatigue further support construct validity of the measure. Evangelista et al. (2008) found strong correlations between fatigue (Profile of Mood States-Fatigue, POMS-F) and MLHFQ total and subscale scores. Stephen (2008) found that HRQoL as measured by the MLHFQ was positively correlated with fatigue intensity (POMS-F), and that HRQoL scores were significantly higher (worse) in subjects who attributed their fatigue to aging. MLHFQ was strongly correlated with NYHA, and measures of physical symptom status (Dyspnea and Fatigue Index), perceived control (Cardiac Attitudes Index), and anxiety and depression (Brief Symptom Inventory) in a study by Heo et al. (2008).

Construct validity of the MLHFQ was partially supported in two studies of patient preferences for end-of-life care in HF. Weak correlations, consistent with other measures, were found between MLHFQ scores and patient preferences in a study by MacIver et al. (2008), although there was a positive trend towards patients with worse MLHFQ scores preferring more aggressive (life-shortening) treatment. In a similar study (Lewis E. et al., 2001), poor MLHFQ scores were strongly associated with utility scores (Standard Gamble and Time Trade-Off).

The ability of the MLHFQ to discriminate known groups was illustrated in several studies.

In a study comparing the HRQoL of older adults with and without HF (Heo et al., 2007a), the MLHFQ overall summary and subscales all discriminated older adults with HF from healthy older adults. An exploration of the relationships among age, sex and QoL in HF (Hou et al., 2004) found the MLHFQ discriminated between older and younger patients, and between men and women; Majani et al. (2005) found that MLHFQ scores discriminated men and women, but not age-group. In a study examining the association between body weight and self-reported HRQoL and depression in HF (Evangelista et al., 2006), MLHFQ was significantly correlated with body mass index (BMI); MLHFQ also discriminated men and women, and NYHA functional class. However, a study of gender differences in symptoms and HRQoL among HF patients (Heo et al., 2007b) found the MLHFQ failed to discriminate men and women, although there were significant correlations between depression (Brief Symptom Inventory) and MLHFQ in men, and between physical symptom status (Dyspnea and Fatigue Index) and MLHFQ in women.

The MLHFQ also discriminated patients with different severities of illness.

Significant correlations were found between MLHFQ scores and NYHA classification, comorbidities, and COPD (but not LVEF) for male patients in a study by Balashov et al. (2008); for women and men taken together, the results were similar but the correlations were not statistically significant. MLHFQ discriminated patients in NYHA functional class III or IV, and those in classes I or II in the Digitalis Intervention Group (DIG) QoL substudy (Lader et al., 2003). A pilot study of CPAP vs BPAP in patients with HF and obstructive sleep apnoea (OSA) (Khayat et al., 2008) found a small to moderate correlation between MLHFQ and improvement in LVEF. In a study of HRQoL in HF patients with low vs preserved ejection fraction (EF) (Lewis E. et al., 2007), MLHFQ scores were not influenced by EF but there were strong correlations with age, sex, dyspnoea, NYHA class and other clinical variables. A study of the characteristics of diastolic and systolic HF (Kitzman et al., 2002) found that MLHFQ scores were worse in patients with SHF than in those with DHF; both groups had substantially worse QoL ratings than benchmark values for the MLHFQ derived from the SOLVD Prevention trial (Rector et al., 1993). All components of the MLHFQ discriminated patients with DHF and those with SHF (higher, i.e. worse, scores in SHF) in a small trial of exercise training (Smart et al., 2007).

Responsiveness of the MLHFQ has been demonstrated in a large number of trials, with wide variety of interventions.

A RCT of cardiac rehabilitation (Austin J. et al., 2005) found both the experimental (CR) group and controls had improved MLHFQ scores at 8 and 24 weeks, but the improvement in the CR group was much greater. A five-year follow-up of this study (Austin J. et al., 2008a, 2008b) found a small improvement in MLHFQ scores to have been maintained in the CR group, whilst the EQ-5D found both groups to have deteriorated. It is suggested that this discrepancy between the two measures is due to the EQ-5D measuring symptoms and impairments of more advanced age not particularly related to HF, such as pain, which is not detectable by the MLHFQ.

The MLHFQ showed significant improvement at six months in a trial of the Paracor HeartNet device (Klodell et al., 2008), reflected in better 6MWT. A trial of enhanced external counterpulsation (EECP) (Feldman et al., 2006) showed significant improvement in MHLFQ at three months, consistent with improvement in NYHA class and increased exercise tolerance, but this was not sustained at six months - a result possibly attributable to the placebo effect. Significant and sustained improvement in MLHFQ scores in both treatment group and controls, consistent with change in NYHA classification, was found in a small trial of a cardiac device (Neelagaru et al., 2006); this again suggests the MLHFQ reflected a powerful placebo effect. A small trial of CPAP (Ferrier et al., 2008) showed no change in MLHFQ scores, consistent with results for exercise capacity.

Responsiveness of the MLHFQ was also supported in the ESCAPE trial of PAC effectiveness (Binanay et al., 2005; Shah et al., 2001) where MLHFQ scores showed greater improvement in the PAC group at one month, and in a small trial of biventricular pacing (Conti & Sears, 2007) which showed significant improvements in MLHFQ scores three months after implantation of the device, consistent with improvements in cognitive functioning and LVD-36 scores. Responsiveness for the MLHFQ was partly supported in a RCT of long-term use of a LVAD in end-stage HF

(Rose et al., 2001), with non-significant improvement in scores for the intervention group at one year; however, the difference in mean scores was 17 points, greatly exceeding the 5-point score change regarded as the MCID for the MLHFQ. MLHFQ scores showed significant improvement in the treatment group compared with controls in a trial of the CorCap cardiac support device (Mann et al., 2007), and in a small study using the Jarvik 2000 as a bridge-to-transplant (Miller et al., 2004).

Responsiveness of the MLHFQ was further supported in a large RCT of cardiac resynchronisation (Abraham et al., 2002); compared with controls, the intervention group had significantly improved MLHFQ scores at six months, with changes of similar magnitude in the other primary endpoints (NYHA class, 6MWT). A pilot study of the impact of cardiac resynchronisation therapy on OSA in HF patients (Stanchina et al., 2007) showed a small decrease (improvement) in MLHFQ scores; this was not reflected in improvement in measures of sleep quality. Significant change in overall MLHFQ and both dimensions was found at one week, but in the emotional dimension only at six months, in a small feasibility study of EECP (Soran et al., 2002).

Responsiveness of the MLHFQ has been demonstrated in numerous drug trials.

In a large international trial of the angiotensin-receptor blocker valsartan (Cohn & Tognoni, 2001; Majani et al., 2005), mean MLHFQ scores (overall and domain) worsened in the placebo group, the score change being statistically significant although below the 5-point change generally regarded as clinically meaningful in applications of the MLHFQ. This positive trend was reflected in improvements in LVEF, NYHA functional class, and signs and symptoms of HF (Cohn & Tognoni, 2001). A substudy of the ELITE trial of losartan and captopril in patients with symptomatic HF (Cowley et al., 2000) found significant score changes in both treatment groups; better HRQoL scores were reflected in improvements in NYHA functional class and HF hospitalisation rates. Responsiveness of MLHFQ was also indicated in a pilot study (Gottlieb et al., 2007) examining the effect of paroxetine on depression and QoL in HF, with significant improvements on MLHFQ emotional function and total scores for the treatment group.

Significant improvement in MLHFQ scores was found in a small trial of the impact of sildenafil on exercise capacity (Lewis G. et al., 2007), and in a pilot study on the impact of GLP-1 (glucagon-like peptide-1) on LVEF and functional capacity (Sokos et al., 2006). A trial of spironolactone in mild HF produced significantly worse MLHFQ and HADS scores in the treatment group, although cardiac function improved as hypothesised (MacDonald et al., 2004); this finding suggests the two instruments detected adverse side-effects of the drug. Responsiveness of the MLHFQ was also indicated in a trial of intravenous iron for CHF patients with anaemia (Bolger et al., 2006), with a significant fall (improvement) in MLHFQ scores, consistent with 6MWT results and NYHA class, and strongly correlated with increased haemoglobin levels.

The MLHFQ has also demonstrated responsiveness in trials of various disease management approaches.

MLHFQ and SF-36 scores showed general improvement in QoL over time among survivors, with significant between-group differences at three months, in a RCT examining the impact of a home-based intervention with chronic CHF patients (Stewart et al., 1999). A study of telehome monitoring in patients with cardiac disease (HF or angina) (Woodend et al., 2008) found both intervention group and controls had significantly improved MLHFQ scores over time, with significant between-group differences at three months (though not at one year), consistent with SF-36 results. Responsiveness of the MLHFQ was also indicated by a statistically non-significant trend towards improvement in the intervention group, consistent with findings on the SF-12, in a 6-month trial of an electronic home monitoring device for patients with advanced HF (Goldberg et al., 2003).

Further evidence of responsiveness was seen in a RCT comparing usual care (UC) with transitional care (TC) in CHF patients discharged from hospital (Harrison et al., 2002), with statistically significant improvements on all components of the MLFHQ for the TC group. Similar results were found in a RCT of a disease management programme designed for use with HF outpatients in primary care, the Oregon model (Hershberger et al., 2005). A RCT of TC for older adults hospitalised with HF (Naylor et al., 2004) showed significant improvement for the intervention group on MLHFQ physical subscale and overall scores at two weeks, sustained at 12 weeks on the overall score.

Responsiveness of the MLHFQ was also supported in a pre-post study examining the impact of a self-management programme (Partners in Care for CHF, PCCHF; Harrison et al., 2007), with statistically significant improvements in emotional and physical dimension, and total scores. Responsiveness of the MLHFQ physical functioning (PF) score was supported in a RCT of behavioural management versus usual care (Shively et al, 2005), where the PF score showed better outcome in the intervention group. A pilot study of a computerised programme to educate patients about medications for heart failure (PUMP UP[6]; Bennett et al., 2006) demonstrated significant improvements over time in the physical dimension and overall scores of the MLHFQ (average change score 14 points, where 5 points is considered clinically meaningful).

Mean MLHFQ scores improved significantly in both intervention groups in a study comparing two modes of home health-care delivery (Benatar et al., 2003), with non-significant between-group differences. In a trial of community case management (CCM) in HF patients with and without preserved ejection fraction (EF), the intervention group had significantly better MLHFQ scores than those receiving usual care, regardless of EF status (Moser et al., 2000). Responsiveness of the MLHFQ was also supported in a RCT comparing a multidisciplinary disease management programme with usual care for elderly women with HF (Azad et al., 2008) where, consistent with other measures (Physical Self-Maintenance Scale, Geriatric Depression Scale), the MLHFQ failed to show a difference between intervention group and controls.

Several trials of exercise interventions show responsiveness of the MLHFQ.

[6] Promoting Understanding and Management through Partnering: 'U' and your Physician

There were significant reductions (improvement) in scores for the intervention group in two randomised trials of T'ai Chi with HF patients (Barrow et al., 2007; Yeh et al., 2008); in the latter, MLHFQ score changes were reflected in significant improvements in 6MWT and serum BNP levels. A study evaluating the impact of an intervention to promote physical activity (Brodie et al., 2003) found significant improvements in MLHFQ score for all groups over time - i.e. MLHFQ did not support effectiveness of the intervention. In the Exercise Rehabilitation Trial (EXERT) (McKelvie et al., 2002), MLHFQ showed statistically non-significant improvement in the intervention group, consistent with 6MWT results. A small trial of exercise training comparing the impact on systolic dysfunction (SD) and diastolic dysfunction (DD) (Smart et al., 2007) found MLHFQ scores discriminated SD and DD subjects, and showed statistically significant improvement post-training for SD, but only on the emotional subscale for DD.

A number of trials found ambiguous results regarding responsiveness to change of the MLHFQ.

Conflicting results for responsiveness of the MLHFQ and the EQ-5D were found in the HeartMed trial to evaluate a community pharmacist intervention (Holland et al., 2007): MLHFQ scores favoured the control group whilst EQ-5D scores favoured the intervention group, but neither difference was statistically significant.

Partial support for responsiveness of the MLHFQ was demonstrated in the African-American Heart Failure Trial (A-HeFT) (Taylor et al., 2006) where the measure was incorporated into a composite score weighting mortality, first hospitalisation for HF, and QoL. The composite score showed significant improvement with treatment, although MLHFQ score changes considered separately did not achieve significance, probably due to reduced sample size by subgroup analysis.

In the study by Hou et al. (2004), MLHFQ scores did not change significantly over time, in contrast to the CHQ, suggesting that the MLHFQ may be less sensitive than the CHQ in detecting change; however, the study was statistically underpowered to detect the 5-point MCID for the MLHFQ. Sensitivity to change of the MLHFQ was not supported in the DIG QoL substudy (Lader et al., 2003), although given the absence of significant change in the other measures used, this may be due to modest effects of the drug intervention. In the study by Tate et al. (2007) both the MLHFQ and SDHFQ, unlike the two single-item measures used concurrently, failed to differentiate between treatment groups, suggesting that they may be insufficiently responsive to change.

Acceptability of the MLHFQ was supported in a comparison of the HRQoL of patients with low, and patients with preserved, ejection fraction (EF) using data from the CHARM trial (Lewis E. et al., 2007). 88% of enrolled patients answered all 21 items of the MLHFQ, and 98% completed over 75% of items, indicating high acceptability to patients. In the QoL substudy of the DIG trial (Lader et al., 2003), 90% and 78% of responses were complete at 4 and 12 months, respectively.

New heart failure-specific instruments

In carrying out the current update of evidence, four further heart failure-specific instruments were identified, and are reported briefly:

e) Chronic Heart Failure Assessment Tool (CHAT)

According to its UK-based developers (Dunderdale et al., 2008), the CHAT is the first and only HF-specific measure to be derived from the patient's perspective, as opposed to being 'expert driven', and captures a wider range of psychosocial concerns than existing instruments. Having conducted a literature review (Dunderdale et al., 2005), the authors developed and tested the instrument in three phases. Questionnaire items were generated by in-depth, semi-structured interviews with CHF patients (n = 11), yielding a 51-item questionnaire (Dunderdale et al., 2007). This version of the questionnaire was tested in a postal survey of 345 CHF patients discharged from hospital; 223 responses were returned (a 65% response rate), suggesting moderate acceptability of the original CHAT.

Following this initial survey, factor analysis was conducted. Four factors were identified – namely, Symptoms, Activity levels, Psychosocial aspects, and Emotions – which explained 49% of the total variance. Internal consistency was calculated using Cronbach's alpha, which was >0.80 for each of these domains, indicating high internal consistency reliability.

Five redundant items were removed and the validity of a 46-item version of the CHAT was tested with a sample of 100 patients, using the SF-36 and MLHFQ as comparators. 68 responses were returned (68% response rate). There were significant correlations between the CHAT and each of the SF-36 domains, except V and MH, and between the CHAT and all aspects of the MLHFQ, indicating construct validity.

The CHAT includes six transition items asking respondents to compare their overall health and other elements with those at specific time in the past; respondents are asked to compare themselves with other healthy adults of the same age. For most items, a five-point response scale is used. Nine items use phrase completion scales with an 11-point scale.

The developers recognise the need for further research to determine the significance of changes in score (interpretability), and acknowledge that the generalisability of their findings is limited by the fact that the patient sample was drawn from one geographical region in the UK, and included few women and no members of ethnic minorities. The sample was also confined to patients following a hospital admission. However, the instrument may merit consideration following further testing.

f) HeartQoL

A research project supported by the European Society of Cardiology is underway to develop a core heart disease-specific HRQoL questionnaire for use among patients with myocardial infarction, angina pectoris, or heart failure (McGee et al, 2005; Oldridge et al., 2005). The instrument, to be known as the HeartQoL, will be translated into 13 European languages, allowing comparison of HRQoL outcomes across different cardiac populations and across European countries (somewhat comparable to the core cancer EORTC scale). The HeartQoL was expected to be available for use from 2007; however, at the time of writing this review, no results have been reported.

g) San Diego Heart Failure Questionnaire

This measure was not identified in the previous review. The current update found one study (Tate et al., 2007) citing Shabetai (1983) as the source for the instrument. The article by Shabetai is a general discussion of cardiomyopathy (definition, assessment and treatment) and gives the content of the SDHFQ but no details regarding its development and testing. The questionnaire includes 14 items on dyspnoea or fatigue in relation to specific physical activities and dysnpoea/fatigue at rest; there are nine questions concerning chest pain, and five items covering other symptoms, alcohol consumption, and weight.

The study by Tate et al. (2007) supported construct validity of the SDHFQ and three other measures applied (see MLHFQ above) but found that both the SDHFQ and the MLHFQ, unlike two single-item measures, failed to differentiate between treatment groups, suggesting that they may be insufficiently responsive to change.

Other specific instruments

Some additional evidence was found for several heart failure- or cardiovascular-specific instruments which were identified in the previous review but had limited evidence of performance – i.e. few studies and/or restricted scope.

h) Cardiac Depression Scale (CDS)

Two studies were identified in the previous review. The current update identified one study (Smart et al., 2007), a small trial of exercise training, which illustrated both construct validity and responsiveness of the measure. Consistent with the MLHFQ and some subscales of the SF-36, the CDS discriminated patients with systolic dysfunction and those with diastolic dysfunction (construct validity), and showed significant improvement in both groups over the course of the intervention (responsiveness).

i) Duke Activity Status Index (DASI)

Although it is not heart failure-specific, the DASI, which measures seven aspects of activity performance, is frequently used in studies of patients with cardiovascular disease. Six studies were identified in the previous review[7]; the current update found no further examples of use with HF patients.

j) Heart Failure Functional Status Inventory (HFFSI)

The previous review does not refer to this instrument; however, like the DASI, it is confined to physical functioning. The measure originally comprised a list of 25 activities (Dracup et al., 1992); for each, the patient is asked to respond with one of the following options: 1) 'Yes, I can do this'; 2) 'Yes, I can do this, but only slowly'; or 3) 'No, I can't do this'. If the response is 2) or 3), patients are asked to indicated whether the activity is limited by shortness of breath (SOB), weakness without SOB, fatigue, chest pain, or some other reason. Each activity is assigned a value according to the metabolic equivalent of the task (MET) – the exertion predicted for each level of activity, based on previous research; the HFFSI score is derived from the average of the three highest MET levels assigned to the activities patients report themselves as being able to perform. Dracup et al. (1992) show evidence of content validity, inter-rater reliability, and construct validity for the instrument.

[7] See Table 8.12 of the 2006 review

The current update identified three studies using the HFFSI. Internal consistency reliability (Cronbach's alpha 0.85) and construct validity (significant correlations between MET and NYHA, EF) were found for a 12-item version of the instrument (Porter et al., 1994). Construct validity was also indicated in a study of mood disturbance amongst patients awaiting ICD implantation (Dunbar et al., 1996), where women had significantly lower HFFSI scores than men. Inter-rater reliability was found to be high (98%) in a study using the 12-item version of the HFFSI, amongst other measures, to examine fatigue in older adults with stable HF (Stephen, 2008). However, no relationship was found between HFFSI score and fatigue intensity (VAS-Fatigue, POMS-F), indicating that the measure lacks sensitivity to this aspect of HF patient experience.

k) LVD-36

The previous review included the original development study for the LVD-36 (O'Leary & Jones, 2000). The present update identified one further study (Conti & Sears, 2007). Responsiveness of the instrument was supported in this small trial of biventricular pacing, with significant improvements in LVD-36 scores three months after implantation of the device, consistent with improvements in cognitive functioning and MLHFQ scores.

l) Memorial Symptom Assessment Scale-Heart Failure (MSAS-HF)

The previous review included one study relating to the MSAS-HF (Zambroski et al., 2005). The present update identified the original study to develop the MSAS-HF (Zambroski et al., 2004). Three other studies applied the MSAS with HF patients, but these used the original scale or its short form (Chang et al., 2000) rather than the specific adaptation.

In the study describing the adaptation of the MSAS for HF (Zambroski et al., 2004), the authors argue that existing instruments are inadequate to assess the full symptom experience for HF patients at the end of life. The original MSAS, developed by Portenoy et al. (1994) and widely used with cancer patients receiving palliative care, was felt to lack key HF symptoms, so the following were added: palpitations, chest pain, waking up breathless at night, weight gain, and difficulty breathing when lying flat. The revised scale, the MSAS-HF, has three subscales: HF symptoms (HFS), physical symptoms (PHYS), and psychological symptoms (PSYCH).

The MSAS-HF was administered to 36 HF patients and 133 healthy volunteers (HV), and found to discriminate between the two groups on each subscale. Of the items added, difficulty breathing lying flat was amongst the most distressing for this sample of HF patients, and some symptoms were detected which are absent from other HF-specific questionnaires (e.g. itching, pain), suggesting that the MSAS-HF may have greater content validity than existing measures. All patients completed the instrument without difficulty, indicating acceptability.

Construct validity of the original MSAS with HF patients was confirmed in a study by Bekelman et al. (2007), with strong associations between depression (Geriatric Depression Scale) and symptom distress, and in an observational study of patients with end-stage CHF (Blinderman et al., 2008) which found a strong correlation between the Global Distress Index of the MSAS and MILQ scores. Responsiveness of

the symptom control items was illustrated in a study of intensive home-based case management for seriously chronically ill patients (COPD and CHF) by Aiken et al. (2006).

Although there are relatively few studies supporting the use of the MSAS, or its condition-specific adaptation the MSAS-HF, in heart failure, the instrument(s) would appear to have particular relevance in advanced or end-stage HF, including items which are missing from other HF-specific measures.

m) Multidimensional Index of Life Quality (MILQ)

The previous review found a single study relating to this cardiovascular-specific instrument – namely, the original development study (Avis et al., 1996). The current update identified one further study (Blinderman et al., 2008) which examined symptom distress and QoL in patients with advanced CHF using the MILQ as the primary outcome measure. Construct validity of the MILQ was indicated by strong associations with symptom distress (MSAS), poor psychological well-being (Mental Health Inventory, MHI-5) and poor functional mobility (SIP).

n) Quality of Life in Heart Failure Questionnaire (QLQ-CHF)

The previous review identified the original development study for this measure (Wiklund et al., 1987). The current update identified one further study (Wiklund et al., 1996), a RCT of metoprolol in dilated cardiomyopathy, which supported construct validity (correlation with NYHA class) and responsiveness of the instrument. However, the authors appear to suggest the measure has been superseded by the MLHFQ.

o) Quality of Life Index – Cardiac Version

The previous review identified two studies using this instrument; the current update identified a further two studies. Internal consistency reliability and responsiveness were supported in a study of nursing interventions with HF patients (Scott et al., 2004). Responsiveness was also supported in a study comparing two modes of home health care delivery (Benatar et al., 2003).

4: SUMMARY, DISCUSSION & RECOMMENDATIONS

SUMMARY & DISCUSSION – generic instruments

The findings of the present update reflect those of the previous review (Fitzpatrick et al., 2006), in that the SF-36 is by far the most frequently used generic measure in heart failure. Twenty-three new studies were identified, with most providing evidence for construct validity or responsiveness. As observed in the previous review, there is less evidence for reliability of the SF-36 with the HF population.

There are mixed findings concerning the discriminative validity of the SF-36 with HF patients. Some studies suggest that the measure lacks sensitivity to small clinical change. This may be of the essence, particularly in the management of advanced heart failure, where the goal of treatment is generally to alleviate symptoms and limit progression of the disease, rather than to improve cardiac function. The previous review concluded that the SF-36 was more subject to floor and ceiling effects than the shorter SF-12.

The SF-12, which had the next highest number of evaluations with ten studies in the present update, may be an acceptable alternative to the longer instrument. However, it does not fully capture the psychological domain, an important consideration given the widely recognised reciprocal relationship between heart failure and depression.

Whilst there is some evidence for validity and responsiveness of the SIP, the relative paucity of information (four studies in the present update) and the length of the instrument argue against its use.

If a utility measure is required, the EQ-5D may be considered. However, there is less evidence to support its use in HF (seven studies in the present update) than for the MOS instruments. Although there is some good evidence for validity, it has less supportive results for responsiveness.

Table 2: **Appraisal of psychometric and operational performance of generic PROMs for people with heart failure**

	EQ-5D	SF-36	SF-12	SIP
Reproducibility	0	+	0	+
Internal consistency	0	++	++	+
Validity: content	0	+	0	0
Validity: construct	+++	+++	+++	++
Responsiveness	++	++	++	++
Interpretability	0	0	0	0
Floor/ceiling/precision	0	−	+	0
Acceptability	0	−	−	0
Feasibility	0	+	0	0

Key
– = evidence does not support criteria
0 = not reported or no evidence in favour
+ = some limited evidence in favour
++ = some good evidence in favour, but some aspects do not meet criteria or some aspects not reported
+++ = good evidence in favour

SUMMARY & DISCUSSION – heart failure-specific instruments

The previous review found that the MLHFQ was by far the most commonly used condition-specific instrument in heart failure. This is more than confirmed by the findings of the present update, with 63 studies identified, most supporting construct validity and/or responsiveness. Predominance of the MLHFQ is reflected in a recent major review of measures used in heart failure interventions (Morgan et al., 2007).

However, there is also strong evidence in favour of the KCCQ (19 studies), which may be more sensitive than the MLHFQ.

Despite some good psychometric evidence for the CHQ, in particular its sensitivity to changes in the individual, the relatively limited volume of evidence and complexity of administration of the measure make it difficult to recommend for routine use.

Evidence for the MacNew in heart failure is even more limited, with no studies found in the present update, though the measure may well have a role in other aspects of heart disease (e.g. CHD).

Given the clinical spectrum covered by 'heart failure', from asymptomatic LVD to chronic and severe symptomatic HF, it may be inappropriate to recommend a single condition-specific instrument. From the available evidence, it is clear that the MLHFQ has the most evidence to support its use, followed by the KCCQ. However, both these instruments may lack the ability to detect small but significant changes in functional status and QoL, essential to the management of chronic HF.

For patients with end-stage heart failure, MLHFQ and KCCQ do not cover the range of symptoms experienced – notably pain. Although both instruments include a psychological domain, they may not be adequate to measure depression and anxiety, which are relatively common in HF and strongly associated with increased morbidity and mortality as well as adverse QoL. Many of the studies reviewed for this update have included a dimension-specific measure to assess psychological well-being; examples include the Beck Depression Inventory, Brief Symptom Inventory, Hospital Anxiety and Depression Scale, and Geriatric Depression Scale.

Newly emerging instruments, such as the CHAT and the MSAS-HF, which attempt to cover the full range of HF patient experience, may merit consideration when more evidence to support their use is available. The general cardiovascular HeartQoL, intended for Europe-wide application, may also be of interest in future.

Table 3: **Appraisal of psychometric and operational performance of heart failure-specific PROMs**

	CHQ	KCCQ	MacNew	MLHFQ
Reproducibility	+	+	0	+
Internal consistency	++	+	+	++
Validity: content	0	0	0	+
Validity: construct	+++	+++	++	+++
Responsiveness	++	+++	++	++
Interpretability	+	+	0	0
Floor/ceiling/precision	++	+	0	+
Acceptability	+	+	+	+
Feasibility	−	0	0	0

Key
– = evidence does not support criteria
0 = not reported or no evidence in favour
+ = some limited evidence in favour
++ = some good evidence in favour, but some aspects do not meet criteria or some aspects not reported
+++ = good evidence in favour

RECOMMENDATIONS

Based on the evidence found, the previous review submitted in 2006 recommended the SF-36 and SF-12 as generic measures for use with people with heart failure, and the MLHFQ as the most appropriate heart failure-specific measure.

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).

Generic measures:

1. SF-36
2. SF-12
3. EQ-5D

HF-specific measures:

4. MLHFQ
5. KCCQ

The multi-disciplinary panel were favourable toward the SF-36 as a generic measure of health status and the MLHFQ as a heart failure-specific measure. They felt that neither instrument should be used in isolation if the full range of patient experience is to be captured. 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 MLHFQ for use in potentially large scale population studies. The multidisciplinary panel commented on the ease of use of EQ-5D. 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 MLHFQ.

APPENDIX A

i. 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.

ii. PROM Bibliography search strategy[8]

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$)))))

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

APPENDIX B: Psychometric criteria

Appraisal of PROMs

The methods that will be used for assessing the performance of PROMs were developed and tested against multidisciplinary 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).

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 independently by two reviewers using the following rating scale, and inter-rater reliability calculated.

Psychometric evidence:
– = evidence does not support criteria
0 = not reported or no evidence in favour
+ = some limited evidence in favour
++ = some good evidence in favour, but some aspects do not meet criteria or some aspects not reported
+++ = good evidence in favour

PROMs for which there are strong psychometric properties will be judged in terms of operational characteristics and clinical credibility.

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

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

This appendix provides a brief description of the four generic health status instruments reviewed in the update on patient-reported instruments for heart failure, and summarises their origins, development, and content. Content, format, and health status domains included are set out in table form at the end of this appendix.

i) 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 HRQoL, 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 4. 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.

ii) 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 & 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 & Sherbourne, 1992; Ware & Gandek, 1998; Jenkinson & 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 4. 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.

iii) 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 4). The same eight domains as the SF-36 are assessed and categorical response scales are used. A computer-based scoring algorithm is used to calculate scores: Physical Component Summary (PCS) and Mental (MCS) Component Summary scales are generated using norm-based methods. Scores are transformed to have a mean value of 50, standard deviation (SD) 10, where scores above or below 50 are above or below average physical or mental well-being, respectively.

Completion by UK city-dwellers reporting the absence of health problems yielded a mean PCS score of 50.0 (SD 7.6) and MCS of 55.5 (SD 6.1) (Pettit et al., 2001). Although not recommended by the developers, Schofield & 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 & Nahm, 2001; Resnick & Parker, 2001).

iv) 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 & 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 & 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 & Peach, 1989; McDowell & 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).

Table 4: **Generic PROMS used in heart failure - content and scoring**

Instrument	***Domains (no. items)***	***Response options***	***Score***	***Completion (time in minutes)***
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
SF-36: MOS 36-item Short Form Health Survey (36)	Bodily pain (BP) (2), General health (GH) (6), Mental health (MH) (5), Physical functioning (PF) (10), Role limitation-emotional (RE) (3), Role limitation-physical (RP) (4), Social functioning (SF) (2), Vitality (4) *plus a health transition question not included in final score*	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, s.d. 10)	Interview (mean values 14-15) Self (mean 12.6)
SF-12: MOS 12-item Short Form Health Survey (12)	Bodily pain (BP) (1), Energy/Vitality (V) (1), General health (GH) (1), Mental health (MH) (2), Physical functioning (PF) (2), Role limitation-emotional (RE) (2), Role limitation-physical (RP) (2), Social functioning (SF) (1)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, s.d. 10)	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)

Table 5: **Generic PROMs used in heart failure - summary of health status domains**

Instrument	***Instrument domains***							
	Physical function	**Symptoms**	**Global judgement**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**
EQ-5D (5+1)	x	x	x	x	x		x	
SF-36 (36)	x	x	x	x	x		x	
SF-12 (12)	x	x	x	x	x		x	
SIP (136)	x	x		x	x	x	x	

APPENDIX D: HEART FAILURE-SPECIFIC PROMs

This appendix provides a brief description of the four condition-specific health status instruments reviewed in the update on patient-reported instruments for heart failure, and summarises their origins, development, and content. Content, format, and health status domains included are set out in table form at the end of this appendix.

i) Chronic Heart Failure Questionnaire (CHQ) (Guyatt et al., 1989)

This 16-item instrument aims to measure subjective health status in heart failure patients, and is complex to administer as open-ended questions are used to yield score weights. It covers dyspnoea, fatigue, and emotional functions; it has a time recall period of two weeks. It was developed by presenting 123 items to a sample of 88 patients, who rated their importance. Item selection was based on frequency and importance ratings. A section of the CHQ is individualised, and patients are asked to nominate those activities associated with shortness of breath and that affect them most often/importantly. It requires a trained interviewer. Administration takes 10-20 minutes.

ii) Kansas City Cardiomyopathy Questionnaire (KCCQ) Green et al., 2000)

This instrument aims to describe HRQoL over the previous two weeks in patients with congestive heart failure (CHF). It contains 23 items, covering physical function, clinical symptoms, social function, self-efficacy and knowledge and QoL ('enjoyment'), each with different Likert scaling wording, including limitations, frequency, bother, change in condition, understanding, levels of enjoyment and satisfaction. It is self-administered. A change of 5 points on the scale scores, either as a group mean or an intra-individual change is regarded as clinically important (Rumsfeld et al., 2003).

iii) MacNew (ex-QLMI: Quality of Life after Myocardial Infarction Questionnaire) (Lim et al., 1993; Valenti et al., 1996)

While not solely heart failure-specific, MacNew measures HRQoL in heart disease (myocardial infarction, coronary disease and heart failure) in the previous two weeks. This instrument is a modification of the earlier Quality of Life after Myocardial Infarction (QLMI) Questionnaire, which had questionable validity (see review by Höfer et al., 2004). MacNew contains 27 items in three domains (Emotional, Physical, and Social). It takes up to 10 minutes to complete, and respondent burden is low.

iv) Minnesota Living with Heart Failure Questionnaire (MLHF/MLHFQ/LHFQ/LiHFe) (Rector et al., 1987)

This contains 21 items that ask about patients' perceptions of the effects of heart failure and its treatment on physical, socioeconomic and psychological aspects of their life, rated on a 6-point Likert scale. Subscale scores for emotional and physical domains can be obtained. It is easy to administer by self-administration or interview. The items were drawn from the SIP. Patients with congestive heart failure were asked to select 21 items from the SIP, and these formed the MLHFQ. Some concern has been expressed about its content validity and whether all relevant items have been included (Dunderdale et al., 2005; O'Leary & Jones, 2000).

Table 6: **Heart failure-specific PROMs - content and scoring**

Instrument	***Domains (no. items)***	***Response options***	***Score***	***Administration Completion time***
Chronic Heart Failure Questionnaire (CHQ) **(Guyatt et al., 1989)**	*16 items in 3 domains:* Dyspnoea (5) Fatigue (4) Emotional function (7) *Plus open-ended probes (3) for most important activities causing symptoms*	1-7 response scales of frequency or severity	Summed to yield subscale scores Weighting based on open-ended responses Minimum (worse function) to maximum (best function) scores in the 3 domains are: dyspnoea 5-35; fatigue 4-28; emotional 7-49	Interview 10-20 mins
Kansas City Cardiomyopathy Questionnaire (KCCQ) **(Green et al., 2000)**	*23 items in 5 domains:* 1. Physical limitation (6) 2. Symptoms (8) 3. Self-efficacy and knowledge (2) 4. QoL/mood (3) 5. Social limitation (4)	6-point Likert scales, including severity and frequency	Summation of physical limitation, symptoms, social limitation and QoL domains. 0-100, higher scores represent fewer symptoms/better function/better QoL	Self-administered 4-6 mins
MacNew (ex-QLMI – Quality of Life after Myocardial Infarction) **(Lim et al., 1993)**	*23-27 items in 3 overlapping domains:* Emotional Physical Social In previous 2 weeks	Item scores 1 = poor, to 7 = high	Summation; domain scores calculated by taking the average of responses to items in each domain; averaging all items gives a global score.	Self-administered (modification of original interviewer-administered QLMI instrument) 5-10 minutes to complete
Minnesota Living with Heart Failure Questionnaire (MLHFQ) **(Rector et al., 1987)**	*21 items on impact of heart failure on:* Physical aspects of daily life (9) Emotional/psychological (5) Social/economic (7) In previous 4 weeks	6-point Likert scales (0 = not at all, to 5 = very much)	Summation; range 0 (best) to 105 (worst QoL). Physical and emotional domains can also be summed.	Self-administered or interview

Table 7: **Heart failure-specific PROMs - health status domains**

Instrument	***Instrument domains***							
	Physical function	**Symptoms**	**Global judgement of health**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**
CHQ		x		x				
KCCQ	x	x		x	x		x	x
MacNew	x	x		x	x		x	x
MLHFQ	x	x		x	x		x	

APPENDIX E: Licensing information for recommended HF-specific instruments

KCCQ
Copyright is owned by Dr. John Spertus, University of Missouri-Kansas City spertusj@ukmc.edu. Licensing information can be found at http://www.cvoutcomes.org/licenses/. It appears necessary to submit a specific request before the fee is determined. However, it is possible to try out the questionnaire online; five repetitions are permitted before a license must be obtained.

MLHFQ
Copyright is held by the University of Minnesota. Health care organisations wishing to use the instrument in the care of patients or to evaluate in-house services are charged a license fee of USD 500. Not-for-profit researchers are charged USD 500 per project (USD 2,500 for for-profit entities).

The MLHFQ website - http://www.mlhfq.org/ provides an overview document with referenced information concerning instrument development, etc; instructions for data collection and scoring; a copy of the License Agreement; and a copy of the questionnaire.

APPENDIX F: Methods of working, membership, and conclusions of the multidisciplinary panel

Members of the multidisciplinary panel were invited to participate based on their clinical or research experience of heart failure and special interest in patient-reported outcome measures.

The panel were sent the following documents:

- A structured review of patient-reported outcome measures for people with heart failure: an update 2009;
- A structured review of patient-reported measures in relation to heart failure (2006);
- copies of the PROMs short-listed for discussion.

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

Notes of electronic discussion: March-April 2009

Generic measures

SF-36

Whilst it was accepted that the SF-36 had been extensively evaluated and applied with patients with heart failure, this was mainly in a research context. It was thought to contain clinically relevant domains and to be a good generic measure for use in conjunction with a disease-specific instrument.

However, several issues were raised regarding its application in the NHS. The scoring system and cost were considered potential barriers. Several alternative domains were suggested to be applicable to patients with heart failure. Several panel members commented on the difficulties older people with heart failure have in completing the SF-36. Missing data has been observed specifically relating to questions about work and physical functioning as they seem not to be relevant. Its length was also considered a barrier.

It was also thought the SF-36 would not reflect small changes in limitation which may be of considerable importance to patients. It was suggested that it may be suitable for patients with chronic decline rather than those with a turbulent disease trajectory and not appropriate for very advanced heart failure.

The SF-12 was considered to be an alternative based on patient acceptability, and feasibility of administration and scoring.

SF-12
Overall, patient acceptability and ease of administration were considered the most attractive features. Issues of specificity were raised which were consistent with those for the SF-36. Some preferred to recommend the SF-36 rather than the SF-12 based on comprehensiveness of items and domains.

EQ-5D
The main attraction of the EQ-5D was ease of completion and administration. Several commented on the potential usefulness of the thermometer as a snapshot of self-perceived wellbeing which might be applicable to respondents with impaired cognition or ethnic groups not having English as a first language. A similar thermometer scale has been used successfully with heart failure patients in supportive and palliative care.

Members who have used the EQ-5D with patients found it easy to complete. It was the preferred questionnaire above the SF-36 and SF-12 for one member who was unable to provide ratings.

However, the response options for the items were deemed to be too broad for heart failure patients and insufficient evidence existed to recommend its use with heart failure patients. The SF-36 or SF-12 were considered to be better alternatives.

Generic total

FIRST RATING	**'not at all suitable' (score 1)**	**'to some extent unsuitable' (score 2)**	**'uncertain' (score 3)**	**'to some extent suitable' (score 4)**	**'very suitable' (score 5)**	
SF-36	-	2	-	12	15	**29**
SF-12	1	4	3	4	10	**22**
EQ-5D	4	2	3	4	-	**13**

Heart failure-specific PROMs
MLHFQ
A vast volume of evidence supports the use of the MLHFQ and the fact that it has been applied widely in a variety of clinical settings added weight to its validity. Most of the items were thought to be relevant to people with heart failure and the response options sensitive. The scoring system for this measure was thought to be straightforward and doesn't require special software, enhancing its utility. However, several panel members commented on whether the instrument would account for co-morbidity or distinguish between symptoms of heart failure and other conditions, making it difficult for patients to complete.

This instrument asks the patient to make judgements about the extent to which 'heart failure' contributed to a list of possible difficulties. This was considered to be problematic and usually inappropriate for elderly people with multiple co-morbidities.

An open-ended questionnaire would be much better. Experience for members indicates that items are often not completed if deemed to be irrelevant to patients.

The MLHFQ was the preferred questionnaire of one member who was unable to provide ratings. The one-page format was considered to be an attractive feature as compared with the KCCQ.

The most important barrier to implementation and endorsement by the group were the cost and licensing implications.

KCCQ
The questions and the layout seemed to be preferable to those of the MLHFQ. Despite less evidence presented in the reviews for the KCCQ as compared with the MLHFQ, the range of questions was superior.

Issues about distinguishing between symptoms of heart failure and co-morbidities were raised with particular reference to older people. The wording of some items was considered too americanized. There was concern about whether it was sensitive to minor changes in heart failure status. One member commented that some of the items were not relevant to people with heart failure, such as issues relating to education.

Overall, it was considered that the KCCQ is good but on balance the MLHFQ was more suitable and has more evidence underpinning its use.

Specific Total

FIRST RATING	**'not at all suitable' (score 1)**	**'to some extent unsuitable' (score 2)**	**'uncertain' (score 3)**	**'to some extent suitable' (score 4)**	**'very suitable' (score 5)**	
MLHFQ				8	25	**33**
KCCQ				20	10	**30**

General comments were made regarding the lack of patient input in the derivation of items and that patient perspectives of their health were very different from what clinicians and experts believed to be important. There is a need for an instrument which covers a wider range of the most relevant limitation and emotional issues.

Major weaknesses of both heart failure-specific instruments were that co-morbidities are not accounted for. One panel member drew the group's attention to the EuroHeart Failure Questionnaire (Cleland et al., 2008) which does address this issue. However, there is as yet little published evidence regarding this instrument.

Recommendations
Based on ratings and comments from the panel, the SF-36 is the preferred generic measure of health status and the MLHFQ for the measurement of heart failure-specific quality of life. 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.

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 heart failure? (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

HEART FAILURE OUTCOMES CONSENSUS GROUP MEMBERS.

Dr James Beattie
Consultant Cardiologist, Heart of England NHS Foundation Trust, Birmingham

Ann Bowling
Professor, Department of Primary Care and Population Health, University College London

John Cleland
Professor of Academic Cardiology, University of Hull

Dr Karen Dunderdale
Lead Cardiac Specialist Nurse, Cardiac Rehabilitation Team, Scunthorpe General Hospital

Dr Richard Lehman
General Practitioner, Banbury, Oxfordshire

Annie MacCallum
Head of CVD and Heart Failure, Gloucestershire PCT

Richard Mayou
Professor of Psychiatry, Warneford Hospital, University of Oxford

David Thompson
Professor of Cardiovascular Nursing, University of Leicester

PROM Group members

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

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

Anne Mackintosh
Research Officer, Department of Public Health, University of Oxford.

REFERENCES

Abraham WT, Fisher WG, Smith AL, Delurgio DB, Leon AR, Loh E et al. Cardiac resynchronization in chronic heart failure. *NEJM: New England Journal of Medicine* 2002; 346(24):1845-1853.

Aiken LS, Butner J, Lockhart CA, Volk-Craft BE, Hamilton G, 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(1):111-126.

Asadi-Lari M, Rao A, Gray D. Health-related quality-of-life tools in heart failure. *Expert Review of Pharmacoeconomics and Outcomes Research* 2005; 5(3):267-270.

Austin J, Williams R, Ross L, Moseley L, Hutchison S. Randomised controlled trial of cardiac rehabilitation in elderly patients with heart failure. *European Journal of Heart Failure* 2005; 7(3):411-417.

Austin J, Williams WR, Hutchison S. Multidisciplinary management of elderly patients with chronic heart failure: five year outcome measures in death and survivor groups. *European Journal of Cardiovascular Nursing* 2008; in press.

Austin J, Williams WR, Ross L, Hutchison S. Five-year follow-up findings from a randomized controlled trial of cardiac rehabilitation for heart failure. *European Journal of Cardiovascular Prevention and Rehabilitation* 2008; 15(2):162-167.

Avis NE, Smith KW, Hambleton RK, Feldman HA, Selwyn A, Jacobs A. Development of the Multidimensional Index of Life Quality: a quality of life measure for cardiovascular disease. *Medical Care* 1996; 34(11):1102-20.

Azad N, Molnar F, Byszewski A. Lessons learned from a multidisciplinary heart failure clinic for older women: a randomised controlled trial. *Age and Ageing* 2008; 37(3):282-287.

Balashov K, Feldman DE, Savard S, Houde S, Frenette M, Ducharme A et al. Percent predicted value for the 6-minute walk test: using norm-referenced equations to characterize severity in persons with CHF. *Journal of Cardiac Failure* 2008; 14(1):75-81.

Barnes S, Gott M, Payne S. Prevalence of symptoms in a community-based sample of heart failure patients. *Journal of Pain and Symptom Management* 2006; 32(3):208-216.

Barrow DE, Bedford A, Ives G, O'Toole L, Channer KS. An evaluation of the effects of Tai Chi Chuan and Chi Kung training in patients with symptomatic heart failure: a randomised controlled pilot study. *Postgraduate Medical Journal* 2007; 83(985):717-721.

Bekelman DB, Havranek EP, Becker DM, Kutner JS, Peterson PN, Wittstein IS et al. Symptoms, depression, and quality of life in patients with heart failure. *Journal of Cardiac Failure* 2007; 13(8):643-648.

Benatar D, Bondmass M, Ghitelman J, Avitall B. Outcomes of chronic heart failure. *Archives of Internal Medicine* 2003; 163(3):347-352.

Bennett SJ, Perkins SM, Lane KA, Deer M, Brater DC, Murray MD. Social support and health-related quality of life in chronic heart failure patients. *Quality of Life Research* 2001; 10(8):671-682.

Bennett SJ, Litzelman DK, Wright A, Perkins SM, Wu J, Meyer L et al. The PUMP UP tailored computerized program for heart failure care. *Nursing Outlook* 2006; 54(1):39-45.

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(3):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(8):787-805.

Berry C, Murphy NF, De Vito G, Galloway S, Seed A, Fisher C et al. Effects of aldosterone receptor blockade in patients with mild-moderate heart failure taking a beta-blocker. *European Journal of Heart Failure* 2007; 9(4):429-434.

Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G et al. Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. *JAMA: Journal of the American Medical Association* 2005; 294(13):1625-1633.

Blinderman CD, Homel P, Billings JA, Portenoy RK, Tennstedt SL. Symptom distress and quality of life in patients with advanced congestive heart failure. *Journal of Pain and Symptom Management* 2008; 35(6):594-603.

Bolger AP, Bartlett FR, Penston HS, O'Leary J, Pollock N, Kaprielian R et al. Intravenous iron alone for the treatment of anemia in patients with chronic heart failure. *Journal of the American College of Cardiology* 2006; 48(6):1225-1227.

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.

Brodie DA, Inoue A, Shaw DG. Motivational interviewing to change quality of life for people with chronic heart failure: a randomised controlled trial. *International Journal of Nursing Studies* 2008; 45(4):489-500.

Bulpitt CJ, Fletcher AE, Dössegger l, Neiss A, Nielsen t, Viergutz S. Quality of life in chronic heart failure: cilazapril and captopril versus placebo. Cilazapril-Captopril Multicentre Group. *Heart* 1998; 79(6):593-598.

Chang VT, Hwang SS, Feuerman M, Kasimis BS, Thaler HT. The Memorial Symptom Assessment Scale Short Form (MSAS-SF). *Cancer* 2000; 89(5):1162-1171.

Chin MH, Goldman L. Gender differences in 1-year survival and quality of life among patients admitted with congestive heart failure. *Medical Care* 1998; 36(7):1033-1046.

Cleland JG, Freemantle N, Daubert JC, Toff WD, Leisch F, Tavazzi L. Long-term effect of cardiac resynchronisation in patients reporting mild symptoms of heart failure: a report from the CARE-HF study. *Heart* 2008; 94(3):278-283.

Cohn JN, Tognoni G, Valsartan Heart Failure Trial Investigators. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. *NEJM: New England Journal of Medicine* 2001; 345(23):1667-1675.

Conti JB, Sears SF. Cardiac resynchronization therapy: can we make our heart failure patients smarter? *Transactions of the American Clinical and Climatological Association* 2007; 118:153-164.

Cowley AJ, Wiens BL, Segal R, Rich MW, Santanello NC, Dasbach EJ et al. Randomised comparison of losartan vs. captopril on quality of life in elderly patients with symptomatic heart failure: the losartan heart failure ELITE quality of life substudy. *Quality of Life Research* 2000; 9(4):377-384.

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

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

De Bruin AF, De Witte LP, Stvene F, Diederiks JPM. Sickness Impact Profile: the state of art of a generic functional status measure. *Social Science and Medicine* 1992; 35(8):1003-14.

De Jong M, Moser DK, Chung ML. Predictors of health status for heart failure patients. *Progress in Cardiovascular Nursing* 2005; 20(4):155-162.

Department of Health. Guidance on the routine collection of Patient Reported Outcome Measures (PROMs) for the NHS in England 2009/10. Department of Health, London, December 2008.

Dracup K, Walden JA, Stevenson LW, Brecht ML. Quality of life in patients with advanced heart failure. *Journal of Heart and Lung Transplantation* 1992; 11(2 Pt.1):273-279.

Dunbar SB, Jenkins LS, Hawthorne M, Porter LS. Mood disturbance in patients with recurrent ventricular dysrhythmia before insertion of implantable cardioverter defibrillator. *Heart and Lung* 1996; 25(4):253-261.

Dunderdale K, Thompson DR, Miles JNV, Beer SF, Furze G. Quality-of-life measurement in chronic heart failure: do we take account of the patient perspective? *European Journal of Heart Failure* 2005; 7(4):572-82.

Dunderdale K, Furze G, Thompson DR, Beer SF, Miles JNV. Health-related quality of life from the perspective of patients with chronic heart failure. *British Journal of Cardiology* 2007; 14(4):207-212.

Dunderdale K, Thompson DR, Beer SF, Furze G, Miles JN. Development and validation of a patient-centered health-related quality-of-life measure: the Chronic Heart failure Assessment Tool. *Journal of Cardiovascular Nursing* 2008; 23(4):364-370.

Earnest MA, Ross SE, Wittevrongel L, Moore LA, Lin C. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. *Journal of the American Medical Informatics Association* 2004; 11(5):410-417.

Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health and Quality of Life Outcomes* 2006; 4:89, 2006.:89, 2006.

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

Evangelista LS, Dracup K, Moser DK, Westlake C, Erickson V, Hamilton MA et al. Two-year follow-up of quality of life in patients referred for heart transplant. *Heart-and-Lung* 2005; 34(3):187-193.

Evangelista LS, Moser DK, Westlake C, Hamilton MA, Fonarow GC, Dracup K. Impact of obesity on quality of life and depression in patients with heart failure. *European Journal of Heart Failure* 2006; 8(7):750-755.

Evangelista LS, Moser DK, Westlake C, Pike N, Ter Galstanyan A, Dracup K. Correlates of fatigue in patients with heart failure. *Progress in Cardiovascular Nursing* 2008; 23(1):12-17.

Feldman AM, Silver MA, Francis GS, Abbottsmith CW, Fleishman BL, Soran O et al. Enhanced external counterpulsation improves exercise tolerance in patients with chronic heart failure. *Journal of the American College of Cardiology* 2006; 48(6):1198-1205.

Ferrier KA, Neill AM, O'Meeghan T, Richards M, Campbell AJ. Continuous positive airway pressure in heart failure patients with obstructive sleep apnoea. *Internal Medicine Journal* 2008.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14).

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 Care, and Carer Impact. Oxford: National Centre for Health Outcomes Development, November 2006.

Friedman MM. Gender differences in the health related quality of life of older adults with heart failure. *Heart and Lung* 2003; 32(5):320-327.

Friedmann E, Thomas SA, Inguito P, Kao CW, Metcalf M, Kelley FJ et al. Quality of life and psychological status of patients with implantable cardioverter defibrillators. *Journal of Interventional Cardiac Electrophysiology* 2006; 17(1):65-72.

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

Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: the Weight Monitoring in Heart Failure (WHARF) trial. *American Heart Journal* 2003; 146(4):705-712.

Gottlieb SS, Kop WJ, Thomas SA, Katzen S, Vesely MR, Greenberg N et al. A double-blind placebo-controlled pilot study of controlled-release paroxetine on depression and quality of life in chronic heart failure. *American Heart Journal* 2007; 153(5):868-873.

Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City cardiomyopathy questionnaire: a new health status measure for heart failure. *Journal of the American College of Cardiology* 2000; 35(5):1245-55.

Greenberg B, Czerska B, Delgado RM, Bourge R, Zile MR, Silver M et al. Effects of continuous aortic flow augmentation in patients with exacerbation of heart failure Inadequately responsive to medical therapy. Results of the Multicenter Trial of the Orqis Medical Cancion System for the Enhanced Treatment of Heart Failure Unresponsive to Medical Therapy (MOMENTUM). *Circulation* published online Sept 2, 2008.

Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. *Journal of General Internal Medicine* 1989; 4(2):101-7.

Harrison MB, Browne GB, Roberts J, Tugwell P, Gafni A, Graham ID. Quality of life of individuals with heart failure: a randomized trial of the effectiveness of two models of hospital-to-home transition. *Medical Care* 2002; 40(4):271-282.

Harrison MB, Graham ID, Logan J, Toman C, Friederg E. Evidence to practice: pre-post-implementation study of a patient/provider resource for self-management with heart failure. *International Journal of Evidence-Based Healthcare* 2007; 5(1):92-101.

Havranek EP, Spertus JA, Masoudi FA, Jones PG, Rumsfeld JS. Predictors of the onset of depressive symptoms in patients with heart failure. *Journal of the American College of Cardiology* 2004; 44(12):2333-2338.

Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS et al. Health status identifies heart failure outpatients at risk for hospitalization or death. *Journal of the American College of Cardiology* 2006; 47(4):752-756.

Heo S, Moser DK, Lennie TA, Zambroski CH, Chung ML. A comparison of health-related quality of life between older adults with heart failure and healthy older adults. *Heart and Lung* 2007a; 36(1):16-24.

Heo S, Moser DK, Widener J. Gender differences in the effects of physical and emotional symptoms on health-related quality of life in patients with heart failure. *European Journal of Cardiovascular Nursing* 2007b; 6(2):146-152.

Heo S, Doering LV, Widener J, Moser DK. Predictors and effect of physical symptom status on health-related quality of life in patients with heart failure. *American Journal of Critical Care* 2008; 17(2):124-132.

Hershberger RE, Nauman DJ, Byrkit J, Gillespie G, Lackides G, Toy W et al. Prospective evaluation of an outpatient heart failure disease management program designed for primary care: the Oregon model. *Journal of Cardiac Failure* 2005; 11(4):293-298.

Höfer S, Lim L, Guyatt GH, Oldridge NB. The MacNew Heart Disease health-related quality of life instrument: a summary. *Health and Quality of Life Outcomes* 2004; 2:3.

Holland R, Brooksby I, Lenaghan E, Ashton K, Hay L, Smith R et al. Effectiveness of visits from community pharmacists for patients with heart failure: HeartMed randomised controlled trial. *BMJ* 2007; 334(7603):1098.

Hou N, Chui MA, Eckert GJ, Oldridge NB, Murray MD, Bennett SJ. Relationship of age and sex to health-related quality of life in patients with heart failure. *American Journal of Critical Care* 2004; 13(2):153-161.

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

Jerant A, Kravitz R, Moore-Hill M, Franks P. Depressive symptoms moderated the effect of chronic illness self-management training on self-efficacy. *Medical Care* 2008; 46(5):523-531.

Khayat RN, Abraham WT, Patt B, Roy M, Hua K, Jarjoura D. Cardiac effects of continuous and bi-level positive airway pressure for patients with heart failure and obstructive sleep apnea: a pilot study. *Chest* published online July 18, 2008.

Kitzman DW, Little WC, Brubaker PH, Anderson RT, Hundley WG, Marburger CT et al. Pathophysiological characterization of isolated diastolic heart failure in comparison to systolic heart failure. *JAMA: Journal of the American Medical Association* 2002; 288:2144-2150.

Klein DM, Turvey CL, Pies CJ. Relationship of coping styles with quality of life and depressive symptoms in older heart failure patients. *Journal of Aging and Health* 2007; 19(1):22-38.

Klodell CT, Jr., Aranda JM, Jr., McGiffin DC, Rayburn BK, Sun B, Abraham WT et al. Worldwide surgical experience with the Paracor HeartNet cardiac restraint device. *Journal of Thoracic and Cardiovascular Surgery* 2008; 135(1):188-195.

Koenig HG, Johnson JL, Peterson BL. Major depression and physical illness trajectories in heart failure and pulmonary disease. *Journal of Nervous and Mental Disease* 2006; 194(12):909-916.

Konstam MA, Gheorghiade M, Burnett JC, Jr., Grinfeld L, Maggioni AP, Swedberg K et al. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST Outcome Trial. *JAMA: Journal of the American Medical Association* 2007; 297(12):1319-1331.

Kosiborod M, Soto GE, Jones PG, Krumholz HM, Weintraub WS, Deedwania P et al. Identifying heart failure patients at high risk for near-term cardiovascular events with serial health status assessments. *Circulation* 2007; 115(15):1975-1981.

Kosiborod M, Krumholz HM, Jones PG, Pitt B, Spertus JA. The relationship between anemia, change in hematocrit over time and change in health status in patients with heart failure after myocardial infarction. *Journal of Cardiac Failure* 2008; 14(1):27-34.

Lader E, Egan D, Hunsberger S, Garg R, Czaijkowski S, McSherry F. The effect of digoxin on the quality of life in patients with heart failure. *Journal of Cardiac Failure* 2003; 9(1):4-12.

Lewis EF, Johnson PA, Johnson W, Collins C, Griffin L, Stevenson LW. Preferences for quality of life or survival expressed by patients with heart failure. *Journal of Heart and Lung Transplantation* 2001; 20(9):1016-1024.

Lewis EF, Lamas GA, O'Meara E, Granger CB, Dunlap ME, McKelvie RS et al. Characterization of health-related quality of life in heart failure patients with preserved versus low ejection fraction in CHARM. *European Journal of Heart Failure* 2007; 9(1):83-91.

Lewis GD, Shah R, Shahzad K, Camuso JM, Pappagianopoulos PP, Hung J et al. Sildenafil improves exercise capacity and quality of life in patients with systolic heart failure and secondary pulmonary hypertension. *Circulation* 2007; 116(14):1555-1562.

Lim LLY, Valenti LA, Knapp JC, Dobson AJ, Plotnikoff R, Higginbotham N et al. A self-administered quality-of-life questionnaire after acute myocardial infarction. *Journal of Clinical Epidemiology* 1993; 46(11):1249-56.

Macdonald JE, Kennedy N, Struthers AD. Effects of spironolactone on endothelial function, vascular angiotensin converting enzyme activity, and other prognostic markers in patients with mild heart failure already taking optimal treatment. *Heart* 2004; 90(7):765-770.

MacIver J, Rao V, Delgado DH, Desai N, Ivanov J, Abbey S et al. Choices: a study of preferences for end-of-life treatments in patients with advanced heart failure. *Journal of Heart and Lung Transplantation* 2008; 27(9):1002-1007.

Majani G, Giardini A, Opaisch C, Glazer R, Hester A, Tognoni G et al. Effect of valsartan on quality of life when added to usual therapy for heart failure: results from the Valsartan Heart Failure Trial. *Journal of Cardiac Failure* 2005; 11(4):253-259.

Mann DL, Acker MA, Jessup M, Sabbah HN, Starling RC, Kubo SH. Clinical evaluation of the CorCap Cardiac Support Device in patients with dilated cardiomyopathy. *Annals of Thoracic Surgery* 2007; 84(4):1226-1235.

Mansfield DR, Gollogly NC, Kaye DM, Richardson M, Bergin P, Naughton MT. Controlled trial of continuous positive airway pressure in obstructive sleep apnea and heart failure. *American Journal of Respiratory and Critical Care Medicine* 2004; 169(3):361-366.

Martin M, Blaisdell-Gross B, Fortin EW, Maruish ME, Manocchia M, Sun X et al. Health-related quality of life of heart failure and coronary artery disease patients improved during participation in disease management programs: a longitudinal observational study. *Disease Management* 2007; 10(3):164-178.

Masoudi FA, Rumsfeld JS, Havranek EP, House JA, Peterson ED, Krumholz HM et al. Age, functional capacity, and health-related quality of life in patients with heart failure. *Journal of Cardiac Failure* 2004; 10(5):368-373.

Mårtensson J, Dracup C, Canary C, Fridlund B. Living with heart failure: depression and quality of life in patients and spouses. *Journal of Heart and Lung Transplantation* 2003; 22(4):460-467.

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

McGee HM, Oldridge N, Hellemans IM. Quality of life evaluation in cardiovascular disease: a role for the European Society of Cardiology? *European Journal of Cardiovascular Prevention and Rehabilitation* 2005; 12(3):191-192.

McKelvie RS, Teo KK, Roberts R, McCartney N, Humen D, Montague T et al. Effects of exercise training in patients with heart failure: the Exercise Rehabilitation Trial (EXERT). *American Heart Journal* 2002; 144(1):23-30.

Miller K, Myers TJ, Robertson K, Shah N, Delgado RMI, Gregoric ID. Quality of life in bridge-to-transplant patients with chronic heart failure after implantation of an axial flow ventricular assist device. *Congestive Heart Failure* 2004; 10(5):226-229.

Morgan K, McGee H, Shelley E. Quality of life assessment in heart failure interventions: a 10-year (1996-2005) review. *European Journal of Cardiovascular Prevention and Rehabilitation* 2007; 14(5):589-607.

Moser DK, Macho MJ, Worster P. Community case management decreases rehospitalization rates and costs, and improves quality of life in heart failure patients with preserved and non preserved left ventricular function: a randomized controlled trial. *Circulation* 2000; 102(Suppl.II):II-749.

Nanda U, Andresen EM. Health-related quality of life: a guide for the health professional. *Evaluation and the Health Professions* 1998; 21(2):179-215.

Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. *Journal of the American Geriatrics Society* 2004; 52(5):675-684.

Neelagaru SB, Sanchez JE, Lau SK, Greenberg SM, Raval NY, Worley S et al. Nonexcitatory, cardiac contractility modulation electrical impulses: feasibility study for advanced heart failure in patients with normal QRS duration. *Heart Rhythm* 2006; 3(10):1140-1147.

O'Leary CJ, Jones PW. The Left Ventricular Dysfunction questionnaire/LVD-36: reliability, validity, and responsiveness. *Heart* 2000; 83(6):634-40.

O'Mahony MS, Sim MF, Ho SF, Steward JA, Buchalter M, Burr M. Diastolic heart failure in older people. *Age and Ageing* 2003; 32(5):519-524.

Oldridge NB, Saner H, McGee HM. The Euro Cardio-QoL Project: an international study to develop a core heart disease health-related quality of life questionnaire, the HeartQoL. *European Journal of Cardiovascular Prevention and Rehabilitation* 2005; 12(2):87-94.

Park CL, Malone MR, Suresh DP, Bliss D, Rosen RI. Coping, meaning in life, and quality of life in congestive heart failure patients. *Quality of Life Research* 2008; 17(1):21-26.

Patrick D, Peach H (eds). Disablement in the Community: A Sociomedical Press Perspective. Oxford: Oxford University Press, 1989.

Paul S, Sneed NV. Patient perceptions of quality of life and treatment in an outpatient congestive heart failure clinic. *Congestive Heart Failure* 2002; 8(2):74, 77-76, 79.

Pettit T, Livingston G, Manela M, Kitchen G, Katona C, Bowling A. Validation and normative data of health status measures in older people: the Islington study. *International Journal of Geriatric Psychiatry* 2001; 16(11):1061-1070.

Piotrowicz K, Noyes K, Lyness JM, McNitt S, Andrews ML, Dick A et al. Physical functioning and mental well-being in association with health outcome in patients enrolled in the Multicenter Automatic Defibrillator Implantation Trial II. *European Heart Journal* 2007; 28(5):601-607.

Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation

of symptom prevalence, characteristics and distress. *European Journal of Cancer, Part A: General Topics* 1994; 30A(9):1326-1336.

Porter L, Davis PG, Dunbar SB, Jenkins L, Hawthorne M. The Heart Failure Functional Status Inventory as a measure of functional status in patients with recurrent ventricular disrhythmia. *Circulation* 1994; 90(1):591.

Post MWM, de Bruin AF, de Witte et al. The SIP68: a measure of health related functional status in rehabilitation medicine. *Archives of Physical Medicine and Rehabilitation* 1996; 77(5):440-5.

Pugh LC, Havens DS, Xie S, Robinson JM, Blaha C. Case management for elderly persons with heart failure: the quality of life and cost outcomes. *Medical and Surgical Nursing* 2001; 10:71-78.

Quinn Griffin MT, Lee YH, Salman A, Seo Y, Marin PA, Starling RC et al. Spirituality and well being among elders: differences between elders with heart failure and those without heart failure. *Clin Interv Aging* 2007; 2(4):669-675.

Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure Questionnaire as a measure of therapeutic response to enalapril or placebo. *American Journal of Cardiology* 1993; 71(12):1106-1107.

Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure - Part 2: content, reliability and validity of a new measure, the Minnesota Living with Heart Failure Questionnaire. *Heart Failure* 1987; Oct-Nov:198-209.

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.

Rich MW, Beckham V, Wittenberg C, Leven CL, Freedland KE, Carney RM. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. *NEJM: New England Journal of Medicine* 1995; 333(18):1190-1195.

Rumsfeld JS, Havranek E, Masoudi FA, Peterson ED, Jones P, Tooley JF et al. Depressive symptoms are the strongest predictors of short-term declines in health status in patients with heart failure. *Journal of the American College of Cardiology* 2003; 42(10):1811-7.

Rose EA, Gelijns AC, Moskowitz AJ, et al, REMATCH Study Group. Long-term mechanical left ventricular assistance for end-stage heart failure. *NEJM: New England Journal of Medicine* 2001; 345(20):1435-1443.

Sayers SL, Riegel B, Goldberg LR, Coyne JC, Samaha FF. Clinical exacerbations as a surrogate end point in heart failure research. *Heart and Lung* 2008; 37(1):28-35.

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.

Scott LD, Setter-Kline K, Britton AS. The effects of nursing interventions to enhance mental health and quality of life among individuals with heart failure. *Applied Nursing Research* 2004; 17(4):248-256.

Shabetai R. Cardiomyopathy: how far have we come in 25 years, how far yet to go? *Journal of the American College of Cardiology* 1983; 1(1):252-263.

Shah MR, O'Connor CM, Sopko G, Hasselblad V, Califf RM, Stevenson LW. Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness (ESCAPE): design and rationale. *American Heart Journal* 2001; 141(4):528-535.

Shively M, Kodiath M, Smith TL, Kelly A, Bone P, Fetterly L et al. Effect of behavioral management on quality of life in mild heart failure: a randomized controlled trial. *Patient Education and Counseling* 2005; 58(1):27-34.

Shuldham C, Theaker C, Jaarsma T, Cowie MR. Evaluation of the European Heart Failure Self-care Behaviour Scale in a United Kingdom population. *Journal of Advanced Nursing* 2007; 60(1):87-95.

Sirker A, Thomas M, Baker S, Shrimpton J, Jewell S, Lee L et al. Cardiac resynchronization therapy: left or left-and-right for optimal symptomatic effect - the LOLA ROSE study. *Europace* 2007; 9(10):862-868.

Smart N, Haluska B, Jeffriess L, Marwick TH. Exercise training in systolic and diastolic dysfunction: effects on cardiac function, functional capacity, and quality of life. *American Heart Journal* 2007; 153(4):530-536.

Smith S, Cano S, Lamping D, et al. Patient reported outcome measures (PROMs) for routine use in treatment centres: recommendations based on a review of the scientific literature. Final Report to Department of Health, 2005.

Sokos GG, Nikolaidis LA, Mankad S, Elahi D, Shannon RP. Glucagon-like peptide-1 infusion improves left ventricular ejection fraction and functional status in patients with chronic heart failure. *Journal of Cardiac Failure* 2006; 12(9):694-699.

Soran O, Fleishman B, Demarco T, Grossman W, Schneider VM. Enhanced external counterpulsation in patients with heart failure: a multicenter feasibility study. *Congestive Heart Failure* 2002; 8(4):204, 227-208, 227.

Soto GE, Jones P, Weintraub WS, Krumholz HM, Spertus JA. Prognostic value of health status in patients with heart failure after acute myocardial infarction. *Circulation* 2004; 110(5):546-551.

Spertus JA, Conrad MW, Rinaldi J, Tsuyuki R, Krumholz H, Pina I et al. The Kansas City Cardiomyopathy Questionnaire is sensitive to clinical change in congestive heart failure. *Journal of the American College of Cardiology* 2002; 39(Suppl.):460A.

Spertus JA, Jones PG, Kim J, Globe D. Validity, reliability, and responsiveness of the Kansas City Cardiomyopathy Questionnaire in anemic heart failure patients. *Quality of Life Research* 2008; 17(2):291-298.

Stanchina ML, Ellison K, Malhotra A, Anderson M, Kirk M, Benser ME et al. The impact of cardiac resynchronization therapy on obstructive sleep apnea in heart failure patients: a pilot study. *Chest* 2007; 132(2):433-439.

Stephen SA. Fatigue in older adults with stable heart failure. *Heart and Lung* 2008; 37(2):122-131.

Stewart S, Marley JE, Horowitz JD. Effects of a multidisciplinary, home-based intervention on unplanned readmissions and survival among patients with chronic congestive heart failure: a randomised controlled study. *Lancet* 1999; 354(9184):1077-1083.

Subramanian U, Hopp F, Mitchinson A, Lowery J. Impact of provider self-management education, patient self-efficacy, and health status on patient adherence in heart failure in a Veterans Administration population. *Congestive Heart Failure* 2008; 14(1):6-11.

Sullivan MD, Levy WC, Russo JE, Crane B, Spertus JA. Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. *Journal of Cardiac Failure* 2007; 13(7):560-568.

Tate CWI, Robertson AD, Zolty R, Shakar SF, Lindenfeld J, Wolfel EF et al. Quality of life and prognosis in heart failure: results of the Beta-Blocker Evaluation of Survival Trial (BEST). *Journal of Cardiac Failure* 2007; 13(9):732-737.

Taylor AL, Lindenfeld J, Ziesche S, Walsh MN, Mitchell JE, Adams K et al. Outcomes by gender in the African-American Heart Failure Trial. *Journal of the American College of Cardiology* 2006; 48(11):2263-2267.

Todero CM, LaFramboise LM, Zimmerman LM. Symptom status and quality-of-life outcomes of home-based disease management program for heart failure patients. *Outcomes Management* 2002; 6(4):161-168.

Turvey CL, Klein DM, Pies CJ. Depression, physical impairment, and treatment of depression in chronic heart failure. *Journal of Cardiovascular Nursing* 2006; 21(3):178-185.

Valenti L, Lim L, Heller RF, Knapp J. An improved questionnaire for assessing quality of life after acute myocardial infarction. *Quality of Life Research* 1996; 5(1):151-61.

Walker DR, Landis DL, Stern PM, Vance RP. Disease management positively affects patient quality of life. *Managed Care Interface* 2003; 16(4):56-60.

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

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

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

Westlake C, Evangelista LS, Stromberg A, Ter Galstanyan A, Vazirani S, Dracup K. Evaluation of a Web-based education and counseling pilot program for older heart failure patients. *Progress in Cardiovascular Nursing* 2007; 22(1):20-26.

Wiklund IK, Lindvall K, Swedberg K, Zupkis RV. Self-assessment of quality of life in severe heart failure: an instrument for clinical use. *Scandinavian Journal of Psychology* 1987; 28(3):220-5.

Wiklund I, Swedberg K. Some methodological problems in analyzing quality of life data in severe congestive heart failure patients. *Journal of Clinical Research and Pharmacoepidemiology* 1991; 5(3):265-273.

Wiklund I, Waagstein F, Swedberg K, Hjalmarsson A. Quality of life on treatment with metoprolol in dilated cardiomyopathy: results from the MDC trial. Metoprolol in Dilated Cardiomyopathy trial. *Cardiovascular Drugs and Therapy* 1996; 10(3):361-368.

Witham MD, Crighton LJ, McMurdo ME. Using an individualised quality of life measure in older heart failure patients. *International Journal of Cardiology* 2007; 116(1):40-45.

Woodend AK, Sherrard H, Fraser M, Stuewe L, Cheung T, Struthers C. Telehome monitoring in patients with cardiac disease who are at high risk of readmission. *Heart and Lung* 2008; 37(1):36-45.

Wyrwich KW, Metz SM, Kroenke K, Tierney WM, Babu AN, Wolinsky FD. Triangulating patient and clinician perspectives on clinically important differences in health-related quality of life among patients with heart disease. *Health Services Research* 2007; 42(6 Pt 1):2257-2274.

Yeh GY, Wayne PM, Phillips RS. T'ai Chi exercise in patients with chronic heart failure. *Medicine and Sport Science* 2008; 52:195-208.

Zambroski CH, Lennie T, Chung ML, Heo S, Smoot T, Ziegler C. Use of the Memorial Symptom Assessment Scale-Heart Failure in heart failure patients. *Circulation* 2004; 110(Suppl.III):739.

Zambroski CH, Moser DK, Bhat G, Ziegler C. Impact of symptom prevalence and symptom burden on quality of life in patients with heart failure. *European Journal of Cardiovascular Nursing* 2005; 4(3