

**SELF-MANAGEMENT PROGRAMMES FOR
CANCER SURVIVORS:
A STRUCTURED REVIEW OF OUTCOME
MEASURES**

Amended Version

Macmillan Cancer Support,
April, 2009

Nicola Davies

CONTENTS

Contents and Acknowledgements	2
Executive Summary	3
Background.....	5
Aims and Objectives.....	7
Methods and Search Strategy	8
Results of Search Strategy.....	9
Quality of Life PROMs for Cancer Survivors.....	11
Recommended QoL PROMs (including Table of Appraisal Criteria).....	23
Outcome Measures for Health Behaviour Change.....	25
Recommended Behaviour Change PROMs (including Table of Appraisal Criteria)	34
Self-Efficacy PROMs.....	36
Recommended Self-Efficacy PROMs (including Table of Appraisal Criteria)	50
Outcome Measures for Patient/Provider Interactions.....	52
Recommended Patient/Provider Outcome Measures (Table of Appraisal Criteria)	67
Instruments outside the Scope of this Review	69
Final Discussion and Recommendations.....	71
Appendix A: Search Strategy and Sources.....	73
Appendix B: Appraisal Criteria.....	76
Appendix C: Instrument Items and Domains.....	77
References	89

Amendment

The original review of March, 2009 has been amended to accommodate the clarified aims and objectives of the self-management programme for cancer survivors. Specifically, the Quality of Life in Adult Cancer Survivors (QLACS) questionnaire has been recommended instead of the Quality of Life – Cancer Survivors (QOL-CS) questionnaire. Both demonstrate evidence of psychometric validity, but the multidimensionality of the QLACS is anticipated to provide greater precision in measuring the desired outcomes.

Acknowledgements

Acknowledgement is extended to the Patient-Reported Outcome Measures (PROMs) Group of the University of Oxford, who provided the framework for this review in terms of appraisal criteria. In particular, gratitude is extended to Professor Ray Fitzpatrick, Professor of Public Health and Primary Care, who offered his expert feedback and guidance on initial instrument recommendations emerging from this review. The appraisal criteria utilised from the PROMs Group, along with Professor Fitzpatrick's input, have facilitated the development of this comprehensive report.

Executive Summary

Based on the seventh recommendation of Macmillan’s ‘Self-Management Support: A review of the Evidence: Summary Document to Support the National Cancer Survivorship Self-Management Workstream,’ this review covers recommendation seven: that any changes in survivorship practice, such as self-management programmes, be fully evaluated. Since there is little evidence available to support cancer self-management programmes, it is fundamental to adequately assess such interventions prior to their implementation. Indeed, there needs to be clearly defined criteria for the expected benefits of a self-management programme and a rigorous evaluation of the most appropriate patient-reported outcome measures (PROMs) for use with cancer survivorship self-management.

In consolidating the literature on survivorship and evaluating current self-management strategies for long-term conditions, six areas of outcome were identified (Table 1). Instruments measuring these outcomes were evaluated for psychometric properties and operational characteristics.

Table 1: Outcome Measures Reviewed

Health-Related Quality of Life	Cancer-Specific PROMs
Health Behaviours (i.e. lifestyle change)	PROMs
Self-Efficacy	PROMs
Symptom Management	PROMs
Patient(Survivor)-Centred Interactions for Self-Management	PROMs Professional-Reported
Utilisation of Services	Outside of scope of this brief

Six cancer QoL PROMs specific to cancer survivors were evaluated. The Quality of Life in Adult Cancer Survivors (QLACS) questionnaire demonstrated the most relevance to survivorship self-management, including a number of generic and cancer-specific domains that would facilitate interpretation of self-management outcomes.

Four behaviour change PROMs were evaluated. The stronger evidence of operational characteristics demonstrated by evidence pertaining to the General Practice Physical Activity Questionnaire (GPPAQ) contributed to the recommendation of this instrument.

Seven self-efficacy PROMs were evaluated, three of which were cancer-specific and four of which were specific to self-management. A vast amount of evidence was available for Lorig’s 6-item Self-Efficacy for Managing Chronic Disease instrument, which has been tested with a variety of chronic illnesses and self-management programmes. This instrument was recommended on this basis.

Nine patient-provider interaction outcome measures were evaluated. The Physician Achievement Review (PAR) Instruments, designed to provide doctors with information about their medical practice from the patients’ perspective, have proved successful. The Episodic Care PAR Instruments for patient and provider were recommended in order to strengthen the collaborative relationship found to support self-management practices.

Two instruments were outside the scope of this review: The Patient Partnership in Care Questionnaire (PPiC) and the Health Education and Impact Questionnaire (HeiQ). Further review of these instruments might assist final instrument selection.

Demographic data (e.g. age, education, marital status, gender, ethnicity etc.) relevant to the survivorship self-management work stream will need to be added to the recommended instruments.

It is advised that this package of outcome measures is pilot tested with cancer survivors involved in a self-management programme prior to wider utilisation.

Background

Macmillan Cancer Support defines a cancer survivor as “*someone who is living with or beyond cancer.*” This definition encapsulates those who have completed initial cancer management and have no apparent evidence of active disease, those living with progressive disease who may be receiving cancer treatment, but is not in the terminal phases of illness, and those who have had cancer in the past (Macmillan, 2008). In the UK, there are an estimated two million people living with or beyond cancer, and this is expected to rise annually by more than 3% (KCL, 2008). As cancer patients live longer, research has moved towards exploring the long-term consequences of cancer and its treatment. Such research has shown that many survivors experience significant negative effects on their quality of life (QoL), often long after the completion of treatment (Ferrell, Dow, and Grant, 1995; Gotay and Muraoka, 1999; Institute of Medicine, 2001; Sepulveda et al., 2002). Depending on the type of cancer, treatment, and individual circumstances, survivors can experience a multitude of psychosocial post-treatment effects (Welch-McCaffrey et al., 1989).

Physically, survivors might incur other illnesses as a result of treatment, such as heart disease, renal impairment, hypertension, and osteoporosis (IMNRCNA, 2007). Furthermore, treatments can result in adverse physical complaints such as pain, cognitive impairments (Tannock et al., 2004), chronic fatigue (Loge et al., 2000), early menopause (Ganz, 2005), infertility (Kim, 2006; Maduro et al., 2003; Penson et al., 2003; Ganz et al., 1998), and lymphoedema, which can be persistent or permanent and can occur a long time after treatment. Psychologically, fear of recurrence (Lee-Jones et al., 1997), anxiety associated with hospital discharge (Thomas et al., 1997), and uncertainties about the future (Holland and Reznik, 2005) are frequently reported by survivors (Bloom, 2002; Massie, 2004; Holland and Reznik, 2005). Social implications include being unable to return to work (Short, Vasey, and Belue, 2007; Short, Vasey, and Tunceli, 2005) and reduced social integration (Schag et al., 1994). In contrast, many survivors do report enhanced appreciation of life, personal growth, and positive lifestyle changes (Weiss, 2004; Thornton, 2002). Indeed, post cancer therapy has been identified as a time to intervene and promote lifestyle modification. Completing breast cancer treatment has been referred to as a ‘*teachable moment*’ for lifestyle change (Damush, Perkins, and Miller, 2006).

This ‘teachable moment’ could be a fundamental period for self-management programmes, and a time to increase ‘patient activation,’ defined as “*a person’s ability to manage their health and healthcare*” (Hibbard and Cunningham, 2008). Indeed, chronic disease self-management has been shown in the US, Canada, and the UK to be effective in maintaining and improving patients’ health behaviour and health status, while lowering health care utilisation through improved self-management skills, ‘self-efficacy’ (i.e. a belief in one’s ability, in this case, to self-manage), and better communication between patients and healthcare providers (Lorig et al., 1999; Holroyd et al., 1986; Lorig, Mazonson, and Holman, 1993; Watson et al., 1997). The leading authority in this field is Professor Kate Lorig of Stanford University, California, USA, who developed the Chronic Disease Self-Management Program (CDSMP), a community-based generic programme applicable to all chronic conditions (Lorig et al., 1999).

The CDSMP is designed around self-efficacy theory (Sieving et al., 1997; O’Leary, 1985), with confidence in one’s abilities to perform specific behaviours being one of the key factors

in successful health behaviour change (O'Leary, 1985). In a five-year research project, the CDSMP was evaluated in a randomised study involving more than 1,000 subjects. This study found that people who took the program, when compared to people who did not, improved their healthy behaviours (i.e. exercise, cognitive symptom management, coping, and communications with physicians), improved their health status (i.e. self-reported health, fatigue, disability, social/role activities, and health distress), and decreased their days in the hospital (The British Liver Trust, 1999).

The UK has followed suit with self-management initiatives, the Expert Patient Programme (EPP) being set up in 2001 with guidance from Kate Lorig of the CDSMP (DH, 2001). Like the CDSMP, the EPP is a lay-led self-management programme for all long-term conditions (LTCs). The programme was motivated by the 'Your Health, Your Care, Your Say' consultation showing that people with LTCs, for which cancer has now become, desire more control of their health via self-management (DH, 2006). Evaluations of the EPP have provided support for its utility, but the programme provides little engagement from health professionals, highlighting a potential limitation for cancer survivors (Wilson, 2008). Whilst self-management is fundamentally a personal and independent journey, interactions between healthcare professionals and the 'expert patient' are critical for the exchange of information and decision-making. The Health Foundation uses the term 'co-creating health' to describe an active and collaborative partnership between patients and health professionals (Coulter and Ellins, 2004). For guided self-management to be successful, a positive patient/physician relationship has been shown to be a key factor (Coulter, 1997; Clark and Gong, 2000; Holman and Lorig, 2000). In particular, it has been reported to enhance patient activation (Hibbard and Cunningham, 2008) and increase self-efficacy (Cimprich et al., 2005), both of which are implicated in self-management.

Cancer support groups are a frequently accessible resource available to individuals living with and beyond cancer, both face-to-face and via the Internet. Whether professionally-led or patient-led, support groups have demonstrated efficacy in terms of enhancing coping and increasing psychological well-being (Coulter and Ellins, 2006; Whatley and Milne, 1998; Bottomley, 1997; Montazeri, 1996; Cella and Yellen, 1993). The same has been found for educational programmes, especially in terms of increasing self-management (Stiegelis et al., 2004) and knowledge (Bloom et al., 2008), as well as improving symptom management (Jacobsen et al., 2002; Yates et al., 2004; Baggot et al., 2004; Sherwood et al., 2005) and uptake of healthy behaviours (Pierce et al., 2007; Denmark-Wahnefried et al., 2003, 2006; Stull, Snyder, and Denmark-Wahnefried, 2007). Self-management programmes inevitably encompass support and education via information provision within a friendly group environment. Thus, it could be argued that with cancer even greater benefits might result from the combination of these interventions; an educational self-management programme within a supportive environment.

An example of such an intervention is 'Taking CHARGE,' a USA-based self-management intervention designed to facilitate successful transitions to survivorship after breast cancer treatment. CHARGE is an acronym for: Choose a concern; Have the information; Assess the situation; Record the plan; Gain confidence and insight; and Evaluate your progress. The intervention offers skill-building activities aimed at enhancing self-efficacy for self-regulatory (i.e. personal health management behaviours). Both peer support and the

guidance of a professional were valued among the majority of attendees (Cimprich et al., 2005).

Similarly, in the UK, Macmillan is one of the first to provide a self-management programme for cancer survivors, in the form of the 'Living with Cancer' course. This is a free six-week course for people living with and beyond cancer, facilitated by trained tutors who have also had cancer themselves. Attendees meet each week for two and a half hours to learn new skills and techniques, with sessions including: Overview of Self-Management and Cancer; Making an Action Plan; Relaxation/Cognitive Symptom Management; Feedback/Problem-Solving; Fatigue Management; Getting a Good Night's Sleep; Difficult Emotions; Regaining Fitness during and after Cancer Treatment; Managing Pain; Living with Uncertainty; Making Decisions; Healthy Eating; Future Plans for Health Care; Communication; Managing Depression; Cancer and Changes to your Body; Positive Thinking; Making Decisions about Treatment and Complementary Therapies; Cancer and Relationships; Working with your Health Care Professional; and Looking Back - Looking Forward.

With the advent of cancer survivors self-management programmes, patient-reported outcome measures (PROMs) offer enormous potential in evaluating, monitoring, and improving the quality and success of these programmes. They can provide validated evidence of health from the point of view of the survivor. Not only can PROMs be utilised to assess levels of health and need in the survivor population, but they can also provide valuable 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 survivors.

Lord Darzi's recent review of the NHS recommended that PROMs should have a greater role in the NHS (Darzi 2008). Furthermore, the 'Standard NHS Contract for Acute Hospital Services' (DH, 2008) includes a requirement under section 4 for health providers to report on evidence from PROMs as of April 2009. In light of government strategies for improving the care of people with LTCs, together with recent policy to include PROMs as an important quality indicator, the utility of PROMs in cancer survivor self-management is vast. Indeed, the Cancer Reform Strategy sets out a programme of action across ten areas: six areas of action to improve cancer outcomes and four areas of action to ensure delivery. Collecting and using improved information on different aspects of cancer services and outcomes is central to delivering this strategy, illustrating an important role for PROMs. Better information via patient-reported outcomes could enhance quality of care, inform commissioning, and promote patient choice.

Aims and Objectives

Based on the seventh recommendation of Macmillan's 'Self-Management Support: A review of the Evidence: Summary Document to Support the National Cancer Survivorship Self-Management Workstream,' this review covers recommendation seven: that any changes in survivorship practice, such as self-management programmes, be fully evaluated. Since there is little evidence available to support cancer self-management programmes, it is fundamental to adequately assess such interventions prior to their implementation. Indeed, there needs to be clearly defined criteria for the expected benefits of a self-management

programme and a rigorous evaluation of the most appropriate outcome measure for use with cancer survivorship self-management.

In consolidating the literature on survivorship and evaluating current self-management strategies for LTCs, the proposed six areas of outcome have been identified, all of which will require a valid and reliable measurement instrument if they are to be assessed in cancer survivorship self-management programmes:

Table 1: Outcome Measures to be Reviewed

Health-Related Quality of Life	Cancer-Specific PROMs
Health Behaviours (i.e. lifestyle change)	PROMs
Self-Efficacy	PROMs
Symptom Management	PROMs
Patient(Survivor)-Centred Interactions for Self-Management	PROMs Professional-Reported
Utilisation of Services	Outside of scope of this brief

Methods and Search Strategy

The literature was scoped in terms of each of the identified areas of required outcome measures. This facilitated a comprehensive yet pragmatic search of the literature.

The methods for searching were conducted using four main sources:

- 1) The University of Oxford PROMs database, containing 16,000+ records (up to December 2005) (<http://phi.uhce.ox.ac.uk/>) (Appendix A)
- 2) PubMed (from January 2006 to the present 2009)
- 3) Hand-searching of key journals, from January 2006 to the present 2009: Cancer; Journal of Clinical Oncology; Health and Quality of Life Outcomes; Medical Care; Quality of Life Research.
- 4) The reference list of reviewed studies was scanned for potential articles of relevance.

The abstract of each record was assessed for relevance, as well as for whether the record satisfied the following inclusion and exclusion criteria:

Table 2: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Adult cancer survivors	Children/adolescents
Adults with LTCs	Palliative care
Self-management programmes	Proxy outcome measures
Support groups	Non-English speaking populations
Educational interventions	
The use and evaluation of PROMs	
The use and evaluation of professionally-reported outcomes of patient-centred interactions for self-management.	
Articles from the period of 1998 - 2009	

For those records for which the abstract did not provide enough information to assess its relevance, the full record was accessed. Where possible, the developers of instruments were contacted regarding information that could not be obtained via the articles reviewed.

Assessment and evaluation of the psychometric properties and operational characteristics of outcome measures identified in the selected records was performed independently by an expert in PROMs and cancer. Psychometric and operational properties were evaluated from an adapted version of the London School of Hygiene appraisal criteria (Appendix B) outlined in their review (*Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence*, Smith et al., 2004). Inter-rater reliability was achieved via another expert within the field rating 50% the same instruments whilst remaining blind to how the instrument had already been rated. Any discrepancies were resolved via a consensus meeting, rendering 100% agreement on the final assessments.

Results of Search Strategy

The primary search strategy using the term 'cancer' in the keyword search generated 17,759 records. Supplementary searches on PubMed and via hand-searching key journals generated 92 records.

When assessed against the review inclusion criteria, 85 articles were retrieved and reviewed in full.

Table 3: Articles Identified and Included in the Review

	N Studies Identified	N Included
Bibliography (up to December 2005)	17,759	69
Pubmed/Key Journals	92	85
TOTAL		154

The outcome measures identified from these retrieved records are reviewed according to the categories identified in Table 1.

**QUALITY OF LIFE PATIENT-REPORTED
OUTCOME MEASURES FOR CANCER
SURVIVORS**

Results: Quality of Life PROMs for Cancer Survivors

On scoping the literature for QoL instruments specifically designed for cancer survivors, nine instruments were identified. Of these nine instruments, five can be used with any type of cancer and one can be used with breast cancer survivors. These six instruments are shown in Table 4, along with the country in which they were developed and the populations they have been used with (i.e. cancer type and duration of survivorship).

Table 4: QoL PROMs for Cancer Survivors

Outcome Measure	Country Developed	Cancer Type	Length of Survivorship
Cancer Survivors' Unmet Needs (CaSUN)	Australia	Breast; Gynecological; Prostate; Colorectal.	1 – 15 years
Impact of Cancer (IOC)	US	Breast; Colorectal; Lymphoma; Prostate.	5-10 years post diagnosis
Long-term Quality of Life (LTQL)	US	Breast; Uterine; Cervical.	>5 years post diagnosis
Quality of Life in Adult Cancer Survivors (QLACS)	US	Breast; Bladder; Head and Neck; Gynaecological; Prostate; Colorectal.	2-months to 18 years
Quality of Life – Cancer Survivors (QOL-CS)	US	Breast; Lymphoma; Ovarian; Hodgkin's; Cervical; Leukemia; Colon.	5-11 years
Survivors Module	Australia	Breast	6-24 months post-treatment

As can be seen in Table 4, one of these instruments has been utilised with survivors of 8-18 years (i.e. QLACS), three with survivors of five or more years (i.e. QOL-CS, LTQL, IOC), one with survivors of one or more years (i.e. CaSUN), and one with survivors one to two years post-diagnosis (i.e. Survivors Module). The most appropriate measure out of these six instruments will be determined via an analysis of the psychometric properties and operational characteristics of each instrument. A brief description of instrument domains and scoring has been provided; a more comprehensive description can be viewed in Appendix C, along with a summary of the health domains included within each of the evaluated instruments.

a) Cancer Survivors' Unmet Needs (CaSUN) (Hodgkinson et al., 2006)

Summary of PROM domains, items and scoring

The CaSUN consists of 35 needs over five domains: Existential Survivorship (14-items; e.g., reduce stress in my life; move on with my life); Comprehensive Cancer Care (6-items; e.g.,

best medical care; complaints addressed); Information (3-items; e.g., up to date information; understandable information); QoL (2-items; e.g., manage side-effects; changes in QoL); and Relationships (3-items; e.g., impact on my relationship, problems with sex life).

Respondents indicate whether they have no need (0); met need (1); weak unmet need (2); moderate unmet need (3); or strong unmet need (4) within the last month. The number of met (scores of 1) and unmet (scores 2–4) needs are summed and Total needs calculated from the sum of met and unmet needs. Items can be scored in terms of *items* or *domains of Met, Unmet and Total need*, and/or *strength* of need. Total scores consist of the sum of all need items. Higher scores indicate greater needs (range 0-35). The instrument also includes 6 positive change items (e.g. I have grow as a person) and an open-ended question designed to elicit any needs not identified within the instrument.

Results of Review:

A total of 4 articles evaluating the CaSUN were reviewed. These included breast, gynaecological, prostate, and colorectal disease-free cancer survivors diagnosed between 1 and 15 years earlier.

Reproducibility

Two studies report disappointingly low 3-week test-retest reliability (Hodkinson et al., 2006a; Hodkinson et al., 2007).

Internal Consistency

The Total Cronbach alpha has been reported as being a high 0.96, with the five domains ranging from 0.78 to 0.93 (Hodkinson et al., 2006a; Hodkinson et al., 2007). The majority of items yielded an item-total correlation of above 0.40 (Hodkinson et al., 2006a).

Content Validity

The CaSUN items were generated from the literature, previous measures, and a qualitative study with survivors and their carers. Validity was verified initially by a research panel and then via feedback from respondents. Evidence for the face and content validity of the CaSUN has also been provided via patient evaluation sheets and feedback supporting the relevance and comprehensiveness of the measure (Hodkinson et al., 2007).

Construct Validity

The five domains of the CaSUN have been confirmed via a factor analysis (Hodkinson et al., 2007).

Construct validity has been supported via correlations with expected psychological outcomes. Unmet needs significantly correlated with levels of anxiety and depression, as would be expected (Hodkinson et al., 2007).

The CaSUN has demonstrated discriminative validity in terms of age and disease stage. It has also discriminated between those with clinical or non-clinical anxiety and depression, as measured via the Hospital Anxiety and Depression Scale (HADS) (Hodkinson et al., 2007).

Responsiveness

Needs and unmet needs were responsive to changes in emotional well-being; levels increased with higher levels of anxiety and depression (Hodgkinson et al., 2007).

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

Floor effects have been reported in two studies, with a large number of respondents expressing no unmet need (Hodgkinson et al., 2006a; Hodgkinson et al., 2007).

Acceptability

Good acceptability has been demonstrated in a sample of cancer survivors who had been diagnosed between 1 and 15 years earlier (Hodgkinson, 2006). However, in a mailed study with gynaecological survivors of 1-8 years, the response rate was a low 56% (Hodgkinson et al., 2006b). A similarly low response rate of 67% has also been reported (Hodgkinson et al., 2007). Reading ease is satisfactory, with the CaSUN taking approximately 10 minutes to complete (Hodgkinson et al., 2007).

Feasibility

Feasibility was not reported in the studies reviewed.

Brief Summary

The CaSUN has demonstrated some good psychometric properties with various types of cancer, as well as with both short- and long-term survivors. However, all of the evidence has been reported from the developers of the instrument and thus the instrument has yet to prove itself more widely in terms of reproducibility. Of greatest concern is the low test-retest reported in these studies. Of note, however, is the existence of the Cancer Survivors Partners Unmet Needs measure (CaSPUN), which might prove beneficial in interventions aimed at carers (Hodgkinson et al., 2007).

b) Impact of Cancer (IOC) (Zebrack et al., 2006)

Summary of PROM domains, items and scoring

The IOC consists of 41-items distributed between ten domains: Health Awareness (4-items); Body Changes (5-items); Health Worries (3-items); Positive Self-Evaluation (8-items); Negative Self-Evaluation (4-items); Positive Outlook (3-items); Negative Outlook (4-items); Social Life Interferences (3-items); Value of Relationships (2-items); and Meaning of Cancer (5-items). Responses are dichotomous (i.e. 'yes' or 'no') and on a five-point Likert scale of 'strongly disagree' to 'strongly agree.' Scores are obtained by calculating the mean of each domain, with higher scores indicating stronger endorsement of the content within a particular domain.

Results of Review:

A total of 4 articles evaluating the IOC were reviewed. These included breast, prostate, colorectal cancers and lymphoma survivors of 5-10 years. Sample size was high, one study consisting of 1188 survivors.

Reproducibility

Crespi et al. (2008) report high reproducibility among breast cancer survivors of 5-10 years; split-sample cross-validation yielded a total congruence of 0.98.

Internal Consistency

The Cronbach's alpha of all subscales is reported to be in the acceptable range of 0.67 to 0.89 (Zebrack et al., 2006; Crespi et al., 2008).

Content Validity

The IOC was developed via qualitative interviews with 47 long-term cancer survivors, followed by a content analysis of responses. Expert review and item reduction prior to pilot testing with other established health status and QoL instruments provided further content validity (Zebrack et al., 2006).

Construct Validity

IOC domains were confirmed via factor analysis (Zebrack et al., 2006).

Convergent validity with the Centre of Epidemiological Studies – Depression Questionnaire (CES-D) and Breast Cancer Prevention Trial (BCPT) symptom scores are reported for the IOC, indicating good construct validity (Crespi et al., 2008). A higher negative IOC score was associated with worse physical functioning and worse mental health, whilst a higher positive IOC score was associated with better mental health (Zebrack et al., 2007). Expected associations within and among the IOC subscales and standardised measures of health status and QoL were observed (Zebrack et al., 2006). Subscales have demonstrated discriminative validity between males and females, among age groups and cancer types, as well as presence of co-morbidities (Zebrack et al., 2006).

Responsiveness

Responsiveness was not reported in the studies reviewed.

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

Precision was not reported in the studies reviewed.

Acceptability and Feasibility

Operational characteristics were not reported in the studies reviewed.

Brief Summary

Evidence cited for the IOC is relatively positive, but is also limited to a few studies. Greater utilisation of the IOC is required in order to provide evidence of reproducibility, responsiveness, and operational characteristics. Of note, however, is the availability of a 16-item Brief Cancer Impact Assessment (BCIA) (Alfano et al 2005), which is likely to offer greater acceptability and feasibility. This shorter instrument has not been evaluated in this review as any evidence will be implicated by the limited evidence available for the larger version.

c) Long-term Quality of Life (LTQL) (Wyatt et al., 1996)

Summary of PROM domains, items and scoring

The LTQL is a 34-item QoL measure for long-term female cancer survivors, which consists of four domains: Somatic Concerns (14-items); Spiritual/Philosophical Views of Life (11-items); Fitness (5-items); and Social Support (4-items). Responses are provided on a 5-point Likert scale of 'Not at All' (0) to 'Very Much' (4), scores being calculated via summing the responses to each domain and dividing the total by the number of items in that domain. Higher scores indicate that a particular concern or change applies to the respondent 'very much.'

Results of Review:

A total of 2 articles evaluating the LTQL were reviewed. These included breast, uterine, and cervical cancer survivors.

Reproducibility

Reproducibility was not reported in the studies reviewed.

Internal Consistency

Internal consistency has been reported as being high for the four subscales, ranging from a Cronbach's alpha of 0.86 to 0.92 (Wyatt et al., 1996).

Content Validity

Items were generated through focus groups of survivors (n = 11) and verified with experts and one survivor (Wyatt et al., 1996). Content validity was further supported through inter-rater agreement of subscale items (Wyatt et al., 1996).

Construct Validity

A factor analysis confirmed the four domains of the LTQL (Wyatt et al., 1996).

Significant correlations between the LTQL and the Cancer Rehabilitation Evaluation System (CARES) support the concurrent validity of the LTQL (Gotay and Muraoka, 1998).

Construct validity has also been supported by differential subscale scores according to demographic and health status data (Wyatt et al., 1996). The LTQL has demonstrated discriminative validity between mastectomy and lumpectomy; recurrence and no recurrence.

Responsiveness

Responsiveness was not reported in the studies reviewed.

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

Precision was not reported in the studies reviewed.

Acceptability and Feasibility

Operational characteristics were not reported in the studies reviewed.

Brief Summary

There is not enough supporting evidence to recommend this instrument for utilisation within the proposed project.

d) Quality of Life in Adult Cancer Survivors (QLACS) (Avis et al., 2005)

Summary of PROM domains, items and scoring

The 47-item QLACS is designed to measure HRQoL among long-term cancer survivors (i.e. those who are five or more year's post-diagnosis). The items are distributed between two domains and twelve subscales. The Generic domain consists of seven subscales: Physical Pain; Negative Feelings; Positive Feelings; Cognitive Problems; Sexual Problems; Social Avoidance; and Fatigue. The Cancer-Specific domain consists of five subscales: Financial Problems; Distress about Family; Distress about Recurrence; Appearance Concerns; Benefits of Cancer. The Generic Summary Score (GSS) is calculated by adding the seven constituent subscale scores (reversing the score for Positive Feelings). The Cancer-Specific Summary Score (CSS) is calculated by adding all constituent subscales except Benefits of Cancer. The Benefits of Cancer domain is reported separately. Responses are on a 7-point Likert scale whereby 1 = Never; 2 = Seldom; 3 = Sometimes; 4 = About as often as not; 5 = Frequently; 6 = Very Often; 7 = Always. Higher scores indicate decreased QoL.

Results of Review:

A total of 4 articles evaluating the QLACS were reviewed. These included breast, bladder, head and neck, gynaecological, prostate, and colorectal survivors of 2-months to 18 years.

Reproducibility

When administered to breast cancer survivors (n = 94), 8-18 years post-treatment, at baseline (i.e. minimum of 8 years post-treatment), 2-weeks, and at 1 year follow-up, all test-retest correlations exceeded the standard of 0.70 for testing reproducibility. The inter-

correlation coefficients (ICCs) were 0.95 and 0.98 for the GSS and CSS summary scores, respectively (Avis et al., 2005).

Internal Consistency

The Cronbach's alpha of each domain has been reported to range from 0.72 to 0.95, demonstrating a high level of internal consistency (Avis, Ip, and Long Foley, 2006).

Content Validity

The QLACS was developed from semi-structured interviews with long-term survivors (n = 59), pilot testing with 242 survivors, and then verified by healthcare professionals and consumer advocacy groups (Avis et al., 2005).

Construct Validity

A factor analyses confirmed a seven factor solution for the Generic domain (Avis et al., 2005). A separate factor analysis for the Cancer-Specific domain showed that Benefits of Cancer did not load with the other subscales, thus the CSS is calculated by adding all constituent subscales except Benefits of Cancer.

Measures used at baseline (8 years earlier) demonstrated the CSS to be predictive of days work/usual activity missed, wishful thinking, and preparedness, supporting retrospective validity with cancer survivors (Avis, Ip, and Long Foley, 2006).

All of the Generic subscales demonstrated strong concurrent validity with the equivalent subscales of the Medical Outcomes 36-Item Short-Form Health Survey (SF-36). The QLACS GSS was highly correlated with the SF-36 Physical and Mental Components Summary Scores (PCS; MCS) and the Functional Assessment of Cancer Therapy – General Form (FACT-G). As would be expected, the CSS had a weaker correlation with the PCS and MSC but showed a greater correlation with the FACT Breast Cancer Specific Concerns subscale (FACT-B). This demonstrates concurrent validity of the GSS and discriminative validity of the CSS with measures designed to assess generic QoL (Avis, Ip, and Long Foley, 2006).

Responsiveness

The Total QLACS has demonstrated high responsiveness to life change when administered 8 years' post-treatment and again 1 year later. This was consistent with changes in perceived health. However, the subscales of Cognitive Problems, Sexual Problems, Family Distress, and Recurrent Distress were not significantly responsive even when self-reported change was ascertained (Avis, Ip, and Long Foley, 2006). Over all, the QLACS has demonstrated greater responsiveness to negative as opposed to positive change.

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

The QLACS has demonstrated low floor and ceiling effects. Financial Problems was the only domain to demonstrate a floor effect, where 61% of the sample reported no financial problems (Avis, Ip, and Long Foley, 2006).

Acceptability and Feasibility

Operational characteristics were not reported in the studies reviewed.

Brief Summary

Evidence in support of the majority of psychometric properties of the QLACS has been reported, although evidence of operational characteristics are rarely evaluated. The QLACS was designed for long-term survivors, but unpublished data demonstrates utility in breast cancer survivors two months post-treatment (Davies et al., in press). Furthermore, a recent review of outcome measures in oncology recommends the QLACS for measuring long-term outcomes (Davies, 2009), as will be required for survivorship initiatives.

e) Quality of Life – Cancer Survivors (QOL-CS) (Ferrell, Dow, and Grant, 1995)

Summary of PROM domains, items and scoring

The QOL-CS is a 41-item rating scale consisting of four QoL domains: Physical (8-items), Psychological (18-items), Social (8-items), and Spiritual (7-items). Each item is rated on a scale of 0 to 10, with 0 representing the worst possible outcome and 10 the best possible outcome.

Results of Review:

A total of 6 articles evaluating the QOL-CS were reviewed. These primarily included moderate sample sizes ($n = 425; 549; 687$) of breast, lymphoma, cervical, gestational, ovarian, and Hodgkin's disease survivors.

Reproducibility

Support for the two-week test-retest reliability of the QOL-CS has been reported as being 0.89 for the Total scale, with subscales ranging from 0.81 for Social Well-Being and 0.90 for Spiritual Well-Being (Ferrell et al., 1995).

Internal consistency

The Cronbach's alpha has been reported to be high for the Total score and Psychological subscale score ($\alpha = >0.80$) and moderate for the Physical, Social, and Spiritual subscale scores ($\alpha = >0.63$) (Ferrell et al., 1995; Ahles et al., 2005). In survivors of cervical and gestational cancer, as well as lymphoma, the Cronbach's alpha has been reported as being above 0.90, with subscales ranging from 0.59 to 0.93 (Wenzel et al., 2005; Wenzel et al., 2005b). All but two of the items have demonstrated an item-total correlation above 0.20 (Ferrell et al., 2005).

Content Validity

A one-year pilot period was dedicated to revising the original Quality of Life Questionnaire (QOL) to reflect the concerns and needs of cancer survivors (Ferrell, Dow and Grant, 1995). A literature review and in-depth qualitative interviews were conducted, followed by pilot testing with a sample of 687 cancer survivors (43% breast cancer, 9% lymphoma, 8% ovarian cancer, and 8% Hodgkin's disease). Content validity is also supported via

assessment by a panel of QoL researchers and nurses with expertise in oncology (Ferrell et al., 1995).

Construct Validity

Using stepwise multiple regressions to determine factors most predictive of overall QoL in cancer survivors, 17 variables within the QOL-CS were statistically significant, accounting for 91% of the variance in overall QoL. Variables accounting for the greatest percentage were Control, Aches/Pain, Uncertainty, Satisfaction, Future, Appearance, and Fatigue (Ferrell, Dow, and Grant, 1995).

Concurrent validity has been supported by moderate to strong correlations between the QOL-CS subscales and associated subscales of the FACT-G, including Physical and Psychological subscales (Ferrell et al., 1995; Wenzel et al., 2005). The overall QOL-CS correlation with the FACT-G has been reported as being 0.78 (Ferrell et al., 1995).

Evidence of discriminative validity has been mixed. The QOL-CS has demonstrated discriminative validity between active disease respondents and those in remission (Ferrell et al., 1995), those who are satisfied with their social support versus those who are not (Lim and Zebrack, 2006), and between survivors and healthy controls (Trudeau et al., 2004). The QOL-CS has not demonstrated discriminative validity between those survivors who had received tamoxifen and those who had not, either on the Total score or the subscale scores (Ahles et al., 2005), but expected differences were found between breast cancer and lymphoma survivors who had received chemotherapy, with the latter scoring significantly lower on the Physical subscale.

Long-term survivors of breast cancer who had received diagnoses at an older age (> 65 years) showed significantly worse QoL outcomes in the Physical domain, while those who had received diagnoses at a younger age (27–44 years) showed worse QoL outcomes in the Social domain (Cimprich, Ronis, and Martinez-Ramos, 2002).

Responsiveness

Responsiveness was not reported in the studies reviewed.

Interpretability

The QOL-CS does not yet have clinically meaningful cut-off scores that would allow for an evaluation of the clinical significance of scores.

Precision

Floor or ceiling effects were not reported in the studies reviewed.

Acceptability

Response rates vary from 57% - 81% in mailed format (Ferrell et al., 1995; Wenzel et al., 2005b) and 76% over the telephone (Ahles et al., 2005), demonstrating moderate to high acceptability.

Feasibility

Feasibility was not reported in the studies reviewed.

Brief Summary

The QOL-CS has demonstrated strong psychometric properties with reasonably large samples of survivors of various types of cancer. The instruments' strengths are in terms of reproducibility, internal consistency, and construct validity. However, more research is required to test the responsiveness of the instrument.

f) Survivors Module (Thewes, 2000)

Summary of PROM domains, items and scoring

The Survivors Module was designed specifically for breast cancer survivors and as a adjunct for the Supportive Care Needs Survey (SCNS). The module consists of 46-items distributed over five domains: Information and Medical Communication (18-items); Coping (14-items); Access to Services and Resources (7-items); Fertility (3-items); and Impact of Pain (4-items). Domain total scores are divided by the number of items in each factor to form an average domain score. In order to create a dichotomous dependent variable, average domain total scores are recoded into a dichotomous level of need (i.e. low, moderate or high need).

Results of Review:

A total of 1 article evaluating the Survivors Module and a thesis provided by the instrument developer were reviewed.

Reproducibility

Reproducibility has been supported by a 14-day test-retest ranging from 0.40 to 0.76 for all domains (Thewes et al., 2004). All items with the exception of Item 23 (Coping with changes in other people's behaviour and attitude towards you) and Item 32 (Being informed about the impact of the cancer or treatment on your ability to breast feed children) were significantly correlated between Time 1 and Time 2.

Internal Consistency

Internal consistency has been supported by reports of the Cronbach's alpha ranging from 0.76 to 0.82 for each domain and all item-total correlations being above 0.20 (Thewes et al., 2004).

Content Validity

The Survivors Module was developed via a review of the literature and interviews with 18 breast cancer survivors. Kappa values (i.e. the level of agreement between items and semi-structured interviews with 10 survivors) ranged from good (48% of items) and fair (19%) to poor (23%) (Thewes et al., 2004).

Construct Validity

A factor analysis has confirmed the five domains of the module (Thewes, 2000).

Predictive validity has been supported for the Information and Medical Communication Needs, which has been shown to predict levels of depression, anxiety and worry about

future health. Further support for the predictive validity of the module has been demonstrated with the Access to Services and Resources Needs domain, which has been shown to predict increases in depression and anxiety. In turn, the Coping domain demonstrated convergent validity with depression, anxiety, worry about future health, menopausal symptoms, side effects of treatment, and poor body image. In contrast, the Fertility domain was not significantly related to any of these variables (Thewes, 2000).

Discriminative validity has been demonstrated for the Impact of Pain domain, with unmet needs discriminating between individuals with higher rates of depression, anxiety, menopausal symptoms, side effects of treatment, and breast and arm symptoms (Thewes, 2000).

Responsiveness

Responsiveness was not reported in the studies reviewed.

Interpretability

A cut-off score of 3 is used to determine 'no need' versus 'some need' on a given domain. An average domain score of less than 3 indicates no need (i.e. needs in this area did not exist, or the needs were met), whilst an average domain score greater than or equal to 3 indicates some degree of need.

Precision

Precision was not reported in the studies reviewed.

Acceptability

An acceptable response rate of 79% has been reported (Thewes, 2000).

Feasibility

Feasibility was not reported in the studies reviewed.

Brief Summary

There is not enough supporting evidence to recommend this instrument for utilisation within the proposed project. The majority of evidence has been collated from the thesis of the Survivors Module developer, which has not been published. The module has undergone limited testing.

Discussion

Reliability: Internal consistency was greater than 0.70 for all but one of the QoL measures, this being the IOC. Item-total or item-domain correlations were examined in three of the instruments (i.e. QOL-CS, CaSUN, and Survivors Module), with only the Survivors Module meeting the accepted criteria of at least 0.20. Reproducibility has been examined in two of the instruments (i.e. QLACS, QOL-CS), with both demonstrating good test-retest reliability.

Validity: All six instruments explicitly examined content validity by gaining patient and provider feedback. Tests of construct validity were carried out for all instruments, the QLACS, QOL-CS, LTQL, and IOC demonstrating significant evidence of discriminative validity between known groups.

Acceptability: Acceptability and completion rates were rarely reported. Readability was only examined for the CaSUN and QLACS, whereby readability was measured as an equivalent to a Flesch-Kincaid reading level of 5.6 (78% reading ease) for the CaSUN and 4.8 (88% reading ease) for the QLACS.

Feasibility: Missing data, completion rates, scoring time, and cost of administration were not reported for any of the instruments.

The QOL-CS has demonstrated the most extensive psychometric evidence, but some important considerations are noteworthy. Validation of the scale involved newly diagnosed patients, hence some items focusing on the diagnosis period, and thus raising limitations in terms of content validity. Furthermore, some items elicit change, but fail to elicit the direction of change. No distinction has been made between issues related to shorter and longer-term survival.

The Survivors Module is the only reviewed measure designed specifically for breast cancer survivors. Nevertheless, the other instruments were utilised most often with breast cancer as opposed to other cancer types. Only the QLACS and IOC have demonstrate development via input from samples that represent the larger population of cancer survivors, including a mix of cancer types, gender, and ages. However, both of these instruments were developed with long-term cancer survivors (over five years post-diagnosis).

Although there is still no standard definition of QoL, it has been established that adequate conceptualisation would include physical, psychological, social, and spiritual domains (Williams and Naylor, 1992; Donovan, Sanson-Fisher, and Redman, 1989). The CaSUN measures QoL from a different perspective to most QoL instruments, with the primary focus being on unmet needs. The instrument does not include physical or psychological domains. The LTQL fails to include a psychological domain. Three of the reviewed instruments cover all four recommended domains (IOC; QLACS; QOL-CS). Indeed the QLACS and the QOL-CS include additional domains of symptoms, cognitive functioning, and role activities, the latter two being an often neglected area within QoL assessment.

Recommendations

Table 5 summarises the evidence of psychometric and operational performance applying the appraisal criteria outlined in Appendix B. Based on this appraisal, the QLACS is recommended due to its multidimensional approach to QoL and its domain-specific sensitivity to a number of potential self-management outcomes.

Table 5: Appraisal of Psychometric and Operational Performance of Condition-Specific QoL PROMs

	CaSUN	IOC	LTQL	QLACS	QOL-CS
Reproducibility	—	+	0	++	+
Internal consistency	++	+	+	+	++
Validity: Content	+++	+++	++	+++	+++
Construct	+	++	++	+++	+++
Responsiveness	+	0	0	+	0
Interpretability	0	0	0	0	0
Precision	—	0	0	+	0
Acceptability	+	0	0	+	+
Feasibility	0	0	0	0	0

Psychometric and operational criteria
0 not reported
— no evidence in favour
+ *some limited evidence in favour*
++ *some good evidence in favour*
+++ *good evidence in favour*

OUTCOME MEASURES FOR HEALTH BEHAVIOUR CHANGE

Results: Health Behaviour Change Outcome Measures

On conducting a general scoping of health behaviour change interventions, the main outcome measures are physical activity and self-efficacy. Measures pertaining to diet are often too long and time-consuming to utilise and thus other areas of lifestyle are measured, such as the aforementioned physical activity and self-efficacy. Indeed, Kate Lorig has advised that outcome measures for diet be excluded from the survivors' self-management programme due to acceptability and feasibility implications surrounding the vast number of items required to accurately measure nutrition. Therefore, this section examines physical activity and lifestyle outcome measures, whilst the next section will examine self-efficacy.

In terms of physical activity and lifestyle, this review provides an evaluation of government-commissioned instruments as well as those used in large-scale cancer studies. A brief description of instrument domains and scoring has been provided; a more comprehensive description can be viewed in Appendix C.

a) European Prospective Investigation into Cancer and Nutrition (EPIC) Physical Activity Questionnaire (PAQ) (EPAQ2) (Wareham et al., 2002)

Summary of PROM domains, items and scoring

The EPAQ2 is a self-completed questionnaire that collects past-year self-reported physical activity behaviours in a dis-aggregated way such that the information may be re-aggregated according to the dimensions of physical activity. The questionnaire consists of three sections: Activity at Home; Activity at Work; and Recreational Activity. Questions are closed rather than open-ended, to make them easy to complete and to facilitate large-scale data entry. For occupational activity, both current employment status and the level of physical activity carried out at work (non-worker; sedentary; standing; manual; heavy manual) are recorded. For recreational and household activities, participants report the duration of activities during a typical week in the past year, in summer and winter. Household activities include housework, home repair, gardening and stair climbing. Recreational activities include walking, cycling and sports activities. Time spent in vigorous non-occupational activity is measured in two ways: in a separate question about self-reported time in activities causing sweating or faster heartbeat, and using the sum of time spent in activities with MET values (the rate at which adults burn calories) ≥ 6 (i.e. cycling, sports and stair climbing). Time spent in light-moderate non-occupational activity is estimated using the sum of time spent in activities with MET values < 6 (i.e. housework, walking, gardening, and home repair). A simple four-level physical activity index (PAI) is derived by combining occupational physical activity together with time participating in cycling and other physical exercise (such as keep fit, aerobics, swimming and jogging): Inactive = Sedentary job and no physical exercise or cycling; Moderately inactive = Sedentary job and some but < 1 hour physical exercise and/or cycling per week OR Standing job and no physical exercise or cycling; Moderately active = Sedentary job and 1-2.9 hours physical exercise and/or cycling per week OR Standing job and some but < 1 hour physical exercise and / or cycling per week OR Physical job and no physical exercise or cycling; Active = Sedentary job and ≥ 3 hours physical exercise and/or cycling per week OR Standing job and 1-2.9 hours physical exercise and/or cycling per week OR Physical job and some but < 1 hour physical exercise and/or cycling per week OR Heavy manual job.

Results of Review:

A total of 13 articles evaluating the EPAQ2 were reviewed.

Reproducibility

Strong reproducibility of the EPAQ2 is demonstrated via 5, 10, and 11 months test-retest correlations being moderate to strong, ranging from 0.58 to 0.89 (Pols et al., 2007; Cust et al., 2008). The repeatability of the PAI was high (weighted kappa = 0.6) when administered four times over a one-year period (Wareham et al., 2003).

Internal Consistency

Internal consistency was not reported in the reviewed studies.

Content Validity

The EPAQ2 was developed after a review of the literature and other physical activity measures before being subjected to nine iterations. The Recreation domain was derived from the well-validated Minnesota Leisure Time Activity questionnaire and the Occupational domain from the Modified Tecumseh Occupational Activity Questionnaire (Wareham et al., 2003).

Construct Validity

Cust et al., 2008 demonstrate convergent validity between accelerometer MET-hours/week and categorical measures of physical activity derived from the EPAQ2, including the total PAI. Convergence has also been demonstrated between the PAI and objective measures of daytime energy expenditure (Wareham et al., 2003), waist circumference (Rana et al., 2009), weaker heel bone density (Jakes et al., 2001), and increased breast cancer risk among post-menopausal women (Lahmann et al., 2007). As an indirect test of validity, a positive association has been demonstrated between the PAI and the ratio of energy intake, assessed by 7-day food diaries (Wareham et al., 2003). In a 10-week exercise programme for cancer patients, increased exercise was associated with perceived improvements in fitness, reduced fatigue, enjoyment, enhanced mood, and a sense of achievement (Stevenson and Fox, 2006). Partial correlations between PAI and self-reported behaviours provided further support for construct validity; however, PAI and weight were weakly correlated (Jakes et al., 2004). Non-workers have demonstrated similar levels of activity to participants in sedentary occupations, as might be expected (Cust et al., 2008).

Concurrent validity has been demonstrated with the Leisure Time Physical Activity Questionnaire (LTPAQ) (Cust et al., 2008) and activity diaries (Pols et al., 1997).

Discriminative validity has been demonstrated between active and inactive cohorts and younger and older respondents (Cust et al., 2008; McMunn et al., 2004), as well as according to post-menopausal risk of breast cancer (Lahmann et al., 2007) and occupation (McMunn et al., 2004). Furthermore, discriminative validity has been demonstrated with men and women of lower and higher risk of Coronary Artery Disease (CAD) (Boekholdt et al., 2006). Over all, validity of the EPAQ2 for discriminating respondents according to activity level has been confirmed (Pols et al., 1997).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Interpretation is via a simple 4-level Physical Activity Index (PAI) reflecting an individual's current physical activity: Inactive; Moderately inactive; Moderately active; Active.

Precision

No floor or ceiling effects have been reported in the reviewed studies.

Acceptability

The PAI is simple and easy to understand (Wareham et al., 2003). Response rates as high as 89% and 90% have been reported (Mciza, Goedecke, and Lamber, 2008; Kaptoge et al., 2004). However, postal questionnaires have been less successful, one study rendering a 45% response rate (Panter, Jones, and Hillsdon, 2008).

Feasibility

The utility of the EPAQ2 in ranking participants in terms of their physical activity in large epidemiological studies has been reported (Wareham et al., 2003).

Brief Summary

The EPAQ2 has stood the test of time and proved itself to be a psychometrically strong instrument, especially in terms of reproducibility and construct validity. Further support is needed in terms of internal consistency. Nevertheless, the extensive utility of the EPAQ2, especially with cancer patients, suggests it to be a serious consideration for the survivorship and self-management work stream. The categorisation of respondents into one of four levels of physical activity is a feasible method of monitoring positive changes in physical activity whilst incorporating all modes of physical activity, from activity around the house to more vigorous exercise.

b) General Practice Physical Activity Questionnaire (GPPAQ)

Summary of PROM domains, items and scoring

In 2002 the DH commissioned researchers to produce a short measure of physical activity, which could be used in routine general practice to assist PCTs to meet the National Service Framework (NSF) recommendations regarding patient physical activity. The GPPAQ consists of two domains: Work-Related Physical Activity; Recreation and Leisure (i.e. household chores; cycling; walking). The occupational domain requires respondents' to select the level of activity involved in their work. The recreational domain requires respondents to rate their involvement in five activities over the last week, from 'none' to '3-hours or more.' Based on the EPAQ2, the GPPAQ provides a simple, 4-level PAI reflecting an individual's current physical activity: Inactive; Moderately inactive; Moderately active; Active. The National Institute of Clinical Excellence has recommended the instrument (NICE, 2006).

Results of Review:

A total of 5 articles evaluating the GPPAQ were reviewed.

Reproducibility

The GPPAQ has demonstrated good reproducibility (NICE, 2006).

Internal consistency

Internal consistency was not reported in the reviewed studies.

Content Validity

Content validity is supported in reports of pilot studies with both patients and healthcare providers, as well as with patients whose first language is not English. Furthermore, the PAI was taken from the well-validated EPAQ2. However, implications with content have been noted, namely that the GPPAQ implies unemployed and retired people are inactive (Abdallah et al., 2008).

Construct Validity

The GPPAQ has demonstrated good construct validity. Indeed, the PAI derived from the GPPAQ has published convergent validity with positive associations with both daytime energy expenditure and cardio-respiratory fitness. However, another study has demonstrated that GPPAQ scores do not correlate with life satisfaction or affect (Abdallah et al., 2008). The PAI has demonstrated predictive validity for all-cause and cardiovascular mortality in men and women (Khaw et al, 2006).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

The GPPAQ includes a 4-level Physical Activity Index (PAI) reflecting an individual's current physical activity: Inactive = Sedentary job and no physical exercise or cycling; Moderately inactive = Sedentary job and some but < 1 hour physical exercise and/or cycling per week OR Standing job and no physical exercise or cycling; Moderately active = Sedentary job and 1-2.9 hours physical exercise and/or cycling per week OR Standing job and some but < 1 hour physical exercise and / or cycling per week OR Physical job and no physical exercise or cycling; Active = Sedentary job and ≥ 3 hours physical exercise and/or cycling per week OR Standing job and 1-2.9 hours physical exercise and/or cycling per week OR Physical job and some but < 1 hour physical exercise and/or cycling per week OR Heavy manual job.

Precision

In a study where over 50% of respondents were unemployed, floor effects were evident in levels of physical activity (Abdallah et al., 2008).

Acceptability

The GPPAQ has been well received by nurses, patients and GPs. The patients did not experience any problems in completing the questionnaire, even when English was not their first language. When compared to the International Physical Activity Questionnaire (IPAQ)

and General Practice Assessment Questionnaire (GPAQ), the GPPAQ was preferred by respondents due to its shorter length (McCormick et al., 2006). The instrument takes approximately 30 seconds to complete.

Feasibility

Reports are that healthcare providers find the GPPAQ to be a simple and efficient way of assessing physical activity (NICE, 2006).

Brief Summary

The GPPAQ has demonstrated evidence of validity and certainly seems acceptable to both the respondent and the administrator. Its shortness and ease of interpretability are appealing, as are recommendations made by NICE.

c) Health Habits and History Questionnaire (HHHQ) (Block et al., 1990)

Summary of PROM domains, items and scoring

The 115-item HHHQ assesses dietary intake, providing estimates of nutrient intake from reported frequency of consumption and portion size of a variety of foods.

Results of Review:

A total of 8 articles evaluating the HHHQ were reviewed.

Reproducibility

The reproducibility of HHHQ for estimating "usual past-year" nutrient intake over three administrations produced ICCs ranging from 0.56 to 0.82, with a median of 0.72. Thus, reliability was moderate to good (Hartman et al., 1996). Another study reports test-retest reliability ranging from 0.5 to 0.9.

Internal Consistency

Internal consistency was not reported in the reviewed studies.

Content Validity

Most questions within the HHHQ were selected, usually with the exact wording, from large-scale or continuing national studies. Both the food list and the nutrient values to be associated with it were developed using dietary data from 11,658 adult respondents to the Second National Health and Nutrition Examination Survey (NHANES II). Portion sizes to be associated with each food item were derived from observed portion size distributions in NHANES II, based on three-dimensional models (Block et al., 1990).

Construct Validity

Individuals who currently smoked cigarettes were at elevated risk of disease (Farrow and Davis, 2006) and Total calcium intake (from diet and supplements) was associated with modestly increased risk of prostate cancer, this association being strongest among men who reported not having prostate-specific antigen testing before 1992 (Our results support the hypothesis that very high calcium intake, above the recommended intake for men, may

modestly increase risk of prostate cancer (Rodriguez et al., 2003). Consuming fruit and vegetables ≥ 3 times/d compared with < 1 time/d was associated with a 27% lower stroke incidence, a 42% lower stroke mortality, a 24% lower ischemic heart disease, a 27% lower cardiovascular disease mortality, and a 15% lower all-cause mortality (Bazzano et al., 2002).

Construct validity is further supported by correlations between the HHHQ and food records ranged (Marles-Perlman et al., 1992; Block et al., 1990; Block et al., 1992).

Discriminative validity was demonstrated between seasons (Hartman et al., 1996).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Since most questions within the HHHQ were derived from national surveys, representative national data is available for interpretative purposes.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

The full questionnaire takes 35-40 minutes to self-administer. A 98% response rate has been reported.

Feasibility

Coding of the full HHHQ takes approximately 15 minutes and does not require any special expertise.

Brief Summary

The HHHQ has good reproducibility and is comprehensive in content, but it is too time-consuming for both respondents and administrators.

d) International Physical Activity Questionnaire (IPAQ) (Craig et al., 1999)

Summary of PROM domains, items and scoring

The IPAQ contains four domains: Leisure Time Physical Activity; Domestic and Gardening Activities; Work-Related Physical Activity; and Transport-Related Physical Activity. The IPAQ is available in long-form (27-items) and short-form (7-items), both of which can be self-administered or conducted over the telephone. The IPAQ short-form (IPAQ-SF) is recommended for national monitoring and the long-form for research requiring more detailed assessment (Craig et al., 2003). Computation of the total score for the short-form requires summation of the duration (minutes) and frequency (days) of these three types of activity. Domain specific estimates cannot be estimated.

Results of Review:

A total of 11 articles evaluating the IPAQ were reviewed.

Reproducibility

The IPAQ has demonstrated a high reproducibility of 0.80 (Craig et al., 2003). Percent agreement scores for activity status are reported to be very good (79%), confirming the reliability of the IPAQ for assessing both activity status and sedentary behaviour (Brown et al., 2004). However, test-retest is reported to be low in an international study of physical activity (Rütten et al., 2003).

Internal consistency

Internal consistency was not reported in the reviewed studies.

Content Validity

The development of an international measure for physical activity commenced in Geneva in 1998 and was followed by extensive reliability and validity testing undertaken across 12 countries (14 sites) during 2000 (Craig et al., 2003). The results suggest that these measures have acceptable measurement properties for use in many settings and in different languages, and are suitable for national population-based prevalence studies of participation in physical activity. A study carried out by the Big Lottery Fund concluded the IPAQ to be better than the GPAQ due to its clear definitions of vigorous and moderate activities (Abdallah et al., 2008).

Construct Validity

Construct validity has been supported via positive correlations between inactivity and age and socioeconomic status, as well as inverse associations with self-reported health status (Hallal et al., 2003). Lower educational attainment, female gender, advancing age, non-U.S. birthplace, poorer self-perceived health status, self-perceived depression, smoking, leisure-time television watching/computer use, and receiving a diabetes diagnosis have also been found to be significantly related to sedentary behaviour (Yancey et al., 2004). Both Total and Vigorous Activity scores have demonstrated predictive validity for reduced risk of breast cancer (Slattery et al., 2007). The validity correlation of the IPAQ has been reported to be a low 0.33 in women with breast cancer, along with evidence that the IPAQ can overestimate Total physical activity by as much as 247% (Johnson-Kozlowet al., 2006).

Responsiveness

The responsiveness of the IPAQ has been supported by a number of studies. Responsiveness has been demonstrated with an exercise-counselling group at 6-months, with the counselling group increasing their physical activity (Hardcastle et al., 2008), as well as with participants involved in a postal exercise information intervention (Marshall et al., 2003). The IPAQ has also been responsive to a work-based health promotion initiative (Dugdill et al., 2007).

Interpretability

The IPAQ allows categorisation of respondents into three levels: low, moderate, and high levels of physical activity. The instrument also provides an assessment of whether the

respondent meets government recommendations on physical activity, and an estimate of weekly metabolic energy use.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

The IPAQ is relatively short and easy to understand (Craig et al., 2003).

Feasibility

The IPAQ can be administered via telephone or self-administered, with no loss of reliability (Craig et al., 2003). The utility of the instrument is limited by the age range required of respondents (15-69 years) and the fact that the developers do not recommend the instrument for use in small-scale interventions.

Brief Summary

The IPAQ has many weaknesses in comparison to the other reviewed physical activity questionnaires. Evidence supporting the responsiveness of the instrument enhances its utility, as does its shortness and ease of use and interpretability. However, the highest age for which the IPAQ has been deemed appropriate is 69 years and since cancer increases with age, the instrument would not facilitate representative data collection. Furthermore, the instrument is specifically not recommended for interventions, with the proposed self-management programme being just that.

Discussion

Reliability: Internal consistency was not reported for any of the instruments. Reproducibility has been examined in all four of the instruments, with the EPAQ and IPAQ performing the highest test-retest reliability.

Validity: All of the instruments demonstrated strong evidence of content validity. Indeed, three of the instruments (i.e. EPAQ2, GPPAQ, IPAQ) included similar content due to evidence within the literature pertaining to this content. Tests of construct validity were carried out for all instruments, with reports of high convergent and discriminative validity for all instruments. Concurrent validity was only reported for one of the instruments (i.e. EPAQ2).

Acceptability: Acceptability and completion rates were reported for all of the instruments. Response rates were high for the EPAQ and completion time low for the GPPAQ. In comparison, the HHHQ was reported to take an extensive 35-40 minutes to complete. The IPAQ demonstrated favourable evidence in terms of length and completion time.

Feasibility: Three of the instruments are easy to score and interpret (i.e. EPAQ2, GPPAQ, IPAQ) and two have been commended by healthcare professionals in terms of their utility (i.e. EPAQ2, GPPAQ).

On evaluating the evidence available for the psychometric and operational properties of four exercise and lifestyle PROMs, the strongest evidence was obtained for the EPAQ2. Indeed, the psychometric and operational strength of the EPAQ2 is perhaps further demonstrated by the fact that two of the other evaluated instruments (i.e. GPPAQ and IPAQ) utilised the EPAQ2 for their development. Nevertheless, discussions with Professor Ray Fitzpatrick of the University of Oxford PROMs Group achieved consensus that the EPAQ2 is too extensive for the desired survivorship self-management work stream outcomes. A more precise and focused outcome is considered appropriate to this work stream, which can be achieved with the GPPAQ. Indeed, the GPPAQ is considerably shorter than the EPAQ2 and has been recommended by NICE (2006) for utilisation within general practice.

Recommendations

Table 6 summarises the evidence of psychometric and operational performance applying the appraisal criteria outlined in Appendix B. Based on this appraisal and expert guidance from Professor Ray Fitzpatrick, the GPPAQ is recommended due to its brevity, acceptability, and feasibility. Furthermore, it was developed from the less feasible yet highly validated EPAQ2.

Table 6: Appraisal of Psychometric and Operational Performance of Behaviour Change PROMs

	EPAQ2	GPPAQ	HHHQ	IPAQ
Reproducibility	+++	+	++	++
Internal consistency	0	0	0	0
Validity: Content	+++	++	+++	+++
Construct	+++	++	+++	++
Responsiveness	0	0	0	++
Interpretability	+++	+++	+	+++
Precision	0	—	0	0
Acceptability	++	+++	+	+
Feasibility	+	++	+	+

Psychometric and Operational Criteria

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

SELF-EFFICACY PATIENT-REPORTED OUTCOME MEASURES

Results: Self-Efficacy Outcome Measures

On conducting a general scoping of the self-efficacy literature, various self-efficacy categories emerged: Illness Self-Efficacy; Self-Management Self-Efficacy; and General Self-Efficacy. For the purpose of this review, the categories of interest are cancer self-efficacy and self-management self-efficacy. Outcome measures that encompass these two categories will be reviewed. A brief description of instrument domains and scoring has been provided; a more comprehensive description can be viewed in Appendix C.

Cancer-Specific Self-Efficacy Outcome Measures

a) The Cancer Behavior Inventory (CBI) (Merluzzi et al., 1997)

Summary of PROM domains, items and scoring

The Cancer Behavior Inventory is a measure of coping self-efficacy, with the primary focus being on coping with treatment. The instrument comes in a 33-item Long Form (CBI-L) and a 14-item Brief Form (CBI-B) and includes seven subscales: Maintenance of Activity and Independence (5-items); Seeking and Understanding Medical Information (5-items); Stress Management (5-items); Coping with Treatment-Related Side-Effects (5-items); Accepting Cancer/ Maintaining Positive Attitude (5-items); Affective Regulation (5-items); Seeking Support (3-items). Respondents rate on a scale of 1 to 9 their level of confidence in accomplishing various behaviours during their treatment. The higher the number, the greater the level of confidence.

Results of Review:

A total of 12 articles evaluating the CBI were reviewed, including studies with newly diagnosed patients, those undergoing treatment, and those involved in supportive interventions. Sample size ranged from 24 to 352. The majority were breast cancer patients, but samples of head and neck and BMT patients were also utilised.

Reproducibility

A one-week test-retest reliability coefficient of 0.74 has been reported (Merluzzi et al., 2001), demonstrating good reproducibility.

Internal Consistency

Evidence supports the internal consistency of the CBI. A Cronbach's alpha of 0.94 has been reported for the CBI-L, with the subscales all being > 0.80 (Merluzzi and Nairn, 1999). Little is reported on the internal consistency of the CBI-B, but the developers estimate an alpha of 0.85 (Merluzzi and Nairn, 1999).

Content Validity

The initial development of the CBI involved patient input (Merluzzi and Martinez Sanchez, 1997; Merluzzi et al., 2001). Evidence is reported for content validity, with both the long and short forms being amended in 1999, including the addition of a newly identified subscale (Merluzzi and Nairn, 1999).

Construct Validity

A factor analysis confirmed the seven subscales of the CBI, but one item loaded less than the recommended >0.40 on the Seeking Social Support subscale (Merluzzi and Martinez Sanchez, 1997).

Responses on the CBI have demonstrated predictive validity for emotional and physical well-being one year post-BMT (Hochhausen et al., 2007).

Further evidence has been reported for the construct validity of the CBI, especially in terms of convergent validity with measures of disease impact, adjustment, and QoL (Merluzzi and Nairn, 1999). Concurrent validity with the Psychosocial Adjustment to Illness Scale (PAIS), the Mental Health Index, the Satisfaction with Life Scale, and the FACT QoL Instrument, has been evidenced. These measures were significantly related to all but one of the CBI subscales: Affective Regulation (Merluzzi and Nairn, 1999). Further concurrent validity has been demonstrated between the CBI-B and the Fatigue Management Barriers Questionnaire (FMBQ) (Passik et al., 2002).

Discriminative validity of the CBI has been demonstrated between those who do and do not have difficulty interacting with their healthcare provider (Collie et al., 2005); lower self-efficacy correlates with greater interaction difficulties. Furthermore, responses on the difficulty of various coping subscales were significantly divergent when rated by cancer patients or healthcare providers (Merluzzi and Sanchez, 1998).

Responsiveness

The CBI has demonstrated responsiveness to a 6-week peer-counselling programme (Giese-Davis et al., 2006) and a DVD intervention aimed at improving self-efficacy for seeking social support (Schofield et al., 2008). In contrast, the CBI lacked responsiveness to a 12-week, web-based, social support group intervention (Winzelberg et al., 2003), an 8-week QoL intervention based on Social Cognitive Theory (Graves et al., 2003), and a DVD information intervention (Dunn et al., 2004). In the latter intervention, many variables were measured that also did not, which is likely to reflect a weak intervention as opposed to a weak instrument.

Interpretability

Norms and standard scores are in the process of being developed (Merluzzi and Nairn, 1999).

Precision

No floor or ceiling effects were reported in the studies reviewed.

Acceptability

The developers report that some patients have found it difficult to understand the CBI and thus have required assistance with questionnaire completion (Merluzzi and Nairn, 1999). A response rate of 89% has been reported in breast cancer patients (Collie et al., 2005).

Feasibility

Feasibility was not reported in the studies reviewed.

Brief Summary

The CBI is a valuable tool, with evidence of psychometric properties. However, the primary focus of the instrument is in terms of coping with treatment, which does limit its utility. Furthermore, data on responsiveness is inconsistent, raising concerns about the instruments utility in measuring self-management outcomes.

b) Communication and Attitudinal Self-Efficacy Scale for Cancer (CASE-Cancer) (Wolf et al., 2005)

Summary of PROM domains, items and scoring

The 12-item CASE-Cancer measures patient self-efficacy within the context of productive communication and positive attitude for cancer patients. The instrument consists of three 4-item domains: Understanding and Participating in Care; Maintaining a Positive Attitude; and Seeking and Obtaining Information.

Results of review:

One study evaluating the CASE-Cancer was reviewed.

Reproducibility

Reproducibility was not reported in the study reviewed.

Internal Consistency

The CASE-cancer proved to have high internal consistency (Wolf et al., 2005).

Content Validity

A set of 19 potential items for the CASE-Cancer was pilot tested with 50 cancer patients. Based on the pilot test, item valence was made consistent (i.e., all items worded positively) and the response scale was simplified. The CASE-Cancer was then administered to 127 patients receiving cancer treatment at general oncology clinics (Wolf et al., 2005).

Construct Validity

Psychometric analyses revealed three 4-item factors: Understanding and Participating in Care; Maintaining a Positive Attitude; and Seeking and Obtaining Information (Wolf et al., 2005).

The CASE-cancer proved to have high construct validity (Wolf et al., 2005).

Responsiveness

Responsiveness was not reported in the study reviewed.

Interpretability

Interpretability was not reported in the study reviewed.

Precision

No floor or ceiling effects were reported in the study reviewed.

Acceptability

Scale items performed similarly across literacy levels, demonstrating acceptability with varying levels of respondent literacy (Wolf et al., 2005).

Feasibility

Feasibility was not reported in the study reviewed.

Brief Summary

One study is not sufficient to make recommendations for this instrument.

c) Stanford Inventory of Cancer Patient Adjustment (SICPA) (Telch and Telch, 1986).***Summary of PROM domains, items and scoring***

The SICPA is a 38-item scale designed to assess patients' beliefs about their ability to cope with their cancer. Six domains are covered in the questionnaire: Coping with Medical Procedures; Communication with Physicians, friends and family; Participation in Vocational, Social and Physical Activities; Personal Management; Affective Management; and Self-Satisfaction.

Results of Review:

A total of 6 articles evaluating the SICPA were reviewed. The reviewed studies included relatively small samples (30 – 308) of breast and prostate cancer patients and survivors. Survivorship length varied from 6-weeks post-radical prostatectomy (RP) to 1-year post-surgery for breast cancer.

Reproducibility

Reproducibility was not reported in the studies reviewed.

Internal Consistency

Internal consistency was not reported in the studies reviewed.

Content Validity

Information relating to the development and initial psychometric properties of the scale has never been published. Very few studies have utilised the SICPA and thus, as far as is known, the scale has been subjected to a minimal degree of testing.

Construct Validity

In a heterogeneous sample of 273 cancer patients, a strong positive correlation was found between self-efficacy and QoL and between self-efficacy and mood. Improvements in all three measures brought about by a brief, group program teaching coping skills were also highly correlated (Cunningham, Lockwood, and Cunningham, 1991). Personal Management Self-Efficacy was associated with higher relationship satisfaction, higher Communication Self-Efficacy was associated with less functional impairment, and higher Affective Management Self-Efficacy was associated with higher self-esteem 1-year post-surgery for breast cancer (Manne et al., 2006). Divergent validity has been demonstrated with depression, supporting the literature (Weber et al., 2007).

Responsiveness

Cancer-related self-efficacy was relatively stable over 1 year post-surgery for breast cancer, with only two domains evidencing significant increases over the 1-year time period: Activity Management and Self-Satisfaction (Manne et al., 2006). Lack of responsiveness was also found during an information provision intervention for prostate cancer, but a number of other outcome measures also did not respond to this intervention (Feldman-Stewart et al., 2006). Support for the responsiveness of the SICPA has been demonstrated after an 8-week one-to-one support intervention (Weber et al., 2004).

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

No floor or ceiling effects were reported in the studies reviewed.

Acceptance

One study reported a high response rate of 81% when administered to prostate cancer patients (Feldman-Stewart et al., 2006).

Feasibility

Feasibility was not reported in the studies reviewed.

Brief Summary

The SICPA has very little data to support its utilisation, especially in terms of its content validity and initial development. This is unfortunate as the few studies reviewed have been with cancer survivors, whereby good construct validity has been evidenced.

Self-Management Specific Self-Efficacy Outcome Measures

d) Self-Efficacy for Managing Chronic Disease 6-Items (Lorig et al., 2001)

Summary of PROM domains, items and scoring

The Self-Efficacy for Managing Chronic Disease 6-Items was adapted from the original 33-items Chronic Disease Self-Efficacy Scale in order to reduce respondent burden. The instrument consists of the following domains that are common across many chronic

diseases: Symptom Control; Role Function; Emotional Functioning; and Communicating with Physicians. The score for each item is the number circled. Responses are recorded on a scale of 1 (not at all confident) to 10 (totally confident). The score for the entire scale is the mean of the items. A higher number indicates higher self-efficacy.

Results of Review:

A total of 10 articles evaluating the Self-Efficacy for Managing Chronic Disease 6-Items were reviewed. These included samples of between 280 and 605 participants with a variety of chronic diseases, including Diabetes, Parkinson's disease, Heart Failure, Chronic Respiratory Disease, Chronic Pain, Hypertension, and Schizophrenia. Older people, different ethnicities, and breast and prostate cancer patients and survivors of up to 9 months post-treatment were also included in these studies. Participants were primarily in their 40s, with two studies focusing on the elderly (> 70 years).

Reproducibility

Strong evidence of reproducibility has been reported. The original scale has demonstrated a test-retest reliability ranging from 0.72 to 0.89 (Lorig et al., 1996). In an internet-administered methodology, test-retest after a few days was 0.91 using Pearson's Correlation and 0.87 using Spearman's Correlation (Ritter et al., 2004).

Internal Consistency

Strong evidence of internal consistency has been reported. The original scale has demonstrated a Cronbach's alpha ranging from 0.77 to 0.92 (Lorig et al., 1996). The alpha for the Total 6-item version has consistently been above 0.90 (Lorig et al., 2001; Lester, Stepleman, and Hughes, 2007; Ritter et al., 2004), even when administered over the internet (Ritter et al., 2004). In another study only using three items from the 6-item scale, the Cronbach's alpha remained a high 0.89 (Krein et al., 2007).

Content Validity

The instrument is behaviour-specific and thus content validity is demonstrated via the behaviour itself.

Construct Validity

Concurrent validity has been demonstrated with the SF-36 PCS and MCS 36 (Twitchell, 2007).

Convergent validity has been demonstrated via changes in self-efficacy correlating with changes in a number of health behaviours (e.g. exercise, cognitive symptom management, and communication with physicians) as well as with health status (e.g. fatigue, shortness of breath, pain, role function, depression, and health distress, and fewer emergency department visits (Lorig et al., 2001). In contrast, higher self-efficacy itself was not associated with reported difficulty taking prescribed medications. There was also no significant association between self-efficacy and following a recommended eating plan (Twitchell, 2007).

Predictive validity has been demonstrated with reports of difficulty exercising (Lorig et al., 1999) and illness self-management (Gallagher et al., 2008).

Discriminative validity of the instrument has been demonstrated with individuals who did and did not have chronic pain (Krein et al., 2007), as well as with depressed patients (Jerant et al., 2008). On the other hand, discriminative validity was not demonstrated between MS patients with and without anxiety or depression (Lester, Stepleman, and Hughes, 2007), which is not consistent with the literature (Thornton et al., 2006). Discriminative validity was also not present according to gender (Lorig, 1999), duration of illness, genetic susceptibility, or health insurance status (Shead, 2005).

Responsiveness

Responsiveness has been demonstrated with a 7-week chronic disease self-management programme based on self-efficacy theory (Lorig et al., 2001); statistically significant improvements in self-efficacy were evident at 1-year post-intervention. Responsiveness to time has also been demonstrated after a self-management intervention for Heart Failure (Twitchell, 2007). When administered over three points in time during and after a professional-patient empowerment intervention, responsiveness was not demonstrated at any time point (Pearl-Kraus, 2007). However, no dependent variables included in the study were responsive to the intervention, and thus the weakness is likely to be with the intervention rather than the outcome measure.

Interpretability

No cut-off scores are provided; the higher the score, the higher the self-efficacy.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

A repeated-measures methodology demonstrated a high response rate of 97% (Pearl-Kraus, 2007). When comparing administration via internet and mail, follow-up prompts were required for 64% of the mailed instruments versus 27% of the internet instruments. Overall response rates were 83% for mailed instruments and 88% for Internet instruments (Ritter et al., 2004).

Feasibility

Comparisons between mailed and Internet administered questionnaires suggests less recruitment effort for Internet questionnaires (Ritter et al., 2004).

Brief Summary

Lorig's self-efficacy instruments have withstood a number of psychometric and operational tests. The Self-Efficacy for Managing Chronic Disease 6-Items scale demonstrates impressive reproducibility, internal consistency, and construct validity. The acceptability of this one-page instrument enhances its utility within a self-management package of outcome measures.

e) The Patient Activation Measure (PAM) (Hibbard et al., 2004)

Summary of PROM domains, items and scoring

The 22-item Patient Activation Measure (PAM) assesses patient knowledge, skill, and confidence for self-management. The measure was developed using Rasch analyses and is an interval level, one-dimensional, Guttman-like measure. The instrument comprises a number of statements covering various dimensions of patient involvement in healthcare. Responses are provided on a scale of 'Strongly Agree; Agree; Disagree; Strongly Disagree; or Not Applicable.' The PAM produces activation scores ranging from 0 to 100, which can be used to categorise respondents according to their 'activation' level. To calculate PAM scores, the 22 responses are added: 'Strongly disagree' = 1; 'Disagree' = 2; 'Agree' = 3; 'Strongly Agree' = 4. The exception to this is question 22, which is scored oppositely. If respondents have given 'Not Applicable' answers, their score is divided by the number of items completed and then multiplied by 22. A 13-item version is available.

Results of Review:

A total of 14 articles evaluating the PAM were reviewed. These included a range of methodologies, including a national USA sample of 1,515, a national UK sample of 3,000, a cohort of individuals with mental health issues, as well as with samples living with a variety of chronic conditions (i.e. Diabetes, Arthritis, Hypertension). A longitudinal study with outcome measures being collected at baseline, 6-weeks, and 6 months (n = 479) was also included in the reviewed studies. The studies reviewed demonstrated utilisation of the PAM via face-to-face interview, telephone, and patient completion. Sample age was primarily reported to be 45 years or older.

Reproducibility

Reproducibility was not reported in the studies reviewed.

Internal Consistency

Extensive evidence has been reported in support of the internal consistency of the PAM. Rasch analysis demonstrated the PAM items have infit values between 0.71 and 1.44, which is within the range required for a one-dimensional measure. The consistency of performance of the measure has been demonstrated in the reliability coefficients across subsamples of respondents with Chronic Pain, Depression, Hypertension, Diabetes, Lung Disease, Cancer, and High Cholesterol. A Cronbach's alpha of 0.91 has been reported (Hibbard et al., 2005). An intraclass coefficient (ICC) of 0.87 provides further evidence of instrument consistency (Skolasky et al., 2008). The 13-item version has demonstrated slightly lower reliability for some subgroups, including those with no chronic illness, those 85 years or older, those with self-rated poor health, and those with lower income and education. Thus, there is some loss of precision with the shorter version of the PAM, however, these lower reliabilities still fall within an acceptable range (Hibbard et al., 2005b).

Content Validity

The PAM-22 was developed via four rigorous stages of literature reviews, expert consultations, focus groups, item pool generation, instrument refinement, pilot testing, and testing of psychometric properties (Hibbard et al., 2004). The PAM-13 was developed via

iterative use of Rasch analysis to identify items that could be eliminated without loss of precision and reliability (Hibbard et al., 2005b).

Construct Validity

The mean PAM has demonstrated convergent validity with measures of asthma control (Saft et al., 2008), health literacy (Dubow, 2005), health status (Ellins and Coulter, 2005), and supportive interactions among practitioners and staff within primary care teams (Becker and Roblin, 2008).

Concurrent validity has been demonstrated with the Medical Outcomes 8-item Short-Form Health Survey (SF-8) health status (Hibbard et al., 2005) and the Patient Assessment of Chronic Illness Care (PACIC) (Glasgow et al., 2005).

The PAM has demonstrated predictive validity for a range of behaviours, including: healthy behaviours (e.g., exercise, diet); disease-specific self-management behaviours; and consumerist type behaviours (e.g., reading about risks with a new drug) (Hibbard et al. 2004, 2005; Fowles et al., 2007).

Discriminative validity of the PAM has been demonstrated in terms of exercise frequency, fruit and vegetable intake, and between smokers versus non-smokers. Among those with chronic conditions, discriminative validity has been demonstrated in terms of frequency of undertaking specific self-management tasks (except use of a glucose journal among diabetics in the UK) (Dixon et al., 2006). There are also reports of important differences in PAM scores between demographic groups in both the USA and UK. Lower levels of knowledge, confidence and skills for self-management have been observed among respondents who were elderly, of lower socioeconomic status, and who had finished their education by the age of 16 (Dixon et al., 2006; Fowles et al., 2007).

Responsiveness

The PAM has demonstrated responsiveness to time, activation levels increasing over time. Time has been found to be a statistically significant predictor of activation levels (Hibbard et al., 2007). The PAM has also demonstrated responsiveness to an intervention, the intervention group presenting with increased activation scores significantly above those in the control group by 6-weeks (Hibbard et al., 2007). Another education intervention also elicited responsiveness from the PAM (Alegria et al., 2008).

Interpretability

Cut-off scores have been determined according to the four proposed stages of patient activation:

Stage 1: believes patient role important PAM score 47.0 or lower

Stage 2: confidence and knowledge to take action PAM score 47.1 to 55.1

Stage 3: taking action PAM score 55.2 to 67.0

Stage 4: staying the course under stress PAM score 67.1 and above

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

A high response rate of 76% has been reported in one study (Glasgow et al., 2005).

Feasibility

The developers highlight that the shorter version of the PAM is more feasible for informing patient care plans (Hibbard et al., 2005b). When tested with nurses in a call-centre, the mean time reported to administer the PAM was 5 minutes, but ease of interpretation was only supported by 58% of the nurses (Osmick, 2007).

Brief Summary

The PAM has undergone rigorous psychometric testing both as part of its development and when utilised within research. The instrument provides consistent evidence of psychometric strength. Better reporting of operational characteristics would be beneficial.

f) Self-Rated Abilities for Health Practices Scale (SRAHP) (Becker et al., 1993)

Summary of PROM domains, items and scoring

The 28-item SRAHP measures self-efficacy specific to health promotion. Items are distributed over four domains: Exercise; Nutrition; Health Responsibility; and Psychological Well-Being. Respondents indicate the extent to which they are able to perform each health behaviour on a 5-point scale of 0 (not at all) to 4 (completely). Ratings for each domain are summed to yield domain scores and domain scores are summed to obtain a total score. Total scores range from 0-112, with higher scores indicating greater self-efficacy for health practices.

Results of Review:

A total of 9 articles evaluating the SRAHP were reviewed. These studies were with relatively small samples, no greater than 379. The SRAHP has been utilised with disabled individuals, and those with multiple sclerosis (MS), cerebral palsy, spinal cord injury, and healthy controls. Age range was diverse, with one study specifically examining the younger population (i.e. 18-21 years) and another the elderly (i.e. > 60 years).

Reproducibility

Becker et al. (1993) report a test-retest correlation coefficient of 0.75 over a two-week period, demonstrating good reproducibility.

Internal Consistency

Becker, et al. (1993) report a Cronbach alpha of 0.94 and Callaghan (2003; 2005) reports coefficients ranging from 0.81 to 0.94, with all subscales being >0.80.

Content Validity

Pilot testing and expert review support content validity (Becker et al., 1993).

Construct Validity

Concurrent validity has been demonstrated with the Health Promoting Lifestyle Profile (Ennis et al., 2009) and the Perceived General Self-Efficacy Scale (Becker and Schaller, 1995). When compared to the Multidimensional Health Locus of Control (MHLC), the SRAHP was reported to be more reliable (Peters and Carlson, 1999).

Convergent validity has been demonstrated between scores on the SRAHP and level of need for physical assistance (Becker and Schaller, 1995).

Discriminative validity has been demonstrated between respondents with cerebral palsy and healthy controls, as well as perceived educational status (Jones, Renger, and Kang, 2007).

Predictive validity of the SRAHP has been demonstrated with disabled individuals. Adults with disabilities have been shown to be more likely to engage in a health-promoting lifestyle if they had higher specific self-efficacy for health behaviours and higher general self-efficacy (Stuifbergen and Becker, 2007). Furthermore, the effects of severity of illness on QoL have been found to be partially mediated by self-efficacy (Stuifbergen, Seraphine, and Roberts, 2000). Predictive validity has also been demonstrated in terms of identifying people most likely to achieve exercise maintenance (Kinne, Patrick, and Maher, 1999).

Responsiveness

Responsiveness has been demonstrated to an 8-week multidisciplinary outpatient health promotion education programme (Ennis et al., 2009), a six half-day wellness workshops over 3-months (Zemper et al., 2003), and to the Deaf Heart Health Intervention (DHHI) (Jones, Renger, and Kang, 2007).

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

No floor or ceiling effects were reported in the studies reviewed.

Acceptability and Feasibility

Acceptability and feasibility were not reported in the studies reviewed.

Brief Summary

The SRAHP shows promise in terms of psychometric properties, but does require further testing and greater reporting of operational characteristics.

g) Strategies Used by People to Promote Health (SUPPH)

Summary of PROM domains, items and scoring

The SUPPH is a 29-item self-report measure of self-care self-efficacy, consisting of three domains: Positive Attitudes; Stress Reduction; Making Decisions. The respondent is asked to rate their degree of confidence in carrying out specific self-care behaviours. Each item of the SUPPH is rated on a 5-point scale of confidence from 1 = "very little" to 5 = "quite a lot."

Higher scores indicate greater self-care self-efficacy. The instrument is scored by calculating mean response across all items for each subscore.

Results of Review:

A total of 9 articles evaluating the SUPPH were reviewed. The reviewed studies were primarily with prostate cancer patients both during and post-treatment. Sample size was small to modest, ranging from 95 to 398.

Reproducibility

Two-week test-retest estimates yielded 0.94, demonstrating evidence of reproducibility (Lev and Owen, 2002).

Internal Consistency

In one study the Cronbach's alpha reliability for the Positive Attitude, Stress Reduction and Making Decisions subscales were reported to be 0.94, 0.88 and 0.77, respectively (Lev et al., 2004). Other studies have confirmed high internal consistency reliability above 0.90 (Lev and Owen, 2002; Lev et al., 2008; Kidd, 2007). However, in the latter study, Making Decisions ranged from an unacceptable 0.42 to 0.50 throughout three measurement time points (Kidd, 2007).

Content Validity

Items for the SUPPH were empirically generated, validated by an expert panel, and tested ($n = 275$) for psychometric properties, factor composition, and convergent and discriminate evidence with existing scales (Lev and Owen, 1996).

Construct Validity

The number of factors identified within the SUPPH has been questioned. Initially, four factors emerged: Coping; Stress Reduction; Making Decisions; and Enjoying Life. These factors are consistent with the underlying self-efficacy theory upon which the scale is based (Lev and Owen, 1996; Owen and Lev, 2001, Lev and Owen, 2002; Lev et al., 2008). Subsequent SUPPH data sets, gathered from patients undergoing treatment for various forms of cancer, were combined for a confirmatory factor analysis of the four-factor model ($n = 398$). A four-factor measurement model was created, showing which SUPPH items corresponded to each hypothesised factor. The four-factor model showed acceptable fit indices, some redundancies were revealed, with the Coping factor, demonstrating correlations in the 0.70s with two of its companion factors, Stress and Optimism. A revised confirmatory measurement model was built, combining two of the highly related factors (Coping and Optimism became Positive Attitude). This three-factor model showed fit indices not substantially different from the four-factor model. However, another study identified two factors within the SUPPH: Physiological Efficacy Information; and Performance Efficacy Information (Eller et al., 2006). Further research is needed to clarify the factors within the SUPPH.

Concurrent validity has been demonstrated with measures of QoL (Lev et al., 2008).

The predictive validity of the SUPPH has been demonstrated in terms of QoL outcomes (Lev et al., 2008).

Responsiveness

Using analysis of variance, a series of one-way repeated measures used to investigate changes in cancer patients' self-care self-efficacy and measures of adjustment revealed significant decreases in patients' self-care self-efficacy and QoL over time (Lev, Paul, and Owen, 2001). Further support for the responsiveness of the SUPPH has been demonstrated with treatment, particularly in terms of increased self-efficacy between the beginning and middle of treatment (Kidd, 2007).

Interpretability

Interpretability was not reported in the studies reviewed.

Precision

No floor or ceiling effects were reported in the studies reviewed.

Acceptability

Missing data is reported to be minimal and more likely during the first cycle of chemotherapy than during later cycles or post-chemotherapy (Kidd, 2007).

Feasibility

The SUPPH is reported to be brief and easily administered (Kidd, 2007).

Brief Summary

The SUPPH is the only scale to assess self-efficacy in relation to self-care and has been used with cancer patients. However, the continued inconsistency pertaining to the number of factors within the instrument raise concerns over the validity of evidence reported in the reviewed studies.

Discussion

Reliability: Internal consistency was greater than 0.70 for five of the seven self-efficacy measures. One domain of the SUPPH did not meet the required criteria for internal consistency and internal consistency was not reported for the SICPA. Rasch analysis was conducted with one of the instruments (i.e. PAM), with infit values being within the required range of reliability. Intra-class correlation coefficients (ICCs) were reported for one of the instruments (i.e. PAM), which demonstrated a high ICC of 0.87. Reproducibility has been examined in four of the instruments (i.e. CBI, SUPPH, SRAHP, Lorig's 6-Item Scale), with all demonstrating good test-retest reliability scores above 0.70.

Validity: All but one (i.e. SICPA) of the instruments explicitly examined content validity by gaining patient and provider feedback. Tests of construct validity were carried out for all instruments, with reports of high convergent, concurrent, discriminative, and predictive validity of all instruments except for the CASE-Cancer.

Acceptability: Acceptability and completion rates were briefly reported for six of the seven instruments. Missing data was reported for the SUPPH at certain times during treatment. Response rates were high for four of the instruments (i.e. PAM, SICPA, CBI, and Lorig's 6-item scale). Literacy was only reported for one instrument (i.e. CASE-Cancer), which has demonstrated similar performance across literacy levels.

Feasibility: Only three of the instruments reported on feasibility. The recruitment effort involved in different methods of administering Lorig's 6-item scale was reported to be lower for internet-administered instruments. Effort of administration is also reported for the SUPPH and PAM as being feasible, although evidence is less supportive of PAM interpretation feasibility.

On evaluating the evidence available for the psychometric and operational properties of three cancer-specific self-efficacy instruments and four instruments specific to self-management, the strongest evidence was obtained for the latter. The CBI has limited utility within the broad range of outcomes to be measured during self-management, whilst there is very little evidence available for the CASE-Cancer and the SICPA.

Of the four self-efficacy instruments specific to self-management, the two strongest contenders are the PAM and Lorig's 6-item scale. A vast amount of evidence is available for Lorig's 6-item scale, which has been tested with a variety of chronic illnesses and self-management programmes. Reproducibility and internal consistency are persistently high and the instrument has strong evidence of responsiveness to self-management programmes. Furthermore, the acceptability and feasibility of a 6-item instrument cannot be underestimated when administered alongside accompanying instruments over a number of times. Evidence for the PAM is just as strong, with the two instruments offering similar strengths in terms of psychometric properties.

Recommendations

Table 7 summarises the evidence of psychometric and operational performance applying the appraisal criteria outlined in Appendix B. Based on this appraisal, Lorig's 6-items Self-Efficacy for Managing Chronic Disease is recommended for the purpose of the survivorship self-management work stream. The PAM is comparable in terms of quality, but Lorig's 6-item instrument requires no licensing fee and is much shorter. Both of these are important considerations in terms of patient acceptability and administration feasibility.

Table 7: Appraisal of Psychometric and Operational Performance of Self-Efficacy PROMs

	CBI	CASE-Cancer	SICPA	CDSES	PAM	SRAHP	SUPPH
Reproducibility	+	0	0	++	0	+	+
Internal consistency	++	+	0	+++	+++	+++	++
Validity: Content	++	++	0		+++	0	++
Construct	++	++	++	++	+++	+++	++
Responsiveness	+	0	+	++	++	++	+
Interpretability	+	0	0	++	+++	0	0
Precision	0	0	0	0	0	0	0
Acceptability	+	+	+	+++	+	0	+
Feasibility	0	0	0	++	++	0	+

Psychometric and Operational Criteria

0 *not reported*

— *no evidence in favour*

+

some limited evidence in favour

++

some good evidence in favour

+++

good evidence in favour

**PATIENT/SURVIVOR-CENTRED
INTERACTIONS FOR SELF-
MANAGEMENT**

**PATIENT AND PROVIDER OUTCOME
MEASURES**

Results: Patient-Centred Interactions for Self-Management

On conducting a general scoping of measures pertaining to patient-centred interactions, a number of outcome measures used by regulatory health and medical bodies emerged. These measures were the primary focus of this review due to their already extensive use within the health sector in terms of service improvement. Ten such instruments were identified, one of which was excluded (SHEFFPAT) as it is yet to be validated with adults. Therefore, nine potential measures were reviewed. A brief description of instrument domains and scoring has been provided; a more comprehensive description can be viewed in Appendix C.

a) The American Board of Internal Medicine Questionnaire (ABIMQ) (Lipner et al 2002)

Summary of domains, items and scoring

The 18-item ABIMQ was designed to assess the following aspects of clinical competence over the telephone: Truthfulness; Manner; Not being Patronising; Listening; Showing Interest; Information during Physical Examination; Involvement in Decision-Making; Encouraging and Answering Questions; Giving Information about Problems; Clear Explanations; Kind of Doctor; Length of Time under Doctor's Care; Number of Visits; Whether the Patient would Recommend this Doctor to others. The aim is to provide feedback for use in Continuing Professional Development (CPD) and practice improvement.

Results of Review:

One article evaluating the ABIMQ was reviewed.

Reproducibility

Reproducibility was not reported in the reviewed study.

Internal consistency

Internal consistency was not reported in the reviewed study.

Content Validity

Content validity was not reported in the reviewed study.

Construct Validity

Construct validity was not reported in the reviewed study.

Responsiveness

Responsiveness was not reported in the reviewed study.

Interpretability

Interpretability was not reported in the reviewed study.

Precision

No floor or ceiling effects were reported in the reviewed study.

Acceptability

The ABIMQ is reported to take 8 minutes to complete via telephone (Lipner et al., 2002).

Feasibility

Feasibility was not reported in the reviewed study.

Brief Summary

Very little is known about the development of the ABIMQ and no evidence is available regarding reliability or validity. The instrument does include items pertaining to patient engagement and empowerment, two very important aspects of self-management, but this alone is not sufficient for a recommendation to be made.

b) Consumer Assessment of Health Plans (CAHPS 2.0) (Hargraves et al., 2003)**Summary of domains, items and scoring**

CAHPS is a USA-developed family of surveys widely used to elicit experiences with, and evaluations of, ambulatory care received from health care professionals. CAHPS surveys are designed to evaluate how health plans compare with one another, the CAHPS 2.0 being the Adult Core Survey, designed for administration by post or by telephone. The instrument contains 43-items, 19 of which are core items. These include two global ratings of care, and also produce domain scores for five areas of care: Getting Care Quickly; Doctors who Communicate Well; Courteous/Helpful Office Staff; Getting Needed Care; Health Plan Customer Service. The remaining questions pertain to health plan usage, demographics and screening. Each global item is rated on a scale of 0 to 10, where 0 represents the worse possible care and 10 represent the best possible care. All questions in the five domains use the response options of 'never' to 'always' or 'not a problem' to 'a big problem.'

Results of Review:

A total of 6 articles evaluating the CAHPS 2.0 were reviewed, most studies involving large samples of 49,327 to 166,074, including a USA study with cancer survivors of 2-5 years (n = 408).

Reproducibility

Reproducibility was not reported in the reviewed studies.

Internal Consistency

The two domains relating to how well doctors communicate and the helpfulness and courtesy of the doctor's office staff have demonstrated high internal consistency (Cronbach's alpha of 0.86 and .075, respectively). However, the remaining three domains produced alpha's of less than the recommended 0.70 (Hargraves, Hays, and Cleary, 2003). Other studies have supported the internal consistency of the CAHPS, reporting alpha's above 0.70 (Weech-Maldonado et al., 2003; Zaslavsky et al., 2000).

Content validity

The face validity of the CAHPS 2.0 survey was sought by undertaking thorough field studies and cognitive interviews (Hargraves, Hays, and Cleary, 2003).

Construct Validity

Construct validity has been mixed, but mainly demonstrative of good validity. The item asking about wait times at the doctor's office was only weakly correlated with the "getting care quickly" scale score and the global rating of care. However, all five domains have been reported to be positively associated with global ratings of health care and health plans (Hargraves, Hays, and Cleary, 2003). Concurrent validity has been demonstrated with a similar instrument, the DoctorGuide (Hays et al., 2003).

Predictive validity of CAHPS scores have been demonstrated in terms of patient satisfaction, as well as discriminative validity in terms of gender (Weisman et al., 2001) and ethnic minorities (Weech-Maldonado et al., 2003).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Interpretability was not reported in the reviewed studies.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

It has been reported that the five domains might be overwhelming or of cognitive burden to respondents (Hargraves, Hays, and Cleary, 2003). Indeed, one study reports a response rate as low as 38% (Weech-Maldonado et al., 2003) and another study with non-Hodgkin's Lymphoma (NHL) survivors of 2-5 years was a low 55%. In the latter study, the length of the questionnaire when administered alongside a number of other outcome measures was cited as a key reason for non-response (Arora et al., 2007).

Feasibility

The number of responses needed to achieve sufficient reliability was highest for the "doctors who communicate well" and "helpful/courteous office staff" items (i.e., more than 140 responses). Fewer than 90 responses would be needed for the other three domains to achieve adequate reliability (Hargraves, Hays, and Cleary, 2003).

Brief Summary

The CAHPS 2.0 has undergone strong development and testing of reliability and validity. Furthermore, the instrument includes important items pertaining to patient engagement and empowerment. The utilisation of this instrument with cancer survivors provides further instrument strength in terms of the current review criteria. It is, nevertheless, important to take into consideration the low levels of patient acceptability of this instrument, especially in terms of patient burden and response rates.

c) The Consultation and Relational Empathy (CARE) Measure (Mercer et al., 2004)

Summary of domains, items and scoring

The 10-item CARE Measure is a UK-developed instrument designed as a patient-assessed measure of communication and empathy for use in general practice. The CARE has also proved successful in a secondary care setting. Items are rated on a scale from 1 to 5 (poor to excellent), with a minimum possible score of 10 and a maximum of 50.

Results of Review:

A total of 7 articles evaluating the CARE Measure were reviewed.

Reproducibility

Evidence of reproducibility is lacking; test-retest correlations over 3 months were reported as being low ($\rho = 0.572$) (Mercer et al., 2004).

Internal Consistency

The internal reliability of the instrument has been demonstrated by a high Cronbach's alpha of 0.93 in a series of pilot studies (Mercer et al., 2004). In another study, inter-item correlations were between 0.52 and 0.73 and the Cronbach's alpha remained a high 0.94, the latter being reduced if any item was deleted (Mercer et al., 2007).

Content Validity

Face and content validity have been demonstrated through evidence of a thorough, iterative process of interviews and revisions with a range of patients who endorsed the clarity and importance of the items in the final instrument (Mercer et al., 2004).

Construct Validity

Scaling assumptions have been made without undertaking any factor analysis.

Concurrent validity has been demonstrated with instruments measuring similar concepts, including the Patient Enablement Inventory (PEI) ((Mercer et al., 2004; Price, Mercer, and MacPherson, 2006; MacPherson et al., 2003). In addition, convergent validity has been demonstrated via evidence that measured empathy was related to other aspects of the patient experience (Bikker, Mercer, and Reilly, 2005; Mercer et al, 2005; Mercer and Howie, 2006), including levels of satisfaction (Mercer et al., 2007).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Interpretability was not reported in the reviewed studies.

Precision

There have been reports of possible ceiling effects, with patient ratings tending to agree on high doctor empathy scores more than on low empathy scores (Mercer et al., 2004). It is unclear whether such skewed ratings are related to measurement properties or merely a reflection of genuine perceived high doctor empathy among patients.

Acceptability

Tested in both high and low deprivation settings, 76% of patients and 78% of doctors rated the items in the instruments as 'very important' to their current consultation (Mercer et al 2005). Further evidence of the acceptability of the CARE Measure has been demonstrated in response rates of 70% in two other studies (Mercer et al, 2005, 2007).

Feasibility

Feasibility was not reported in the reviewed studies.

Brief Summary

The CARE Measure was developed with extensive input from patients, but the construct validity and factor structure are yet to be tested sufficiently.

d) Consultation Satisfaction Questionnaire (CSQ) (Baker, 1990)

Summary of domains, items and scoring

The CSQ is an 18-item UK-developed instrument for use in general practice and is recommended for use in the NHS appraisal toolkit. The instrument is designed to be completed by the patient immediately post-consultation and aims to measure patients' satisfaction with their GP consultation. Items focus on the most recent consultation, such as: General Satisfaction; Professional Care; Depth of Relationship; and Perceived Time spent with Doctor. Technical Competence is included via questions on thoroughness of physical examinations. Responses are provided on a five-point Likert scale of 'strongly agree' to 'strong disagree.' Higher scores indicate higher levels of satisfaction.

Results of Review:

A total of 13 articles evaluating the CSQ were reviewed. Respondents were both male and female, the mean age being mid-40s and the central setting being primary care.

Reproducibility

Reproducibility has been demonstrated by a test-retest correlation of 0.82 (Baker and Whitfield, 1992).

Internal Consistency

Internal reliability has been supported by a Cronbach's alpha of 0.91 overall, with subscales all being above 0.80 (Baker, 1990). Item-total correlations have been reported as satisfactory (Baker, 1990).

Content Validity

Content validity has been demonstrated by evidence of a literature review to identify relevant patient issues, discussion with GPs and consensus groups to discuss questionnaire items. A statistical approach to content validity has also been reported, with any editing being guided by principal components analysis (Baker, 1990). Content validity has, however, been questioned due to the ambiguity of some items and the paucity of direct patient involvement in the development of the instrument (Chisholm and Askham, 2006).

Construct Validity

Principal components analysis has demonstrated three CSQ factors: Professional Care; Depth of Relationship; and Perceived Time (Baker, 1990). This is in contrast to the suggested four domains included in the instrument.

Concurrent validity has been demonstrated with the Medical Interview Satisfaction Scale (Bower and Roland, 2003; Howie et al., 1998). Only low correlations have been demonstrated with the PEI (Howie et al., 1998) and between patient satisfaction and physician perception of patient satisfaction (McKinstry, Colthart, and Walker, 2006).

Convergent validity has been demonstrated with measures of relaxation and lowered anxiety in cancer patients after their first outpatient visit (Gallant and Coutts, 2004), but not with exam performance (Stewart, Brown, and Weston, 2003). Another study has shown that GP patient-centred behaviour is not predictive of CSQ outcomes (Mead, Bower, and Hann, 2002).

Discriminative validity has been demonstrated in the follow-up care of individuals with prostate cancer, the Depth of Relationship and Perceived Time domains being significantly different between telephone and outpatient follow-up; scores were lower in the telephone group (Shaida et al., 2006). Discriminative validity has also been demonstrated between General Satisfaction in a group who had changed their GP practice without changing their home address, with a group who had not (Baker & Whitfield 1992c).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Published aggregate means are available for interpretation purposes (Baker, 1996, 1997; Howie et al., 1998).

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

Response rates of 75% and 77% have been reported (Howie et al., 1998; Baker, 1996). Both doctor questionnaires (97%) and patient questionnaires (87%) were completed sufficiently (McKinstry, Colthart, and Walker, 2006).

Feasibility

Doctors who have taken part in CSQ studies have found the process easy to manage, although three found it intrusive. Those who gave an explanation for not taking part expressed concerns regarding time issues (McKinstry, Colthart, and Walker, 2006). The CSQ is reported to provide reliable data with 40 patients (Mercer et al., 2004).

Brief Summary

The reliability and validity of the CSQ have been rigorously tested, but there are reported concerns regarding the content validity of the instrument. Evidence for construct validity is also very mixed. Furthermore, there is little focus on patient engagement and empowerment, key concepts within self-management.

e) Doctors' Interpersonal Skills Questionnaire (DISQ) (Greco, 1999)

Summary of domains, items and scoring

The 12-item DISQ is a UK-developed instrument for use in primary care, which aims to assess the quality of doctors' interpersonal skills within the consultation. The following domains are covered: Greeting; Listening; Explanation; Eliciting Concerns; Reassurance; Time; Consideration of Personal Situation; Respect; Privacy and Dignity; Recommendation to Friends. DISQ asks patients to rate their doctor's performance on a five-point scale from "poor" to "excellent" on these areas.

Results of Review:

A total of 7 articles evaluating the DISQ were reviewed. Respondents' were primarily recruited from primary care, ages ranging from 21-60 years. Sample size was generally high, ranging from 1342 to 55,687.

Reproducibility

Evidence has been reported for a high test-retest reliability of 0.75 (Greco et al., 1998).

Internal Consistency

Internal consistency has been demonstrated to be high (Cronbach's alpha =0.96). Item-correlations have also been reported to be 0.78, providing further support of internal consistency (Greco et al., 1998).

Content Validity

Content validity has been achieved by involving focus groups of patients and GPs in the development of the questionnaire, as well as a review of the literature (Greco et al., 1998).

Construct Validity

Principal components analysis has been carried out, demonstrating the DISQ to be a one-dimensional instrument (Greco et al., 1998).

Concurrent validity has been established with the Falvo-Smith Interaction Scale (Falvo & Smith 1983; Greco and Carter, 2002).

Discriminative validity has been demonstrated between GPs who were and were not vocationally trained (CFEP 2000).

Responsiveness

Two studies evaluating the effect of DISQ feedback and communication training on GP registrars demonstrated significant improvements in intervention groups (Greco et al., 1998; Greco, Brownlea, and McGovern, 2001).

Interpretability

Published means are available, with 95% confidence intervals, for interpretative purposes (Greco et al., 1998).

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

A high response rate of 90% has been reported, as has a completion time of approximately 3 minutes (Greco et al., 2002).

Feasibility

Thirty to fifty patients has been recommended for reliable data (Greco et al., 1998).

Brief Summary

Although the DISQ offers evidence of reasonable psychometric properties and operational characteristics, given the aims of this review there is concern regarding the lack of items pertaining to patient engagement and empowerment.

f) General Practice Assessment Questionnaire (GPAQ) (Ramsay et al., 2000)

Summary of domains, items and scoring

The 25-item GPAQ is a UK-developed instrument for use in primary care and is used for gathering feedback about GPs and to measure aspects of quality of primary care under the new General Medical Services (GMS) contract in the UK. It can be administered by post or immediately after consultation. Five domains are measured: Access (11-items); Continuity of Care (2-items); GP's Communication Skills (8-items); Practice Nurse (3-items); and Patient Enablement (derived from the six-item PEI) (3-items). In addition, there are 7 items collecting a range of health and socio-demographic information from respondents, and a free text section for patient comments. Responses are provided on a scale of 'very poor' to 'excellent.' Domain scores are expressed as a percentage, 100 being the best possible care.

Results of Review:

A total of 2 articles evaluating the GPAQ were reviewed. Respondents' were mainly recruited from primary care, ages ranging from 18-60 years.

Reproducibility

High test–retest correlations of 0.81 and 0.92 have been reported for the GPAQ, demonstrated high reproducibility (Ramsay et al., 2000).

Internal Consistency

The Communication and Interpersonal domains have demonstrated high Cronbach's alpha's of >0.90. All other domains have ranged from 0.69–0.95. Estimates of internal reliability were uniformly high for postal versions (0.88 to 0.97) and consultation versions (0.86 to 0.97) (Ramsay et al., 2000).

Content Validity

The GPAQ was developed from its parent General Practice Assessment Survey (GPAS) questionnaire as part of the 'Patient Experience' domain of the Quality and Outcomes Framework (QoF). Statistical analyses of a dataset of over 20,000 GPAS responses identified those items that were poorly discriminating or potentially redundant due to high inter-item correlations. These were subsequently incorporated into a short survey before being sent to clinical governance leads of 100 randomly selected English Primary Care Trusts (PCTs) for feedback (Ramsay et al., 2000).

Construct Validity

Three factors have been identified via scree plot (Ramsay et al., 2000), in contrast to the suggested five domains.

Responsiveness

Responsiveness was not reported in the studies reviewed.

Interpretability

National benchmarks are available on the GPAQ website.

Precision

Ceiling effects have been reported, with patients tending to rate highly on the Communication domain (Ramsay et al., 2000).

Acceptability

A response rate of 66% has been reported (Ramsay et al., 2000). Completion time is about 10-15 minutes.

Feasibility

Fifty patients are sufficient for reliable data (Mead, Bower, and Roland, 2008).

Brief Summary

Despite utilisation within the Patient Experience section of the QoF, the validity of the GPAQ has not been rigorously demonstrated. Evidence of construct validity is especially lacking.

g) Patient Enablement Instrument (PEI)

Summary of domains, items and scoring

The 6-item PEI is a UK-developed instrument designed for use in primary care and designed to be completed by the patient immediately after consultation in attempts to capture the concept of 'enablement,' which the authors believe reflects patients' ability to understand and cope with their illness. The instrument is most often administered alongside another instrument, such as the CSQ or GPAQ. The PEI focuses specifically on outcomes related to themes of patient-centeredness and empowerment. Response options for the PEI are 'much better/Much more,' 'better/more or same or less,' and the items ask, 'As a result of your visit to the doctor today, do you feel you are...'

Results of Review:

A total of 5 articles evaluating the PEI were reviewed.

Reproducibility

Reproducibility was not reported in the reviewed studies.

Internal Consistency

Internal reliability has been assessed rigorously, offering a high alpha of 0.93 (Howie et al 1998)

Content Validity

It has been reported that the PEI would benefit from a 'not applicable' response option since not all items are relevant to all patients (Howie et al., 1998).

Construct validity

Divergent validity has been demonstrated with the CSQ and MISS, demonstrating the PEI to measure a distinct concept (Howie et al., 1998). Convergent validity has been demonstrated with longer consultation waiting times (Howie, Heaney, and Maxwell, 1994). Another study has shown that patient-centred GP behaviour is not predictive of PEI outcomes (Mead , Bower, and Hann, 2002).

Discriminative validity has been demonstrated between patients who reported more positive compared to those reporting less positive evaluations of involvement in decision-making (Wensing et al., 2008).

Responsiveness

The PEI has been reported to be sensitive and specific in measuring changes in self-concept as a result of acupuncture (Paterson, 2006).

Interpretability

Interpretability was not reported in the reviewed studies.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptance

A response rate of 75% has been reported (Howie et al., 1998). Both doctor questionnaires (97%) and patient questionnaires (91%) were sufficiently completed (McKinstry, Colthart, and Walker, 2006). One study reports that respondents did not like the term 'illness' used in the questionnaire (Paterson, 2006).

Feasibility

Doctors who took part generally found the process easy to manage, although three found it intrusive. Those who gave an explanation for not taking part expressed concerns about the time required (McKinstry, Colthart, and Walker, 2006).

Brief Summary

The PEI has undergone rigorous testing, but offers limited scope in terms of patient-provider interaction beyond patient enablement.

h) Patient Satisfaction Questionnaire (PSQ)

Summary of domains, items and scoring

The 18-item PSQ was developed for patient feedback for monitoring the delivery of medical care. It is designed to gather feedback on medical services as a whole rather than just on individual doctors. Seven dimensions of satisfaction with medical care are measured: General Satisfaction; Technical Quality; Interpersonal Manner; Communication; Financial Aspects; Time with Doctor; Accessibility and Convenience. Responses are provided on a scale of 'very strongly agree' to 'very strongly disagree,' with higher scores representing higher satisfaction.

Results of Review:

One article evaluating the PSQ was reviewed.

Reproducibility

Reproducibility was not reported in the reviewed study.

Internal Consistency

The Cronbach's alpha of all seven PSQ subscales are reported to range from 0.64 to 0.77 (Marshall and Hays, 1994), demonstrating moderate to high internal consistency.

Content Validity

Content validity was not reported in the reviewed study.

Construct Validity

The PSQ subscales have demonstrated concurrent validity with the Patient Satisfaction III (PSQ III) (Marshall and Hays, 1994).

Responsiveness

Responsiveness was not reported in the reviewed study.

Interpretability

Interpretability was not reported in the reviewed study.

Precision

No floor or ceiling effects were reported in the reviewed study.

Acceptability

Acceptability was not reported in the reviewed study.

Feasibility

Feasibility was not reported in the reviewed study.

Brief Summary

Due to the very common term 'patient satisfaction questionnaire,' records pertaining to this instrument were difficult to identify and it was not always clear whether presented data was related to this instrument or another instrument measuring patient satisfaction. Better reporting of the PSQ is required to facilitate reviews such as this if the instrument is to withstand scrutiny.

i) Physician Achievement Review (PAR) Program Questionnaires (Page et al., 1995)

Summary of domains, items and scoring

The PAR programme was set up as a formative performance assessment to provide a multidimensional view of performance through structured feedback to physicians from the physician themselves, their peers, patients and colleagues. It is an element of the Alberta College of Physicians and Surgeons' relicensing process. The programme routinely assesses the performance of physicians, drawing particular attention to physician-patient communication, and is intended primarily to improve the quality of medical practice, but also to identify physicians for whom more detailed assessment of practice performance of medical competence may be needed. The precise content of the PAR programme questionnaires varies depending on the specialism with which each instrument is designed as well as the respondent group (patient, co-worker and peer) the instrument is aimed at. In general, the questionnaires cover five attributes of the physician's performance: Clinical Knowledge and Skills; Communication Skills; Psychosocial Management; Office Management; Collegiality. Patient questionnaires have been adapted for the following specialties: primary care, surgery, anaesthesiology, medical specialties, and episodic care. They are designed to be administered on paper immediately post-consultation to 25 consecutive patients. Responses are provided on a scale of 'strongly agree' to 'strongly disagree.'

Results of Review:

A total of 6 articles evaluating the PAR instruments were reviewed. Respondents' were mainly recruited from primary care, ages ranging from 18-60 years.

Reproducibility

Reproducibility was not reported in the reviewed studies.

Internal Consistency

Internal consistency has been supported by a high Cronbach's alpha co-efficient of >0.90 for all PAR instruments (Hall et al 1999). Another study has reported that all instruments have an alpha of >0.95, providing further support of internal consistency (Lockyer, Violato, and Fidler, 2006).

Content Validity

A pilot study with 308 physician volunteers in order to evaluate the psychometric and statistical properties of the questionnaires and to develop operating policies provides evidence of content validity (Hall et al., 1999).

Construct Validity

Factor analysis has confirmed seven factors that account for 70% of the variance in questionnaire outcomes. These factors were reported to be cohesive, meaningful whilst also providing a multidimensional assessment of the physician (Lockyer, Violato, and Fidler, 2006).

Convergent validity has been evidenced in correlations between patient and colleague responses (Violato and Hall, 2000; Williams et al., 2008). Correlations have also been found between physician performance and declining age (Norton and Faulkner, 1999).

Responsiveness

Responsiveness was not reported in the reviewed studies.

Interpretability

Interpretability was not reported in the reviewed studies.

Precision

No floor or ceiling effects were reported in the reviewed studies.

Acceptability

Response rates of 56% for patients, 95% for co-workers, and 95% for medical colleagues demonstrates high acceptability (Lockyer, Violato, and Fidler, 2006). Furthermore, Physicians have reported the utility of feedback obtained from the questionnaires (Lockyer, 2003).

Feasibility

Lockyer, Violato, and Fidler (2006) report the feasibility of developing multi-source feedback instruments that are valid and reliable. Indeed, in one study no major obstacles to the implementation of the peer assessment questionnaire were encountered (Elwyn, Lewis and Evans, 2005).

Brief Summary

The PAR demonstrates rigorous development techniques and provides forms for both the patient and the provider. Furthermore, patient engagement and empowerment are core topics, thus enhancing the collaborative interaction that is part of many self-management strategies. On closer examination of the PAR instruments, the most relevant to the aims of this review are those designed for episodic care.

Discussion

Reliability: Internal consistency was greater than 0.70 for six of the nine patient-provider interaction instruments (i.e. CARE Measure, CSQ, DISQ, GPAQ, PEI, and PAR).

Satisfactory item-total correlations were reported for one instrument (i.e. CSQ) and satisfactory inter-item correlations for another instrument (i.e. CARE Measure).

Reproducibility has been examined in four of the instruments, with three demonstrating good test-retest reliability scores above 0.70 (i.e. CSQ, DISQ, GPAQ) and one demonstrating low reproducibility (i.e. CARE Measure).

Validity: Scaling assumptions have been made without undertaking any factor analysis for one of the instruments (i.e. CARE Measure). Content validity was minimally reported for many of the instruments, with the exception of the GPAQ and PAR. Tests of construct validity were carried out for all but one of the instruments (i.e. ABIMQ). Evidence was limited for the GPAQ and PSQ, with all other instruments demonstrating good construct validity.

Acceptability: Acceptability and completion rates were reported for the majority of patient-provider interaction measures. Response rates were reported for seven of the nine reviewed instruments. Response rates were generally high (i.e. PEI, CARE Measure, CSQ, DISQ, and PAR). Moderate response rates were reported for the GPAQ and low response rates for the CAHPS, the latter also raising implications surrounding patient cognitive burden.

Feasibility: Feasibility factors were minimally reported, with three instruments having normative data to facilitate interpretation (i.e. CSQ, DISQ, and GPAQ) and one instrument being reported as providing clinical utility without presenting obstacles (i.e. PAR).

Of the nine instruments reviewed, two questionnaires are strongest overall in terms of psychometric properties and operational characteristics: PAR (Canada) and CAHPS 2.0 (USA). The CAHPS 2.0 has, however, encountered a number of problems regarding missing data and patient acceptability. The PAR, on the other hand, has gained positive feedback from both patients and physicians. While systems for gathering feedback from patients at the organisation level are well developed, mechanisms for doing so at physician level are less well established and thus the PAR instruments offer a novel approach to collaborative healthcare.

Recommendations

Table 8 summarises the evidence of psychometric and operational performance applying the appraisal criteria outlined in Appendix B. Based on this appraisal, Episodic Care PAR Instruments for Patient and Provider are recommended. These instruments are psychometrical and operationally ideal for the purposes of the survivorship self-management work stream, especially in terms of providing outcomes for both the patient/survivor and the healthcare provider.

Table 8: Appraisal of Psychometric and Operational Performance of Patient-Provider Interaction Outcome Measures

	ABIMQ	CAHPS 2.0	CARE Measure	CSQ	DISQ	GPAQ	PEI	PSQ	PAR
Reproducibility	0	0	—	+	+	+	0	0	0
Internal consistency	0	++	++	++	++	++	+	+	++
Validity: Content	0	+++	+++	+	+++	+++	+	0	+++
Construct	0	++	++	++	++	—	++	+	++
Responsiveness	0	0	0	0	++	0	+	0	0
Interpretability	0	0	0	+++	+++	+++	0	0	0
Precision	0	0	—	0	0	—	0	0	0
Acceptability	+	—	++	+++	+	+	+	0	+++
Feasibility	0	+	0	++	+	+	+	0	+++

Psychometric and Operational Criteria

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

Instruments outside the Scope of this Review

One instrument relevant to this review was not detected via the search strategy (i.e. Patient Partnership in Care Questionnaire) and another was excluded from appraisal due to limited evidence (i.e. Health Education and Impact Questionnaire). These instruments are briefly discussed in terms of their potential utility within the survivorship self-management work stream.

(a) The Patient Partnership in Care Questionnaire (PPiC) (CFEP)

The 16-item PPiC was not identified via the search strategy for this review due to it being much more practice-based than research-based. The instrument, developed as part of the Health Foundation's 'Co-creating Health' initiative, was designed for clinicians with an interest in supporting patients in self-management. Along with the 16-items, the instrument includes free text comments, relevant demographic information, and benchmark scores. It also provides an evaluation of a health professional's ability to work in partnership with patients with LTCs, improvement of partnership skills, and assessment of patients' confidence to manage their conditions. Personal reports are returned to clinician's, highlighting areas of strength and those requiring potential development.

From briefly scoping the PPiC, the instrument appears consistent with the outcomes Macmillan are assessing. However, there are a number of concerns regarding the feasibility of the instrument. Individual application forms are required per health professional being assessed, which could prove extremely time-consuming. Furthermore, the questionnaire includes a fee of £88 per health professional being assessed (inclusive of everything required to utilise the questionnaire). At least 40 completed questionnaires are required per health professional being assessed in order to achieve sufficient reliability. These various implications might cause future problems at different stages of data collection. Having the questionnaires analysed does enhance feasibility, but the total cost and potential implications pertaining to missing data, etc. are likely to outweigh the benefits of this. Recommending the PPiC above reviewed instruments that do not require a licensing fee and also have evidence supporting their utility cannot be justified.

(b) Health Education and Impact Questionnaire (HeiQ) (Nolte et al., 2006)

The 51-item HeiQ was designed to measure the effectiveness of health education programs and to inform health professionals and researchers on the outcomes of chronic disease health education programs. The instrument consists of eight scales, each being an independent questionnaire providing one part of a comprehensive profile of health education outcomes: Health Directed Behaviour; Positive and Active Engagement in Life; Emotional Well-Being; Self -Monitoring and Insight; Constructive Attitudes and Approaches; Skill and Technique Acquisition; Social Integration and Support; and Health Service Navigation.

The number of studies utilising the instrument are scarce in comparison to those that have been reviewed, which is likely to be due to the novelty and infancy of the instrument. The initial scoping of the instrument was promising, but on closer inspection, a number of

feasibility concerns were raised. The primary concern is the sheer length of the instrument. The instrument consists of eight individual surveys and 51-items. A number of outcomes required by the survivorship self-management work stream are not included within these scales, thus additional necessary questionnaires are likely to exasperate respondent burden. In turn, this is likely to increase non-response or missing data. A key concern would be losing other important outcomes at the expense of the HeiQ.

The licensing fee for the HeiQ is approximately \$478.40 annually for 100 participants, although a collaborative reduction is available to Macmillan.

Recommending this instrument is difficult without sufficient support for its utility. Ideally, each of the eight scales require independent and aggregate appraisal in terms of their relevance to the survivorship self-management work stream. A possible option is to select those scales of relevance and omit those that are not required. This would enhance feasibility in terms of other required outcomes to be assessed. Unfortunately, appraising each of these eight scales so extensively has been beyond the scope of this review and would have resulted in similarly important instruments being omitted.

Final Discussion

Quality of Life PROMs: Six cancer QoL PROMs specific to cancer survivors were evaluated. The Quality of Life in Adult Cancer Survivors (QLACS) questionnaire demonstrated the most relevance to survivorship self-management, including a number of generic and cancer-specific domains that would facilitate interpretation of self-management outcomes. The QLACS also encapsulates all recommended health domains (i.e. physical, psychological, and social) whilst including the additional domains of symptoms, cognitive functioning, and role activities. These three additional domains provide data on symptom management, a key outcome within the survivorship self-management work stream.

Behaviour Change PROMs: Four behaviour change PROMs were reviewed. The European Prospective Investigation into Cancer and Nutrition (EPIC) Physical Activity Questionnaire (EPAQ2) demonstrated consistent evidence of reliability and validity. This instrument, which has been extensively utilised with cancer patients and survivors, was utilised in the development of another reviewed instrument, the General Practice Physical Activity Questionnaire (GPPAQ). The stronger evidence of operational characteristics demonstrated by evidence pertaining to the GPPAQ contributes to the recommendation of this instrument.

Self-Efficacy PROMs: Seven self-efficacy PROMs were evaluated, three of which were cancer-specific and four of which were specific to self-management. A vast amount of evidence was available for Lorig's 6-item Self-Efficacy for Managing Chronic Disease instrument, which has been tested with a variety of chronic illnesses and self-management programmes. Evidence for the Patient Activation Measure (PAM) was equally strong, but Lorig's instrument has been recommended on the basis of length and costs of utilisation.

Patient-Provider Interactions: Nine patient-provider interaction outcome measures were evaluated. The Physician Achievement Review (PAR) Instruments, designed to provide doctors with information about their medical practice from the patients' perspective, have proved successful in Alberta. Indeed, the PAR instruments are used every five years for quality assurance purposes. The Episodic Care PAR Instruments for patient and provider are recommended for the survivorship self-management work stream in order to strengthen the collaborative relationship found to support self-management practices.

Recommendations

Based on the appraisal of the psychometric and operational properties of a number of outcome measures suitable for cancer survivors and chronic disease self-management, the following package is recommended for the survivorship self-management work stream:

1. Quality of Life in Adult Cancer Survivors (QLACS)
2. General Practice Physical Activity Questionnaire (GPPAQ)
3. Lorig's 6-item Self-Efficacy for Managing Chronic Disease Instrument
4. Episodic Care PAR Instruments – Patient Version
5. Episodic Care PAR Instruments – Provider Version

These recommendations are made in the fulfilment of the seventh recommendation of Macmillan's 'Self-Management Support: A review of the Evidence: Summary Document to Support the National Cancer Survivorship Self-Management Workstream': that any changes in survivorship practice, such as self-management programmes, be fully evaluated. The recommended outcome measures are all free to use and can be accessed via the following links:

QLACS:

<http://www.psy.miami.edu/faculty/ccarver/sciQLACS.html>

GPPAQ:

http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_063812

Lorig's 6-item Self-Efficacy for Managing Chronic Disease Instrument:

<http://patienteducation.stanford.edu/research/secd6.pdf>

Episodic Care PAR Instruments – Patient Version:

http://www.par-program.org/EP_Patient%20Form.pdf

Episodic Care PAR Instruments – Provider Version:

http://www.par-program.org/EP_Self%20Assessment%20Form.pdf

Demographic data (e.g. age, education, marital status, gender, ethnicity etc.) relevant to the survivorship self-management work stream will need to be added to the recommended instruments.

It is advised that this package of outcome measures is pilot tested with cancer survivors involved in a self-management programme prior to wider utilisation.

APPENDIX Ai: Sources for Oxford PROM Bibliography

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

APPENDIX Aii: PROM Bibliography Search Strategy

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

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

and

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

or

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

b. records from January 2006 (download 13)

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

and

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

instruments or measure or measures or questionnaire\$ or profile\$ or scale\$ or score\$ or status or survey\$))))

or

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

APPENDIX Bii: Appraisal Criteria

The methods that will be used for assessing the performance of PROMs were developed and tested against multi-disciplinary consensus and peer review. They focus on explicit criteria to assess reliability, validity, responsiveness, precision, acceptability and feasibility.

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

Appraisal Component	Definition/Test	Criteria for Acceptability
<i>Reliability</i>		
Reproducibility/Test-Retest Reliability	The stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and re-test scores	Test re-test reliability correlations for summary scores 0.70 for group comparisons
Internal Consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of items in a scale); assessed by Cronbach's alpha's and item-total correlations	Cronbach's alphas for summary scores ≥ 0.70 for group comparisons Item-total correlations ≥ 0.20
<i>Validity</i>		
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	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation

	correlations between the measure and other similar measures	
<i>Operational</i>		
Acceptability	Evidence of patients' willingness to complete the instrument, as demonstrated through response rates and missing data.	High response rates and low missing data. A reading level equivalent to that of a 12 year-old has been recommended for questionnaires applicable to the general population.
Feasibility	Evidence pertaining to the ease of instrument administration.	Ease of administration, scoring, and interpretation. Low costs.

Each of these appraisal components will be assessed according the extent of the evidence supporting them, the following scoring criteria being utilised:

Psychometric and Operational Criteria	
<i>0</i>	<i>not reported</i>
<i>—</i>	<i>no evidence in favour</i>
<i>+</i>	<i>some limited evidence in favour</i>
<i>++</i>	<i>some good evidence in favour</i>
<i>+++</i>	<i>good evidence in favour</i>

APPENDIX C: INSTRUMENT DOMAINS AND SCORING

QoL PROMs: Domains and Scoring

Instrument	Domains (No. Items)	Response Options	Score	Completion Time	Licensing
CaSUN	Five domains: Existential Survivorship (14-items); Comprehensive Cancer Care (6-items); Information (3-items); QoL (2-items); and Relationships (3-items).	Respondents indicate whether they have no need (0), met need (1), weak unmet need (2), moderate unmet need (3), or strong unmet need (4) within the last months.	The number of met (scores of 1) and unmet (scores 2–4) needs are summed and Total needs calculated from the sum of met and unmet needs. Items can be scored in terms of <i>items</i> or <i>domains of Met, Unmet and Total need</i> , and/or <i>strength</i> of need. Total scores consist of the sum of all need items. Higher scores indicate greater needs (range 0-35).	Approximately 10 minutes.	Free
IOC	41-items distributed between ten domains: Health Awareness (4-items); Body Changes (5-items); Health Worries (3-items); Positive Self-Evaluation (8-items); Negative Self-Evaluation (4-items); Positive Outlook (3-items); Negative Outlook (4-items); Social Life Interferences (3-items); Value of Relationships (2-items); and Meaning of Cancer (5-items).	Responses are dichotomous (i.e. 'yes' or 'no') and on a five-point Likert scale of 'strongly disagree' to 'strongly agree.'	Scores are calculated by calculating the mean of each domain, with higher scores indicating stronger endorsement of the content within a particular domain.	Not reported.	Free

LTQL	34-items consisting of four domains: Somatic Concerns; Spiritual/Philosophical Views of Life; Fitness; and Social Support.	Responses are provided on a 5-point Likert scale of 'Not at All' to 'Very Much.'		Not reported	Free
QLACS	47-items and 12 domains; 7 Generic domains include: Physical Pain (4-items), Negative Feelings (4-items), Positive Feelings (4-items), Cognitive Problems (4-items), Sexual Problems (4-items), Social Avoidance (4-items), Fatigue (4-items). 5 Cancer-Specific domains include: Financial Problems (4-items), Distress about Family (3-items), Distress about Recurrence (4-items), Appearance Concerns (4-items), Benefits of Cancer (4-items).	7-point Likert scale: 1 = Never; 2 = Seldom; 3 = Sometimes; 4 = About as often as not; 5 = Frequently; 6 = Very Often; 7 = Always.	The Generic Summary Score (GSS) is calculated by adding the seven constituent domain scores (reversing the score for Positive Feelings). The Cancer-Specific Summary Score (CSS) is calculated by adding all constituent domains except the Benefits of Cancer. The Benefits of Cancer domain is reported separately.	Not reported.	Free from http://www.psy.miami.edu/faculty/carver/sciQLACS.html
QOL-CS	41-item rating scale consisting of four QoL domains: Physical (8-items), Psychological (18-items), Social (8-items), and Spiritual (7-items).	Each item is rated on a scale of 0 to 10, with 0 representing the worst possible outcome and 10 the best possible outcome.	Some items require reverse scoring. Higher scores indicate better outcomes.	Not reported.	Free from http://prc.coh.org/QOL-CS.pdf

Survivors Module	46-items over 5 domains: Information and Medical Communication; Coping; Access to Services and Resources; Fertility; and Impact of Pain.		Average domain total scores are recoded into a dichotomous level of need: “no need” or “some need” (ie. low, moderate or high need). A cut-off score of 3 is used to determine if there is none versus some need on a given factor. An average factor score less than 3 are determined to have no need greater than or equal to 3 were determined to have some degree of need.	Not reported.	\$150 for administration, scoring, and analysis guide.
-------------------------	--	--	--	---------------	--

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

Instrument	Instrument domains							
	Physical function	Symptoms	Global judgement	Psychol. Well-being	Social well-being	Cognitive functioning	Role activities	Personal construct
CaSUN		X			X			X
IOC	X			X	X			X
LTQL	X				X			X
QLACS	X	X		X	X	X	X	X
QOL-CS	X	X		X	X	X	X	X
Survivors Module	X	X		X				

Physical Activity PROMs: Domains and Scoring

Instrument	Domains (No. Items)	Response Options	Score	Completion Time	Licensing
EPAQ2	Three domains: Activity at Home; Activity at Work; and Recreational Activity.	Respondents select duration of a number of physical activities. Response options vary from 'once a day' to 'less than once a month' etc.	4-level Physical Activity Index (PAI): Inactive; Moderately inactive; Moderately active; Active	Not reported.	Free: http://www.sdprc.org/lhn-tools/epaq2.pdf
GPPAQ	Two domains: Work-Related Physical Activity; Recreation and Leisure (i.e. household chores; cycling; walking).	Work-Related Physical Activity: Respondents select the level of activity involved in their work (i.e. Unemployed; Manual Labour); Recreation and Leisure: Respondents rate various activities on a scale of 'None' to '3 hours or More per Week'.	4-level Physical Activity Index (PAI): Inactive; Moderately inactive; Moderately active; Active	30 seconds.	Free: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_063812
HHHQ	115-items assessing dietary intake.	Responses are provided according to quantity of nutrient and food intake.	Estimates of nutrient intake are determined from reported frequency of consumption and portion size of a variety of foods.	35-40 minutes.	Licensing fee required via contacting Gladys Block.

<p>IPAQ</p>	<p>7-items over 4 domains: Leisure Time Physical Activity; Domestic and Gardening Activities; Work-Related Physical Activity; and Transport-Related Physical Activity. The IPAQ-SF asks about three specific types of activity undertaken in the four domains: Walking; Moderate-Intensity Activities; and Vigorous-Intensity Activities.</p>	<p>Respondents provide information on minutes and hours per day involved in various levels of physical activity over the past 7 days.</p>	<p>Computation of the total score requires summation of the duration (in minutes) and frequency (days) of the three activity types. Domain specific estimates cannot be estimated.</p> <p>Respondents can be categorised into one of three levels of physical activity: low, moderate, and high levels of physical activity. It also provides an assessment of whether they meet government recommendations on physical activity, and a calculation of estimates of weekly metabolic energy use.</p>	<p>Not reported.</p>	<p>Free</p>
--------------------	---	---	--	----------------------	-------------

Self-Efficacy PROMs: Domains and Scoring

Instrument	Domains (No. Items)	Response Options	Score	Completion Time	Licensing
CBI	The instrument comes in a 33-item Long Form (CBI-L, Version 2.0) and a 14-item Brief Form (CBI-B, Version 2.0) and includes seven subscales: Maintenance of Activity and Independence (5-items); Seeking and Understanding Medical Information (5-items); Stress Management (5-items); Coping with Treatment-Related Side-Effects (5-items); Accepting Cancer/ Maintaining Positive Attitude (5-items); Affective Regulation (5-items); Seeking Support (3-items).	Respondents rate on a scale of 1 to 9 their level of confidence in accomplishing various behaviours during their treatment.	The higher the number, the greater the level of confidence.	Not reported.	Free
CASE-Cancer	18-items measuring patient self-efficacy within the context of productive communication.	Not reported.	Higher scores indicate higher self-efficacy.	Not reported.	Not reported.

SICPA	38-items in six domains: Coping with Medical Procedures; Communication with Physicians, friends and family; Participation in Vocational, Social and Physical Activities; Personal Management; Affective Management; and Self-Satisfaction.	4-point Likert scale of agreement.	Higher scores indicate higher self-efficacy.	Not reported.	Free
Lorig's 6-Item Self-Efficacy for Self-Management Scale	<p>9 subscales of 33-items: Exercise; Information; Obtaining Help; Communication with Physician; Manage Disease; Chores; Social/Recreational Activities; Manage Symptoms; Control Depression.</p> <p>The 6-item scale contains items taken from these 9 subscales. It covers several domains that are common across many chronic diseases: Symptom Control; Role Function; Emotional Functioning; and Communicating with Physicians.</p>	Responses are recorded on a scale of 1 (not at all confident) to 10 (totally confident).	The score for each scale is the mean of the items. Higher number indicates higher self-efficacy.	Seconds.	Free from http://patienteducation.stanford.edu/research/secd6.pdf

PAM	22-item measure that assesses patient knowledge, skill, and confidence for self-management. The measure comprises a number of statements covering various dimensions of patient involvement in healthcare.	Responses are provided on a scale of Strongly Agree, Agree, Disagree, Strongly Disagree or Not Applicable.	The measure produces activation scores ranging from 0 to 100, which can be used to categorise respondents according to their 'activation' level. To calculate PAM scores, the 22 responses are added: 'Strongly disagree' = 1; 'Disagree' = 2; 'Agree' = 3; 'Strongly Agree' = 4. The exception to this is question 22, which is scored oppositely). If respondents have given 'Not Applicable' answers, their score is divided by the number of items completed and then multiplied by 22.	Not reported.	Licensing fee required via contacting Insignia Health.
SRAHP	28-item, summated rating scale, designed to measure self-efficacy specific to health promotion: Exercise; Nutrition; Health Responsibility; and Psychological Well-Being.	Respondents indicate the extent to which they are able to perform each health behaviour on a 5-point scale, from "0—Not at all" to "4—Completely."	Higher scores indicate greater self-efficacy in health practice.	Not reported.	Not reported.
SUPPH	29-item self-report measure of self-care self-efficacy, consisting of three domains: Positive Attitudes; Stress Reduction; Making Decisions.	Each item of the SUPPH is rated on a 5-point scale of confidence from 1 = "very little" to 5 = "quite a lot."	Higher scores indicate greater self-care self-efficacy. The instrument is scored by calculating mean response across all items for each subscore.	Not reported.	Not reported.

Patient-Provider Interaction Outcome Measures: Domains and Scoring

Instrument	Domains (No. Items)	Response Options	Score	Completion Time	Licensing
ABIMQ	18-items assessing clinical competence in: Truthfulness; Manner; Not being Patronising; Listening; Showing Interest; Information during Physical Examination; Involvement in Decision-Making; Encouraging and Answering Questions; Giving Information about Problems; Clear Explanations; Kind of Doctor; Length of Time under Doctor's Care; Number of Visits; Whether the Patient would Recommend this Doctor to others.	5-point Likert scale of 'poor' to 'excellent'	Higher scores indicate better competence.	8 minutes.	Not reported.
CAHPS 2.0	43-items, 19 of which are core items. These include two global ratings of care, and also produce domain scores for five areas of care: Getting Care Quickly; Doctors who Communicate Well; Courteous/Helpful Office Staff; Getting Needed Care; Health Plan Customer Service.	Each global item is rated on a scale of 0 to 10. All questions in the five domains use the response options of 'never' to 'always' or 'not a problem' to 'a big problem.'	Higher scores indicates better care.	Approximately 15 minutes.	Free

Care Measure	10-items assessing doctor communication and empathy.	Items are rated on a scale from 1 to 5 (poor to excellent).	A minimum possible score of 10 and a maximum of 50; higher scores indicate better care.	Not reported.	Free
CSQ	18-item focusing on the most recent consultation: General Satisfaction, Professional Care, Depth of Relationship, and Perceived Time spent with Doctor. Technical Competence is included via questions on thoroughness of physical examinations.	Responses are provided on a five-point Likert scale of 'strongly agree' to 'strong disagree.'	Higher scores indicate high satisfaction.	Approximately 10-15 minutes.	Free
GPAQ	25-item over 5 domains: Access (11-items); Continuity of Care (2-items); GP's Communication Skills (8-items); Practice Nurse (3-items); and Patient Enablement (derived from the six-item PEI) (3-items). In addition, there are 7 items collecting a range of health and socio-demographic information from respondents, and a free text section for patient comments.	Responses are provided on a scale of 'very poor' to 'excellent.'	Domain scores are expressed as a percentage, 100 being the best possible care.	10-15 minutes.	Free

PEI	6-items assessing the concept of 'enablement' and patient-centeredness.	Response options for the PEI are 'much better/Much more,' 'better/more or same or less,' and the items ask, 'As a result of your visit to the doctor today, do you feel you are...'	Higher scores indicate better coping or 'enablement.'	Not reported.	Not reported.
PSQ	18-item over 7 domains: General Satisfaction; Technical Quality; Interpersonal Manner; Communication; Financial Aspects; Time with Doctor; Accessibility and Convenience.	Responses are provided on a scale of 'very strongly agree' to 'very strongly disagree.'	Higher scores indicate higher satisfaction.	Not reported.	Not reported.
PAR	Number of items vary, but all instruments assess 5 attributes of the physician's performance: Clinical Knowledge and Skills; Communication Skills; Psychosocial Management; Office Management; Collegiality.	Responses are provided on a scale of 'strongly agree' to 'strongly disagree.'	Higher scores indicate better physician performance.	All instruments vary in length.	Free for private study, research or review: Patient Version – http://www.par-program.org/EP_Patient%20Form.pdf Provider Version – http://www.par-program.org/EP_Self%20Assessment%20Form.pdf

References

1. Abdallah S, Steuer N, Marks N, Page N.(2008) Well-being evaluation tools: A research and development project for the big lottery fund. Centre for Well-Being, New Economics Foundation, May 2008.
2. Ahles TA, Saykin AJ, Furstenberg CT, Cole B, Mott LA, Titus-Ernstoff L, Skalla K, Bakitas M, Silberfarb PM. Quality of Life of Long-Term Survivors of Breast Cancer and Lymphoma Treated With Standard-Dose Chemotherapy or Local Therapy. *Journal of Clinical Oncology* 2005; Vol 23, No 19 (July 1), 2005: pp. 4399-4405.
3. Alfano CM, McGregor BA, Kuniyuki A, Reeve BB, Bowen DJ, Smith AW, Baumgartner KB, Bernstein L, Ballard-Barbash R, Malone KE, Ganz PA, McTiernan A. Psychometric Evaluation of the Brief Cancer Impact Assessment among Breast Cancer Survivors *Oncology* 2005;70:190-202 (DOI: 10.1159/000094320).
4. Avis NE, Smith K, McGraw S, Smith RG, Petronis VM, Carver CS. Assessing quality of life in adult cancer survivors (QLACS). *Qual Life Res* 2005;14(4): 1007–1023.
5. Avis NE, Edward I, Long Foley K. Evaluation of the Quality of Life in Adult Cancer Survivors (QLACS) scale for long-term cancer survivors in a sample of breast cancer survivor. *Health and Quality of Life Outcomes* 2006; 4:92.
6. Bandura A. Social foundations of thought and action. Englewood Cliffs (NJ): Prentice Hall; 1986.
7. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology*. 1990;1:58-64.
8. Block G, Thompson FE, Hartman AM, et al. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *J Am Diet Assoc*. 1992;92:686-93.
9. CFEP Surveys (2008) Self-management. <http://www.cfep.co.uk/selfmanagement.aspx> [Last accessed 19 April 2009]
10. Chisolm A and Askham J. What do you think of your doctor? A review of questionnaires for gathering patients' feedback on their doctor. Picker Institute Europe, Sept 2006.
11. Cimprich B, Ronis DL, Martinez-Ramos G. Age at Diagnosis and Quality of Life in Breast Cancer Survivors. *Cancer Practice* 2002; Volume 10 Issue 2, Pages 85 – 93.

12. Cimprich B, Janz NK, Northouse L, Wren PA, Given B, Given CW. Taking CHARGE: A self-management program for women following breast cancer treatment. *Psycho-Oncol.* 2005;14:704–17
13. Coulter, A. & Ellins, J. (2006) *The quality enhancing interventions project: patient-focused interventions*. London: The Health Foundation.
14. Crespi CM, Ganz PA, Peterson L, Castillo A, Caan B (2008) Refinement and Psychometric Evaluation of the Impact of Cancer Scale. *Journal of the National Cancer Institute* 2008, 100(21):1530-1541; doi:10.1093/jnci/djn340.
15. Cunningham AJ, Lockwood GA and Cunningham JA (1991) A relationship between perceived self efficacy and quality of life in cancer patients *Patient Education and Counseling* 17:71-78.
16. Cust et al. (2008) Validity and repeatability of the EPIC physical activity questionnaire: a validation study using accelerometers as an objective measure *Int J Behav Nutr Phys Act.* 2008; 5: 33. Published online 2008 June 2. doi: 10.1186/1479-5868-5-33.
17. Darzi A, Professor Lord (2008) *High Quality Care for All: NHS Next Stage Review*. Department of Health, London.
18. Davies NJ, Kinman G, Thomas R, Bailey TA. Predicting quality of life post-chemotherapy: the role of health baseline comparisons in breast cancer, in press, 2009.
19. Davies NJ Measuring health-related quality of life in cancer patients. *Nursing Standard*, 23, 30, 42-49.
20. Department of Health (2001) *The Expert Patient: A New Approach to Chronic Disease Management for the 21st Century*. DH, August 2001.
21. Department of Health (2006) *Your Health, Your Care, Your Say*. DH Consultation, 2006.
22. *Department of Health (2008) Health, Your Way – A Guide to Long-Term Conditions and Self-Care.* DH, November 2008.
23. Department of Health (2008) *The Standard NHS Contract for Acute Hospital Services*. Department of Health, London.
24. Dunn J, Steginga SK, Rose P, Scott J, Allison R. Evaluating patient education materials about radiation therapy. *Patient Education and Counseling* 52 (2004) 325–332.

25. Ferrell BR, Dow KH, Leigh S, Ly J, Gulasekaram P. Quality of life in long-term cancer survivors. *Oncology Nursing Forum* 1995;22:915-22.
26. Ferrell BR, Dow KH, Grant M. Measurement of the quality of life in cancer survivors. *Qual Life Res* 1995;4(6):523–531.
27. Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technology Assessment*. 2, 14, i-iv, 1-74. <http://www.ncchta.org/fullmono/mon214.pdf>. (Last accessed: January 30 2009.).
28. Gallagher, Robyn BA (Psych) MN PhD¹; Donoghue, Judith RN PhD²; Chenoweth, Lynn RN DRec BA MA (Hons) GCert Tch/Lrn MAdEd PhD³; Stein-Parbury, Jane RN PhD⁴ Self-management in older patients with chronic illness. *International Journal of Nursing Practice:Volume 14(5)October 2008p 373-382*.
29. Ganz PA. Breast cancer, menopause, and long-term survivorship: critical issues for the 21st century. *Am J Med* 2005;118(Suppl. 12B):136–141.
30. Giese-Davis J, Bliss-Isberg C, Carson K, Star P, Donaghy J, Cordova MJ *et al*. The Effect of Peer Counseling on Quality of Life Following Diagnosis of Breast Cancer: An Observational Study. [References]. *Psycho-Oncology* 2006;15:1014-22.
31. Gotay CC and Muraoka MY (1996) Quality of Life in Long-Term Survivors of Adult-Onset Cancers. *Cancer Nursing:Volume 19(1)February 1996pp 1-7*.
32. Hall W, Violato C, Lewkonja, R, Lockyer J, Fidler H, Toews J, Jennett P, Donoff M, Moores D, Assessment of physician performance in Alberta: the Physician Achievement Review Project, *Canadian Medical Association Journal*, 1999; 161:52-57.
33. Hibbard JH, Stockard J, Mahoney ER, Tusler M. (2004) Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. *Health Service Research*. Aug; 39(4 Pt 1):1005-26.
34. Hibbard JH, Mahoney ER, Stockard J, Tusler M. (2005)_Development and testing of a short form of the patient activation measure. *Health Service Research*. Dec; 40(6 Pt 1):1918-30.
35. Hibbard jh and peter j. Cunningham How Engaged Are Consumers in Their Health and Health Care, and Why Does It Matter? NO. 8, Center for Studying Health System Change, Oct 2008.

36. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Lo SK, Wain G. *Psycho-Oncology* 2006; published on-line 19 Dec ahead of print.
37. Holland JC and Reznik I (2005) Pathways for psychosocial care of cancer survivors. *Cancer*. 1, 104, 11, 2624-37.
38. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Lo SK, Wain G. (2006). The development and evaluation of a measure to assess cancer survivors' unmet needs: The CaSUN (Cancer Survivors' Unmet Needs Measure). *Psycho-Oncology*. vol. 16, 1 – 9.
39. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Wain G. Breast cancer survivors' supportive care needs 2–10 years after diagnosis. *Supportive Care in Cancer*.
40. Hodgkinson K, Butow P, Fuchs A, Hunt GE, Stenlake A, Hobbs KM, Brand A, Wain G. Long-term survival from gynecologic cancer: Psychosocial outcomes, supportive care needs and positive outcomes *Gynecologic Oncology* 2007; Volume 104, Issue 2, Pages 381-389.
41. Hodgkinson K, Butow P, Hunt GE, Wyse R, Hobbs KM, Wain G. Life after cancer: couples' and partners' psychological adjustment and supportive care needs. *Supportive Care in Cancer* 2007; Volume 15, Number 4.
42. Holroyd Kenneth A, Creer Thomas L. *Self-management of chronic disease: handbook of clinical interventions and research*. Orlando (FL): Academic Press;1986.
43. Jacobsen PB, Meade CD, Stein KD, Chirikos TN, Small BJ and Ruckdeschel JC (2002) Efficacy and costs of two forms of stress management training for cancer patients undergoing chemotherapy *Journal of Clinical Oncology* 20:12:2851-2862.
44. Khaw KT, Jakes R, Bingham S, Welch A, Luben R, Day N, Wareham N. Work and leisure time physical activity assessed using a simple, pragmatic, validated questionnaire and incident cardiovascular and all-cause mortality in men and women: The European Prospective Investigation into Cancer in Norfolk prospective population study. *Int J Epidemiol*. (2006, in press).
45. Kidd L (2007) An exploration of patients' perceived control, self efficacy and involvement in self care during chemotherapy for colorectal cancer, thesis, <https://dspace.stir.ac.uk/dspace/bitstream/1893/307/1/LK%20Thesis%2007.pdf>.
46. Kim SS. Fertility preservation in female cancer patients: current developments and future directions. *Fertil Steril* 2006;85(1):1–11.
47. King's College London, Macmillan Cancer Support and National Cancer Intelligence Network, *Cancer Prevalence in the UK, 2008*.

48. Krein SL, Heisler M, Piette JD, Butchart A, Kerr EA. Overcoming the influence of chronic pain on older patients' difficulty with recommended self-management activities. *Gerontologist*. 2007 Feb;47(1):61-8.
49. Lev EL, Owen SV: A measure of self-care self-efficacy. *Res Nurs Health* 1996, 19:421-429.
50. Lev EL, Paul D and Owen SV (1999) Age, self efficacy and change in patients' adjustment to cancer *Cancer Practice* 7:4:170-176.
51. Lev EL, Eller LS, Gejerman G, Lane P, Owen SV, White M and Nganga N (2004) Quality of life of men treated with brachytherapies for prostate cancer *Health and Quality of Life Outcomes* 2:28 ><http://www.hqlo.com/content/2/1/28>>, last accessed March 2007.
52. Lim JW and Zebrack B. Social networks and quality of life for long-term survivors of leukemia and lymphoma. *Support Care Cancer* (2006) 14: 185–192.
53. Lockyer J, Multi source feedback in the assessment of physician competencies. *Journal of Continuing Education in the Health Professions*, 2003; 23(1): 23: 4-12.
54. Lockyer JM, Violato C, Fidler H, A multisource feedback program for anesthesiologists (Un programme de rétroaction multisources pour les anesthésiologistes), *Can J of Anesthesia*, 2006, 53:33-39.
55. Loge JH, Abrahamsen AF, Ekeberg, Kaasa S. Fatigue and psychiatric morbidity among Hodgkin's disease survivors. *J Pain Symptom Manage* 2000;19(2):91–99.
56. Lorig KR, Sobel, DS, Ritter PL, Laurent, D, Hobbs, M. Effect of a self-management program for patients with chronic disease. *Effective Clinical Practice*, 4, 2001, pp. 256-262.
57. Lorig KR, Mazonson PD, Holman HR. Evidence suggesting that health education for self-management in patient with chronic arthritis has sustained health benefits while reducing health care costs. *Arthritis and Rheumatism* 1993;36:439-46.
58. Lorig K, Stewart A, Ritter P, González V, Laurent D and Lynch J. Outcome measures for health education and other health care interventions. Thousand Oaks, CA: Sage Publications; 1996.

59. Lorig K, Gonzalez V, Laurent D. The chronic disease self-management course leader's manual. Stanford (CA): Stanford University; 1997.
60. Lorig K, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, González VM, Laurent DD, Holman HR. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. *Med Care* 1999; 37(1):5-14.
61. Macmillan Cancer Support (2008) Two Million Reasons: The Cancer Survivorship Agenda. Macmillan Cancer Support, September 2008.
62. Maduro JH, Pras E, Willemse PH, de Vries EG (2003) Acute and long-term toxicity following radiotherapy alone or in combination with chemotherapy for locally advanced cervical cancer. *Cancer Treatment Review*. 29, 6, 471-88.
63. Manne SL, Ostroff JS, Norton TR, Fox K, Grana G and Goldstein L (2006) Cancer-specific self efficacy and psychosocial and functional adaptation to early stage breast cancer *Annals of Behavioural Medicine* 31:2:145-154.
64. Merluzzi TV, Martinez Sanchez M. Assessment of self efficacy and coping with cancer: development and validation of the Cancer Behaviour Inventory. *Health Psychol* 1997;16:163–70.
65. Merluzzi TV, Nairn RC, Hedge K, Martinez Sanchez MA and Dunn L (2001) Self efficacy for coping with cancer: revision of the Cancer Behaviour Inventory (Version 2.0) *Psycho-Oncology* 10:206-217.
66. McCormick C, Hooper J, Harrison R, Timpson C, Pearce D, Tocque K. (2006) Lifestyle surveys core questions and methods. Northwest Public Health Observatory, 2006. Penson DF, Litwin MS. Quality of life after treatment for prostate cancer. *Curr Urol Rep*. 2003;4(3):185–195.
67. National Institute for Health and Clinical Excellence. Four commonly used methods to increase physical activity: brief interventions in primary care, exercise referral schemes, pedometers and community-based exercise programmes for walking and cycling – Public health Intervention Guidance no.2. London: National Institute for Health and Clinical Excellence; 2006.
68. Nolte S., Elsworth G.R., Sinclair A.J., Osborne R.H. The extent and breadth of benefits from participating in chronic disease self-management courses: A national patient-reported outcomes survey (2007) *Patient Education and Counseling*, 65 (3), pp. 351-360.
69. O'Leary A. Self-efficacy and health. *Behaviour Research and Therapy* 1985;23:437-51.

70. Passik SD, Kirsh KL, Donaghy K, Holtsclaw E, Theobald D, Cella D, Breitbart W, for the Fatigue Coalition. Patient-Related Barriers to Fatigue Communication: Initial Validation of the Fatigue Management Barriers Questionnaire. *Journal of Pain and Symptom Management* - November 2002 (Vol. 24, Issue 5, Pages 481-493).
71. Pearce NJM, Rob Sanson-Fisher, H. Sharon Campbell Measuring quality of life in cancer survivors: a methodological review of existing scales. *Psycho-Oncology*, Volume 17, Issue 7, Pages 629 – 640
72. Pearl-Kraus, Lorraine Effects of two Parkinson's disease patient education models on disease self-management and quality of life. Dissertation, <http://gradworks.umi.com/32/65/3265921.html>.
73. Pierce JP, Stefanick ML, Flatt SW, Natarajan L, Sternfeld B, Madlensky L, Al-Delaimy WK, Thomson CA, Kealey S, Hajek R, Parker BA, Newman VA, Caan B, Rock CL. Greater survival after breast cancer in physically active women with high vegetablefruit intake regardless of obesity. *J Clin Oncol* 2007;25:2345-2351.
74. Ream E, Richardson A and Alexander-Dann C (2006) Supportive intervention for fatigue in patients undergoing chemotherapy: a randomised controlled trial *Journal of Pain and Symptom Management* 31:2:148-161.
75. Ritter, P; Kate Lorig; Diana Laurent; Katy Matthews. Internet Versus Mailed Questionnaires: A Randomized Comparison. (*J Med Internet Res* 2004;6(3):e29).
76. Rodriguez, Carmen, Jacobs, Eric J., Mondul, Alison M., Calle, Eugenia E., McCullough, Marjorie L., Thun, Michael J. Vitamin E Supplements and Risk of Prostate Cancer in U.S. Men. *Cancer Epidemiol Biomarkers Prev* 2004 13: 378-382
77. Schag CA, Ganz PA, Wing DS, Sim MS, Lee JJ (1994) Quality of life in adult survivors of lung, colon and prostate cancer. *Quality of Life Research*. 3, 2, 127-41.
78. Schofield P, Jefford M, Carey M, Thomson K, Evans M, Baravelli C, Aranda S. Preparing patients for threatening medical treatments: effects of a chemotherapy educational DVD on anxiety, unmet needs, and self-efficacy. *Support Care Cancer* (2008) 16:37–45.
79. Sherwood P, Given BA, Given CW, Champion VL, Doorenbos AZ, Azzouz F, Kozachik S, Wagler ZK, Monahan PO. A cognitive behavioral intervention for symptom management in patients with advanced cancer. *Oncology-Nursing-Forum* (2005) 32 : 1190-1198.
80. Sieving R, Resnick MD, Bearinger L, Remafedi G, Taylor BA, Harmon B. Cognitive and behavioral predictors of sexually transmitted disease risk behavior among sexually active adolescents. *Archives of Pediatrics and Adolescent Medicine* 1997;151:243-51.

81. Smith SC, Cano S, Lamping DL, Staniszewska S, Browne J, Lewsey J, van der Meulen J, Cairns J, Black N. Patient-reported outcome measures (PROMs) for routine use in treatment centres: recommendations based on a review of the scientific evidence. London School of Hygiene and Tropic Medicine, London, UK, and the Department of Health, 2004.
82. Tannock IF, Ahles TA, Ganz PA, Van Dam FS. Cognitive impairment associated with chemotherapy for cancer: report of a workshop. *J Clin Oncol* 2004;22(11): 2233–2239.
83. Telch CF and Telch MJ (1986) Group coping skills instruction and supportive group therapy for cancer patients: a comparison of strategies *Journal of Consulting and Clinical Psychology* 54:6:802-808.
84. Telford K, Kralik D and Koch T (2006) Acceptance and denial: implications for people adapting to chronic illness: literature review *Journal of Advanced Nursing* 55:4:457-464.
85. Thewes B, Butow P, Girgis A, Pendlebury S. Assessment of unmet needs among survivors of breast cancer. *J Psychosoc Oncol* 2004;22(1):51–71.
86. Thomas SF, Glynne-Jones R, Chait I, Marks DF (1997) Anxiety in long-term cancer survivors influences the acceptability of planned discharge from follow-up. *Psycho-oncology*. 6, 3, 190-6.
87. Thornton A. Perceiving benefits in the cancer experience. *J Clin Psychol Med Settings* 2002;9(2):153–165.
88. Trudeau, M. E., Lickley, L., Narod, S., Mergler, S., Hawker, G., Sawka, C., Lee, D. Ten-year breast cancer survival - at what cost in human and physical terms? A case-control study of long-term survivors of breast cancer and their physical and functional well-being. *J Clin Oncol (Meeting Abstracts)* 2004 22: 8010.
89. Wareham NJ, Jakes RW, Renni KL, Schuit J, Mitchell J, Hennings S, Day NE. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition* 2002; 6:407-413.
90. Watson PB, Town GI, Holbrook N, Dwan C, Toop LJ, Drennan CJ. Evaluation of a self-management plan for chronic obstructive pulmonary disease. *European Respiratory Journal* 1997;10:1267-71.
91. Weber BA, Roberts BL, Resnick M, Deimling G, Zauszniewski J, Musil C and Yarandi HN (2004) The effects of dyadic intervention on self efficacy, social support and depression for men with prostate cancer *Psycho-Oncology* 13:47-60.

92. Weiss T (2004) Correlates of posttraumatic growth in husbands of breast cancer survivors. *Psycho-oncology*. 13, 4, 260-8.
93. Wenzel L, DeAlba I, Habbal R, Kluhsman BC, Fairclough D, Krebs LU, Anton-Culver H, Berkowitz R, Aziz N. Quality of life in long-term cervical cancer survivors. *Gynecologic Oncology* Volume 97, Issue 2, May 2005, Pages 310-317.
94. Wenzel L, Dogan-Ates A, Habbal R, Berkowitz R, Goldstein DP, Bernstein M, Coffey B, Kluhsman, Osann K, Newlands E, Seckl MJ, Hancock B, Cella D. defining and Measuring Reproductive Concerns of Female Cancer Survivors *JNCI Monographs* 2005 2005(34):94-98; doi:10.1093/jncimonographs/lgi017.
95. Wilson P. The UK Expert Patients Program: Lessons learned and implications for cancer survivors' self-care support programs . *Journal of Cancer Survivorship Research and Practice*, 2007.
96. Winzelberg AJ, Catherine Classen, Georg W. Alpers, Heidi Roberts, Cheryl Koopman, Robert E. Adams, Heidemarie Ernst, Parvati Dev, C. Barr Taylor. Evaluation of an internet support group for women with primary breast cancer. *Cancer*; 97(5), PP. 1164-1173, 2003
97. Wolf MS, Chang C-H, Davis T, Makoul G. (2005) Development and validation of the Communication and Attitudinal Self-Efficacy scale for cancer (CASE-cancer). *Patient Education and Counselling*, Volume 57, Issue 3, Pages 333-341 (June 2005).
98. Wyatt G, Kurtz ME, Freidman LL, Given B, Given C. Preliminary testing of the Long-Term Quality of Life (LTQL) Instrument for female cancer survivors. *J Nurs Meas* 1996;4(2):153–170.
99. Zebrack BJ, Ganz P, Bernaards CA, Petersen L, Abraham L. Assessing the impact of cancer: development of a new instrument for long-term survivors. *Psycho-Oncology* 2006;15(5):407–421.
100. Zebrack BJ, Ganz PA, Bernaards CA, Petersen L, Abraham L. (2007) Assessing the impact of cancer: development of a new instrument for long-term survivors. *Psycho-Oncology*, 17(9), pp. 891-900.