PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES USED IN ELECTIVE PROCEDURES FOR CORONARY REVASCULARISATION

Report to the Department of Health
2010

patient-reported

O
U
T
C
O
M
E

measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES USED IN ELECTIVE PROCEDURES FOR CORONARY REVASCULARISATION, 2010

Anne Mackintosh
Elizabeth Gibbons
Carolina Casañas i Comabella
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

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES USED IN ELECTIVE PROCEDURES FOR CORONARY REVASCULARISATION, 2010

CONTENTS

Executive summary 1

1. Introduction and methods 3

2. Preference-based measures 8

3. Generic PROMs 12

4. Cardiovascular-specific PROMs 22

5. Summary and recommendations 33

Appendix A – Methodological criteria 37

Appendix B – Content and scoring of instruments 39

Appendix C – Cardiovascular-specific, dimension-specific measures 44

Appendix D – Methods of working, membership and conclusions of the multidisciplinary panel 48

Appendix E – Abbreviations and acronyms 54

References 55

EXECUTIVE SUMMARY

Aims and structure of the report

The aims of this report are to identify Patient-reported Outcome Measures (PROMs) which have been evaluated with patients undergoing an elective coronary revascularisation procedure (ECRP) and to provide recommendations to the Department of Health (DH) of PROMs for ECRP 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. The methods of the review and the results of the search are described, including sources and search terms used to identify relevant published research. Details of this evidence are presented for preference-based measures, generic health status measures, and condition or procedure-specific PROMs evaluated with people with coronary heart disease undergoing elective procedures for coronary revascularisation. The review resulted in the identification of a short-list of PROMs which were presented to a multidisciplinary panel for comment. The review of the literature-based evidence and the comments of the multidisciplinary panel underpin final recommendations to the DH.

Results

PREFERENCE-BASED MEASURES

Four preference-based measures were identified:

a. 15D
b. EQ-5D
c. Health Utilities Index
d. SF-6D

GENERIC MEASURES

Six generic measures were identified:

a. Functional Status Questionnaire
b. Nottingham Health Profile
c. SF-36
d. SF-20
e. SF-12
f. Sickness Impact Profile

CARDIOVASCULAR-SPECIFIC QUESTIONNAIRES

Nine multidimensional cardiovascular-specific measures were identified:

a. Cardiac Symptom Survey
b. Coronary Revascularisation Outcome Questionnaire, CROQ
c. Duke Activity Status Index
d. Heart Surgery Symptom Inventory, HSSI
e. Kansas City Cardiomyopathy Questionnaire
f. MacNew/QLMI
g. Quality of Life Index-Cardiac Version
h. Seattle Angina Questionnaire, SAQ
i. Symptoms of Illness Score, SOIS

Fifteen cardiovascular-specific measures focussing on a single symptom or dimension were also identified:

a. Angina Questionnaire
b. Barnason Efficacy Expectation Scale
c. Cardiac Adjustment Scale
d. Cardiac Depression Scale
e. Cardiac Event Threat Questionnaire
f. Cardiac Self-Efficacy Scale
g. Cardiac Surgery Symptom Inventory
h. Cardiac Symptoms Scale
i. Control Attitudes Index
j. ENRICHD Social Support Index
k. Rose Angina Questionnaire
l. Rose Dyspnoea Questionnaire
m. Specific Activity Scale
n. Symptom Inventory
o. Symptom Scale

Short-listed PROMs for discussion

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

1. SF-36
2. EQ-5D
3. SAQ
4. CROQ

Recommendations

Based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the following measures have the strongest evidence supporting use with patients undergoing elective procedures for coronary revascularisation:

a. Preference-based measure: EQ-5D
b. Generic, multidimensional measure: SF-36
c. Cardiovascular-specific, multidimensional measure: SAQ

In the third category, with further evidence, the CROQ would merit consideration in the future.

1. INTRODUCTION

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 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, included 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 procedures. 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 selected elective procedures (Department of Health, 2008). Since April 2009, the routine collection of PROMs for the selected elective procedures has been implemented and is ongoing. This review extends that programme of work and considers PROMs for other elective procedures.

There are three broad categories of PROMs: generic health status, preference-based, and condition- or population-specific-measures. Generic instruments comprise items intended to be relevant to the widest range of patient conditions and the general population. Preference-based measures are also broad in content but additionally provide utilities or values regarding health (for use in, for example, cost-utility analyses of interventions). Condition-specific instruments are often more focused on a particular disease or health condition (for example, diabetes), a patient population (for example, older people), a specific problem or symptom (for example, pain), or a described function (for example, activities of daily living). For any given area of health, condition-specific instruments may have greater clinical appeal due to the inclusion of content specific to particular conditions, and the likelihood of increased responsiveness to interventions.

It has been recommended that a combination of a generic or utility measure with a specific measure be used in the assessment of patient-reported health outcomes, on the grounds that the complementary content of the two types of measure, when combined, should assess a full range of aspects of health relevant to the particular population concerned. However, consensus is often lacking as to which instrument to use for specific purposes and contexts (Garratt et al., 2002). Structured reviews of PROMs for specific health conditions or populations can provide guidance for selection. An evidence-based approach strengthens recommendations from these reviews.

Selection criteria have been defined for assessing the quality of existing PROMs (McDowell, 2006; Fitzpatrick et al., 1998; Streiner and Norman, 2003). These include measurement issues, such as reliability, validity, responsiveness and precision, as well as practical issues, such as acceptability and feasibility.

Heart disease in the UK

Coronary heart disease (CHD) is the commonest cause of premature death in the UK; it is therefore a major priority for the UK government both to prevent the condition from developing, and to reduce the suffering and risk of death entailed by CHD (Department of Health, 2000). The National Strategic Framework for Coronary Heart Disease has set a number of standards for combating the impact of CHD; these include increasing the number of surgical revascularisation procedures carried out, and a reduction in waiting times, entailing major investment in services (Department of Health, 2000). Measuring patient-reported outcomes should clearly be a key component in assessing the effectiveness of this strategy.

Coronary revascularisation

The term 'coronary revascularisation' encompasses both medical and surgical treatment for coronary artery disease (CAD). This review will examine the use of PROMs in relation to the most common elective procedures for CAD, namely coronary artery bypass grafting (CABG) and percutaneous coronary intervention (PCI) – also known as balloon angioplasty, coronary angioplasty, or percutaneous transluminal coronary angioplasty (PTCA).

Coronary artery bypass grafting

Coronary artery bypass grafting (CABG) is an elective surgical procedure to relieve angina and reduce the risk of death in patients with CAD whose coronary arteries are severely narrowed or blocked, restricting blood supply to the heart muscle. A vein or artery from elsewhere in the body (often one of the internal thoracic arteries, or the great saphenous vein from the leg) is used to form a graft, creating an alternative route around the damaged area. CABG is open heart surgery, usually performed with the heart stopped, necessitating the use of cardiopulmonary bypass (CPB). Since the mid-1990s, there has been a steady increase in the number of CABG procedures performed on a beating heart; this is known as 'off-pump' surgery. In recent years, a less invasive procedure has been developed, namely, minimally invasive or direct coronary artery bypass (MIDCAB), suitable for patients with stenosis of the left anterior descending coronary artery (LAD). This procedure involves making an incision in the left chest, retracting the ribs and harvesting the left internal thoracic artery (LITA) to make a graft for the LAD; MIDCAB can be performed 'on-' or 'off-pump'. Just over 20,000 CABG procedures using thoracic artery grafts were performed in the UK in 2007-8[1].

Percutaneous coronary intervention

In the past 20 years, the use of percutaneous coronary intervention (PCI) to treat patients with CAD in the UK has increased steadily, and the number of procedures carried out is now more than double the number of CABGs – over 50,000 in 2007-2008[2]. PCI can be an elective or an emergency procedure, and is used primarily to relieve symptoms. It involves the insertion of a catheter into a narrowed or stenotic section of a coronary artery, usually via a femoral artery; a balloon is then inflated to widen the stenosed area. A wire mesh tube or stent is often left in place to maintain the patency of the vessel; in suitable cases, medically coated or 'drug-eluting' stents may be used to prevent restenosis and/or an inflammatory reaction to the stent. PCI is far less invasive and traumatic for the recipient than the open heart surgery of CABG; however, it is unclear whether long-term results offer any advantage over optimal

[1] See Hospital Episode Statistics, Headline figures 2007-8 at http://www.hesonline.org.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=193
[2] As above.

medical therapy for patients with chronic stable CAD (Boden et al., 2007), and the procedure would appear to be inappropriate for asymptomatic patients, or those with minimal impairment (Curtis & Krumholz, 2004). It may also be less effective in terms of quality-of-life and cost benefits if repeat revascularisation is required, and in patients with multi-vessel disease (Poulin et al., 2007; Pusca & Puskas, 2007).

Aims of the report

The aims of this report are to identify Patient-reported Outcome Measures (PROMs) which have been evaluated with patients undergoing an elective coronary revascularisation procedure (ECRP) and to provide recommendations to the Department of Health (DH) of PROMs for ECRP 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.

Structure of the report

The methods of the review and the results of the search are described, including sources and search terms used to identify relevant published research. Details of this evidence are presented for preference-based measures, generic health status measures, and condition or procedure-specific PROMs evaluated with people with coronary heart disease undergoing elective procedures for coronary revascularisation. The full body of evidence and a short-list of PROMs were presented to a multidisciplinary panel for comment; details of their comments and ratings are reported in Appendix D. The PROMs Group considered the combination of the review of evidence and the views of the multidisciplinary panel before reaching its own conclusions and recommendations (Section 5).

Methods

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 evidence of measurement performance and operational characteristics abstracted for each PROM identified.

a) Search sources and terms

Several sources were searched to identify relevant articles.

The primary source of evidence was the bibliographic database compiled by the PROM group in 2002 with funding from the Department of Health and hosted by the University of Oxford. In 2005, it became the property of the NHS Information Centre for Health & Social Care. The PROM database comprises over 16,000 records (available online at http://phi.uhce.ox.ac.uk). The titles and abstracts of these, as well as a further 14,000 records identified as potential inclusions, were searched using the terms 'cardiovascular OR cardiac OR coronary OR heart' AND 'surgery OR revascularisation OR revascularization OR coronary artery bypass OR CABG OR CABS' OR 'angioplasty OR percutaneous OR PCI OR PCTA OR PTCA OR stent* OR stenting'.

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

Hand-searching of titles of key journals from 2007 to July 2009 was conducted. The following journals were selected:

- European Journal of Cardiothoracic Surgery
- Health and Quality of Life Outcomes
- Heart
- Heart and Lung
- Journal of Cardiopulmonary Rehabilitation and Prevention
- Journal of Thoracic and Cardiovascular Surgery
- Quality of Life Research

The National Institute for Health Research: Health Technology Assessment Programme, published research was also searched.

In addition, English-language PubMed records for the period 2007-9 (to 18 September 2009) were searched using a modified version of the cardiovascular-specific terms listed above, combined with a search strategy to identify PROMs devised by the PROM Group in collaboration with the Knowledge Centre of the University of Oxford (available on request).

b) Inclusion criteria

Published articles were included if they provided evidence of measurement and/or practical properties of relevant PROMs (Fitzpatrick et al., 1998).

Population

- patients undergoing coronary artery bypass graft surgery (CABGS) or percutaneous coronary intervention (PCI);
- English-speaking populations.

Study design selection

- studies where a principal PROM is being evaluated;
- studies evaluating several PROMs concurrently;
- trials or studies evaluating the effectiveness of interventions; where a PROM is used as an endpoint;
- prospective studies measuring patient-reported outcomes where data is available for a PROM in terms of measurement performance or operational characteristics.

Specific inclusion criteria for generic, preference-based and condition-specific instruments

- the instrument is patient-reported;
- there is published evidence of measurement reliability, validity or responsiveness following completion in the specified patient population;
- evidence is available from English-language publications, and instrument evaluations conducted in populations within the UK, North America, or Australasia;
- the instrument will ideally be multi-dimensional. It is at the reviewer's discretion to include PROMs which are specific to a health condition but have a narrow focus, for example, a specific dimension of health, such as symptoms.

Exclusions

- studies using clinician-rated instruments;
- studies evaluating the performance of non-patient reported measures of functioning or health status where a PROM is used as a comparator;
- studies with very small samples, i.e. fewer than 50 participants (except in the case of instrument development studies);
- studies using substantially incomplete versions of instruments.

c) Data extraction

For all PROMs included in the review, evidence is reported for the following measurement criteria:

- reliability
- validity
- responsiveness
- precision

Operational characteristics, such as patient acceptability and feasibility of administration for staff, are also reported.

d) Assessment of methodological quality of PROMs

Assessment and evaluation of the PROMs identified was performed using the criteria described in Appendix A. Searches identified nearly 4,000 potentially relevant records; of these, 259 papers were retrieved and reviewed in full. When assessed against the inclusion criteria, 128 studies were included in the review (Table 1).

Table 1: Number of articles identified by the literature review

Source	***Results of search***	***Number of articles included in review***
PROM bibliography: 30,350	252	33
PubMed 2007-September 2009	3699	18
Hand searching	-	76
TOTAL		**127**

2. PREFERENCE-BASED MEASURES

Four preference-based measures were identified:

a. 15D
b. EQ-5D
c. Health Utilities Index
d. SF-6D

See Appendix B, Table iii for a summary of content and scoring of these instruments.

a. 15D-Measure of Health-Related Quality of Life, 15D (Sintonen, 2001)

The 15D was developed as a generic multidimensional measure of HRQoL which provides both a single index score and a comprehensive health profile. Content was based on the WHO definition of health, and modified in light of contributions from health professionals, patient surveys, and factor analysis. The measure comprises 15 items encompassing physical, mental, and social well-being, with five graded responses for each item; there is no summed score. A set of preference weights elicited from the general population is used to generate a single index score between 0 and 1, where 0 is the worst (death) and 1 the best possible health. Change in score greater than 1 point per dimension, or 0.02-0.03 in the summary index, appears to reflect a clinically important difference.

Two studies from 2006 were identified which evaluated the 15D with Australian patients undergoing CABG or valve surgery.

Face validity of the measure is supported in a study of patients undergoing CABG or valve surgery, although the authors suggest having a single item for each dimension may limit sensitivity (Elliott et al., 2006a).

Internal consistency of the 15D was reported in a study comparing patient and proxy ratings, with a median Cronbach's alpha of 0.81 (Elliott et al., 2006b).

The 15D discriminated between patient groups, with those scheduled to receive valve surgery having significantly lower pre-operative scores than those due to undergo CABG (Elliott et al., 2006a).

Responsiveness of the instrument was supported (Elliott et al., 2006a), with significant improvements in score at six months post-surgery compared with baseline in several dimensions of the 15D.

Response rates ranging 65%-94% for postal administration of the 15D have been reported by the developer (Sintonen, 2001), indicating moderate to good acceptability to patients. However, despite commending the brevity and ease of completion of the 15D, Elliott and colleagues report a substantial loss to follow-up (28% drop-out rate) due to patient burden in their study, where the 15D and the SF-36 were administered concurrently (Elliott et al., 2006a).

b. EQ-5D (EuroQol Group, 1990)

The European Quality of Life instrument (EuroQol), now generally known as the EQ-5D, was developed by researchers in five European countries as a measure with a core set of generic health status items, based on existing PROMs (EuroQol Group, 1990; Brazier et al.,

1993). It was intended that, in application, the EQ-5D would be supplemented by disease-specific instruments. The developers recommend the EQ-5D for use in evaluative studies and policy research; it can also be used for economic evaluation. The measure can be self or interview-administered.

There are two sections to the EuroQol: the five-dimensions index and the EQ 'thermometer'. The EQ-5D assesses health across five domains, namely Anxiety/Depression (AD), Mobility (M), pain/discomfort (PD), Self-Care (SC), and Usual Activities (UA); each domain has one item and a three-point categorical response scale. 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 maximum well-being; a score profile can also 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.

Seven studies were identified supporting the use of the EQ-5D with patients undergoing CABG or PCI. Five of these were with UK samples (Ascione et al.; 2004, Denvir et al., 2006; Dunning et al., 2008; Kim et al., 2005; Leslie et al., 2007). Five studies examined outcomes of CABG, three examined PCI outcomes, with one including both.

Construct validity of the EuroQol was supported by high correlation (0.67) between the two components of the measure, EQ-5D and EQ-VAS, in a study of QoL at ten years post-CABG (Dunning et al., 2008, UK). In the same study, poor EQ-5D scores were significantly related to worse angina, as reflected in CCS grade (Dunning et al., 2008, UK). Construct validity of measure was supported by moderate to strong correlations between EQ-VAS and the eight SF-36 domains (range: 0.48 for RE to 0.77 for GH) in a study of long-term survival post-CABG (Bradshaw et al., 2006).

Discriminative validity of a six-item version of the EuroQol, including a global health transition question, was demonstrated in a study of CABG in octogenarians (Sollano et al., 1998) where patients who underwent surgery had significantly better outcomes by comparison with a medical cohort. The EQ-VAS also discriminated treatment groups in a trial of early interventional (IS) versus conservative (CS) treatment (including PCI or CABG) for patients with angina or NSTEMI (Kim et al., 2005, UK; RITA-3 trial). However, between-group differences were less marked for the five domain scores, with only UA showing significantly greater improvement in the IS group at both four months and one year (Kim et al., 2005, UK). The global utility score showed a significant difference between the IS and CS groups at four months, but this was no longer significant at one year (Kim et al., 2005, UK).

EQ-5D discriminated between known groups in a study examining the impact of socio-economic status (SES) on outcomes after PCI (Denvir et al., 2006, UK), where patients with lower SES had significantly lower scores at baseline and one year. In the same sample, unemployed patients also had significantly less improvement in QoL at one year (Leslie et al., 2007, UK).

Responsiveness of the EQ-VAS and EQ-5D global utility score was supported in the RITA-3 trial (Kim et al., 2005, UK), with both treatment groups experiencing a significant increase in score at both time points.

Response rates of 93% and above were reported for in-hospital administration of a battery of instruments including the EQ-5D in the RITA-3 trial (Kim et al., 2005, UK), indicating good acceptability to patients. A return rate of 87% for postal administration of a battery of instruments including the EQ-5D was reported by the BHACAS trial comparing OPCAB and CABG-CPB (Ascione et al., 2004, UK). However, there were low rates of return for postal administration in three other studies – 64%, 52%, and 68%, respectively (Sollano et al., 1998; Denvir et al., 2006, UK; Dunning et al., 2008, UK).

Interpretability of the EQ-VAS and EQ-5D global utility score was supported in the RITA-3 trial (Kim et al., 2005, UK), with a 5-unit decline in VAS and a 0.068 decline in global utility score, respectively, corresponding to a 1-unit increase in angina grade, according to the CCS classification.

c. Health Utilities Index (Feeny et al., 1995)

The Health Utilities Index (HUI) was designed as a comprehensive framework for describing health status and health-related quality of life for use in clinical studies, population health surveys, and economic evaluations; the original HUI has been largely superseded by HUI2 and HUI3 (Feeny et al., 1995). The Health Utilities Index Mark 3 (HUI3) consists of eight dimensions, rated by members of the general population as the most important attributes or dimensions of health status. For each attribute, there are five or six levels of functioning, ranging from highly impaired to normal, defined in terms of capacity rather than performance, to avoid confounding abilities with preferences. A combination of levels across the attributes constitutes a health state; utility scores, based on community preferences, can be obtained for each health state using an algorithm, with 0 representing death and 1 perfect health.

Population norm data have been obtained from several large general population surveys. Over 15 different language versions of the HUI are available, and it has been used in more than 25 countries.

Two studies were identified using HUI3 with patients undergoing elective coronary procedures, the most recent being from 2004; neither was conducted in the UK. One of the studies examined outcomes of PCI, the other feasibility of outcome measurement in CABG or PCI.

In a head-to-head comparison between the HUI3 and the SF-6D with a sample of patients undergoing PCI (Hatoum et al., 2004; EXCITE study), both measures demonstrated discriminative validity, with significant differences in score distribution compared with the general population samples from whom preference weights were originally obtained, and significant between-group differences according to severity of angina (Hatoum et al., 2004). However, unlike the SF-6D, the HUI did not discriminate women and men, and patients requiring a longer stay in CCU (Hatoum et al., 2004).

Concurrent validity was demonstrated by significant correlations between comparable domains of the HUI3 and SF-6D (Hatoum et al., 2004).

Responsiveness of the HUI3 to change was also supported in the EXCITE study, with significant score changes between baseline (pre-discharge) and six months (Hatoum et al., 2004). However, HUI scores showed large ceiling effects, with 50% of the sample scoring at the top level in five of the eight dimensions, suggesting that the measure may lack precision,

particularly with non-hospital samples, i.e. patients with a generally higher level of functioning, samples (Hatoum et al., 2004).

Acceptability of both the HUI3 and SF-6D was high, with a response rate of 89% and a completion rate for the HUI3 of 97% (in-hospital administration at baseline, mode of administration at six months unclear) (Hatoum et al., 2004).

Feasibility of the HUI as part of a survey package for the routine collection of baseline data on patients undergoing CABG or PCI was explored in a study by Spertus and colleagues (2001), with mixed results. Although the survey was acceptable to the majority of patients (92% agreed to participate), it proved difficult to convince nursing staff to integrate such data collection into routine care.

d. SF-6D (Brazier et al., 1998, 2002)

The SF-6D is a preference-based measure derived from the SF-12 and the SF-36 (Brazier et al., 2002). The eight dimensions of the SF-36 were reduced to six by omitting General Health (GH), and combining the two role limitation dimensions (RE and RP). For each of the dimensions (Physical Functioning, Role Limitations, Social Functioning, Pain, Mental Health, Vitality), there are between four and six levels of functioning. The SF-6D health state classification is calculated from responses to the SF-12 or SF-36, using preference weights obtained from a general population sample.

One study was identified which compared the performance of the SF-6D and HUI3 in North American patients undergoing PCI (Hatoum et al., 2004; EXCITE study). Both measures demonstrated discriminative validity, with significant differences in score distribution compared to the general population samples from which preference weights were originally obtained, and significant between-group differences according to severity of angina. However, only the SF-6D discriminated women and men, and patients requiring a longer stay in CCU (Hatoum et al., 2004).

Concurrent validity was demonstrated by significant correlations between comparable domains of the SF-6D and HUI3 (Hatoum et al., 2004).

Responsiveness of the SF-6D to change was also supported in the EXCITE study, with significant score changes between baseline (pre-discharge) and six months (Hatoum et al., 2004). However, the SF-6D showed significant floor effects, particularly in the Role Limitations dimension, with 40% of respondents scoring at the lowest level, suggesting that the it may be inappropriate for assessing a hospital sample, i.e. patients with a (generally) lower level of function (Hatoum et al., 2004).

Acceptability of both the SF-6D and HUI3 was high, with a response rate of 89% and a completion rate for the SF-6D of 91% (in-hospital administration at baseline, mode of administration at six months unclear) (Hatoum et al., 2004).

3. GENERIC PROMs

Six generic measures were identified:

a. Functional Status Questionnaire
b. Nottingham Health Profile
c. SF-36
d. SF-20
e. SF-12
f. Sickness Impact Profile

See Appendix B, Table iii for a summary of content and scoring of these instruments.

a. Functional Status Questionnaire, FSQ (Jette et al., 1986)

The US-developed FSQ was originally designed to assess physical, psychological, social, and role functioning in ambulatory patients. It has since been validated with hospital patients and in pharmaceutical trials. Questions were adapted from existing instruments, including the SIP. The measure comprises 34 items in four sections: physical function (Basic and Intermediate ADL scales), psychological function (MH); social/role function scales (Work Performance WP, Social Activity SA, Quality of Interaction QI), and six single items. Responses to subscale items are averaged and transformed into a 0-100 scale, where 100 denotes maximum function. A computerised report can be generated, displaying scores in the form of VASs, with 'warning zones' to indicate important disability. The FSQ is self-administered and takes approximately 15 minutes to complete.

Three US studies were identified supporting the use of the FSQ with patients undergoing elective coronary procedures, the most recent being from 2004. Two of the studies examined outcomes of CABG; the third compared outcomes of CABG and PCI. Several additional studies were found which used only two or three FSQ subscales (fewer than 50% of FSQ items); these studies were excluded from the review.

Internal consistency reliability of the FSQ was supported in a study comparing outcomes of patients undergoing either CABG or PTCA, with Cronbach's alpha coefficients ranging 0.71 to 0.99 (Allen et al., 1990). However, values greater than 0.95 suggest there may be some redundancy of items. Cronbach's alpha ranged 0.69-0.87 for the CABG group in a study of patients undergoing one of four elective surgical procedures (Cleary et al., 1991; Six Hospital study).

The IADL and SA subscales discriminated between treatment groups in the study by Allen et al. (1990), with PTCA patients having significantly better function in these domains at baseline and one month. By one year, however, IADL was similar in both groups, while CABG patients had significantly higher SA scores (Allen et al., 1990). MH and QI scores were not significantly different at baseline, but at six months, CABG patients had significantly higher scores in both domains; for MH, this difference was sustained at one year (Allen et al., 1990). These findings would appear to show that the FSQ reflected the different trajectories of recovery following the two procedures, with CABG patients having a longer post-operative recovery period and perhaps more realistic expectations with respect to the regaining of function (Allen et al., 1990).

ADL, IADL, and SA subscales of the FSQ discriminated between older women and other demographic groups, and between patients who did and did not exercise, in a study of

functional recovery and exercise behaviour 5-6 years post-CABG (Treat-Jacobson & Lindquist, 2004).

Responsiveness of five subscales of the FSQ (ADL subscale and single items not used) was supported in the study by Allen et al. (1990), with statistically significant score changes at one year for both treatment groups, the one exception being QI score for the PTCA patients, which was high at baseline and remained so. All subscales except WP demonstrated responsiveness to change in the Six Hospital study (Cleary et al., 1991), with statistically significant score changes at six months post-CABG; however, this was based on patient recall of their pre-operative functioning, which may not have been reliable after six months.

Acceptability of the FSQ was supported in the studies by Cleary et al. (1991) and Treat-Jacobson and Lindquist (2004), with response rates of 87% and 91%, respectively, to postal administration of the measure.

b. Nottingham Health Profile, NHP (Hunt et al., 1980, 1985, UK)

The Nottingham Health Profile (NHP) was developed in the UK during the 1970s for use in the evaluation of medical or social interventions (Hunt et al., 1980). Instrument content was derived from 2200 statements given by over 700 patients with a variety of chronic ailments, and other lay people. These were rationalised, tested, and eventually reduced to 38 items found to be reliable, capable of distinguishing different types and levels of disability, sensitive to change, and readily understood (Hunt et al., 1985).

Part I of the instrument has 38 items across six domains: Bodily Pain (BP), Emotional Reactions (ER), Energy (E), Physical Mobility (PM), Sleep (S), and Social Isolation (SI). Each item is a statement referring to a departure from normal functioning and the feelings induced. Respondents answer 'yes' or 'no' according to whether they feel the item applies to them at present. Positive responses are weighted and summed to give six domain scores between 0 and 100, where 100 denotes maximum limitation. Part II of the NHP consists of seven statements relating to areas of daily life most often affected by ill-health, namely, paid employment, housework, social life, personal relationships, sex life, hobbies/interests, and holidays, with 'yes/no' responses as for Part I. The instrument may be self-, interview-, or telephone-administered.

Three studies were identified evaluating the use of the NHP with elective coronary procedures. Two of these were with UK samples (Caine et al., 1991; Pocock et al., 1996). One study examined outcomes of CABG; the two others compared outcomes of CABG and PCI. No recent studies were identified.

Construct validity of the NHP was supported by significant correlations between all parts of the instrument and angina status at baseline and two years in the RITA study comparing the impact of PTCA and CABG (Pocock et al., 1996, UK).

Both parts of the NHP discriminated the PTCA from the CABG group, suggesting slightly but significantly greater impairment in the former (Pocock et al., 1996, UK). Pre-operative NHP ER, E, and PM scores predicted return to work in a small study of male patients undergoing CABG (Caine et al., 1991, UK). In the BARI trial comparing outcomes of CABG versus PTCA, the NHP discriminated treatment groups at initial and one-year follow-up, though QoL outcomes were similar thereafter (BARI Investigators, 1997).

Responsiveness of the NHP was supported in the study by Caine and colleagues (1991, UK) by significant improvements in all domains at three months, sustained at one year. There were also statistically significant score improvements for both treatment groups in all domains in the study by Pocock and colleagues (1996, UK).

Interpretation of NHP scores was facilitated in the study by Pocock and colleagues (1996, UK) by classifying patients into four categories of impairment for each domain, depending on the number of items with 'yes' (i.e. negative) responses.

c. SF-36 (Ware and Sherbourne, 1992; Ware, 1997)

The SF-36 is a generic health status instrument capturing both mental and physical aspects of health; it is intended for application in a wide range of conditions and with the general population. The measure comprises 36 items assessing health across eight domains, 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 (VT: four items). 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, 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 (PCS and MCS, respectively) can also be calculated. The SF-36 can be self-, interview-, or telephone-administered.

Thirty-nine studies were identified supporting the use of the SF-36 with patients undergoing elective coronary procedures; of these, eight studies were carried out in the UK (Agarwal et al., 2009; Ascione et al., 2004; Kim et al., 2005; Lee, 2008; Lindsay et al., 2000; Pocock et al., 2000; Schroter & Lamping, 2006; Thornton et al., 2005). Thirty studies examined outcomes in CABG patients, 13 were concerned with outcomes of PCI; four studies examined both CABG and PCI. About half of the studies were published in the last five years.

Internal consistency reliability of the SF-36 was supported with Cronbach's alpha ranging from 0.73 (GH) to 0.90 (PF) in a large study of patients undergoing CABG and/or valve surgery (McCarthy et al., 1995; PSOCS study). Similar values were found in a study comparing generic and specific measures for measuring HRQoL after PCI (Krumholz et al., 1996). Cronbach's alpha exceeded 0.85 for all subscales in a study comparing outcomes following coronary stenting versus balloon angioplasty (Krumholz et al., 1997), ranged 0.80-0.94 in the Stent-PAMI trial (Rinfret et al., 2001), and exceeded 0.80 in a postal survey assessing QoL in women after either an acute cardiac event or CABG (Worcester et al., 2007). Good internal consistency was also reported in a study examining differences in recovery outcomes for two groups of older adults undergoing cardiac rehabilitation (CR) post-CABG (Dolansky & Moore, 2004), with high alpha coefficients of 0.84 (pre-test) and 0.81 (post-test) for the PCS, and 0.86 (pre-test) and 0.84 (post-test) for the MCS. High Cronbach's alpha values in six of the domains (0.82 and above) were reported in a postal survey of post-CABG patients undergoing CR (Hawkes et al., 2003).

The PF subscale correlated strongly with the CCS classification in the 1997 study by Krumholz et al. (1997), supporting construct validity of the measure. Construct validity of the

SF-36 was further supported by strong correlation of the PCS (0.65) with the EQ-VAS in a study of long-term survival post-CABG (Bradshaw et al., 2006); however, correlation between the MCS and the EQ-VAS was weaker (0.48) (Bradshaw et al., 2006).

Predictive validity of the SF-36 was supported in a prospective cohort study with 1778 CABG patients, examining the extent to which perceived health status predicts mortality and length of stay (Curtis et al., 2002). The PCS score was highly predictive of hospital mortality and greater length of stay, while the MCS score was significantly associated with prolonged hospital stay (Curtis et al., 2002). Lower pre-operative PCS and MCS scores independently predicted lower post-operative scores in a study of post-menopausal women six months after CABG (Hogue et al., 2008).

Predictive validity of the SF-36 was reported in the PSOCS study (Rumsfeld et al., 1999) where a 10-point lower baseline PCS was significantly correlated with mortality at six months. In the same study, baseline PCS and MCS below 38 and 44, respectively, predicted significantly greater post-operative improvement in HRQoL (Rumsfeld et al., 2001). Further evidence for predictive validity of the MCS was reported in a study exploring long-term outcomes of CABG, where patients who had higher MCS scores pre-operatively had better QoL (SF-36, DASI, and a life satisfaction measure) at six weeks and six months after surgery (Sawatzky & Naimark, 2009a). Lower pre-operative PCS scores were predictive of poorer post-operative functioning in a study of CABG outcomes in elderly patients (Mayer et al., 2003).

The SF-36 discriminated between known groups in several studies.

In a comparison of outcomes in fatigued and non-fatigued patients post-CABG (Barnason et al., 2008), the fatigued group had significantly ($p<0.05$) lower scores post-surgery than the non-fatigued group in the PF dimension ($p<0.01$) at six weeks, and in the MH, RE, and SF dimensions at six weeks and three months. There was also a statistically significant difference between the two groups with respect to the summary scores, with the fatigued group having poorer MCS scores at six weeks ($p\leq0.03$) and poorer PCS scores at three months ($p\leq0.003$).

In a cluster analysis by Fukuoka et al. (2007), the SF-36 discriminated between three groups of patients, namely, the Weary group, the Diffuse symptom group, and the Breathless group, 12 months after AMI and/or CABG. Scores in all dimensions except SF were significantly lower in the Weary group than in the Diffuse symptom group (Fukuoka et al., 2007). Scores in the RP, BP, GH, and MH subscales were significantly lower for the Weary group, compared to the Breathless group (Fukuoka et al., 2007).

The SF-36 PCS discriminated between diabetics and non-diabetics in a cross-sectional substudy of patients enrolled in the COURAGE trial (Deaton et al., 2006), with diabetics having significantly lower PCS scores. Similar results were found in a study of insulin-treated and non-insulin-treated patients undergoing CABG, where insulin-treated patients had significantly lower PCS scores than non-insulin treated patients (Deaton et al., 2009). In the same study, patients without complications had better three-month PCS scores ($p=0.025$) and MCS ($p=0.001$) than those with complications (Deaton et al., 2009). Patients with better PCS scores at three months had significantly shorter post-operative length of stay than those whose scores did not improve between baseline and three months ($p=0.024$) (Deaton et al., 2009).

The PF subscale discriminated between patient groups and detected within-group change in a study examining outcomes in patients with advanced CAD; patients undergoing CABG experienced significantly greater improvement at two years than those who did not receive surgery (Kandzari et al., 2001; MOSS study). All SF-36 domains except BP and MH discriminated between treatment groups at four months in a trial of interventional versus conservative treatment for patients with angina or NSTEMI (Kim et al., 2005, UK; RITA-3 trial); however, these differences remained significant at one year in only four of eight dimensions (RP, SF, V, and GH) (Kim et al., 2005, UK). Similar results were reported in the COURAGE trial which compared the use of optimal medical therapy (OMT) alone and OMT with PCI support (Weintraub et al., 2008). SF-36 scores in five domains (PF, RP, V, BP, GH) showed an advantage of PCI over OMT alone up to three months; by 12 months, however, no advantage was apparent (Weintraub et al., 2008).

The SF-36 scores discriminated patient groups by demographic characteristics in several studies. Five of eight SF-36 domains (PF, RF, SF, RE, MH) showed that Black patients had significantly worse outcomes than Whites at six months post-revascularisation (Kaul et al., 2005). All SF-36 domains except MH showed significant differences between age-groups in the study by Bradshaw et al. (2006), with younger respondents generally reporting higher HRQoL. Significant differences between men and women in SF-36 physical component scores (BP, GH, PF, RP) were found at three time-points in a study examining gender differences in patients' experience of CABG surgery (Sawatzky & Naimark, 2009b). Lower PCS and MCS scores were significantly associated with mortality in older (≤65), but not younger, patients following CABG and/or valve surgery ($p=0.01$ for PCS and $p=0.03$ for MCS) (Ho et al., 2005).

In a RCT assessing the QoL of individuals undergoing PTCA (Pocock et al., 2000, UK), the SF-36 discriminated between grades of disease severity, for both the control and the intervention group, in the PF, VT, and GH domains, with greater disease severity resulting in lower scores for each of these three dimensions. In a prospective study of post-CABG patients by Lindsay et al. (2000, UK), SF-36 detected different levels of health in individuals at a single point in time.

The SF-36 also discriminated between patient sample and population norms in several studies. The study by Worcester et al. (2007) reported scores significantly lower than norms on all SF-36 subscales except GH at two, four, and 12 months. In the study by Hawkes et al. (2003), patient scores on all subscales were significantly lower than for age-matched population norms at baseline ($p \leq 0.05$); at six months, RP and BP subscales were significantly lower ($p<0.0005$), while at 12 months only BP was significantly lower ($p<0.0005$) (Hawkes et al., 2003).

Several studies report the responsiveness of SF-36 scores to clinical change.

SF-36 scores were responsive to change in patients undergoing CABG in the study to develop the cardiovascular-specific measure SAQ (Spertus et al., 1994b); however, the SF-36 was less sensitive to small clinical changes than the SAQ, as evidenced by Guyatt's responsiveness statistic (Spertus et al., 1994b). All SF-36 domains except GH showed significant improvement at six months post-PCI in a study comparing the responsiveness of generic and specific measures (Krumholz et al., 1996). Five of eight domains (PF, RP, V, BP, SF) showed significant improvement in a study of outcomes at three months post-CABG (Krumholz et al., 1997). Significant improvement in mean post-operative scores for both PCS

and MCS was reported in the PSOCS study (Rumsfeld et al., 2001). There were statistically significant improvements over 12 months for all domains except GH in the Stent-PAMI trial (Rinfret et al., 2001). In the COURAGE trial (Weintraub et al., 2008), responsiveness of the SF-36 was demonstrated by statistically significant score changes in all domains after the first three months of treatment.

Responsiveness was further demonstrated in a survey by Kiebzak et al. (2002) where the SF-36 was administered pre-operatively and at 12 months post-CABG; scores on six of the eight subscales (PF, RP, BP, RE, SF, and VT) were significantly better ($p<0.05$) at one year (Kiebzak et al., 2002). Significant differences in PCS scores between the two arms of the trial were reported in the study by Dolansky & Moore (2004) at six weeks post-CABG (coinciding with the start of the CR programme), and six months. MCS scores showed statistically significant improvement (mean increase in score of 2.5, $p<0.001$) between in-hospital baseline and three months post-CABG in Type 2 diabetic patients (Deaton et al., 2009). SF-36 detected statistically significant change in three domains (PF, V, and GH) and overall PCS in a quasi-experimental (non-randomised) study by Ballan & Lee (2007) with patients undergoing their first or second CABG.

Several dimensions of the SF-36 showed responsiveness to change in a study examining the impact of a symptom management telehealth intervention (Barnason et al., 2009) in patients undergoing CABG. PF, RP, and VT subscales were used to measure physical functioning, while psychosocial functioning was measured using RE, MH, and SF subscales, at baseline during hospitalisation, and post-operatively at six weeks, three months, and six months (Barnason et al., 2009). Statistically significant change was detected in psychosocial scores at six weeks, and in physical functioning at three months. Overall, RP, VT, and MH scores improved significantly over time ($p<0.01$, $p<0.01$, and $p<0.005$, respectively) (Barnason et al., 2009).

Responsiveness of the SF-36 was further supported in two studies by Elliott and colleagues (Elliott et al., 2006a, 2006b), with significant score changes over time in several dimensions of the SF-36 in patients undergoing cardiac surgery (primarily CABG). It appeared to be more sensitive than the 15D, particularly in the mental health domain (Elliott et al., 2006a). However, it was less sensitive than the SAQ in a study examining outcomes of CABG in a sample of elderly patients, evidenced by lower SRMs in all domains except PF (MacDonald et al., 1998). Responsiveness of the SF-36 was further supported in a study examining QoL after PCI in octogenarians, with significant improvement in RP, BP, and RE subscales at six and 12 months (Agarwal et al., 2009, UK).

BP subscale scores were significantly higher (better) post-operatively in patients who underwent coronary stenting compared to those receiving balloon angioplasty (Krumholz et al., 1997). As there were no other between-group differences, this may have been a chance finding (Krumholz et al., 1997). However, a similar result was found in the Stent-PAMI trial (Rinfret et al., 2001), with patients receiving coronary stenting versus balloon angioplasty having the advantage at one and six months, reflected in significantly improved BP; by 12 months the difference was no longer significant.

Significantly improved scores on the PF, VT, GH, and MH subscales were reported in the study by Pocock et al. (2000, UK), at three months and one year post-procedure. SF-36 detected statistically significant change in the PR, VT, SF, and GH ($p\leq0.01$) domains at two- and six-months follow-up, in a prospective longitudinal study assessing HRQoL,

neuropsychologic deficits, and mood of post-CABG male patients (Thornton et al., 2005, UK). Several studies reported significant improvement in six or more of the subscales following CABG (Hawkes et al., 2003; Kiebzak et al., 2002; Lindsay et al., 2000, UK; Welke et al., 2003).

In a head-to-head comparison with the CROQ in patients undergoing CABG or PTCA, SF-36 subscales were generally less responsive than similar scales of the CROQ (Schroter & Lamping, 2006; UK), although both measures showed large effect sizes (ES) and SRMs in several domains, particularly with the CABG patients. ESs and SRMs for SF-36 SF and MH subscales were significantly lower than for the CROQ Psychosocial scale in both groups; in the CABG group, ES and SRM were also significantly lower for SF-36 PF compared with CROQ PF (Schroter & Lamping, 2006; UK).

CABG: ES and SRM for the CROQ Psychosocial scale were significantly larger than for the SF-36 SF and MH subscales, although PF scales of the two measures were comparable (Schroter & Lamping, 2006; UK). In the PTCA sample, the ES for CROQ PF was significantly larger than for SF-36 PF.

Precision of the SF-36 was assessed in the study by Pocock et al. (2000, UK), which reported a considerable ceiling effect. 29% of PTCA and 21% of medically treated patients scored ≥90 at one year on the PF subscale, 23% of PTCA and 15% of medically treated patients scored ≥80 on the VT subscale at one year, while 28% of PTCA and 19% of medically treated patients scored ≥80 at one year on the GH subscale (Pocock et al., 2000, UK).

High acceptability of the SF-36 to patients has been reported in several studies.

Response rates of 93% and above were reported for in-hospital administration of a battery of instruments including the SF-36 in the RITA-3 trial (Kim et al., 2005, UK) indicating good acceptability to patients. A return rate of 87% for postal administration of the SF-36 as part of a battery of instruments was reported by the BHACAS trial comparing OPCAB and CABG-CPB (Ascione et al., 2004, UK). Response rates were also high (82%) in the postal survey by Bradshaw et al. (2006), with near-complete SF-36 data for 96% of respondents, and in the hospital-based study by Deaton and colleagues (2009) where 85% of participants completed the questionnaire at three months post-CABG. A lower, but still acceptable, rate (80%) was reported for postal administration of the SF-36 at between six and 18 months post-operatively in the Stent Restenosis study (Krumholz et al., 1997). There was a response rate of 72% and 62% for the CABG and PTCA groups, respectively, in the postal survey comparing SF-36, CROQ, and SAQ (Schroter & Lamping, 2006, UK); of those who responded, 89% and 94%, respectively, completed the questionnaires post-revascularisation (Schroter & Lamping, 2006, UK).

However, other studies have reported low response rates. Elliott and colleagues report loss to follow-up because of patient burden in a study where the SF-36 and the 15D were administered concurrently, with only 60% completing postal administration at six months (Elliott et al., 2006a). Only 50% of questionnaires were completed in the study assessing physical and mental health after CABG by Welke et al. (2003).

In a study examining HRQoL at five years post-CABG (Lee, 2008, UK), it was found that patients returning postal questionnaires had significantly lower scores than those who completed the questionnaire at a hospital follow-up appointment, in five of eight domains

(PF, RP, RE, SF, GH) and the PCS. Although the postal administration sample was very small, this finding suggests that use of postal follow-up prevented loss of data from those with poorer physical health, which might otherwise have led to biased results.

d. SF-20 (Ware et al., 1992)

The Medical Outcomes Study (MOS) 20-item Short Form Health Survey (SF-20) is a 20-item abbreviation of the same Rand instrument from which the SF-36 is derived.
SF-20 assesses health across six domains, namely, Bodily Pain (BP), General Health perception (GH), Physical Function (PF), Mental Health (MH), Social Function (SF), and Role Function (RF). Each item has between three and six categorical response options; several items have reversed scoring. Domain item summed scores are transformed into a scale from 0 to 100, where higher values denote better health. The instrument may be self-, interview- or telephone-administered, and takes about five minutes to complete.

The SF-20 has been largely superseded by the SF-12. Nevertheless, two recent studies were identified which used the former with the same sample of US patients undergoing open-heart surgery (CABG and/or valve procedure) (Halpin et al., 2008; Martin et al., 2008).

Adjusted SF-20 GH, SF, and overall scores were predictive of decreased survival at two, three, and five years (Halpin et al., 2008; Martin et al., 2008).

Responsiveness of the SF-20 was supported with significant positive score changes in all domains at one year, most notably in the GH and PF domains, less so in SF and MH (Halpin et al., 2008; Martin et al., 2008).

Findings from these studies should be viewed in light of a large number lost to follow-up, or who declined participation (Halpin et al., 2008; Martin et al., 2008).

e. SF-12 (Ware et al., 1995, 1996)

A shorter version of the SF-36 was developed 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. 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. The SF-12 may be self-, interview-, or telephone-administered.

Four studies (three of which are very recent) were identified evaluating the use of SF-12 with patients undergoing CABG or PCI; all were US samples. Four studies examined outcomes of CABG; one study examined both CABG and PCI.

Construct validity of the SF-12 was supported by significant correlations with measures of comorbidity and depression, both pre- and post-operatively, in a sample of older patients undergoing CABG (Sorensen & Wang, 2009).

SF-12 BP, GH, and PF domains discriminated between obese and non-obese patients, and all domains discriminated the severely obese (BMI >/= 35 kg/m²) from the remainder of the sample, in a study examining the impact of BMI on outcomes in open-heart surgery (Barnett et al., 2010). SF-12 scores discriminated men and women in the study by Sorensen & Wang

(2009), with women having significantly poorer pre- and post-operative functional status, independently of age.

Responsiveness of the SF-12 was supported in the study by Barnett and colleagues (2010), with statistically significant score improvements in all domains at one year, for all BMI groups. Another study of CABG patients showed significant improvements at three months compared with baseline in both MCS and PCS (Sandau et al., 2008).

Acceptability and feasibility of the SF-12 as part of a survey package for the routine collection of baseline data on patients undergoing CABG or PCI were explored in a study by Spertus and colleagues (2001), with mixed results. Although the survey was acceptable to the majority of patients (92% agreed to participate), it proved difficult to convince nursing staff to integrate such data collection into routine care.

f. Sickness Impact Profile (Bergner et al., 1976; 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. It was intended to inform policy decision-making.

Instrument content was based on the concept of 'sickness', defined as reflecting the change in an individual's activities of daily life, emotional status, and attitude as a result of ill-health (McDowell and Newell, 1996). Items were derived from literature reviews and statements from health professionals, carers, patient groups, and healthy subjects describing change in behaviour as a result of illness. The SIP comprises 136 items across 12 domains: Alertness Behaviour (AB), ambulation (A), Body Care and Movement (BCM), Communication (C), Eating (E), Emotional Behaviour (EB), Home Management (HM), Mobility (M), Recreation and Pastimes (RP), Sleep and Rest (SR), Social Interaction (SI), and Work (W).

Each item is a statement; those that best describe a respondent's perceived health state on that day are ticked; items are weighted, with higher weights representing increased impairment. An overall percentage score can be calculated for the total SIP or for each domain, where 0 is better health and 100 is worse health. Two summary scores are calculated: Physical function (SIP-PhysF), comprising A, BCM, and M, and psychosocial function (SIP-PsychF), comprising AB, C, EB, and SI; the five remaining categories are scored independently. The developers state that subscales can be administered separately without compromising reliability or construct validity (Bergner, 1978). The instrument may be self- or interview-administered.

Four studies were identified which evaluated the SIP with patients undergoing CABG; all were US samples. No recent studies were identified.

Internal consistency of a combination of six SIP subscales applied in a study of perceptions of QoL at one year post-CABG was good, with a Cronbach's alpha of 0.86 (King et al., 1992). A slightly different set of subscales was used in two studies examining age and sex differences in patterns of recovery post-CABG; internal consistency was low to acceptable, with Cronbach's alphas ranging 0.53 (RP) to 0.81 (BCM) (Artinian et al., 1993, 1995). The authors note that these relatively low values (except for the BCM subscale) signify that results should be interpreted with caution (Artinian et al., 1993, 1995).

Construct validity of the SIP was supported in a study of psychosocial predictors of post-operative recovery, where SIP was used as the criterion measure, by strong correlations of

appropriate subscales (AB, RP, SI, SR, W) with a depression measure (Kos-Munson et al., 1988). SIP scores correlated significantly with angina severity, need for rehospitalisation, and return to work in the study by King et al. (1992).

SIP scores discriminated men and women, with men having significantly better post-operative function in every dimension except SI (Kos-Munson et al., 1988). In the study of sex differences in recovery post-CABG, the Ambulation subscale discriminated men and women, with women having significantly greater dysfunction than men at three time-points (Artinian et al., 1995).

Responsiveness of the six subscales used in the study of sex differences by Artinian and colleagues was high, with significant post-operative score changes in all groups (Artinian et al., 1995).

Burden on respondents is high. Artinian and colleagues, who used only 80 out of the SIP's 136 items, note that the length of the measure may have challenged the attention span of respondents (Artinian et al., 1995).

4. CARDIOVASCULAR-SPECIFIC PROMs

Cardiovascular multidimensional measures

Nine multidimensional cardiovascular-specific measures were identified:

a. Cardiac Symptom Survey, CSS
b. Coronary Revascularisation Outcome Questionnaire, CROQ
c. Duke Activity Status Index, DASI
d. Heart Surgery Symptom Inventory, HSSI
e. Kansas City Cardiomyopathy Questionnaire, KCCQ
f. MacNew/QLMI
g. Quality of Life Index-Cardiac Version, QLI-CV
h. Seattle Angina Questionnaire, SAQ
i. Symptoms of Illness Score, SOIS

See Appendix B, Table iv for a summary of content and scoring of these instruments.

a. Cardiac Symptom Survey, CSS (Nieveen et al., 2008)

The Cardiac Symptom Survey is a recently reported 40-item scale intended to examine symptoms and evaluate symptom management in patients having undergone CABG. The measure comprises ten symptoms which are assessed in terms of Frequency and Severity (evaluation of symptoms) and Interference with Physical Activity and Enjoyment of Life (response to symptoms) over the previous seven days. There is an additional item inviting respondents to specify any symptoms that have not been included. Each symptom is scored on a scale of 0-10 for each of the four aspects (Frequency, Severity, Interference – Physical Activity, Interference – Enjoyment of Life). A mean is produced for Frequency/Severity of each item; this is added to single item scores for the Interference items to produce an overall total.

Three current studies have been identified which evaluate the properties of the CSS with US patients undergoing CABG (two studies, including one comparing CABG and MIDCAB) or PCI (one study).

Content of the CSS was derived from interviews with patients and nurse researchers, and literature reviews; this was then tested in pilot studies (see Nieveen et al., 2008; Zimmerman et al., 2002) and evaluated by an expert panel for relevance and clarity.

Internal consistency was supported with correlations between the Frequency and Severity components of each item ranging 0.85-0.98 in a sample of CABG patients (Barnason et al., 2006). Test-retest reliability was demonstrated with the same sample, with correlations ranging 0.92-1.00 (Barnason et al., 2006).

Construct validity of the CSS was supported with significant correlations between CSS items and related SF-36 subscales in a small pilot study (Zimmerman et al., 2002). CSS scores also discriminated between treatment groups: patients undergoing MIDCAB had significantly more symptoms than CABG patients, though this unexplained finding was based on a very small sample (Zimmerman et al., 2002).

Responsiveness was tested in a sample of older patients undergoing CABG (Nieveen et al., 2008). Statistically significant score changes were found for Frequency/Severity in all but

two of the symptoms (fluttering in the chest and anxiety), but for only four of the ten symptoms when assessed for Interference.

Further testing is planned, with a particular focus on predictive validity.

b. Coronary Revascularisation Outcome Questionnaire, CROQ (Schroter & Lamping, 2004, UK)

The UK-developed CROQ is intended to compare quality of life and outcomes in patients undergoing CABG and PTCA. Four forms of the instrument have been developed, namely, pre- and post-operative versions for CABG and PTCA, respectively. Each has a core of 32 items in four domains: Symptoms, and Physical, Psychosocial, and Cognitive Functioning; the post-operative versions have added items for Adverse Effects; and Satisfaction with treatment. All forms of the instrument have a free-text item, allowing the patient to add anything which is not covered in the questionnaire but is important to them.

Five studies evaluating the CROQ have been identified (Ascione et al., 2004, UK; Reeves et al., 2004; Schroter & Lamping, 2000, 2004, 2006, UK). All of these are with UK samples. All five examined outcomes of CABG; three of the five also examined PCI.

Content of the CROQ was derived from literature review, existing measures (SF-36 and SAQ), expert opinion, and interviews with patients (Schroter & Lamping, 2004, UK). Preliminary versions of the CROQ-CABG and CROQ-PTCA were field-tested, items were reduced, and the shortened versions evaluated in a second field test (Schroter & Lamping, 2000, 2004, UK).

Internal consistency of the CROQ was reported as high, with Cronbach's alpha values ranging 0.81-0.96, and item-total correlations all >0.20 (Schroter & Lamping, 2004, UK). Test-retest reliability was high, with ICCs 0.80-0.93 (Schroter & Lamping, 2004, UK).

Construct validity of the CROQ was evidenced by moderate to high correlations with similar domains of the SF-36 and SAQ (Schroter & Lamping, 2004, UK). CROQ scores also discriminated patients by pre-operative severity of angina (CCS classification) and dyspnoea (NYHA class) (Schroter & Lamping, 2004, UK).

Significant score changes among subsamples assessed before and after revascularisation was reported for all CROQ scales, with moderate to large effect sizes, supporting responsiveness of the measure (Schroter & Lamping, 2004, UK). Responsiveness of the CROQ was further illustrated in the BHACAS trial comparing on- and off-pump CABG (Ascione et al., 2004, UK). CROQ scores deteriorated significantly over time, in both treatment groups; this trend was not observed with the other measures used, namely, EQ-5D, SF-36 and SAQ, suggesting that the CROQ may have greater sensitivity in this population (Ascione et al., 2004, UK).

In a head-to-head comparison with the SAQ and SF-36 in patients undergoing CABG or PTCA, the CROQ showed comparable responsiveness to the SAQ, and greater responsiveness than the SF-36 (Schroter & Lamping, 2006, UK), though all three measures showed large effect sizes (ES) and SRMs in several domains, particularly with the CABG patients. In the CABG group, the CROQ Psychosocial scale demonstrated a significantly lower ES than the SAQ Disease Perception scale, although there was no significant difference for the symptoms and physical functioning scales of the two instruments (Schroter & Lamping, 2006, UK). ES and SRM for the CROQ Psychosocial scale were significantly

larger than for the SF-36 SF and MH subscales, although PF scales of the two measures were comparable (Schroter & Lamping, 2006, UK). In the PTCA sample, the SRM for CROQ Symptoms was significantly larger than for the SAQ Anginal Frequency scale, but the SRM for CROQ PF was significantly lower than for SAQ Exertional Capacity (Schroter & Lamping, 2006, UK). The ES for CROQ PF was significantly larger than for SF-36 PF and, as in the CABG group, ES and SRM for the CROQ Psychosocial scale were significantly larger than for SF-36 SF and MH (Schroter & Lamping, 2006, UK).

Responsiveness of the CROQ was further illustrated in a trial comparing the clinical effectiveness of MIDCAB versus PTCA (Reeves et al., 2004, UK; HTA report). CROQ scores favoured MIDCAB; however, between-group differences were statistically significant only in the cognitive functioning domain, at three months (Reeves et al., 2004, UK). As the authors of this study have noted, given the number of comparisons carried out (five and six dimensions/scores at three different time points), a single significant finding should be interpreted with caution (Reeves et al., 2004, UK).

Acceptability of the CROQ was supported by response rates of over 80% in the original development study (Schroter & Lamping, 2000, UK), and a return rate of 87% for a battery including the CROQ in the BHACAS trial (Ascione et al., 2004, UK); both studies used postal administration. In the BHACAS trial (Ascione et al., 2004, UK), completion rates for CROQ items were slightly higher than for the SF-36 and SAQ, suggesting that patients in this sample found the measure easier to answer and/or more relevant to their circumstances (Ascione et al., 2004, UK). There was a response rate of 72% and 62% for the CABG and PTCA groups, respectively, in the postal survey comparing CROQ, SAQ, and SF-36 (Schroter & Lamping, 2006, UK); of those who responded, 89% and 94%, respectively, completed the questionnaires post-revascularisation (Schroter & Lamping, 2006, UK).

Time required for completion of the measure is estimated by the developers to be 10 minutes.

c. Duke Activity Status Index, DASI (Hlatky et al., 1989)

The DASI was developed to provide a means of assessing functional capacity in cardiac patients that would be more accurate and feasible to apply than existing measures, namely, the NYHA and CCS classification systems (Criteria Committee of the NYHA, 1994; Campeau, 1976), and the Specific Activity Scale (Goldman et al., 1981). It comprises 12 items representing major aspects of physical function, each weighted according to the known metabolic cost of the activity (MET units); a difference of 2.2 units or greater has been shown to be clinically significant (Hlatky et al., 1997). Although its content is not strictly cardiovascular-specific, and it has been used with other clinical populations (see, for example, Carter et al., 2002), the DASI is generally regarded as a condition-specific measure, and is widely used with cardiovascular populations. An eight-item version has been developed to reduce patient burden, and with a modified scoring system (Phillips et al., 1990; Alonso et al., 1997).

Sixteen studies supporting the use of DASI with patients undergoing CABG or PCI were identified; all were with North American samples. Fourteen studies examined outcomes of CABG; seven studies examined PCI (six studies examined both). Three studies are very recent.

The DASI correlates strongly with the 'gold standard' for functional capacity in cardiac disease, namely, maximal oxygen uptake during exercise testing, and therefore has criterion validity (Hlatky et al., 1989).

Discriminative validity of the DASI was reported in a study of patients undergoing cardiac catheterisation prior to coronary revascularisation (Nelson et al., 1991). Significant correlations were found between DASI scores and demographic factors (age, sex), and between DASI scores and severity of cardiac disease (three-vessel disease, history of MI or heart failure) (Nelson et al., 1991).

Construct validity of the DASI was supported in the Study of Economics and Quality of Life ancillary to the BARI trial (BARI-SEQOL) by significant correlations with angina symptoms (Bourassa et al., 2000), and with TTO scores (Melsop et al., 2003).

The DASI discriminated between treatment groups in a major study comparing the outcomes of coronary angioplasty versus bypass surgery, with significantly greater improvement in the CABG group (Hlatky et al., 1997; BARI trial). Improvement in DASI scores also differed significantly in those with HF compared to those without HF, in diabetics versus non-diabetics, in men versus women, and in older versus younger patients (Hlatky et al., 1997). However, the difference between treatment groups narrowed over time, and was no longer significant at four years (Hlatky et al., 2004; BARI trial).

DASI scores showed significant differences between treatment groups at four months in another large study comparing PCI and medical treatment (Mark et al., 2009a; OAT trial), though the difference was not sustained over time. DASI scores also discriminated between patients with and without prior symptoms (Pilote et al., 1995), and between Canadian and US patients in the SEQOL substudy (Pilote et al., 1995; Bourassa et al., 2000). The latter finding appears to reflect different patterns of care in the two countries.

The DASI has been used in a number of studies investigating gender differences in the experience of cardiac surgery. The measure appears to discriminate between men and women, as illustrated in a study of patients undergoing CABG where women's scores on the DASI were significantly lower at baseline and six months than men's, although the degree of change in score was comparable (Stewart et al., 1999). Women had significantly lower DASI scores at baseline in a study examining the effect of gender on early recovery from CABG and/or valve surgery (King, 2000). At three months post-procedure, this difference was no longer significant; however, women experienced a significantly greater degree of improvement (King, 2000). DASI scores showed significantly poorer outcomes for women at one year post-CABG, independent of pre-existing risk factors (Phillips-Bute et al., 2003). Women undergoing CABG had significantly poorer HRQoL than men as reflected in DASI score, at baseline and follow-up, independent of age, co-morbidity, and post-operative sequelae (Koch et al., 2004). Significant differences in DASI scores between men and women undergoing CABG were found at three time-points in a study by Sawatzky & Naimark (2009b).

Predictive validity of the DASI was reported in two studies. DASI scores taken during the recovery period (at six and 12 months) after CABG and/or valve surgery predicted long-term survival (median follow-up 8.6 years) (Koch et al., 2007). Baseline DASI score predicted early readmission post-discharge in a study of risk and outcomes with a CABG sample (Sawatzky & Naimark, 2009a).

Responsiveness of the DASI was illustrated in the study by King (2000), with significant improvement in functional status for both genders. Significant score changes were reported for the CABG group in a study comparing outcomes among three groups of patients with advanced CAD (Kandzari et al., 2001; MOSS study). Significant improvement in DASI scores post-revascularisation was reported for the Canadian (but not the US) group in the GUSTO-IIb QoL substudy (Kaul et al., 2004). The DASI also demonstrated responsiveness in the survival study by Koch et al. (2007); however, distribution of scores was skewed, with a clustering of scores at the higher end of the range, suggesting the measure lacks precision for patients with better functional capacity.

Interpretability of the DASI was supported in the 1997 study by Hlatky and colleagues where a small change in score (2.7 units) was found to be clinically meaningful.

Ease of application, or feasibility, of the DASI is noted (Nelson et al., 1991).

d. Heart Surgery Symptom Inventory, HSSI (LaPier & Jung, 2002)

The HSSI was developed to provide a measure of functional status and QoL for patients in the subacute stage of recovery from CABG (from two to six months post-surgery). It covers a broader range of symptoms than most cardiac-specific measures, which focus principally on angina and dyspnoea. In its current form it is not appropriate for pre-operative assessments owing to the large number of surgery-specific items (LaPier & Jung, 2002). The developers suggest it could be readily adapted for use in the immediate post-operative period (up to two months) and/or during long-term recovery, i.e. beyond six months (LaPier & Jung, 2002); however, it has not yet been evaluated for this purpose (LaPier, 2006; LaPier & Wilson, 2006).

Three studies (the most recent being from 2007) and a review article were identified which evaluated the HSSI in US patients having undergone CABG.

Content was derived from previously published qualitative studies, clinicians, and patients recovering from CABG; a preliminary list of items was revised and added to by a panel of specialist physical therapists (LaPier & Jung, 2002). The final version of the instrument comprises 76 items in five symptom categories - namely, cardiac; general (includes sleep, fatigue, sexual functioning, cognitive functioning); trunk (includes incision/drain site pain, wound healing); lower extremity (for those receiving grafts harvested from the saphenous vein); upper extremity (for recipients of radial artery grafts). Respondents are asked to assess symptoms over the previous week, selecting from five Likert-type response options. Domain scores are summed to produce a total, with lower scores indicating greater severity of symptoms.

Internal consistency was supported by significant correlations between domain and total scores, and most inter-domain scores, in a study with a small sample of patients who had undergone CABG six months previously (LaPier, 2006). However, individual item-domain and item-total correlations ranged 0.01-0.85, with several values falling well below the norm of 0.20 (LaPier, 2006). Test-retest reliability was supported by significant correlations between HSSI scores on two administrations; however, these were conducted on the same day, and results could therefore be subject to recall bias (LaPier, 2006).

Construct validity of the HSSI was supported by significant correlations between HSSI and SF-36 subscales in the 2006 study by LaPier. In another small study, construct validity of the HSSI was further supported by significant correlations between HSSI scores (domain and total) and psychosocial assessments, as well as performance-based measures (LaPier, 2007).

e. Kansas City Cardiomyopathy Questionnaire, KCCQ (Green et al., 2000)
The KCCQ is a 23-item used to measure the effect of heart failure in five domains, namely, physical limitations, symptoms, self-efficacy, social interference, and quality of life. It is self-administered; respondents are asked to consider each item over the previous two weeks. A change of 5 points on the scale scores is regarded as clinically important (Spertus et al., 2005). The KCCQ has been extensively evaluated in patients with heart failure (Fitzpatrick et al., 2006).

One study was identified which used the KCCQ to compare outcomes in a US sample of patients undergoing CABG with surgical ventricular reconstruction (SVR) versus CABG alone (Mark et al., 2009b; STICH trial).

Responsiveness of the KCCQ was supported with significant score changes at follow-up compared with baseline, namely, a 30-point improvement (where a 5-point change is considered meaningful), though no between-group differences were found (Mark et al., 2009b).

Response rates for administration by telephone interview ranged 82%-95% over the follow-up period, indicating good acceptability to patients (Mark et al., 2009b).

f. MacNew Heart-Disease Health-Related Quality of Life Questionnaire (Valenti et al., 1996)
The MacNew measures HRQoL in heart disease (myocardial infarction, coronary disease and heart failure) over the previous two weeks. This instrument is a modification of the earlier Quality of Life after Myocardial Infarction (QLMI) Questionnaire (Oldridge et al., 1991; Lim et al., 1993). The MacNew contains 27 items in three domains (Emotional, Physical, and Social). The instrument has been reviewed for reliability, validity, and responsiveness (Höfer, 2004). It takes up to ten minutes to complete, and respondent burden is low (Höfer, 2004).

Two studies were identified using the MacNew; one was with a US sample undergoing CABG, the other with an Australian sample receiving PCI. Both studies were published in 2007.

Responsiveness of the measure was supported in a study comparing cardiac rehabilitation (CR) outcomes in patients having undergone on- and off-pump CABG (Aron et al., 2007), with significant score changes after CR, although between-group differences were insignificant.

Acceptability of the MacNew was moderately well supported by a 75% response rate for postal administration in a cross-sectional study examining HRQoL at different time-points post-PCI (Fernandez et al., 2007).

g. Quality of Life Index-Cardiac Version, QLI-CV (Ferrans and Powers, 1985)
The Quality of Life Index (QLI) was developed to measure QoL in terms of satisfaction with life (Ferrans & Powers, 1985). The instrument consists of two parts: the first measures

satisfaction with various aspects of life, while the second measures the importance of those same aspects. Importance ratings are used to weight the satisfaction responses, so that scores reflect satisfaction with those aspects of life most valued by the respondent. Scores are calculated for quality of life overall and in four domains: Health and functioning (HQOL), Psychological/spiritual (PQOL), Social and economic (SQOL), and Family (FQOL); higher scores indicate greater perceived quality of life. Two items relating to dyspnoea and lifestyle changes due to cardiac problems were added to the generic QLI to create the 35-item QLI-CV.

Four nurse-led studies were identified which support the use of the QLI-CV with US patients undergoing elective coronary procedures, the most recent being from 2005. Three of the studies examined outcomes of CABG; one examined PCI.

Internal consistency reliability of QLI-CV subscales was high, with Cronbach's alphas ranging 0.79-0.90 in a study comparing QoL pre- and post-CABG or PTCA (Papadantonaki et al., 1994). Similar values (range 0.76-0.93) were found in a study comparing QoL after CABG or PTCA (Skaggs and Yates, 1999), while alphas of 0.91 and 0.95 were noted in a study of pre- and post-operative QoL in women undergoing CABG (Penckofer et al., 2005).

QLI-CV HQOL and PQOL domains discriminated between PTCA patients who did and did not experience angina post-operatively in a study comparing QoL after CA and CABG (Skaggs & Yates, 1999). HQOL, PQOL, and SQOL subscales discriminated men and women in a study examining gender differences in the experience of CABG (Keresztes et al., 2003).

Responsiveness of the measure was illustrated by statistically significant score changes in overall QoL and the Health and functioning subscale (Papadantonaki et al., 1994; Penckofer et al., 2005), though changes on the other subscales were non-significant. HQOL, FQOL, and SQOL subscales showed responsiveness to change in the study by Keresztes and colleagues (2003).

h. Seattle Angina Questionnaire, SAQ (Spertus et al., 1994b, 1995)

The SAQ measures the physical and emotional effects of CAD over the previous four weeks. It comprises 19 items in five domains, namely Exertional Capacity or Physical Limitation (PL, nine items), Anginal Stability (AS, one item), and Anginal Frequency (AF, two items), patients' satisfaction with their treatment (TS, four items), and Disease Perception/Quality of Life (QL, three items). There are five or six response options for each question, which are assigned ordinal values. Scores within each dimension are summed and transformed into a 0-100 range, with higher scores indicating better function, fewer symptoms, and better quality of life; there is no overall summary score. A change in score of ten points reflects change that is meaningful to patients (Spertus et al., 1995); however, in a later study the developer suggests that a five- to eight-point change is clinically significant (Spertus et al., 2000).

A UK version of the SAQ, reduced to 14 items in three dimensions and with some linguistic modifications, has been evaluated for validity, reliability, and responsiveness with angina patients (Garratt et al., 2001, UK). However, no further evidence for this version of the instrument was found, and it does not appear to have been tested with patients undergoing revascularisation procedures.

Twenty-two studies were identified which used the SAQ to assess outcomes after elective surgical revascularisation. Five of these were with UK samples (Agarwal et al., 2009; Al-

Housni et al., 2009; Ascione et al., 2004; Kim et al., 2005; Schroter & Lamping, 2006). Eleven studies examined outcomes of CABG, 19 studies examined PCI; eight studies examined both procedures. Three studies were from 2009 (one CABG, two PCI).

Content of the scale was derived in part from Feinstein and Wells' classification system for patients with angina (Feinstein & Wells, 1977) and other existing measures, namely, Goldman's Specific Activity Scale (Goldman et al., 1981) and Peduzzi's Angina Questionnaire (Peduzzi & Hultgren, 1979) – see Appendix C for these two measures. A study examining themes identified as relevant by patients undergoing major surgery, and grouped by the researchers into six domains, found that the SAQ included some (physical and emotional well-being, quality of care) but lacked others (social and spiritual well-being, cognitive preparation for surgery) (Morris et al., 2006).

Internal consistency of the SAQ was considered acceptable in the Stent-PAMI trial (Rinfret et al., 2001) with Cronbach's alphas greater than 0.62 for each subscale. However, a Cronbach's alpha of 0.70 is generally considered the minimum for acceptability (Fitzpatrick et al., 1998).

The scale was initially tested in four groups of patients, including 45 who underwent successful PTCA and 130 with stable CAD (Spertus et al., 1994b; Spertus et al., 1995). Reproducibility of the SAQ was tested in the group of patients with stable CAD; intraclass correlation coefficients (ICC) were high (0.76-0.83), except for the Anginal Stability item (0.24) (Spertus et al., 1995).

Construct validity of each of the five scales was supported by moderate to high correlation of scores with criterion measures - for example, correlation between the PL scale and treadmill test duration was 0.42, while the DP/QL scale and the GH scale of the SF-36 were correlated 0.60 (Spertus et al., 1995).

In a later study by the developers to determine predictors of QoL benefit after PCI, the SAQ AF items had the strongest predictive validity, followed by PL (Spertus et al., 2004). The OPUS trial investigators also found that SAQ scores predicted the need for repeat surgery (Weaver et al., 2000).

The SAQ detected between-group differences and within-group change in a number of studies.

In the Stent or Surgery (SoS) trial, which compared functional status and quality of life outcomes at one year for patients undergoing CABG versus stent-assisted PCI, score changes on the PL, AF, and DP/QL scales showed CABG patients had greater improvement than those assigned to PCI (Zhang et al., 2003). However, while men showed significantly greater improvement at one year with CABG, the same was not true for women, suggesting that they may have less to gain from undergoing the more invasive procedure (Zhang et al., 2004). SAQ scores in the SoS trial also discriminated patients with and without acute coronary syndrome (ACS) (Zhang et al., 2005), with similar implications for treatment choices.

Discriminative validity and responsiveness of the SAQ was also supported in the 2004 study by Spertus et al., with patients who were more severely affected pre-procedure having substantially higher changes in score (by 21-35 points) after PCI. In a study examining outcomes of PCI and CABG in patients stratified by risk of restenosis (Spertus et al., 2005),

the SAQ AF and QL scales showed significant differences between treatment groups among patients at higher risk.

The SAQ detected within-group differences in a large trial (n=2287) comparing the use of optimal medical therapy (OMT) alone and OMT with PCI, with patients having the most severe angina at baseline (i.e. angina frequency scores <50) showing significantly greater improvement at three months after PCI than those with less severe angina (Weintraub et al., 2008; COURAGE trial). This was most marked in the PL, AF, and QL domains (Weintraub et al., 2008). The PL domain, but not AF, AS, or QL, discriminated CHD patients with and without diabetes in a cross-sectional substudy of patients enrolled in the COURAGE trial (Deaton et al., 2006).

The SAQ differentiated those who did or did not require repeat target vessel revascularisation (TVR) in the Stent-PAMI trial (Rinfret et al., 2001). SAQ scores adjusted for age and comorbidities indicated clinically significant differences between men and women in the APPROACH study at one and three years post-PCI (Norris et al., 2004a), an important finding given well-established gender differences in presentation and outcomes for elective coronary procedures (Mikhail, 2003).

SAQ scores discriminated patients undergoing different modes of treatment in several studies. The PL, AF, and QL domains discriminated between treatment groups in a trial comparing outcomes of PCI and CABG (Borkon et al., 2002). A study comparing HRQoL outcomes at one year for patients receiving cardiac surgery, PCI, or medical management found significantly higher SAQ scores for cardiac surgery or PCI versus medical management (Norris et al., 2004b). For PCI with stenting versus PCI without stenting, and for CABG versus PCI (with or without stenting), score differences were also statistically significant except in the exertional capacity (PL) domain (Norris et al., 2004b). All components of the SAQ showed significant differences between treatment groups at four months and one year, in a trial of interventional versus conservative treatment for patients with unstable angina and NSTEMI (Kim et al., 2005, UK; RITA-3 trial).

Sensitivity of the SAQ to large and small clinical changes was evidenced in the original development study by comparing baseline and three-month scores in the PTCA and stable CAD groups (Spertus et al., 1994b; Spertus et al., 1995). There were statistically significant score changes at three months for patients undergoing successful PTCA, and amongst those with relatively stable CAD who reported their condition as having worsened or improved (in response to a separate global question); these findings confirm the results of an earlier study (Spertus et al., 1994a). For the PTCA group, responsiveness of all SAQ scales (except TS) was considerably greater than for the SF-36 as indicated by Guyatt's responsiveness statistic (Spertus et al., 1994b). In the COURAGE trial (Weintraub et al., 2008) responsiveness of the SAQ was demonstrated by statistically significant score changes in all domains after the first three months of treatment.

The SAQ was more sensitive to change than the SF-36 in a study examining outcomes of CABG in a sample of elderly patients; this was illustrated by higher SRMs in all dimensions except PL (where the SRM was comparable with SF-36 PF) and TS (MacDonald et al., 1998). In a head-to-head comparison with the CROQ in patients undergoing CABG or PTCA, SAQ subscales showed comparable responsiveness to similar scales of the CROQ (Schroter & Lamping, 2006, UK), though there were some differences. In the CABG group, the SAQ QL scale had a significantly higher ES than the CROQ Psychosocial scale (Schroter

& Lamping, 2006, UK). In the PTCA sample, the SRM for SAQ AF was significantly lower than for CROQ Symptoms, while the SRM for SAQ PL was significantly higher than for CROQ PF (Schroter & Lamping, 2006, UK).

Numerous other studies have supported the responsiveness of the SAQ, as evidenced by statistically significant longitudinal score changes on SAQ scales (Agarwal et al., 2009, UK; Al-Housni et al., 2009, UK; Borkon et al., 2002; Kim et al., 2005, UK; Norris et al., 2004b; Spertus et al., 2004, 2005; Zhang et al., 2003, 2004, 2005).

Patient burden is low, as the measure takes less than five minutes to complete. It is also designed in a machine-readable format to facilitate data entry, enhancing its feasibility of application in the clinical setting. A return rate of 90%, and 87% for postal administration of the SAQ was reported by the OPUS investigators and the BHACAS trial, respectively (Weaver et al., 2000; Ascione et al., 2004, UK) indicating good acceptability to patients. An even higher completion rate (93%) was reported in the study by Borkon et al. (2002), which used a mixture of postal and telephone administration. Response rates of 93% and above were reported for in-hospital administration of a battery of instruments including the SAQ in the RITA-3 trial (Kim et al., 2005, UK). The APPROACH investigators noted a 78% return rate for questionnaires mailed at one year (Norris et al., 2004a; Norris et al., 2004b) suggesting moderate acceptability. There was a response rate of 72% and 62% for the CABG and PTCA groups, respectively, in the postal survey comparing SAQ, CROQ, and SF-36 (Schroter & Lamping, 2006, UK); of those who responded, 89% and 94%, respectively, completed the questionnaires post-revascularisation (Schroter & Lamping, 2006, UK).

Interpretability of the SAQ was supported in the 2004 study by the developer and colleagues (Spertus et al., 2004) who proposed baseline score ranges corresponding to minimal, mild, moderate, and severe limitation (>75, 51-75, 25-60, 0-25, respectively) for the PL dimension, and scores for four levels of AF (100 - no angina, 61-99 - monthly angina, 31-60 - weekly angina, 0-30 - daily angina). Score changes were also categorised as large deterioration (< −20), moderate deterioration (−10 to −20), minimal change (−10 to +10), moderate improvement (+10 to +20), and large improvement (> +20) (Spertus et al., 2004). A different set of score-ranges was adopted in the COURAGE trial (Weintraub et al., 2008), where clinical significance was defined as ≥8, ≥25, ≥20, ≥12, and ≥16 points difference for the PL, AS, AF, TS, and QL domains, respectively.

Feasibility of the SAQ as part of a survey package for the routine collection of baseline data on patients undergoing CABG or PCI was explored in a study by the developer and colleagues (Spertus et al., 2001). A major challenge was the number of different admission portals for these patients. In addition, although the survey was acceptable to the majority of patients (< 8% refused), it proved difficult to convince nursing staff to integrate such data collection into routine care, as the information gained did not appear relevant to their specific role.

i. Symptoms of Illness Score (Jenkins et al., 1990, 1994)

The SOIS was developed by Jenkins and colleagues from the Recovery Study, to quantify and predict recovery after heart surgery. It comprises several existing measures, namely, the Rose Angina Questionnaire (Rose & Blackburn, 1968), the Rose Dyspnoea Questionnaire (Rose & Blackburn, 1968), the Fatigue and Vigour subscales of the Profile of Mood States (McNair et al., 1971), the Sleep Problems Scale (Jenkins et al., 1988), and additional questions on cardiac symptoms, and physical and psychological recovery.

Content of the SOIS was based on factor analysis of 58 items with a US sample of patients having undergone CABG or valve surgery (Jenkins et al., 1990). Internal consistency was high, with a Cronbach's alpha of 0.84 (Jenkins et al., 1994). SOIS scores discriminated women and men, with women reporting significantly more symptoms than men.

No further studies using the SOIS were identified in this review.

Cardiovascular dimension-specific measures

Fifteen cardiovascular-specific measures focussing on a single symptom or dimension were also identified. Except where indicated, these were supported for use with CABG/PCI patients by a single study only. Description of content and evidence for these measures are summarised in Appendix C.

a. Angina Questionnaire (2 studies)
b. Barnason Efficacy Expectation Scale
c. Cardiac Adjustment Scale
d. Cardiac Depression Scale
e. Cardiac Event Threat Questionnaire
f. Cardiac Self-Efficacy Scale
g. Cardiac Surgery Symptom Inventory
h. Cardiac Symptoms Scale (2 studies)
i. Control Attitudes Index (2 studies)
j. ENRICHD Social Support Index (3 studies)
k. Rose Angina Questionnaire (2 studies)
l. Rose Dyspnoea Questionnaire
m. Specific Activity Scale (3 studies)
n. Symptom Inventory (2 studies)
o. Symptom Scale

5. SUMMARY AND RECOMMENDATIONS

Table 2 provides a summary of the evidence of psychometric properties for the instruments identified in this review. Table 3 shows the HRQoL domains covered by each PROM. A summary of their content and scoring method can be found in Appendix B.

Preference-based measures
Amongst the preference-based measures, the EQ-5D was the only instrument with sufficient evidence to warrant consideration (seven studies; five with UK samples). It demonstrated validity and responsiveness with both CABG and PCI samples. However, there was mixed evidence regarding its acceptability to patients, as reflected in response rates. Where a utility is required, the EQ-5D is the most appropriate measure to use.

Generic measures
Amongst the generic, multidimensional instruments reviewed, the SF-36 had by far the most evidence supporting its use with the ECP population (39 studies, eight with UK samples). The SF-36 appears to work well in both types of procedure, although there were many more studies with patients undergoing CABG compared with PCI. Evidence for reliability, validity, responsiveness, and acceptability to patients was substantial; however, one study reported ceiling effects, and some studies had low response rates. If the aim is to compare ECP patients with other disease populations, the SF-36 may be considered; however, if used alone, it may not be sensitive to subtle changes within this particular group.

Condition-specific measures
Of the condition-specific instruments identified, the Seattle Angina Questionnaire had by far the most evidence supporting its use with patients undergoing CABG or PCI (22 studies), demonstrating most of the important psychometric properties. It has been extensively applied in both types of procedure, and with UK patient samples (five studies). However, it was developed for the assessment of CAD in general, and could be said to lack some items that are important for patients undergoing cardiac surgery. In particular, psychological well-being appears to be an important predictor of recovery, and there is evidence that cardiac procedures, particularly CABG, may have long-term effects on cognitive functioning. The SAQ does not include cognitive functioning, and does not fully cover the psychosocial domain.

Although there is as yet relatively little evidence to support its use (five studies, all UK), the Coronary Revascularisation Outcomes Questionnaire, recently developed in the UK, is promising. The CROQ includes both psychosocial and cognitive functioning, as well as surgery-specific items. Some of the other measures reviewed, such as the Cardiac Symptom Survey and the Heart Surgery Symptom Inventory, may also prove valuable if further evidence emerges. Choice of instrument may be influenced by whether short- and medium-, or longer-term outcomes are the focus of investigation. If short- and medium-term outcomes are the main object of attention, one of these surgery-specific measures could be the instrument of choice.

The Duke Activity Status Index has considerable merit as a measure of cardiac patients' capacity to engage in activities which are important to their QoL. It has greater reliability and sensitivity than the NYHA and CCS classification systems, and stronger psychometric properties than the Specific Activity Scale; it has also been widely used with patients

undergoing ECP. However, it is (intentionally) narrow in scope, and does not appear to have been evaluated in UK patients undergoing ECP.

There is little evidence to support the use with patients undergoing ECP of two other well-established PROMs for CAD, namely the Kansas City Cardiomyopathy Questionnaire and the MacNew.

Amongst the large number of cardiovascular-specific, dimension-specific PROMs identified, some may be useful for narrowly-focused studies. However, most of these measures have limited evidence of psychometric properties and, given their limited scope, cannot be recommended for routine use in the NHS.

Recommendations
Based on this appraisal, the following instruments were recommended for consideration by a multidisciplinary panel (see Appendix D):

1. SF-36
2. EQ-5D
3. SAQ
4. CROQ

According to the ratings and comments of the panel, the EQ-5D and the SF-36 scored equally well as generic measures of health status whilst the cardiovascular-specific measures, namely, the SAQ and the CROQ suite of questionnaires, were also rated as equivalent.

The PROM Group concludes that the following measures have the strongest evidence for use with patients undergoing elective procedures for coronary revascularisation:

a. Preference-based measure: EQ-5D
b. Generic, multidimensional measure: SF-36
c. Cardiovascular-specific, multidimensional measure: SAQ

In the third category, with further evidence, the CROQ would merit consideration in the future.

Table 2: Appraisal of PROMs included in the review *(see Appendix B for a guide to these rating scales)*

PROM	Reproducibility	Internal consistency	Validity – content	Validity – construct	Responsiveness	Interpretability	Precision	Acceptability	Feasibility
Preference-based measures									
15D	0	+	+	+	+	0	0	+	-
EQ-5D	0	n/a	0	++	+	+	0	+	0
HUI	0	0	0	+	+	0	-	+	-
SF-6D	0	0	+	+	+	0	-	+	0
Generic measures									
FSQ	0	+	0	+	+	0	0	+	0
NHP	0	0	0	+	+	+	0	0	0
SF-36	0	++	0	+++	+++	0	-	++	0
SF-20	0	0	0	+	+	0	0	0	0
SF-12	0	0	0	++	+	0	0	+	-
SIP	0	+	0	++	+	0	0	-	0
Condition-specific measures									
CSS	+	+	+	+	+	0	0	0	0
CROQ	+	+	+	+	+	0	0	+	0
DASI	0	0	++	+++	++	+	+	0	++
HSSI	+	+	+	+	0	0	0	0	0
KCCQ	0	0	0	0	+	0	0	+	0
MacNew	0	0	0	0	+	0	0	+	0
QLI-CV	0	+	0	+	+	0	0	0	0
SAQ	+	+	++	+++	+++	++	0	++	+
SOIS	0	+	+	+	0	0	0	0	0

Table 3: Summary of instruments: health status domains *(after Fitzpatrick et al., 1998)*

Instrument	*Instrument domains*								
	Physical function	**Symptoms**	**Global judgement of health**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**	**Satisfaction with care**
Preference-based measures									
15-D	x	x		x	x	x	x		
EQ-5D	x	x	x	x	x		x		
HUI3	x	x		x		x			
SF-6D	x	x		x	x		x		
Generic measures									
FSQ	x		x	x	x		x	x	
NHP	x	x		x	x		x		
SF-36	x	x	x	x	x		x		
SF-20	x	x	x	x	x		x		
SF-12	x	x	x	x	x		x		
SIP	x	x		x	x	x	x		
Condition-specific measures									
CSS	x			x				x	
CROQ	x	x		x	x	x			x
DASI	x				x		x		
HSSI		x			x	x			
KCCQ	x	x		x	x		x	x	
MacNew	x	x		x	x		x	x	
QLI-CV	x	x	x	x	x		x	x	x
SAQ	x	x	x					x	x
SOIS	x	x		x					

Appendix A

Appraisal of the methodological quality of PROMs

A simple rating scale (Table i) was used to rate the sum total of evidence available for each dimension or criterion against which PROMs were assessed. The dimensions or criteria are summarised in Table ii.

Table i: Psychometric and operational criteria

0	*not reported (no evaluation completed)*
—	*Evaluation evidence available indicating poor performance of instrument*
+	*Some limited evidence in favour*
++	*Good evidence in favour*
+++	*Excellent evidence in favour*

Table ii: Appraisal criteria

Appraisal component	Definition/test	Criteria for acceptability
Reliability		
Test-retest reliability	The stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and re-test scores	Test-retest reliability correlations for summary scores 0.70 for group comparisons
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of items in a scale); assessed by Cronbach's alpha's and item-total correlations	Cronbach's alphas for summary scores ≥0.70 for group comparisons Item-total correlations ≥ 0.20
Validity		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to cover; assessed qualitatively during the questionnaire development phase through pre-testing with patients. Expert opinion and literature review	Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale represent the construct being measured Patients involved in the development stage and item generation
Construct validity	Evidence that the scale is correlated with other measures of the same or similar constructs in the hypothesised direction; assessed on the basis of correlations between the measure and other similar measures	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation
	The ability of the scale to differentiate known-groups; assessed by comparing scores for sub-groups who are expected to differ on the construct being measured (e.g a clinical group and control group)	Statistically significant differences between known groups and/or a difference of expected magnitude
Responsiveness	The ability of a scale to detect significant change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including t-tests, effect sizes (ES), standardised response means (SRM) or responsiveness statistics	Statistically significant changes on scores from pre to post-treatment and/or difference of expected magnitude
Floor/ceiling effects	The ability of an instrument to measure accurately across full spectrum of a construct	Floor/ceiling effects for summary scores <15%
Practical properties		
Acceptability	Acceptability of an instrument reflects respondents' willingness to complete it and impacts on quality of data	Low levels of incomplete data or non-response
Feasibility/burden	The time, energy, financial resources, personnel or other resources required of respondents or those administering the instrument	Reasonable time and resources to collect, process and analyse the data

Appendix B

Table iii: summary of content and scoring of preference-based and generic instruments

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration (completion time, in minutes)*
PREFERENCE-BASED MEASURES				
15 Dimensions Quality of Life questionnaire, 15-D (15)	*15 items:* Mobility, Vision, Hearing, Breathing, Sleeping, Eating, Speech, Elimination, Usual activities, Mental function, Discomfort & symptoms, Depression, Distress, Vitality, Sexual activity	5 ordinal response options for each item, 1=full function, 5=no/minimal function	1) summation of item scores 2) calculation single index score based on general population weightings 0-1, where 0=dead, 1=no problems in any dimension	Self (5-10)
European Quality of Life instrument, EuroQol/EQ-5D (5+1)	*EQ-5D* Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) *EQ-thermometer* Global health (1) *Health 'today'*	*EQ-5D* Categorical: 3 options *EQ-thermometer* VAS	*EQ-5D* Summation: domain profile Utility index (–0.59 to 1.00) *EQ-Thermometer* VAS (0-100)	Interview or self
Health Utilities Index 3 (8)	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, Pain	Three domains have five response options, five have six response options	Global Utility index and single attribute utility scores for the eight separate dimensions	Interview, telephone or self
SF-6D: MOS 6-dimensional health state classification (6)	Bodily pain (BP) (1), Energy/Vitality (VT) (1), Mental health (MH) (1), Physical functioning (PF) (1), Role limitation (1), Social functioning (SF) (1)	Categorical: 3 options	Algorithm Domain profile (0-100, 100 best health)	Interview or self
GENERIC MEASURES				
Functional Status Questionnaire (34)	Physical function: BADL (3), IADL (6) Psychological function: MH (5) Social/role function: Work Performance (6), Social Activity (3), Quality of Interaction (5) 6 single items: work situation, no. days in bed, activity restriction (no. days), satisfaction with sexual relationships, satisfaction with health, frequency of social activities	Over past month Categorical: 4-6 options	Algorithm 6 scale scores 0-100, 100=maximum function, presented by computerised reports as VASs	Self (15)
Nottingham Health Profile, NHP (38)	Bodily pain (BP) (8), Emotional reactions (ER) (9), Energy (E) (3), Physical mobility (PM) (8), Sleep (S) (5), Social isolation (SI) (5)	Yes/no; positive responses weighted	Algorithm Domain profile 0-100, 100 is maximum limitation	Interview Self (10-15)
SF-36: MOS 36-item Short Form Health Survey (36)	Physical functioning (PF) (10), Role limitation-physical (RP) (4), Bodily pain (BP) (2), General health (GH) (5), Vitality (VT) (4), Social functioning (SF) (2), Role limitation-emotional (RE) (3), Mental health (MH) (5), Health transition (1)	Recall: standard 4 weeks, acute 1 week Categorical: 2-6 options	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean 14-15) Self (mean 12.6)

SF-20: MOS 20-item Short Form Health Survey (20)	Bodily pain (BP) (1), General health (GH) (5), Mental health (MH) (5), Physical functioning (PF) (6), Role functioning (RF) (2), Social functioning (SF) (1)	Recall: standard 4 weeks, acute 1 week Categorical: 3-6 options	Algorithm Summation Domain profile (0-100, 100 best health)	Interview, telephone or self (5-7)
SF-12: MOS 12-item Short Form Health Survey (12)	Bodily pain (BP) (1), Energy/Vitality (VT) (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)	Recall: standard 4 weeks, acute 1 week Categorical: 2-6 options	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview or self
Sickness Impact Profile, SIP (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)	Current health Applicable statements ticked. Items weighted: higher weights indicate increased impairment	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 iv: summary of content and scoring of condition-specific instruments

Instrument name (total items)	*Domains (no. items)*	*Response options*	*Scoring*	*Administration* *Completion time, where reported*
Cardiac Symptom Survey (40)	*10 symptoms:* Angina Shortness of breath Fatigue Depression Sleeping difficulty Incisional pain Leg swelling Palpitations Anxiety Poor appetite *Each rated for:* *- frequency| symptom* *- severity | evaluation* *- interference with PF | symptom* *- interference with QoL| response*	Symptoms over 7 days *Symptom evaluation* 0-10: 0=absence of symptom, 10=very frequent/severe *Symptom response* 0-10: 0=no interference, 10=a great deal of interference	Mean of frequency & severity calculated for each item; response-to-symptom items are single-item scores	
Coronary Revascularisation Outcome Questionnaire (33-52)	*32 core items:* Symptoms (7) Physical functioning (8) Psychosocial functioning (14) Cognitive functioning (3) *Plus, 1 descriptive item (not scored)* *Post-op version, added items:* Satisfaction with treatment (6) *2 descriptive items:* Anxiety - return of symptoms Hospital readmission *CROQ-CABG_Post (52)* Adverse effects (11) *CROQ-PTCA_Post (47)* Adverse effects (6)	3- to 6-point response options Self-administration	Items in each scale summed, transformed 0-100 scale, where 0=worst, 100=best outcome	10 mins

Duke Activity Status Index, DASI (12)	*12 items:* Personal care (1) Ambulation (4) Household tasks (4) Sexual function (1) Recreational activities (2) *Shortened version: 8 items – symptoms over 2 weeks:* Ambulation (4) Household tasks (3) Exercise (1)	Yes/no responses	–ve responses score 0, +ve responses assigned a MET rating between 1.75 and 8.00, and summed. Score range 0-58.2 *Shortened version:* 3-point responses, 1=not done for health reasons, 3=done without difficulty. Score range 8-24, higher scores indicate better functional capacity	Interview- or self-administered
Heart Surgery Symptom Inventory, HSSI (76)	cardiac (13) general (22) trunk (21) lower extremity (10) – *for saphenous vein grafts* upper extremity (10) – *for radial artery grafts*	Symptoms over previous week 5-point Likert scales, from 0=not at all to 4=very much	Higher scores indicate worse QoL Domain scores summed to create a total	Self-complete or interview
Kansas City Cardiomyopathy Questionnaire (KCCQ) (23)	*23 items in 5 domains:* Physical limitation (6) Symptoms (8) Self-efficacy and knowledge (2) QoL/mood (3) Social limitation (4)	Symptoms over previous 2 weeks 5-, 6-, and 7-point Likert scales	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) (23-27)	*23-27 items in 3 overlapping domains:* Emotional Physical Social	Symptoms over 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 mins
Quality of Life Index, Cardiac Version IV (35 x 2)	*35 items in four domains, each rated for satisfaction & importance:* Health & functioning Socioeconomic Psychological/spiritual Family	Likert-type scales 1=very dissatisfied/unimportant to 6=very satisfied/important	Score range 0-30, higher scores reflect better QoL	

Seattle Angina Questionnaire, SAQ (19)	*19 items:* Physical Limitation (9) Anginal Stability (1) Frequency of Angina (2) Treatment Satisfaction (4) Disease Perception (3)	Symptoms over previous 4 weeks 5 or 6 response options	Domain scores transformed into 0-100 scales; higher scores indicate better QoL. No summary score	5 mins
Symptoms of Illness Score	*No. items not specified; covers:* Angina Dyspnoea Other cardiac symptoms Fatigue/vigour Sleep problems Physical recovery Psychological recovery	*Not specified*	Scale scores summed. Higher scores indicate many or more severe symptoms	

Appendix C

Table v: Summary of cardiovascular-specific, dimension-specific measures

Instrument References	***Intended purpose/study population***	***Content and scoring***	***Psychometric properties reported with ECP populations*** *[other cardiac populations]*	***Comment***
Angina Questionnaire Peduzzi & Hultgren, 1979 Peduzzi, Hultgren et al., 1987 VA Co-operative Study of Coronary Artery Surgery	Patients with CAD	1. Severity of angina (frequency of angina, presence of rest/nocturnal angina, type of activity producing angina) 2. Medications used *Severity score + medication score summed to produce total; score range 0-18 where <7=mild angina, >11=severe angina*	Reliability – test-retest Validity - discriminative Responsiveness Interpretability	A source for content of SAQ (Spertus et al., 1994b)
Barnason Efficacy Expectation Scale Barnason et al. (2002)	Patients undergoing CABG – measures self-efficacy relating to recovery	*15 items covering:* Physical functioning Psychosocial functioning CAD risk factor modification Self-care management 4-point response format	Validity – content Reliability – internal consistency Validity – construct	
Cardiac Adjustment Scale **Rumbaugh (1966) - source* Brown & Rawlinson (1979)	Psychological assessment of patients with cardiac disease – potential for rehabilitation Patients (n=51) randomised to surgical or medical management of angina	160 items with yes/no responses Score range: 0=poor adjustment, 156=best adjustment 15-20 mins to administer	*[reported by developers*:* *Validity – content* *Reliability – internal consistency* *Validity – predictive* *Feasibility]* Reliability – test-retest	
Cardiac Depression Scale **Hare & Davis (1996) – source Birks et al. (2004)* King et al. (2009)	Measure of depression in cardiac disease; developed in medical cardiac patients in Australia, further validated in the UK Men recovering from CABS (n=120)	*26 items, five factors:* Sleep (2) Anhedonia (3) Uncertainty (6) Mood (5) Cognition (4) Hopelessness (3) Inactivity (3) 7-point response scales, higher scores indicate more severe depression	*[reported by developers*:* *Validity – content* *Reliability – internal consistency* *Validity – construct* *Feasibility]* Reliability – internal consistency Validity – construct Responsiveness Acceptability	Appears more sensitive to depression in this population than BDI and HADS

Cardiac Event Threat Questionnaire, CTQ Bennett et al. (1996)	Measures threat/stress in relation to cardiac events Patients hospitalised for angina, MI or cardiac surgery (n=270)	*32 items measuring threat in relation to:* Fatigue (10) General health (9) Disease-specific symptoms (6) Work (4) Family (30) 4-point response scales, score range 32-128, with higher scores indicating greater threat	Validity – content Validity – construct Reliability – internal consistency Reliability – test-retest	Significantly correlated with POMS
Cardiac Self-Efficacy Questionnaire **Sullivan et al. (1998) – source* Mark et al. (2009b)	Measures patient confidence in controlling symptoms and maintaining function in CHD Comparison CABG + SVR with CABG (n= 991)	*13 items:* Controlling symptoms, SE-CS (7) Maintaining function, SE-MF (6) 5-point response scales, score range 0-100, higher scores reflect greater confidence	*[reported by developers*:* *Validity – content* *Reliability – internal consistency* *Validity – construct]*	*(no reporting of psychometric properties in Mark study)*
Cardiac Surgery Symptom Inventory Miller & Grindel (2004) Gallagher (2004)	Comparison of symptoms in older (>65) and younger (<65) patients undergoing CABS (n=102)	10 pre- and 16 post-operative symptoms Respondents indicate frequency over past week on a 0-5 scale where 0=none, 5=constantly	Validity – content Reliability – internal consistency	Instrument developed for this study Lacks some important post-op symptoms, e.g. fatigue, sleep disturbance
Cardiac Symptoms Scale Plach & Heidrich (2001) Plach & Heidrich (2002)	Measures symptoms following cardiac procedures. Pilot study: midlife and older women receiving heart surgery or CA (n=58) Midlife and older women receiving CABG, valve surgery, or ASD repair (n=157)	Frequency of physical symptoms following cardiac procedures 8 items, 5-point response format: 1=not at all, 5=several times/day	Validity – content Reliability – internal consistency Validity – construct Validity – discriminative Validity – construct Acceptability	Instrument developed for this study

Control Attitudes Index* **Moser & Dracup (1995) – source* Doering et al. (2005)	Perceived control in cardiac disease; patient and family versions. Evaluation of physical and emotional recovery in patients recovering from CABS (n=72)	4 items with 7-point Likert-scale responses. Items summed to produce total score; higher scores indicate higher perceived control	*[reported by developers*:* *Reliability – internal consistency* *Reliability – test-retest* *Validity – construct* *Validity – discriminative]* Validity – construct	**referred to by Doering et al. as 'Cardiac Attitudes Index' and cites 19 items.* Discriminated depressed/non-depressed patients at two time-points
ENRICHD Social Support Index, ESSI **ENRICHD Investigators (2000)* Vaglio et al. (2004) Mallik et al. (2005) Deaton et al. (2006) COURAGE substudy	Measures social and emotional support in cardiac patients *Patients undergoing:* PCI (n=271) CABG (n=963) CABG or PCI (n=1013)	5 items** assessing perceived social support (PSS) 5-point response categories where 1=none of the time, 5=all of the time. Scores </=3 on 2 or more items, and total </= 18 indicate depression/low PSS	*[reported by developers*:* *Validity – content* *Reliability – internal consistency* *Validity – construct* *Interpretability]* Reliability – internal consistency Reliability – test-retest Validity – construct Validity – predictive Feasibility Acceptability	**Vaglio et al. (2004) list 7 items
Rose Angina Questionnaire **Rose (UK; 1962, 1965, Rose & Blackburn, 1968) – source* Jenkins et al. (1983) Trzieniecka-Green & Steptoe (1996) Mark et al. (2009a)	Measures cardiac-specific pain Patients undergoing CABG (n=318) MI (n=50) and CABG (n=50) patients undergoing stress management training PCI versus medical therapy (n=951)	*8-18 items covering:* Site of pain Severity of pain Quality of pain Duration of attack Activities that provoke pain Presence of rest pain Response to pain Yes/no responses Interview- (18 items) or self-administered (8 items)	*[reported by developers*:* *Validity – content* *Validity – construct* *Feasibility]* Responsiveness	Also known as the London School of Hygiene Angina Scale; is a component of SOIS (Jenkins et al., 1994)

Rose Dyspnoea Questionnaire *Rose & Blackburn (1968) – source* Jenkins et al. (1983) Mark et al. (2009a)	Measures exertional dyspnoea Patients undergoing CABG (n=318) PCI versus medical therapy (n=951)	*4 items* Yes/no responses	Responsiveness	Also known as the London School of Hygiene Dyspnoea Scale; is a component of SOIS (Jenkins et al., 1994)
Specific Activity Scale **Goldman et al. (1981) – source* Keresztes et al. (2003) Krumholz et al. (1996) Penckofer et al. (2005)	Measures ability to perform activities without cardiac symptoms 40 matched M/F pairs undergoing CABG Patients undergoing PCI (n=98) Women undergoing CABG (n=61)	21 activities with METS ratings; respondents assigned to functional class based on highest rated activity where I=highest, IV=lowest Interview-administered	*[reported by developers*:* *Reliability – test-retest* *Validity – construct* *Feasibility]* Validity – discriminative Responsiveness	A source for content of SAQ (Spertus et al., 1994b) Substantial ceiling effect reported by Krumholz et al.
Symptom Inventory Artinian et al. (1993), Artinian & Duggan (1995)	Measures recovery in cardiac surgery Patients undergoing CABS (n=184)	20 items – symptoms over previous week 7-point response scales where 1=not at all, 7=always	Reliability – internal consistency Validity – discriminative Responsiveness	
Symptom Scale **Keresztes et al. (1993) - source* Keresztes et al. (2003)	Assesses cardiac symptoms and interference with functional ability. 40 matched M/F pairs undergoing CABG	*Three subscales(score range):* Angina, AS (0-28) Dyspnoea, SOBS (0-28) Fatigue, FS (0-25) Symptoms rated for frequency, severity, ease of occurrence, interference, methods of relief, on 3- to 6-point response scales. Scores summed to produce subscale scores; total score obtained by summing subscales. Range 0-81, where 0=no symptoms, 81=very severe symptoms	*[reported by developers*:* *Reliability – internal consistency* *Validity – construct]* Validity – discriminative Responsiveness	

Appendix D

Membership, methods of working, and conclusions of the multidisciplinary panel

A search was conducted, using published literature and the websites of professional organisations and health authorities, to identify a multidisciplinary panel with clinical or research experience of elective coronary revascularisation procedures (ECRP) and a special interest in patient-reported outcome measures (PROMs). Eight persons were invited to participate, six of whom returned completed ratings scales. The group of six respondents comprised two consultant cardiac surgeons, a consultant cardiologist, a professor of nursing with particular expertise in cardiovascular disease, a research psychologist with an interest in ECRP, and an expert in clinical governance based in a cardiothoracic centre.

The panel were sent the following documents:

- 'A Structured Review of Patient-Reported Outcome Measures used in Elective Procedures for Coronary Revascularisation', February 2010;
- copies of the PROMs short-listed for discussion (SF-36, EQ-5D, SAQ, CROQ).

The panel were sent by e-mail rating scales to judge the suitability of the questionnaires for use in the NHS for the evaluation of services. The panel was asked whether they could suggest other PROMs and whether they were in overall agreement with the conclusions of the review; there was a section for additional 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.

Five panel members provided scores for instruments; the sixth panel member feeling that no measure was sufficiently appropriate to give a score.

Notes of electronic discussion: June-August 2010

Generic measures

SF-36

Members of the group had experience of using this well-established instrument, widely used in ECRP. The availability of comparative and standardised normative data offered the advantage of allowing for comparisons with other populations, including non-UK populations. Given that recovery from revascularisation may be affected by multiple morbidity, some panel members considered use of a generic measure to be appropriate. However, others warned of the risk of false negatives, and pointed out that the success of revascularisation ought not to be judged by the impact on symptoms which cannot be ameliorated by the procedure.

Uncertainty was expressed as to whether subscale or summary scores (PCS and MCS) should be reported; summary scores were felt to limit Type I errors. There was also a question

regarding the criterion for measuring change: whether a five- or a ten-point change be considered the standard? A clinically important score change might not be statistically significant, and vice versa.

The length of the SF-36 was felt to be a significant limitation. In light of this, whilst acknowledging that it is the weaker instrument, two members of the panel suggested that the SF-12 might be more acceptable. The other major disadvantage of the SF-36 noted by the group was the cost of licensing.

EQ-5D
Amongst the advantages of this measure the group noted its applicability in cost-effectiveness analysis, and the fact that scores can be benchmarked with results for the general population. It was felt to have good evidence based on the review, although not all members of the group were familiar with its use. One member considered it their preferred generic instrument. Another was aware of unpublished evidence supporting good discrimination of the measure in ECRP. The EQ-5D was felt to be short and simple to complete, and apparently acceptable to patients.

On the other hand, one panel member noted the poor completion rates reported in several studies, suggesting that it is not universally acceptable to patients. Another felt that the simplicity of the instrument was also its weakness, with the limited range of questions and response options rendering it insensitive to all but gross change, and likelihood of ceiling effects. Low reliability of the VAS thermometer was noted.

Generic total

RATING	'not at all suitable' (score 0)	'to some extent unsuitable' (score 1)	'uncertain' (score 2)	'to some extent suitable' (score 3)	'very suitable' (score 4)	TOTAL
SF-36	0	-	-	9	8	**17**
EQ-5D	0	1	-	-	16	**17**

Specific PROMs
SAQ
This instrument was regarded as useful for follow-up studies. It was considered to have robust characteristics and, given its widespread use in ECRP, to allow for meaningful comparisons to be made. Members of the group noted the measure's good face validity and sensitivity to change, with a wide range of response options.

Uncertainties expressed by the panel included whether clinically relevant differences in subscale scores had been established. Given the focus on angina, the measure might not be sufficiently responsive to improvements in dyspnoea or fatigue. It was also noted that published studies did not always clarify which domains of the instrument had been used.

The group's main criticism of the SAQ was regarding the measure's limited content. It was felt to be heavily weighted towards physical symptoms, and unable fully to capture emotional or cognitive changes. It was also inappropriate for evaluating ECRP carried out for prognostic purposes.

CROQ
This suite of instruments was felt to be useful for follow-up to ECRP, and in detecting the need for repeat revascularisation. Members of the group noted the questionnaires had good face validity, with high relevance of questions to patients, and wording that was easily understood. Content was more comprehensive than that of the SAQ: physical, social, and psychological aspects were covered, making the addition of a generic measure unnecessary. A further advantage of the CROQ over the SAQ was the inclusion of post-operative recovery and complications, as well as patient satisfaction and experience. This, in the opinion of one panel member, made it possibly the ideal measure from a DH perspective.

Doubt was expressed regarding follow-up administration at three months: this could be too early to detect some outcomes, which might not become apparent until six months post-operatively. Members of the group had no personal experience of using the measure, and noted there were few publications supporting its use. They also observed that the CROQ-PTCA needed some updating: the name should be changed, as PCI (percutaneous coronary intervention) was the more comprehensive term currently in use. Moreover, the procedure was often carried out via the radial (not only the femoral) approach; questions needed adding to consider this. One member of the panel suggested there might be some redundancy between questions 12 and 14 in the post-procedure versions of the questionnaires.

The group noted the length of the CROQ questionnaires as their main disadvantage. This might be burdensome to elderly patients, in particular, and result in low completion rates. One member of the panel felt there was a disadvantage in having separate questionnaires for the two types of procedure, where the intention was to compare their effectiveness.

Specific Total

RATING	'not at all suitable' (score 0)	'to some extent unsuitable' (score 1)	'uncertain' (score 2)	'to some extent suitable' (score 3)	'very suitable' (score 4)	TOTAL
SAQ	0	-	-	12	4	**16**
CROQ-CABG	0	-	4	-	12	**16**
CROQ-PTCA	0	-	4	3	8	**15**

Other measures
Two of the group's members drew attention to the Health Complaints Scale (Denollet 1994), developed in Belgium. However, no evidence of psychometric properties with UK samples was found. Two other members had experience of using the MacNew, but the review found little evidence to support its use in ECRP.

Other comments
Two panel members raised concerns of a general nature. Whilst broadly agreeing with the conclusions of the review, one of the group suggested that it was unsound to recommend instruments without consideration of the specific objectives, given the variety of possible interventions and dimensions of change that could be the subject of evaluation. Measures that

were insensitive to particular interventions could give rise to false negative or positive findings.

A second panel member was of the view that, if the object were to benchmark the performance of one surgical unit against another, then none of the instruments reviewed was suitable for use; outcomes from cardiac surgical units were, in any case, already subject to considerable scrutiny. He commented that the high response rates noted in the study environment were unlikely to be replicated with the routine use of PROMs. There were also practical considerations, including the time required and the logistics of administering questionnaires to patients admitted via different pathways. Revascularisation was often performed to relieve dyspnoea and for prognostic reasons, where the object might be to stabilise but not necessarily improve the patient's condition. Moreover, dyspnoea and other symptoms could be multifactorial, and a lack of improvement would not necessarily reflect the quality of surgery.

Conclusion

Based on ratings and comments from the panel, the EQ-5D and the SF-36 scored equally well as generic measures of health status, as did the specific measures, viz. the SAQ and the CROQ suite of questionnaires. It should be noted that one member of the panel considered that none of the measures identified was suitable for the purpose stated – i.e. measurement of the quality and outcomes of services for people undergoing ECRP. The remainder of the panel was in general agreement with the conclusions of the review.

Patient-reported Outcome Measure Rating Scale

1. On the basis of the review of evidence and your personal experience, and considering the content of the questionnaire, is this questionnaire suitable for the measurement of the quality and outcomes of services for people undergoing elective coronary revascularisation procedures? (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?

Overall, do you agree with the conclusions of the review? ☐ Yes ☐ No: Please explain

Any additional comments

Multidisciplinary panel – list of members

Mr. Thanos Athanasiou
Consultant Cardiothoracic Surgeon and Reader in Clinical Surgery, Imperial College London

Professor Christi Deaton
Professor of Nursing, University of Manchester

Dr. Mark Jackson
Head of Clinical Governance/Quality Assurance Lead, The Cardiothoracic Centre, Liverpool NHS Trust

Mr. James Roxburgh
Consultant Cardiac Surgeon, St Thomas' Hospital

Dr. Rod Stables
Consultant Cardiologist, The Cardiothoracic Centre, Liverpool NHS Trust

Dr. Everard W. Thornton *(retired)*
Research Psychologist, University of Liverpool

PROM Group

Ray Fitzpatrick
Professor of Public Health and Primary Care and Acting 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

Appendix E: Abbreviations and acronyms

ADL	Activities of daily living
ADORE	Aggressive Diagnosis Of REstenosis in high-risk patients trial
APPROACH	Alberta Provincial PRoject for Outcome Assessment in Coronary Heart disease
ASCENT	Advanced Cardiovascular System Multi-Link-Stent System trial
ASD	Atrial-septal defect
BARI	Bypass Angioplasty Revascularization Investigation
BARI-SEQOL	BARI Study of Economics and Quality of Life
BDI	Beck Depression Inventory
BHACAS	Beating Heart Against Cardioplegic Arrest Studies
BMI	Body Mass Index
BOAT	Balloon versus Optimal Atherectomy trial
CA	Coronary angioplasty
CABG	Coronary artery bypass grafting
CABG-CPB	Coronary artery bypass grafting with cardiopulmonary bypass ('on-pump')
CAD	Coronary artery disease
CCS	Canadian Cardiovascular Society
CPB	Cardio-pulmonary bypass
CHD	Coronary heart disease
COURAGE	Clinical Outcomes Utilizing percutaneous coronary Revascularization and Aggressive Guideline-driven drug Evaluation trial
CROQ	Coronary Revascularisation Outcome Questionnaire
CSS	Cardiac Symptom Survey
DASI	Duke Activity Status Index
DES	Drug-eluting stents
ECP	Elective coronary procedure
ENRICHD	Enhancing Recovery In Coronary Heart Disease patients study
EQ-5D	EuroQol 5 Dimensions Index
EQ-VAS	EuroQol Visual Analogue Scale or EQ Thermometer
EXCITE	Evaluation of Xemilofiban in Controlling Thrombotic Events study
GUSTO	Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries trial
HADS	Hospital Anxiety and Depression Scale
HRQoL	Health-related Quality of Life
HSSI	Heart Surgery Symptom Inventory
IADL	Instrumental activities of daily living
ICC	Intraclass correlation coefficient
KCCQ	Kansas City Cardiomyopathy Questionnaire
MCS	Mental Component Summary (of the SF-36 and related measures)
METS	Metabolic equivalents of activity
MIDCAB	Minimally invasive direct coronary artery bypass graft
MI	Myocardial infarction
MOSS	Mediators of Social Support study
NSTEMI	Non-ST Segment Elevation Myocardial Infarction
NYHA	New York Heart Association
OAT	Occluded Artery Trial
OPCAB	Coronary artery bypass surgery on the beating heart ('off-pump')
OPUS	Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial
PCI	Percutaneous coronary intervention
POMS	Profile of Mood States
PTCA	Percutaneous transluminal coronary angioplasty
PCS	Physical Component Summary (of the SF-36 and related measures)
PROM	Patient-reported outcome measure
QLI-CV	Quality of Life Index-Cardiac Version
QLMI	Quality of Life after Myocardial Infarction
QoL	Quality of Life
RCT	Randomised controlled trial
RITA	Randomised Intervention Treatment of Angina trial
SAQ	Seattle Angina Questionnaire
SF-36	Medical Outcomes Study 36-Item Short-Form Health Survey
SIP	Sickness Impact Profile
SoS	Stent or Surgery trial
Stent-PAMI	Stent-Primary Angioplasty for Myocardial Infarction trial
STICH	Surgical Treatment for Ischemic Heart Failure trial
SVR	Surgical ventricular reconstruction

REFERENCES

Agarwal S, Schechter C, Zaman A. Assessment of functional status and quality of life after percutaneous coronary revascularisation in octogenarians. *Age and Ageing* 2009; 38(6):748-751.

Al-Housni MB, Hutchings F, Dalby M, Dubowitz M, Grocott-Mason R, Ilsley CD et al. Does myocardial perfusion scintigraphy predict improvement in symptoms and exercise capacity following successful elective percutaneous coronary intervention? *Journal of Nuclear Cardiology* 2009; 16(6):869-877.

Allen JK, Fitzgerald ST, Swank RT, Becker DM. Functional status after coronary artery bypass grafting and percutaneous transluminal coronary angioplasty. *American Journal of Cardiology* 1990; 66(12):921-925.

Alonso JA, Permanyer-Miralda G, Cascant-Castello P, Brotons C, Prieto L, Soler-Soler J. Measuring functional status of chronic coronary patients: reliability, validity, and responsiveness to clinical change of the reduced version of the Duke Activity Status Index (DASI). *European Heart Journal* 1997; 18(3):414-419.

Aron A, Klinger tA, McConnell TR. Cardiac rehabilitation outcomes no different after on-pump versus off-pump coronary artery bypass surgery. *Journal of Cardiopulmonary Rehabilitation and Prevention* 2007; 27(1):35-41.

Artinian NT, Duggan C, Miller P. Age differences in patient recovery patterns following coronary artery bypass surgery. *American Journal of Critical Care* 1993; 2(6):453-461.

Artinian NT, Duggan CH. Sex differences in patient recovery patterns after coronary artery bypass surgery. *Heart and Lung* 1995; 24(6):483-494.

Ascione R, Reeves BC, Taylor FC, Seehra HK, Angelini GD. Beating heart against cardioplegic arrest studies (BHACAS 1 and 2): quality of life at mid-term follow-up in two randomised controlled trials. *European Heart Journal* 2004; 25(9):765-770.

Ballan A, Lee G. A comparative study of patient perceived quality of life pre and post coronary artery bypass graft surgery. *Australian Journal of Advanced Nursing* 2007; 24(4):24-28.

BARI Investigators. Five-year clinical and functional outcome comparing bypass surgery and angioplasty in patients with multivessel coronary disease. A multicenter randomized trial. *JAMA: Journal of the American Medical Association* 1997; 277(9):715-721.

Barnason S, Zimmerman LM, Atwood JR, Nieveen JL, Schmaderer M. Development of a self-efficacy instrument for coronary artery bypass graft patients. *Journal of Nursing Measurement* 2002; 10(2):123-133.

Barnason SA, Zimmerman LM, Brey B, Catlin S, Nieveen JL. Patterns of recovery following percutaneous coronary intervention: a pilot study. *Applied Nursing Research* 2006; 19(1):31-37.

Barnason S, Zimmerman L, Nieveen J, Schulz P, Milller C, Hertzog M et al. Relationships between fatigue and early postoperative recovery outcomes over time in elderly patients undergoing coronary artery bypass graft surgery. *Heart and Lung* 2008; 37(4):245-256.

Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. *Heart and Lung* 2009; 38(5):364-376.

Barnett SD, Martin LM, Halpin LS, Ad N. Impact of body mass index on clinical outcome and health-related quality of life following open heart surgery. *Journal of Nursing Care Quality* 2010; 25(1):65-72.

Bennett SJ, Puntenney PJ, Walker NL, Ashley ND. Development of an instrument to measure threat related to cardiac events. *Nursing Research* 1996; 45(5):266-270.

Bergner M, Bobbitt RA, Pollard WE, Martin DP, Gilson BS. The Sickness Impact Profile: validation of a health status measure. *Medical Care* 1976; 14(1):57-67.

Bergner M. The Sickness Impact Profile: a brief summary of its purposes, uses, and administration. 1-10. 1978. Department of Health Sciences, University of Washington, WA, USA [report].

Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* 1981; 19(8):787-805.

Birks Y, Roebuck A, Thompson DR. A validation study of the Cardiac Depression Scale/CDS in a UK population. *British Journal of Health Psychology* 2004; 9(1):15-24.

Boden WE. Interpreting the COURAGE trial. It takes COURAGE to alter our belief system. *Cleveland Clinic Journal of Medicine* 2007; 74(9):623, 629-625, 633.

Borkon AM, Muehlebach GF, House J, Marso SP, Spertus JA. A comparison of the recovery of health status after percutaneous coronary intervention and coronary artery bypass. *Annals of Thoracic Surgery* 2002; 74(5):1526-1530.

Bourassa MG, Brooks MM, Mark DB, Trudel J, Detre KM, Pitt B et al. Quality of life after coronary revascularization in the United States and Canada. *American Journal of Cardiology* 2000; 85(5):5548-5553.

Bradshaw PJ, Jamrozik KD, Gilfillan IS, Thompson PL. Asymptomatic long-term survivors of coronary artery bypass surgery enjoy a quality of life equal to the general population. *American Heart Journal* 2006; 151(2):537-544.

Brazier JE, Jones NMB, 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.

Brazier JE, Usherwood TP, Harper R, Thomas KJ. Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology* 1998; 51(11):1115-1128.

Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002; 21(2):271-292.

Brown JS, Rawlinson ME. Psychosocial status of patients randomly assigned to medical or surgical therapy for chronic stable angina. *American Journal of Cardiology* 1979; 44(3):546-554.

Caine N, Harrison SC, Sharples LD, Wallwork J. Prospective study of quality of life before and after coronary artery bypass grafting. *BMJ: British Medical Journal* 1991; 302(6775):511-516.

Campeau L. Letter: Grading of angina pectoris. *Circulation* 1976; 54(3):522-523.

Carter R, Holiday DB, Grothues C, Nwasuruba C, Stocks J, Tiep B. Criterion validity of the Duke Activity Status Index for assessing functional capacity in patients with chronic obstructive pulmonary disease. *Journal of Cardiopulmonary Rehabilitation* 2002; 22(4):298-308.

Cleary PD, Greenfield S, McNeil BJ. Assessing quality of life after surgery. *Controlled Clinical Trials* 1991; 12(4 Suppl.):189S-203S.

Criteria Committee of the New York Heart Association. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. pp 253-6. Boston, Massachusetts, USA: Brown & Co, 1994.

Curtis LH, Phelps CE, McDermott MP, Rubin HR. The value of patient-reported health status in predicting short-term outcomes after coronary artery bypass graft surgery. *Medical Care* 2002; 40(11):1090-1100.

Curtis JP, Krumholz HM. Keeping the patient in view: defining the appropriateness of percutaneous coronary interventions. *Circulation* 2004; 110(25):3746-3748.

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

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

Deaton C, Kimble LP, Veledar E, Hartigan P, Boden WE, O'Rourke R et al. The synergistic effect of heart disease and diabetes on self-management, symptoms, and health status. *Heart and Lung* 2006; 35(5):315-323.

Deaton C, Thourani V. Patients with type 2 diabetes undergoing coronary artery bypass graft surgery: predictors of outcomes. *European Journal of Cardiovascular Nursing* 2009; 8(1):48-56.

Denvir MA, Lee AJ, Rysdale J, Walker A, Eteiba H, Starkey IR et al. Influence of socioeconomic status on clinical outcomes and quality of life after percutaneous coronary intervention. *Journal of Epidemiology and Community Health* 2006; 60(12):1085-1088.

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

Department of Health, London, March 2000. National Strategic Framework for Coronary Heart Disease.

Doering LV, Moser DK, Lemankiewicz W, Luper C, Khan S. Depression, healing, and recovery from coronary artery bypass surgery. *American Journal of Critical Care* 2005; 14(4):316-324.

Dolansky MA, Moore SM. Effects of cardiac rehabilitation on the recovery outcomes of older adults after coronary artery bypass surgery. *Journal of Cardiopulmonary Rehabilitation* 2004; 24(4):236-244.

Dunning J, Waller JR, Smith B, Pitts S, Kendall SW, Khan K. Coronary artery bypass grafting is associated with excellent long-term survival and quality of life: a prospective cohort study. *Annals of Thoracic Surgery* 2008; 85(6):1988-1993.

Elliott D, Lazarus R, Leeder SR. Health outcomes of patients undergoing cardiac surgery: repeated measures using Short Form-36 and 15 Dimensions of Quality of Life questionnaire. *Heart and Lung* 2006a; 35(4):245-251.

Elliott D, Lazarus R, Leeder SR. Proxy respondents reliably assessed the quality of life of elective cardiac surgery patients. *Journal of Clinical Epidemiology* 2006b; 59(2):153-159.

ENRICHD Investigators. Enhancing Recovery In Coronary Heart Disease patients (ENRICHD): study design and methods. *American Heart Journal* 2000; 139(1 Pt 1):1-9.

EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990.

Feeny DH, Furlong WJ, Boyle MH, Torrance GW. Multi-attribute health status classification systems: the Health Utilities Index. *PharmacoEconomics* 1995; 7(6):490-502.

Feinstein A, Wells CK. A new clinical taxonomy for rating change in functional activities of patients with angina pectoris. *American Heart Journal* 1977; 93(2):172-182.

Fernandez RS, Davidson P, Salamonson Y, Griffiths R, Juergens C. The health-related quality of life trajectory in patients after percutaneous coronary intervention. *Journal of Cardiopulmonary Rehabilitation and Prevention* 2007; 27(4):223-226.

Ferrans CE, Powers MJ. Quality of Life Index: development and psychometric properties. *Advances in Nursing Science* 1985; 8(1):15-24.

Fitzpatrick R. Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technology Assessment* 1998; 2(14):74p.

Fitzpatrick R, Bowling A, Mackintosh A, Gibbons, E. A structured review of patient-reported outcome measures for people with heart failure. Department of Health, London, November 2006.

Fukuoka Y, Lindgren TG, Rankin SH, Cooper BA, Carroll DL. Cluster analysis: a useful technique to identify elderly cardiac patients at risk for poor quality of life. *Quality of Life Research* 2007; 16(10):1655-1663.

Gallagher R. 'Comparison of symptoms of younger and older patients undergoing coronary artery bypass surgery' - commentary. *Clinical Nursing Research* 2004; 13:194-198.

Garratt AM, Hutchinson A, Russell IT. The UK version of the Seattle Angina Questionnaire/SAQ-UK: reliability, validity and responsiveness. *Journal of Clinical Epidemiology* 2001; 54(9):907-915.

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.

Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. *Circulation* 1981; 64(6):1227-1234.

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):1245-1255.

Halpin LS, Barnett SD, Martin LM, Hunt SL, Henry L, Ad N. Survival and quality of life following elective open-heart surgery. *Journal of Nursing Care Quality* 2008; 23(4):369-374.

Hare DL, Davis CR. Cardiac Depression Scale: validation of a new depression scale for cardiac patients. *Journal of Psychosomatic Research* 1996; 40(4):379-386.

Hatoum HT, Brazier JE, Akhras KS. Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. *Value in Health* 2004; 7(5):602-609.

Hawkes AL, Nowak M, Speare R. Short Form-36 Health Survey as an evaluation tool for cardiac rehabilitation programs: is it appropriate? *Journal of Cardiopulmonary Rehabilitation* 2003; 23(1):22-25.

Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM et al. A brief self-administered questionnaire to determine functional capacity, the Duke Activity Status Index. *American Journal of Cardiology* 1989; 64(10):651-654.

Hlatky MA, Rogers WJ, Johnstone I, Boothroyd D, Brooks MM, Pitt B et al. Medical care costs and quality of life after randomization to coronary angioplasty or coronary bypass surgery. Bypass Angioplasty Revascularization Investigation (BARI) Investigators. *New England Journal of Medicine* 1997; 336(2):92-99.

Hlatky MA, Boothroyd DB, Melsop KA, Brooks MM, Mark DB, Pitt B et al. Medical costs and quality of life 10 to 12 years after randomization to angioplasty or bypass surgery for multivessel coronary artery disease. *Circulation* 2004; 110(14):1960-1966.

Ho PM, Masoudi FA, Peterson PN, Shroyer AL, McCarthy MJ, Grover FL et al. Health-related quality of life predicts mortality in older but not younger patients following cardiac surgery. *American Journal of Geriatric Cardiology* 2005; 14(4):176-182.

Hogue CW, Jr., Fucetola R, Hershey T, Nassief A, Birge S, Davila-Roman VG et al. The role of postoperative neurocognitive dysfunction on quality of life for postmenopausal women six months after cardiac surgery. *Anesthesia and Analgesia* 2008; 107(1):21-28.

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.

Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *Journal of Epidemiology and Community Health* 1980; 34(4):281-286.

Hunt SM, McEwen J, Mckenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners* 1985; 35(273):185-188.

Jenkins CD, Stanton BA, Savageau JA, Denlinger P, Klein MD. Coronary artery bypass surgery. Physical, psychological, social, and economic outcomes six months later. *JAMA: Journal of the American Medical Association* 1983; 250(6):782-788.

Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. *Journal of Clinical Epidemiology* 1988; 41(4):313-321.

Jenkins CD, Jono RT, Stanton BA, Stroup-Benham CA. The measurement of health-related quality of life: major dimensions identified by factor analysis. *Social Science and Medicine* 1990; 31(8):925-931.

Jenkins CD, Stanton BA, Jono RT. Quantifying and predicting recovery after heart surgery. *Psychosomatic Medicine* 1994; 56(3):203-212.

Jette AM, Davies AR, Cleary PD, Calkins DR, Rubenstein LV, Fink A et al. The Functional Status Questionnaire: reliability and validity when used in primary care. *Journal of General Internal Medicine* 1986; 1(3):143-149.

Kandzari DE, Lam LC, Eisenstein EL, Clapp-Channing N, Fine JT, Califf RM et al. Advanced coronary artery disease: appropriate end points for trials of novel therapies. *American Heart Journal* 2001; 142(5):843-851.

Kaul P, Armstrong PW, Fu Y, Knight JD, Clapp-Channing NE, Sutherland W et al. Impact of different patterns of invasive care on quality of life outcomes in patients with non-ST elevation acute coronary syndrome: results from the GUSTO-IIb Canada-United States substudy. *Canadian Journal of Cardiology* 2004; 20(8):760-766.

Kaul P, Lytle BL, Spertus JA, DeLong ER, Peterson ED. Influence of racial disparities in procedure use on functional status outcomes among patients with coronary artery disease. *Circulation* 2005; 111(10):1284-1290.

Keresztes P, Penckofer S, Holm K, Merritt S. Measurements of functional ability in patients with coronary artery disease. *Journal of Nursing Measurement* 1993; 1(1):19-28.

Keresztes PA, Merritt SL, Holm K, Penckofer S, Patel M. The coronary artery bypass experience: gender differences. *Heart and Lung* 2003; 32(5):308-319.

Kiebzak GM, Pierson LM, Campbell M, Cook JW. Use of the SF-36 general health status survey to document health-related quality of life in patients with coronary artery disease: effect of disease and response to coronary artery bypass graft surgery. *Heart and Lung* 2002; 31(3):207-213.

Kim J, Henderson RA, Pocock SJ, Clayton T, Sculpher MJ, Fox KA et al. Health-related quality of life after interventional or conservative strategy in patients with unstable angina or non-ST-segment elevation myocardial infarction: one-year results of the third Randomized Intervention Trial of unstable Angina (RITA-3). *Journal of the American College of Cardiology* 2005; 45(2):221-228.

King KB, Porter LA, Norsen LH, Reis HT. Patient perceptions of quality of life after coronary artery surgery: was it worth it? *Research in Nursing & Health* 1992; 15(5):327-334.

King KM. Gender and short-term recovery from cardiac surgery. *Nursing Research* 2000; 49(1):29-36.

King KM, Colella TJ, Faris P, Thompson DR. Using the Cardiac Depression Scale in men recovering from coronary artery bypass surgery. *Journal of Clinical Nursing* 2009; 18(11):1617-1624.

Koch CG, Khandwala F, Cywinski JB, Ishwaran H, Estafanous FG, Loop fD et al. Health-related quality of life after coronary artery bypass grafting: a gender analysis using the Duke Activity Status Index. *Journal of Thoracic and Cardiovascular Surgery* 2004; 128(2):284-295.

Koch CG, Li L, Lauer M, Sabik J, Starr NJ, Blackstone EH. Effect of functional health-related quality of life on long-term survival after cardiac surgery. *Circulation* 2007; 115(6):692-699.

Kos-Munson BA, Alexander LD, Hinthorn PA, Gallagher EL, Goetze CM. Psychosocial predictors of optimal rehabilitation post-coronary artery bypass surgery. *Scholarly Inquiry for Nursing Practice* 1988; 2(3):171-199.

Krumholz HM, McHorney CA, Clark L, Levesque M, Baim DS, Goldman L. Changes in health after elective percutaneous coronary revascularization: a comparison of generic and specific measures. *Medical Care* 1996; 34(8):754-759.

Krumholz HM, Cohen DJ, Williams C, Baim DS, Brinker J, Cabin HS et al. Health after coronary stenting or balloon angioplasty: results from the Stent Restenosis Study. *American Heart Journal* 1997; 134(3):337-344.

LaPier TK, Jung C. Development and content validity of the Heart Surgery Symptom Inventory. *Acute Care Perspectives* 2002; 11(2):5-12.

LaPier TK, Wilson B. Outcome measures in cardiopulmonary physical therapy: focus on the Heart Surgery Symptom Inventory (HSSI). *Cardiopulmonary Physical Therapy Journal* 2006; 17(2):77-83.

LaPier TK. Psychometric evaluation of the Heart Surgery Symptom Inventory in patients recovering from coronary artery bypass surgery. *Journal of Cardiopulmonary Rehabilitation* 2006; 26(2):101-106.

LaPier TK. Functional status of patients during subacute recovery from coronary artery bypass surgery. *Heart and Lung* 2007; 36(2):114-124.

Lee GA. Patients reported health-related quality of life five years post coronary artery bypass graft surgery - a methodological study. *European Journal of Cardiovascular Nursing* 2008; 7(1):67-72.

Leslie SJ, Rysdale J, Lee AJ, Eteiba H, Starkey IR, Pell J et al. Unemployment and deprivation are associated with a poorer outcome following percutaneous coronary angioplasty. *International Journal of Cardiology* 2007; 122(2):168-169.

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):1249-1256.

Lindsay GM, Hanlon P, Smith LN, Wheatley DJ. Assessment of changes in general health status using the short-form 36 questionnaire one year following coronary artery bypass grafting. *European Journal of Cardiothoracic Surgery* 2000; 18(5):557-564.

MacDonald P, Stadnyk K, Cossett J, Klassen G, Johnstone D, Rockwood K. Outcomes of coronary artery bypass surgery in elderly people. *Canadian Journal of Cardiology* 1998; 1410(1215):1222.

Mallik S, Krumholz HM, Lin ZQ, Kasl SV, Mattera JA, Roumains SA et al. Patients with depressive symptoms have lower health status benefits after coronary artery bypass surgery. *Circulation* 2005; 111(3):271-277.

Mark DB, Pan W, Clapp-Channing NE, Anstrom KJ, Ross JR, Fox RS et al. Quality of life after late invasive therapy for occluded arteries. *NEJM: New England Journal of Medicine* 2009a; 360(8):774-783.

Mark DB, Knight JD, Velazquez EJ, Howlett JG, Spertus JA, Djokovic LT et al. Quality of life and economic outcomes with surgical ventricular reconstruction in ischemic heart failure: results from the Surgical Treatment for Ischemic Heart Failure trial. *American Heart Journal* 2009b; 157(5):837-844.

Martin LM, Halpin LS, Barnett SD, Speir AM, Hunt SL, Mukherjee N et al. The association between early outcome, health-related quality of life, and survival following elective open-heart surgery. *Jouranl of Cardiovascular Nursing* 2008; 23(5):432-442.

Mayer C, Ergina P, Morin JF, Gold S. Self-reported functional status as a predictor of coronary artery bypass graft surgery outcome in elderly patients. *Canadian Journal of Cardiology* 2003; 19(2):140-144.

McCarthy MJJ, Shroyer AL, Sethi GK, Moritz TE, Henderson WG, Grover FL et al. Self-report measures for assessing treatment outcomes in cardiac surgery patients. *Medical Care* 1995; 33(10 Suppl):OS76-OS85.

McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 2nd ed. ed. New York, NY, USA: Oxford University Press, 1996.

McDowell I. Measuring Health: a guide to rating scales and questionnaires. 3rd ed. Oxford, UK & New York, USA: Oxford University Press, 2006.

McNair DM, Lorr M, Droppelman LF. Profile of Mood States - manual. San Diego, CA, USA: Educational and Industrial Testing Service, 1971.

Melsop KA, Boothroyd DB, Hlatky MA. Quality of life and time trade-off utility measures in patients with coronary artery disease. *American Heart Journal* 2003; 145(1):36-41.

Mikhail GW. Coronary revascularisation in women. *Heart* 2003; 92(Suppl.III):III 19-III 23.

Miller KH, Grindel CG. Comparison of symptoms of younger and older patients undergoing coronary artery bypass surgery. *Clinical Nursing Research* 2004; 13(3):179-193.

Morris DB, Wilson KG, Clinch JJ, Ammerman DJ, Fergusson D, Graham ID et al. Identification of domains relevant to health-related quality of life in patients undergoing major surgery. *Quality of Life Research* 2006; 15(5):841-854.

Moser DK, Dracup K. Psychosocial recovery from a cardiac event: the influence of perceived control. *Heart and Lung* 1995; 24(4):273-280.

Nelson CL, Herndon JE, Mark DB, Pryor DB, Califf RM, Hlatky MA. Relation of clinical and angiographic factors to functional capacity as measured by the Duke Activity Status Index. *American Journal of Cardiology* 1991; 68(9):973-975.

Nieveen JL, Zimmerman LM, Barnason SA, Yates BC. Development and content validity testing of the Cardiac Symptom Survey in patients after coronary artery bypass grafting. *Heart and Lung* 2008; 37(1):17-27.

Norris CM, Ghali WA, Galbraith PD, Graham MM, Jensen LA, Knudtson ML. Women with coronary artery disease report worse health-related quality of life outcomes compared to men. *Health and Quality of Life Outcomes* 2004a; 2(1):21.

Norris CM, Saunders LD, Ghali WA, Brant R, Galbraith PD, Graham M et al. Health-related quality of life outcomes of patients with coronary artery disease treated with cardiac surgery, percutaneous coronary intervention or medical management. *Canadian Journal of Cardiology* 2004b; 20(1):1259-1266.

Oldridge NB, Guyatt GH, Jones N, Crowe J, Singer J, Feeny DH et al. Effects on quality of life with comprehensive rehabilitation after acute myocardial infarction. *American Journal of Cardiology* 1991; 67(13):1084-1089.

Papadantonaki A, Stotts NA, Paul SM. Comparison of quality of life before and after coronary artery bypass surgery and percutaneous transluminal angioplasty. *Heart and Lung* 1994; 23(1):45-52.

Peduzzi P, Hultgren HN. Effect of medical vs surgical treatment on symptoms in stable angina pectoris. The Veterans Administration Cooperative Study of surgery for coronary arterial occlusive disease. *Circulation* 1979; 60(4):888-900.

Peduzzi P, Hultgren H, Thomsen J, Detre K. Ten-year effect of medical and surgical therapy on quality of life: Veterans Administration Cooperative Study of Coronary Artery Surgery. *American Journal of Cardiology* 1987; 59(12):1017-1023.

Penckofer S, Ferrans CE, Fink N, Barrett ML, Holm K. Quality of life in women following coronary artery bypass graft surgery. *Nursing Science Quarterly* 2005; 18(2):176-183.

Phillips-Bute B, Mathew J, Blumenthal JA, Welsh-Bohmer K, White WD, Mark D et al. Female gender is associated with impaired quality of life 1 year after coronary artery bypass surgery. *Psychosomatic Medicine* 2003; 65(6):944-951.

Phillips RS, Goldman L, Bergner M. Patient characteristics in SUPPORT: activity status and cognitive function. *Journal of Clinical Epidemiology* 1990; 43(Suppl.):33S-36S.

Pilote L, Bourassa MG, Bacon C, Bost J, Detre K, Mark DB et al. Better functional status in American than Canadian patients with heart disease: an effect of medical care? *Journal of the American College of Cardiology* 1995; 26(5):1115-1120.

Plach SK, Heidrich SM. Women's perceptions of their social roles after heart surgery and coronary angioplasty. *Heart and Lung* 2001; 30(2):117-127.

Plach SK, Heidrich SM. Social role quality, physical health, and psychological well-being in women after heart surgery. *Research in Nursing & Health* 2002; 25(3):189-202.

Pocock SJ, Henderson RA, Seed P, Treasure T, Hampton JR. Quality of life, employment status, and anginal symptoms after coronary angioplasty or bypass surgery. 3-year follow-up in the Randomized Intervention Treatment of Angina (RITA) Trial. *Circulation* 1996; 94(2):135-142.

Pocock SJ, Henderson RA, Clayton T, Lyman GH, Chamberlain DA. Quality of life after coronary angioplasty or continued medical treatment for angina: three-year follow-up in the RITA-2 trial. Randomized Intervention Treatment of Angina. *Journal of the American College of Cardiology* 2000; 35(4):907-914.

Poulin F, Rinfret S, Gobeil F. Potential shift from coronary bypass surgery to percutaneous coronary intervention for multivessel disease and its economic impact in the drug-eluting stent era. *Canadian Journal of Cardiology* 2007; 23(14):1139-1145.

Pusca SV, Puskas JD. Revascularization in heart failure: coronary bypass or percutaneous coronary intervention? *Heart Failure Clinics* 2007; 3(2):211-228.

Reeves BC, Angelini GD, Bryan AJ, Taylor FC, Cripps T, Spyt TJ et al. A multi-centre randomised controlled trial of minimally invasive direct coronary bypass grafting versus percutaneous transluminal coronary angioplasty with stenting for proximal stenosis of the left anterior descending coronary artery. Health Technology Assessment 8[16], 1-43. 2004. Southampton, UK, HMSO [monograph].

Rinfret S, Grines CL, Cosgrove RS, Ho KK, Cox DA, Brodie BR et al. Quality of life after balloon angioplasty or stenting for acute myocardial infarction. One-year results from the Stent-PAMI trial. *Journal of the American College of Cardiology* 2001; 38(6):1614-1621.

Rose GA. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bulletin of the World Health Organisation* 1962; 27:645-658.

Rose GA. Chest pain questionnaire. *Millbank Memorial Fund Quarterly* 1965; 43:32-39.

Rose GA, Blackburn H. Cardiovascular survey methods. World Health Organization monograph. 1968. Geneva, Switzerland, World Health Organisation [report].

Rumbaugh DM. The coronary spectrum. The psychological aspects. *Journal of Rehabilitation* 1966; 32(2):56-58.

Rumsfeld JS, MaWhinney S, McCarthy MJ, Shroyer AL, VillaNueva CB, O'Brien M et al. Health-related quality of life as a predictor of mortality following coronary artery bypass graft surgery. Participants of the Department of Veterans Affairs Cooperative Study Group on Processes, Structures, and Outcomes of Care in Cardiac Surgery. *JAMA: Journal of the American Medical Association* 1999; 281(12):1298-1303.

Rumsfeld JS, Magid DJ, O'Brien M, McCarthy MJ, MaWhinney S, Shroyer AL et al. Changes in health-related quality of life following coronary artery bypass graft surgery. *Annals of Thoracic Surgery* 2001; 72(6):2026-2032.

Sandau KE, Lindquist RA, Treat-Jacobson DJ, Savik K. Health-related quality of life and subjective neurocognitive function three months after coronary artery bypass graft surgery. *Heart and Lung* 2008; 37(3):161-172.

Sawatzky JA, Naimark BJ. Coronary artery bypass graft surgery: exploring a broader perspective of risks and outcomes. *Journal of Cardiovascular Nursing* 2009a; 24(3):198-206.

Sawatzky JA, Naimark BJ. The coronary artery bypass graft surgery trajectory: gender differences revisited. *European Journal of Cardiovascular Nursing* 2009b; 8(4):302-308.

Schroter S, Lamping DL. Psychometric evaluation of the Coronary Revascularisation Outcome Questionnaire (CROQ-CABG/PTCA). Quality of Life Research 9[3], 311. 2000 [abstract].

Schroter S, Lamping DL. Coronary Revascularisation Outcome Questionnaire (CROQ): development and validation of a new, patient based measure of outcome in coronary bypass surgery and angioplasty. *Heart (London)* 2004; 90(12):1460-1466.

Schroter S, Lamping DL. Responsiveness of the Coronary Revascularisation Outcome Questionnaire compared with the SF-36 and Seattle Angina Questionnaire. *Quality of Life Research* 2006; 15(6):1069-1078.

Sintonen H. The 15-D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 2001; 33(5):328-336.

Skaggs BG, Yates BC. Quality of life comparisons after coronary angioplasty and coronary artery bypass graft surgery. *Heart and Lung* 1999; 28(6):409-417.

Sollano JA, Rose EA, Williams DL, Thornton B, Quint E, Apfelbaum M et al. Cost-effectiveness of coronary artery bypass surgery in octogenarians. *Annals of Surgery* 1998; 228(3):297-306.

Sorensen EA, Wang F. Social support, depression, functional status, and gender differences in older adults undergoing first-time coronary artery bypass graft surgery. *Heart and Lung* 2009; 38(4):306-317.

Spertus JA, Winder JA, Fihn SD, Dewhurst TA, Deyo RA. Assessing quality of life in patients with CAD: an alternative end point in outcome oriented clinical research. Journal of the American College of Cardiology 1A, 434A. 1994a [abstract].

Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Fihn SD. Monitoring the quality of life in patients with coronary artery disease. *American Journal of Cardiology* 1994b; 74(12):1240-1244.

Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell MB et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *Journal of the American College of Cardiology* 1995; 25(2):333-341.

Spertus JA, McDonell M, Woodman CL, Fihn SD. Association between depression and worse disease-specific functional status in outpatients with coronary artery disease. *American Heart Journal* 2000; 140(1):105-110.

Spertus JA, Bliven BD, Farner M, Gillen A, Hewitt T, Jones P et al. Integrating baseline health status data collection into the process of care. *Joint Commission Journal on Quality Improvement* 2001; 27(7):369-380.

Spertus JA, Salisbury AC, Jones PG, Conaway DG, Thompson RC. Predictors of quality-of-life benefit after percutaneous coronary intervention. *Circulation* 2004; 110(25):3789-3794.

Spertus JA, Nerella R, Kettlekamp R, House J, Marso S, Borkon AM et al. Risk of restenosis and health status outcomes for patients undergoing percutaneous coronary intervention versus coronary artery bypass graft surgery. *Circulation* 2005; 111(6):768-773.

Stewart RD, Blair JL, Emond CF, Lahey SJ, Levitsky S, Campos CT. Gender and functional outcome after coronary artery bypass. *Surgery* 1999; 126(2):184-190.

Streiner DL, Norman GR. Health Measurement Scales: a practical guide to their development and use. 3rd ed. Oxford University Press, 2003.

Sullivan MD, LaCroix AZ, Russo J, Katon WJ. Self-efficacy and self-reported functional status in coronary heart disease: a six-month prospective study. *Psychosomatic Medicine* 1998; 60(4):473-478.

Thornton EW, Groom C, Fabri BM, Fox MA, Hallas C, Jackson M. Quality of life outcomes after coronary artery bypass graft surgery: relationship to neuropsychologic deficit. *Journal of Thoracic and Cardiovascular Surgery* 2005; 130(4):1022-1027.

Treat-Jacobson D, Lindquist RA. Functional recovery and exercise behavior in men and women 5 to 6 years following coronary artery bypass graft (CABG) surgery. *Western Journal of Nursing Research* 2004; 26(5):479-498.

Trzcieniecka-Green A, Steptoe A. The effects of stress management on the quality of life of patients following acute myocardial infarction or coronary bypass surgery. *European Heart Journal* 1996; 17(11):1663-1670.

Vaglio J, Conard M, Poston WS, O'Keefe J, Haddock CK, House J et al. Testing the performance of the ENRICHD Social Support Instrument in cardiac patients. *Health and Quality of Life Outcomes* 2004; 2(1):24.

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):151-161.

Ware JE, Sherbourne CD, Davies AR. Developing and testing the MOS 20-Item Short-Form Health Survey: a general population application. In: Stewart AL, Ware JE, editors. Measuring functioning and well-being: the Medical Outcomes Study approach. Durham, NC, USA: Duke University Press, 1992: 277-290.

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 JE, Kosinski M, Keller SD. How to score the SF-12 Physical and Mental Health Summary scales. Boston, MA, USA: The Health Institute, New England Medical Center, 1995.

Ware JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care* 1996; 34(3):220-233.

Ware JE. SF-36 Health Survey. Manual and Interpretation Guide. 2nd ed. Boston, MA, USA: Nimrod Press, 1997.

Weaver WD, Reisman MA, Griffin JJ, Buller CE, Leimgruber PP, Henry T et al. Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial (OPUS-1): a randomised trial. *Lancet* 2000; 355(9222):2199-2203.

Weintraub WS, Spertus JA, Kolm P, Maron DJ, Zhang Z, Jurkovitz C et al. Effect of PCI on quality of life in patients with stable coronary disease. *NEJM: New England Journal of Medicine* 2008; 359(7):677-687.

Welke KF, Stevens JP, Schults WC, Nelson EC, Beggs VL, Nugent WC. Patient characteristics can predict improvement in functional health after elective coronary artery bypass grafting. *Annals of Thoracic Surgery* 2003; 75(6):1849-1855.

Worcester MU, Murphy BM, Elliott PC, Le Grande MR, Higgins RO, Goble AJ et al. Trajectories of recovery of quality of life in women after an acute cardiac event. *British Journal of Health Psychology* 2007; 12(Pt 1):1-15.

Zhang Z, Mahoney EM, Stables RH, Booth J, Nugara F, Spertus JA et al. Disease-specific health status after stent-assisted percutaneous coronary intervention and coronary artery bypass surgery: one-year results from the Stent or Surgery trial. *Circulation* 2003; 108(14):1694-1700.

Zhang Z, Weintraub WS, Mahoney EM, Spertus JA, Booth J, Nugara F et al. Relative benefit of coronary artery bypass grafting versus stent-assisted percutaneous coronary intervention for angina pectoris and multivessel coronary disease in women versus men (one-year results from the Stent or Surgery trial). *American Journal of Cardiology* 2004; 93(4):404-409.

Zhang Z, Spertus JA, Mahoney EM, Booth J, Nugara F, Stables RH et al. The impact of acute coronary syndrome on clinical, economic, and cardiac-specific health status after coronary artery bypass surgery versus stent-assisted percutaneous coronary intervention: 1-year results from the stent or surgery (SoS) trial. *American Heart Journal* 2005; 150(1):175-181.

Zimmerman L, Barnason S, Brey BA, Catlin SS, Nieveen J. Comparison of recovery patterns for patients undergoing coronary artery bypass grafting and minimally invasive direct coronary artery bypass in the early discharge period. *Progress in Cardiovascular Nursing* 2002; 17(3):132-141.