Patient-Reported

OUTCOME

Measures

PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

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

Report to the Department of Health, 2010

UNIVERSITY OF OXFORD

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR WOMEN WITH BREAST CANCER

Nicola Davies
Elizabeth Gibbons
Ray Fitzpatrick
Anne Mackintosh

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

May 2010

CONTENTS

EXECUTIVE SUMMARY

Aims of the report

The aims of this report are to review the evidence of Patient-reported Outcome Measure (PROMs) for people with breast cancer and provide a short-list of the most promising generic and cancer-specific instruments.

The methods of the review are described and the results of the search, including sources and search terms used to identify specific published research. Details of this evidence are presented firstly for generic PROMs evaluated with people with breast cancer, followed by condition-specific PROMs. The report concludes with discussion and recommendations.

Results

Three generic instruments, which have been evaluated with breast cancer, were identified in this review:
1. Medical Outcomes Study 36-Item Health Survey (SF-36)
2. Medical Outcomes Study 8-Item Health Survey (SF-8)
3. Sickness Impact Profile (SIP)

One preference-based measure was identified
1. European Quality of Life Questionnaire (EuroQol; EQ-5D)

Six cancer-specific instruments were identified in this review; two are specific to breast cancer (EORTC BR23, FACT-B):
1. Cancer Rehabilitation Evaluation System – Short-Form (CARES-SF)
2. European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
3. EORTC BR23
4. Functional Assessment of Cancer Therapy – General (FACT-G)
5. Functional Assessment of Cancer Therapy – Breast (FACT-B)
6. Functional Living Index – Cancer (FLIC)

Recommendations

Based on the volume of evaluations and good measurement and operational characteristics, the following are highlighted as promising PROMs for piloting in the NHS:

Generic:
    SF-36
Preference-based:
    EQ-5D as a preference based measure

Cancer-specific:
1. EORTC QLQ-C30
2. FACT-B or FACT-G

Attention may need to be given to longer term issues of survivorship if the full spectrum of cancer in the population is to be included in surveys via PROMs although insufficient evidence was found to high-light measures for this review.

INTRODUCTION

Background

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

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

In light of recent policy to include PROMs as an important quality indicator, the Department of Health now seeks guidance on PROMs which can be applied in patients with cancer and commissioned the Patient-reported Outcome Measurement Group, Oxford, to review the evidence of PROMs for selected cancers. It is proposed that the most common cancers, as identified via the Office for National Statistics, should be the subject of review in terms of most promising PROMS. Breast, lung, colorectal and prostate cancer are highlighted as being the four most common cancers, accounting for half of the 239,000 new cases of malignant cancer (excluding non-melanoma skin cancer) registered in England in 2005 (Figure 1). On scrutinising cumulative incidence data from the cancer registry of the Oxford region, findings support that these four cancers are the most common. According to the Department of Health's Cancer Reform Strategy (2007), which aims to place the patient at the centre of cancer services, a 'vision 2012' has been created for each of these four cancer types, highlighting the progress that it is hoped will be made by 2012 in terms of the cancer pathway. Underlying these visions are the aims to achieve full implementation of improving outcomes guidance. In this context, PROMS are an important resource to monitor cancer outcomes.

Figure 1: Incidence of the major cancers, 2005, England (ONS, 2007)
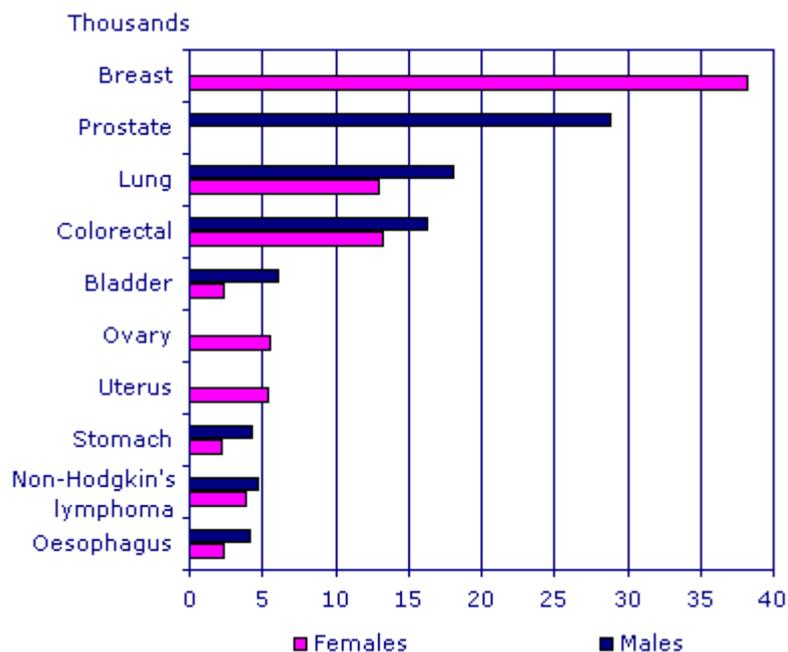
Breast cancer

Breast cancer is the most common cancer in England (Figure 1); one in nine women will develop breast cancer at some point in their lives. In 2004 there were around 36,900 new cases diagnosed, representing 32% of all cancers in women at a rate of 121 cases per 100,000. Four in five new cases are diagnosed in women aged 50 and over, peaking in the 55 to 64 year age-group. Incidence increased by 81% between 1971 and 2004, and by 13% in the ten years to 2004. Earlier detection and improved treatment have resulted in a rise in survival rates; five-year survival was 81% for women diagnosed in 1999-2003 in England. Survival from breast cancer is higher than that for cervical cancer and much higher than for the other major cancers in women - lung, colorectal and ovarian. For women diagnosed in 2001-03, 72% are likely to survive for at least ten years.

Breast cancer is thus a priority on the government health agenda and has been for some time. 'Improving Outcomes in Breast Cancer' (1996, Department of Health) identified healthcare professionals' roles in the treatment, management and care of women with breast cancer. Recommendations focused on how these services should be organised so that women with breast cancer across England and Wales receive high-quality healthcare. The National Institute of Health and Clinical Excellence have since published an updated version of this document (NICE, 2002). Collecting and using improved information on different aspects of cancer services and outcomes is central to delivering this strategy. Better information via patient-reported outcomes could enhance quality of care, inform commissioning, and promote patient choice. The following review provides current information available on PROMs used with breast cancer patients.

METHODS AND SEARCH STRATEGY
Aim of the report

The aim of this report is to identify Patient-reported Outcome Measures (PROMs) which have been evaluated with patients with breast cancer.

Structure of the report

The methods of the review are described, including search strategies and search terms used to identify relevant published research regarding PROMs for people with breast cancer. Details of this evidence are presented for generic PROMs, Preference-based measures and cancer-specific instruments. The report concludes with discussion and recommendations.

Methods for the review

a) Inclusion criteria
Titles and abstracts of all articles were assessed for inclusion/exclusion by one reviewer and a selection agreement was checked by another reviewer. Included articles were retrieved in full. Published articles were included if they provided evidence of measurement and/or practical properties (Fitzpatrick et al., 1998) for multi-item instruments assessing aspects of health status or quality of life in women with breast cancer.

-Study design selection
- studies where a principal PROM is being evaluated;
- studies evaluating several PROMs concurrently;
- applications of PROMs with sufficient reporting of methodological issues.

-Specific inclusion criteria for generic and disease-specific instruments
- the instrument is patient-reported;
- there is published evidence of measurement reliability, validity or responsiveness following completion in the specified patient population;
- the instrument has been recommended for use with patients with breast cancer;
- evidence is available from English language publications, and instrument evaluations conducted in populations within UK, North America, Australasia.

b) Exclusion criteria
- Clinician-assessed instruments

Comprehensive searches were conducted; articles retrieved were assessed for relevance and checked by another reviewer; and evidence of measurement performance and operational characteristics abstracted for each PROM identified.

c) Search terms and results: identification of articles
The searches were conducted using three main sources.

The primary source of evidence was the bibliographic database compiled by the PROM group in 2002 with funding from the Department of Health and hosted by the University of Oxford. In 2005, it became the property of the NHS Information Centre for Health & Social Care. The most recent bibliographic update is current to December 2005. The PROM database comprises 16,054 records (up to December 2005) downloaded from several electronic databases using a comprehensive search strategy (Appendix A). These records had been assessed as eligible for inclusion in the bibliography and assigned keywords. The primary search strategy using the term 'cancer' in the keyword search generated 17,759 records, 272 of which were for breast cancer.

There are also 14,000+ additional records (2006/2007) held by the PROMs group. The primary search strategy using the term 'cancer' and 'breast' or 'breast cancer' in the abstract and title search generated 402 records.

Supplementary searchers which included hand searching of titles from 2006 to 2008 of the following key journals, and PubMed, generated 43 records:
- Cancer
- Journal of Clinical Oncology
- Health and Quality of Life Outcomes
- Medical Care
- Quality of Life Research

When assessed against the review inclusion criteria, 96 articles were included in the review..

Table 1: Number of articles identified by the literature review

Sources	No. studies identified	No. studies included
Bibliography (up to August 2007)	674	52
Supplementary searches		44
TOTAL		**96**

d) Data extraction
Data were extracted on the psychometric performance and operational characteristics of each PROM. Assessment and evaluation of the methodological quality of PROMs was performed independently by two reviewers adapting the London School of Hygiene appraisal criteria outlined in their review (Smith et al., 2005). These criteria were modified for our review (Appendix B).

The final short-listing of promising PROMs to formulate recommendations is based on these assessments and discussion between reviewers.

RESULTS: Generic PROMs evaluated in women with breast cancer

Three generic instruments were identified, which have been evaluated in breast cancer:

1. Medical Outcomes Study 36-Item Health Survey (SF-36)
2. Medical Outcomes Study 8-Item Health Survey (SF-8)
3. Sickness Impact Profile (SIP)

These instruments are briefly described in Appendix C.

1. Medical Outcome Short Form 36-Item Health Survey (SF-36)
A total of 35 articles evaluating the SF-36 with breast cancer patients were included.

High internal consistency of the eight SF-36 subscales is supported by a number of the included studies, with Cronbach's alpha in the range of 0.68 to 0.94 (Bardwell et al., 2006; Byar et al., 2006; Wyatt et al., 2004; Figueiredo et al., 2004; Helgeson and Tomich, 2005; Alfano et al., 2006a & b; Doorenbos et al., 2006). Further support for the internal consistency of the SF-36 has been demonstrated for the MCS and PCS scores, alpha's being 0.89 and 0.94, respectively (Golden-Kreutz et al., 2005). However, inter-item correlations failed to meet the accepted correlation of 0.35 or greater for 20 out of 82 comparisons in one study (Glick et al., 1998).

The SF-36 MCS demonstrated strong convergent validity with the Centre for Epidemiologic Studies Scale Screening Form in Bardwell et al. (2006) and scores discriminated cancer-related traumatic stress (Golden-Kreutz et al., 2005).

Moderate to strong convergent validity has been demonstrated with the Health Utilities Index (Lovrics et al., 2008). Moderate convergent validity has been demonstrated with the Impact of Cancer scale (Zebrack et al., 2006). Moderate to weak convergent validity has been demonstrated with the hormone-related symptoms subscale of the Breast Cancer Prevention Trial (BCPT) Symptom Checklist (Alfano et al., 2006a). Weak convergent validity was demonstrated with the Brief Cancer Impact Assessment (Alfano et al., 2006b), the FLIC (Wilson et al., 2005), and the EORTC QLQ-C30 (Glick et al., 1998).

Discriminant validity has been supported for the SF-36 with differences in scores for cancer type (Claus et al., 2006), breast cancer recurrence and disease-free status (Helgeson and Tomich, 2005), co-morbidities and treatment type (Byar et al., 2006; Bower et al., 2006; Figueiredo et al., 2004; Griggs et al., 2007; Purushotham et al., 2005; Muss et al., 2008). Patients' scores on the SF-36 differed significantly for depressive status as measured via the CES-D (Bardwell et al., 2006; Doorenbos et al., 2006) and cancer-related fatigue (Byar et al., 2006; Andrykowski et al., 2005).

As predicted, the SF-36 Energy/Vitality subscale discriminated according to natural killer cell numbers in an Irish study (Garland et al., 2004). However, the SF-36 has demonstrated weak discriminative power with regard to emotional well-being, failing to demonstrate mental health differences between women with or without lymphoedema secondary to breast cancer (Wilson et al., 2005), as well as breast

cancer patients and healthy post-menopausal women (Yost et al., 2005), breast cancer patients and USA normative data (Bardwell et al., 2004).

The SF-36 has demonstrated responsiveness in a number of intervention studies consistent with other outcome measures in the studies. These include evaluations of surgery (Lovrics et al., 2008), an educational support group intervention (Helgeson and Tomich, 2005), a self-efficacy enhancing intervention (Doorenbos et al., 2006), in-home nursing (Wyatt et al., 2004) and different drug therapies (Muss et al., 2008).

Further support for responsiveness to change has been demonstrated in the PCS and MCS scores at 4 and 12 months post-surgery (Golden-Kreutz et al., 2005).

Two studies reported no change in scores in women with unilateral breast cancer-associated lymphoedema of the arm receiving a massage intervention in patients (Wilburn et al., 2006) or specialist follow-up care (Grunfeld et al., 2006).

Six of the included studies explicitly commented on patient acceptability of the SF-36, one reporting completion time as being less than 10 minutes (Byar et al., 2006) and the other five reporting response rates ranging from 54% to 100% (Zebrack et al., 2006; Bower et al., 2006; Lovrics et al., 2008; Muss et al., 2008; Alfano et al., 2006a & b).

2. Medical Outcome Short Form 8-Item Health Survey (SF-8)
One study evaluating the SF-8 with breast cancer patients was identified.

The SF-8 PCS score has demonstrated convergent validity with the Patient Health Questionnaire (Hegel et al., 2006). The SF-8 did not differentiate newly diagnosed breast cancer patients from US norms (Hegel et al., 2006).

The SF-8 is still in its infancy and thus there is very little evidence supporting its psychometric properties with breast cancer patients.

3. Sickness Impact Profile (SIP)
A total of two articles evaluating the SIP with breast cancer patients were identified.

The home management subscale and recreation and pastimes subscale have demonstrated good internal consistency at three time points, Cronbach's alpha being in the range of 0.63-0.77 and 0.62-0.76, respectively (Allard, 2007).

The subscales measuring adverse impact of illness or its treatment on social activities and recreational activities only had a weak correlation with emotional distress as measured via the Affect Balance Scale in a study by Carver et al. (1998).

The home management and recreation and pastime subscale scores were responsiveness to change in women receiving an Attentional Focus and Symptom Management Intervention (AFSMI) intervention, whereby a significant time effect was identified between time 2 (one week following the first intervention session) and time 3 (one week following the second intervention session) (Allard, 2007).

A response rate of 79% was reported in one study utilising the SIP via telephone interview (Allard, 2007).

Preference-based measures
One measure was identified:

1. European Quality of Life Questionnaire (EuroQol; EQ-5D)
Three articles evaluating the EQ-5D with breast cancer patients were included in the review.

Convergent validity has been demonstrated with the FLIC and a VAS scale measuring perceived QoL (Conner-Spady et al., 2001).

The EQ-5D is responsive to change with significantly large change scores in patients receiving high dose chemotherapy (Conner-Spady et al., 2001) and to those receiving screening interventions (Jeruss et al., 2006).

One study has interpreted EQ-5D scores via conventional interpretations of effect size of small (0.2), medium (0.5), and large (0.8) (Conner-Spady et al., 2001), whilst another has used ECOG Performance Status as an anchor (Pickard et al., 2007). In the latter study, minimal important differences (MIDs) for the overall cohort (n=534), which included breast cancer patients, were 0.10 to 0.12 for EQ-5D dimensions and 7 to 10 for VAS scores.

Ceiling effects have been reported in stage II and III breast cancer patients (n=40) undergoing high dose chemotherapy (Conner-Spady et al., 2001), but not in a cohort of mixed cancer patients, 50 of whom had breast cancer (Pickard et al., 2007).

RESULTS: Cancer-specific PROMs in breast cancer

Six cancer-specific instruments were identified in this review; two are specific to breast cancer (EORTC BR23, FACT-B):

1. Cancer Rehabilitation Evaluation System – Short-Form (CARES-SF)
2. European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
3. EORTC BR23
4. Functional Assessment of Cancer Therapy – General (FACT-G)
5. Functional Assessment of Cancer Therapy – Breast (FACT-B)
6. Functional Living Index – Cancer (FLIC)

These instruments are briefly described in Appendix D.

1. Cancer Rehabilitation Evaluation System – Short Form (CARES-SF)

A total of six articles evaluating the CARES-SF are included in the review.

Test-retest reliability has been demonstrated in one study, with 86% agreement between test and retest at one, seven, and 13 months post-diagnosis (Schag et al., 1991).

The CARES-SF has acceptable internal consistency, supported by three studies (Schag et al., 1991; Dausch et al., 2004; Shelby et al., 2006), with alphas for subscales ranging between 0.61 and 0.85.

Convergent validity with the long-form CARES was high, correlations ranging from 0.90 to 0.98 on the global score and five sub-scores (Schag et al., 1991). The CARES-SF global and five subscale scores had strong correlations with the Functional Living Index – Cancer (FLIC) in two studies of newly diagnosed breast cancer patients (Ganz et al., 1990; Schag et al., 1991). Further evidence of convergent validity was demonstrated with all CARES scores significantly correlating with the clinician-rated Karnofsky Performance Status. As was expected, the physical CARES summary scale showed the strongest correlation with this instrument.

The CARES-SF has demonstrated discriminative validity in a number of studies, with scores discriminating between women (n=216) receiving various types of treatment, including chemotherapy, hormone therapy and systematic adjuvant therapy (Casso et al., 2004). The instrument scores have also demonstrated discriminant validity for women reporting physical problems and women reporting difficulty communicating with friends and relatives (Shelby et al., 2006). However, discriminative validity has been lacking in terms of self-reported psychological problems (Shelby et al., 2006) and depression (Dausch et al., 2004) in women with breast cancer.

The CARES-SF has demonstrated responsiveness over time, with the Global CARES-SF scores significantly improving from 1 month to 7 months, 1 month to 13 months, and 7 months to 13 months post-surgery (Schag et al., 1991). Further support for the responsiveness of the CARES-SF has been demonstrated in a weight training

intervention, with significant changes on the physical and psychosocial subscale scores at 6 months post intervention (Ohira et al., 2006).

High response rates of 75% for a postal survey including the CARE-SF were reported in one study (Casso et al., 2004).

2. European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
A total of 16 articles evaluating the QLQ-C30 are included in the review.

The QLQ-C30 has demonstrated acceptable internal consistency with Cronbach's alpha levels above 0.70 for all domains except role function and cognitive function (Osoba et al., 1994). Further evidence is reported for item level analysis with item-total correlations being above 0.35 on all but three of the comparisons made (Glick et al., 1998). In terms of item discrimination, overlaps are evident between the social function and cognitive function scales, the emotional function and cognitive function scales, and the physical function, role function, and social function scales, alphas for combined scales ranging from 0.51 to 0.87. Except for the relatively low correlation between item 5 (whether the responder needed 'help with eating, dressing, washing yourself or using the toilet') and the physical function domain, the correlations for all other items were much higher within their own domain than with any other domain (ranging from -0.65 to 0.95) (Osoba et al., 1994). The QLQ-C30 has met accepted internal consistency and item discrimination standards, above that of the generic SF-36 (Glick et al., 1998).

Using principal components factor analysis, two factors were found from the 12 psychosocial items, namely, emotional distress and functional ability (McLachlan et al., 1999). Analysis of the factor structure via an orthogonal varimax transformation showed reasonably good agreement with the postulated factor structure (Osoba et al., 1994). The main difference was that items 1 and 5 pertaining to physical function did not load with the other items (2, 3 and 4) on this factor at baseline, pre-chemotherapy (McLachlan et al., 1999).

Convergent validity has been reported for the QLQ-C30 with a number of similar conceptually analogous PROMs. The QLQ-C30 has demonstrated reasonable convergent validity with the FACT-G (Glick et al., 1998) and the Supportive Care Needs Survey (Snyder et al., 2008). Further support for convergent validity is reported for the psychosocial subscale with several subscales from other PROMs (Psychosocial Adjustment to Illness Scale [PAIS], the Profile of Mood States [POMS], the Mental Adjustment to Cancer [MAC] scale, and the Impact of Events Scale [IES]), with correlations ranging from 0.52 (PAIS) to 0.74 (POMS) (McLachlan et al., 1998). There were weak to moderate correlations between QLQ-C30 emotional function and MAC fighting spirit, hopeless/helpless and anxious preoccupation. Furthermore, QLQ-C30 social function did not relate substantially to MAC coping style.

With the IES, the only quantitatively significant correlation occurred between QLQ-C30 emotional function and IES intrusion. Agreement was highest for the QLQ-C30 with the POMS and the PAIS, but moderate with the IES, and lowest for the MAC (McLachlan et al., 1998).

The QLQ-C30 global health/QoL, role function, and social function subscales show evidence of discriminant validity between known groups identified via the clinician-rated Eastern Cooperative Oncology Group (ECOG) Performance Status, but the emotional, cognitive, and psychosocial function subscales did not demonstrate discriminant ability with reference to all anticipated clinical parameters (McLachlan et al., 1998). The QLQ-C30 has also demonstrated discriminant validity in terms of cancer type (including breast cancer), disease severity (Osoba et al., 1994), DSM-IV depression status, age (Grabsch et al., 2006), widowed status (Hack et al., 2006), and supportive care needs (Snyder et al., 2008). The instrument has also demonstrated discriminative ability to disease progression (Lemieux et al., 2007).

Evidence of responsiveness of the QLQ-C30 is reported in a longitudinal study of chemotherapy, where the only domain that did not change significantly between pre- and post-treatment was pain (Osoba et al., 1994).

Minimally Important Differences (MIDs) have been reported for the QLQ-C30 (Osoba et al., 1998). Effect sizes corresponding to a small improvement were between 0.09 and 0.51 for subscales (Osoba et al., 1998). The MID using 0.5 SD was confirmed by Lemieux et al., (2007) in metastasized breast cancer (n=133).

Floor effects for items 5 (help with eating, dressing, washing yourself or using the toilet), 14 (nausea), 15 (vomiting) and 17 (diarrhoea) are reported in Osoba et al. (1994), indicating lower functioning and poorer HRQoL for patients pre-chemotherapy. No floor or ceiling effects are reported for items 14 and 15 post-chemotherapy. Ceiling effects were evident for all five psychosocial subscales in one study (McLachlan et al., 1998).

Acceptability of the QLQ-C30 is demonstrated in a number of studies reporting high response rates and low levels of missing data in both the pencil and pen response format (Aranda et al., 2005; Osoba et al., 1994) and a computerised response format (Taenzer et al., 1997). The average time required to complete the instrument is approximately 11 minutes, with most patients requiring no assistance (Aaronson et al., 1993).

The feasibility of using the QLQ-C30 to collect QoL data was supported by two of the studies reviewed (Coates et al., 1992), one utilising a computerised response format (Taenzer et al., 1997).

3. EORTC BR23

The EORTC BR23 has 23 items in two functional sub-scales (body image and sexual functioning), three symptom sub-scales (breast, arm and systemic side effects), and individual items covering sexual enjoyment, hair loss and future concerns. Very little evidence is reported for this module. One recent study used it to evaluate HRQoL in women following breast re-constructive surgery. The authors report that the EORTC BR23 and the FACT-B were not sensitive to detecting change in this population (Potter et al., 2009 UK).

4. Functional Assessment of Cancer Therapy – General Version (FACT-G)
A total of 13 articles evaluating the FACT-G are included in the review.

Evidence of acceptable reproducibility has been reported in a mixed sample of cancer patients (n=60) with ICC ranges from 0.82 to 0.92 with a test re-test period of 3 to 7 days (Cella et al., 1993).

Cronbach's alpha for the total FACT-G have been reported between 0.89 (Cella et al., 1993) and 0.94 (Romero et al., 2006), with the four subscales ranging from 0.65-0.87 (Cella et al., 1993; Smith et al., 2007; Davies et al., 2008a; Davies et al., 2008b). However, in a study demonstrating internal consistency via inter-item correlations, failure to meet the accepted 0.35 was evident on 31 of the 68 comparisons. The FACT-G fared better than the generic SF-36, but worse than the EORTC QLQ-C30 (Glick et al., 1998).

Factor analysis revealed five subscales to the FACT-G, namely, physical, social, emotional, and functional well-being, and relationship with doctor (Cella et al., 1993). A four-factor structure of the FACT-G was obtained in a UK study, these factors corresponding to version 4 subscales of physical, social, emotional, and functional well-being (Smith et al., 2007). However, in this same study, the social well-being subscale was identified as being multidimensional, suggesting a two-factor scale of family concerns and close personal relationships.

Convergent validity has been demonstrated with strong correlation between the FACT-G and the Functional Living Index – Cancer (FLIC) and measures of mood distress (Taylor Manifest Anxiety Scale, Brief Profile of Mood States). Moderately correlations are reported between the FACT-G and the Eastern Cooperative Oncology Group (ECOG) Performance Status self-report (Cella et al., 1993) and Quality of Life in Cancer Survivors (QOL-CS) (Ferrell et al., 1995), and reasonable with the EORTC QLQ-C30 and SF-36 (Glick et al., 1998). Subscale outcomes have also been found to correlate with personal measures of health status, as assessed via the Health Baseline Comparison Questionnaire (Davies et al., 2008b) and information satisfaction, as assessed via the Information Satisfaction Questionnaire (Davies et al., 2008a). Divergent validity has been supported by low correlations with the shortened Marlowe-Crowne Social Desirability Scale (Cella et al., 1993).

The FACT-G has demonstrated discriminant validity in terms of different stages of disease in a study by Cella et al. (1993) and chemotherapy patients and healthy controls (Mar Fan et al., 2005).

However, evidence has emerged in terms of differential item functioning (DIF), whereby items have been found to perform differently for different groups (i.e. race, ethnicity, education, language, self- vs. interviewer-administered) in a study by Crane et al. (2006).

The FACT-G has demonstrated responsiveness to change in scores from a group of patients receiving chemotherapy (Cella et al., 1993) as well showing time-dependent improvements post-chemotherapy at one and two year follow-up (Mar Fan et al., 2005). However, there were non-significant changes in FACT-G scores in patients

receiving an exercise rehabilitation programme, as measured at baseline and at 12-weeks (Campbell et al., 2005).

Distribution-based and anchor-based methods are available for identifying clinically meaningful differences. Normative data are available for the interpretation of FACT-G scores (Brucker et al., 2005) and the ECOG Performance Status is commonly used as an anchor for FACT-G interpretation (Cella et al., 2002).

Patient burden is minimal and the reading level of the questionnaire is 4[th]grade. Acceptability is supported in a longitudinal study (Mar Fan et al., 2005) and postal methodology (Ferrell et al., 1995), response rates being high in both cases.

Ease of administration and scoring has been supported during questionnaire development (Cella et al., 1993).

5. Functional Assessment of Cancer Therapy – Breast (FACT-B)
A total of 21 articles evaluating the FACT-B are included in the review.

Evidence of acceptable reproducibility has been reported for subscales and aggregate scores with ICC ranges of 0.88 to 0.85 for the breast cancer subscale and FACT-B total score, respectively, with a test re-test period of 3 to 7 days (Brady et al., 1997).

Internal consistency has been supported by a number of studies, with Cronbach's alphas for the entire FACT-B ranging from 0.82 to 0.91 (Hack et al., 2003; Wyatt et al., 2004) and subscales being in the range of 0.63-0.90 (Manning-Walsh, 2005 a); Hack et al., 2003; Cella et al., 1993)

The FACT-B general subscales retain the same content validity as the FACT-G (Cella et al., 1993). The total FACT-B was developed with an emphasis on patients' values, firstly being tested longitudinally (n=47) and then with a larger sample (n=295), supporting its use in oncology clinical trials as well as in clinical practice (Brady et al., 1997).

Evidence of convergent validity is demonstrated with significant correlations with conceptually similar measures, such as the FLIC (Brady et al., 1997). Further support of convergent validity is evident with hypothesised correlations between FACT-B well-being subscales and the POMS subscales (Depression/Tension/Anger, Vigour, and Fatigue). Weak correlations have been reported between the FACT-B and the Marlowe-Crowne Social Desirability Scale, indicating the expected lack of relationship between these differing concepts.

Discriminant validity has been demonstrated with scores differing in patient- and clinician-reported ECOG Performance Status (Brady et al., 1997) and according to chemotherapy status (Shilling and Jenkins, 2007) and use of herbal remedies (Zick et al., 2006). No discriminant validity was established between types of nursing care received (Wyatt et al., 2004). The ability of the FACT-B to discriminate according to demographic variables is mixed. In one study, the questionnaire discriminated by age, disease severity, and socio-economic status (Zick et al., 2006) whilst in another it did not discriminate in terms of age, time since diagnosis, or socioeconomic or marital

status (Manning-Walsh, 2005 b). The latter study is not necessarily generalisable since it was an Internet-based study.

The FACT-B has demonstrated responsiveness to change with significant improvement in scores in a number of intervention studies, including complete decongestive therapy for upper extremity lymphoedema (Mondry et al., 2004), in-home nursing (Wyatt et al., 2004), a dance and movement class at 13 and 26 weeks post-intervention (Sandel et al., 2005), an eight-month self-efficacy counselling intervention at four and eight months (Lev and Owen, 2000), and to a 12-week support-group intervention (Targ and Levine, 2002). Nevertheless, in the latter study, the significant score change was in terms of functional and emotional well-being domains, with no responsiveness from the social well-being domain, as would have been expected.

By combining the results of distribution- and anchor-based methods, MID estimates have been obtained, these being 2-3 points for the breast cancer module, and 7-8 points for the total FACT-B (Eton et al., 2004).

Patient burden is minimal and the reading level of the questionnaire is 4[th]grade. Acceptability is supported by high response rates in a number of the reviewed studies (Brady et al., 1997; Manning-Walsh, 2005a & b; Gordon et al., 2005).

Ease of administration and item brevity has been reported in the development of the FACT-B, as has the appropriateness of the instrument in clinical settings (Brady et al., 1997).

6. Functional Living Index – Cancer (FLIC)

A total of eight articles evaluating the FLIC are included in the review.

One study reports a Cronbach's alpha of 0.89 (Clinch, 1996), demonstrating good internal consistency.

Strong convergent validity has been reported between the FLIC and the FACT-G (Brady et al., 1997), ECOG Performance Status (Elliott et al., 2004), CARES (Ganz et al, 1990), and EQ-5D (Conner-Spady et al., 2001), but weak convergent validity with the generic SF-36 (Wilson et al., 2005).

The FLIC has failed to demonstrate discriminant validity for nodal status, receipt of chemotherapy, and type of surgery (Ganz et al., 1990) and the presence of other health conditions (Elliott et al., 2004.) However, it has demonstrated discriminant validity in terms of concurrent illness, ECOG Performance Status (Elliot et al., 2004) and lymphoedema status (Wilson et al., 2005).

FLIC scores have been found to be responsive to change in patients receiving high dose chemotherapy (Conner-Spady et al., 2001).

To the best of our knowledge and that of other researchers (Wilson et al., 2005), minimally clinically important differences have not been published for the FLIC.

Reported patient acceptability of the FLIC is mixed, with one study reporting a high response rate of 94% (n=405) (Elliott et al., 2004) and another reporting poor compliance rates (Finkelstein et al., 1988).

Results: Other condition-specific PROMs in breast cancer

Several additional cancer-specific PROMs were identified in the search strategy, but there was insufficient evidence to include them in this review. These PROMs are listed in Table 2.

Table 2: Other cancer-specific PROMs not reviewed

Instrument Name	No. records
Breast Cancer Chemotherapy Questionnaire (BCQ)	2
Breast Cancer Prevention Trial (BCPT) Symptom Scales/BCPT Symptom Checklist	4
Brief Cancer Impact Assessment (BCIA)	3
Cancer Behaviour Inventory (CBI)	1
Concerns about Recurrence Scale (CARS)	1
Experience of Breast Cancer Questionnaire (EBCQ)	1
Ferrans and Powers QoL Index – Cancer Version (QLI-CV)	1
Functional Assessment of Cancer Therapy – Breast Symptom Index (FBSI)	1
Health and Activity Limitations Index (HALex)	2
Health Utilities Index (HUI)	1
Impact of Cancer (IOC)	1
Linear Analogue Self-Assessment (LASA)	3
Quality of Life in Adult Cancer Survivors (QLACS)	1
Lymphoedema and Breast Cancer Questionnaire	1
Quality of Life Breast Cancer Version (QOL-BC)	2
Quality of Life – Cancer Survivors (QOL-CS)	3
Supportive Care Needs Survey (SCNS)	2
TTO Instrument	1

SUMMARY OF EVIDENCE

Generic PROMs

Table 3 summarises the evidence of measurement and operational performance applying the adapted appraisal criteria of the generic PROMs identified in this review. Based on the volume and quality of evidence, the SF-36 is clearly the most promising generic health measure. The EQ-5D and SIP both have some evidence of good performance in women with breast cancer although the SIP may be more burdensome due to the large number of items. The EQ-5D has the advantage of the generation of a utility score.

The SF-36, though popular for measuring health status, does have limitations when utilised with breast cancer patients. Support for the psychometric properties of the SF-36 in this cohort is evident, but the evidence also makes it clear that this instrument is not sensitive enough to measure breast cancer-specific health status data. If used with breast cancer patients, it is recommended that an instrument specific to breast cancer is administered alongside the SF-36.

The SF-8 is still in its infancy and thus there is very little evidence supporting its psychometric properties with breast cancer patients.

Cancer-specific PROMs

Table 4 summarises the evidence of measurement and operational performance applying the adapted appraisal criteria of the PROMs identified in these reviews.

The CARES-SF has been evaluated mainly with newly diagnosed breast cancer patients. Rehabilitation needs are likely to vary across age, disease severity, treatment type, and many more generic and disease-specific variables. Therefore, the CARES-SF needs to ideally be utilised across a more heterogeneous sample before its full potential can be accurately assessed.

The EORTC QLQ-C30 has been evaluated mainly with advanced and metastatic breast cancer patients. Furthermore, it is worth noting that site-specific modules can be added to the QLQ-C30 to gain insight into site-specific QoL. In the case of breast cancer, the appropriate module is BR23 although little evidence is reported to date.

The FACT-G has been utilised with a variety of disease and treatment stages as well as in a variety of interventional studies. In terms of breast cancer, a site-specific module can be added to the FACT-G- the FACT-B.

The FACT-B has been utilised with a variety of disease and treatment stages as well as in a variety of interventional studies, proving itself to be especially appropriate within clinical settings due to its ease of administration and completion. A FACT-B+4, including a 4-item arm morbidity subscale can also be used. The FACT-B+4 has demonstrated test-retest reliability, discriminant validity and responsiveness. The four arm morbidity items were selected following consultation with breast cancer health professionals and breast cancer patients, demonstrating content validity. Furthermore, during the validation process patients found the scale quick and easy to complete, with good completion rates for all items within the questionnaire.

The FLIC was designed specifically for cancer patients undergoing treatment and is functionally oriented. Therefore, it might be most useful in clinical practice with patients who are expected to experience functioning difficulties, or in the monitoring of rehabilitation interventions.

The instruments included in this review have been used with patients with recent diagnosis or undergoing treatment, principally to evaluate the effectiveness of interventions. However, as more patients are living longer following treatment, there is increasing interest in measuring quality of life amongst the long-term survivors of cancer. In this context, our recommendations regarding generic instruments will likely remain appropriate, but the content of condition-specific instruments may lose face validity as the predicament of survivorship is different to undergoing treatment. There is evidence that the experiences of individuals who have survived cancer for some time are not completely captured by instruments focused on experiences surrounding diagnosis and treatment (Gotay and Muraoka, 1998). Pearce et al. (2008) identified and appraised several instruments which have been specifically developed for long-term survivors of cancer some of which have been evaluated with women with breast cancer. Currently, from a psychometric perspective, survivorship instruments are deemed to be limited. Avis and Foley (2006) found some positive evidence of responsiveness for a cancer survivor outcome measure, the Quality of Life in Adult Cancer Survivors (QALCAS) Scale. Further consideration would be necessary if the health and well-being of the long-term survivors is to be measured routinely.

RECOMMENDATIONS

There are three main approaches to the measurement of patient-outcomes with people with cancer. The generic approach to health status measurement allows for the comparison across health conditions; the cancer-generic modules which are more focused on dimensions relevant to people with cancer; and add-on modules which focus more on specific domains relative to the type of cancer. This is illustrated in the evidence found in this review.

The EORTC QLQ-C30 and FACT-G or FACT-B are shortlisted as promising cancer-specific PROMs for piloting in the NHS. The more generic FACT-G may be sufficient where the focus is on outcomes for individuals with a broad range of types of cancer, including cancer of the breast. The SF-36 and EQ-5D are recommended as generic measures, the EQ-5D particularly when a utility score is required. Attention may need to be given to longer term issues of survivorship if the full spectrum of cancer in the population is to be included in surveys via PROMs although insufficient evidence was found to high-light specific measures for this review.

Table 3: Appraisal of psychometric and operational performance of generic PROMs for breast cancer

Instrument (n studies)	EQ-5D (3)	SF-36 (35)	SF-8 (1)	SIP (2)
Reproducibility	0	0	0	0
Internal consistency	n/a	++	0	+
Validity: Content	+	+++	+	+
Construct	+	++	+	+
Responsiveness	+	++	0	+
Interpretability	+	+++	+++	0
Floor/ceiling/precision	+/-	+	0	0
Acceptability	0	++	+	+
Feasibility	0	++	+	0

Psychometric and operational criteria

0   *not reported*

—   *no evidence in favour*

+   *some limited evidence in favour*

++   *some good evidence in favour*

++ +   *good evidence in favour.*

Table 4: Appraisal of psychometric and operational performance of condition-specific PROMs for breast cancer

	CARES-SF (6)	EORTC QLQ-C30 (16)	FACT-G (13)	FACT-B (21)	FLIC (8)
Reproducibility	+	0	+	+	0
Internal consistency	+	++	+++	+++	+
Validity: Content	++	+++	+++	+++	++
Construct	++	+++	+++	++	+
Responsiveness	+	+	++	+++	+
Interpretability	0	++	++	+	0
Floor/ceiling/precision	0	-	0	0	0
Acceptability	+	++	++	+++	+
Feasibility	+	++	+	+	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.

APPENDIX A: sources for PROM bibliography

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

PROM Bibliography search strategy[1]
a. records to December 2005 (downloads 1-12)
((acceptability or appropriateness or (component$ analysis) or comprehensibility or (effect size$) or (factor analys$) or (factor loading$) or (focus group$) or (item selection) or interpretability or (item response theory) or (latent trait theory) or (measurement propert$) or methodol$ or (multi attribute) or multiattribute or precision or preference$ or proxy or psychometric$ or qualitative or (rasch analysis) or reliabilit$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or (standard gamble) or (summary score$) or (time trade off) or usefulness$ or (utility estimate) or valid$ or valuation or weighting$)
and
((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY$ or QL or QoL or (qualit$ of life) or (quality adjusted life year$) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))
or
((bibliograph$ or interview$ or overview or review) adj5 ((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY$ or QL or QoL or (qualit$ of life) or (quality adjusted life year$) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))
b. records from January 2006 (download 13)
((((acceptability or appropriateness or component$ analysis or comprehensibility or effect size$ or factor analys$ or factor loading$ or feasibility or focus group$ or item selection or interpretability or item response theory or latent trait theory or measurement propert$ or methodol$ or multi attribute or multiattribute or precision or preference$ or proxy or psychometric$ or qualitative or rasch analysis or reliabilit$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or valid$ or valuation or weighting$)
and
(HRQL or HRQoL or QL or QoL or qualit$ of life or quality adjusted life year$ or QALY$ or disability adjusted life year$ or DALY$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score$ or time trade off or health index or health profile or health status or ((patient or self$) adj (rated or reported or based or assessed)) or ((disability or function$ or subjective or utilit$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$))))
or
((bibliograph$ or interview$ or overview or review) adj5 (HRQL or HRQoL or QL or QoL or qualit$ of life or quality adjusted life year$ or QALY$ or disability adjusted life year$ or DALY$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score$ or time trade off or health index or health profile or health status or ((patient or self$) adj (rated or reported or based or assessed)) or ((disability or function$ or subjective or utilit$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire$ or profile$ or scale$ or score$ or status or survey$)))))

---

[1] Note: the bibliography includes approximately 1,650 hand searched additions.

APPENDIX B: Psychometric criteria

Appraisal of PROMs

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

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

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

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

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

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

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

APPENDIX C: Generic PROMs

a) European Quality of Life instrument, EQ-5D (The EuroQol Group, 1990; revised 1993)

The European Quality of Life instrument (EuroQol), now generally known as the EQ-5D, was developed by researchers in five European countries to provide an instrument with a core set of generic health status items (The EuroQol Group, 1990; Brazier et al., 1993). Although providing a limited and standardized reflection of HRQoL, it was intended that use of the EuroQol would be supplemented by disease-specific instruments. The developers recommend the EuroQol for use in evaluative studies and policy research; given that health states incorporate preferences, it can also be used for economic evaluation. It can be self or interview-administered.

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

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

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

Items were derived from several sources, including extensive literature reviews and existing instruments (Ware and Sherbourne, 1992; Ware and Gandek, 1998; Jenkinson and McGee, 1998). The original Rand MOS Questionnaire (245 items) was the primary source, and several items were retained from the SF-20. The 36 items assess health across eight domains (Ware, 1997), namely bodily pain (BP: two items), general health perceptions (GH: five items), mental health (MH: five items), physical functioning (PF: ten items), role limitations due to emotional health problems (RE : three items), role limitations due to physical health problems (RP: four items), social functioning (SF: two items), and vitality (VT: four items). An additional health transition item, not included in the final score, assesses change in health.

There are between two and six categorical response options for each item. Scoring uses a weighted scoring algorithm and a computer-based programme is recommended.

Eight domain scores give a health profile; scores are transformed into a scale from 0 to 100, where 100 denotes the best health. Scores can be calculated when up to half of the items are omitted. Two component summary scores for physical and mental health (MPS and MCS, respectively) can also be calculated. A version of the SF-36 plus three depression questions has been developed and is variously called the Health Status Questionnaire (HSQ) or SF-36-D.

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

c) Medical Outcome Short Form 8-Item Health Survey (SF-8)

The SF-8 was constructed to provide a shorter alternative to the SF-36 for use in large population-based surveys of general and specific populations. The SF-8 was constructed on the basis of empirical studies linking each item to a comprehensive pool of widely used questionnaire items proven to measure the same concept (Ware et al., 2001). The SF-8 is a multidimensional measure of health-related function derived from the SF-36. The instrument comprises eight subscales - each of which relies on a single item: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health - and two summary scores for physical and mental health-related ability to function (PCS and MCS, respectively). The SF-8 is scored using a normative algorithm, based on a general US population. There are five or six possible responses to each item, answered for the past four weeks, and responses are weighted based on the algorithm.

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

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

Instrument content was informed by the concept of 'sickness', which was defined as reflecting the change in an individual's activities of daily life, emotional status, and attitude as a result of ill-health (McDowell and Newell, 1996). Item derivation was based on literature reviews and statements from health professionals, carers, patient groups, and healthy subjects describing change in behaviour as a result of illness. The SIP has 136 items across 12 domains: alertness behaviour (AB: ten items), ambulation (A: 12 items), body care and movement (BCM: 23 items), communication (C: nine items), eating (E: nine items), emotional behaviour (EB: nine items), home management (HM: ten items), mobility (M: ten items), recreation and pastimes (RP: eight items), sleep and rest (SR: seven items), social interaction (SI: 20 items), and work (W: nine items).

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

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

GENERIC INSTRUMENTS DOMAINS AND SCORING

Table 5: Domains and Scoring

Instrument	Domains (No. Items)	Response Options	Score	Completion
EQ-5D	EQ-5D: Anxiety/depression (1), mobility (1), pain/discomfort (1), self-care (1), usual activities (1). EQ-thermometer: Global health (1)	EQ-5D: Categorical: 3 options - 1 (no problems), 2 (some problems), and 3 (inability/extreme problems), EQ-thermometer: VAS current health.	EQ-5D: Summation: responses are scored for each individual health state and then multiplied by the duration of time in that health state; Utility index (-0.59-100, where -0.59 is a state worse than death and 1.00 is maximum well-being. EQ-thermometer: VAS (0-100, 100 representing best imaginable health).	Interview or self.
SF-36	Bodily pain (BP) (1), General health (GH) (5) Mental health (MH) (5), Physical functioning (PF) (6) Role functioning (RF), Social functioning (SF) (1)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS) and Mental (MCS) (mean 50, sd 10).	Interview or self. 5-10 minutes.
SF-8	8-Items, one-item per 8 subscales: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health; and two summary scores for physical and mental health-related ability to function (PCS and MCS, respectively).	Likert scale representing degree of symptom experience.	The SF-8 is scored using a normative algorithm, based on a general US population. There are 5 or 6 possible responses to each item, answered for the past 4 weeks, and responses are weighted based on the algorithm.	Interview or self.
SIP	Alertness behaviour (AB) (10), Ambulation (A) (12) Body care and movement (BCM) (23), Communication (C) (9) Eating (E) (9), Emotional behaviour (EB) (9) Home management (HM) (10), Mobility (M) (10) Recreation and pastimes (RP) (8), Sleep and rest (SR) (7) Social interaction (SI) (20), Work (W) (9)	Check applicable statements. Items weighted: higher weights indicate increased impairment Recall current health	Algorithm Domain profile (0-100%, 100 worst health); Index (0-100%) Summary: Physical (A, BCM, M), Psychosocial function (AB, C, EB, SI)	Interview (range: 21-33) Telephone: PF only (11.5) Self (19.7)

Table 6: Summary of generic instruments: health status domains (after Fitzpatrick et al., 1998)

Instrument	Instrument domains							
	Physical function	Symptoms	Global judgement of health	Psychological well-being	Social well-being	Cognitive functioning	Role activities	Personal constructs
EQ-5D	x	x	x	x	x		x	
SF-36	x	x	x	x	x		x	
SF-8	x	x	x	x	x		x	
SIP	x	x		x	x	x	x	

APPENDIX D: Cancer-specific PROMs

Cancer Rehabilitation Evaluation System – Short-Form (CARES-SF)
The CARES-SF was derived directly from the 139-item version of the CARES, which has well-documented sound psychometric properties (Meyerowitz et al., 1983; Ganz et al., 1990) Schag et al., 1990). Empirical data and clinical considerations were used to develop an instrument that was clinically useful and valid in documenting rehabilitation problems and QoL. An effort was made to retain representation in all content domains of the original CARES, which was assisted by data from previous research and analyses on the psychometric properties of the long form. Four cancer professionals reviewed this data independently to confirm the content validity of the short-form (Schag et al., 1991). The scale consists of 59 items, which together generate a single global score as well as five sub-scores representing the following domains: physical, psychosocial, medical, marital, and sexual functioning. The CARES-SF rates the degree to which a given problem applied during the four weeks prior to the measure being administered. Scoring is based on a 4-point Likert scale from 0 (not at all) to 4 (very much), with higher scores indicating greater difficulty or impairment.

European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
The EORTC QLQ-C30 (Aaronson et al., 1993) is a 30-item questionnaire composed of five multi-item functional subscales: physical, role, emotional, social, and cognitive functioning; three multi-item symptom scales: fatigue, pain, and emesis; global health/QoL subscale; and six single items to assess financial impact and symptoms such as dyspnoea, sleep disturbance, appetite, diarrhoea, and constipation. The questionnaire employs 28 four-point Likert scales of 'not at all' to 'very much,' and two seven-point Likert scales for the global health and QoL domain, total scores ranging from 0 to 100. For functional and global QoL scales, higher scores represent a better level of functioning. For symptom-oriented scales, a higher score represents more severe symptoms. Extensive patient input during an international field study contributed to the development of the QLQ-C30, initially with lung cancer patients (Aaronson et al., 1993) and subsequently with patients with heterogeneous diagnoses (Osoba et al., 1994). Content validity of the QLQ-C30 has been maintained via modifications to improve the content, specifically in terms of the Role Functioning scale and a conceptual difficulty (undue emphasis on physical functioning) in the global QoL scale (Osoba et al., 1997).

Functional Assessment of Cancer Therapy – General (FACT-G)
The 27-item FACT-G (Cella et al., 1993) measures global health-related QoL and four different dimensions thereof (i.e. physical, social, emotional, and functional well-being). The instrument is considered appropriate for use with any form of cancer and there are a number of scales that can be added to the FACT-G in order to measure disease- and treatment-specific components of the cancer experience. Answers are provided on a scale of Not at all; A little; Somewhat; Quite a bit; Very Much, in the context of 'during the past seven days'. Items are scored from 0-4, with negatively-phrased items requiring reverse response scores. Higher scores represent better well-being on each of the dimensions or better global QoL when combined. Content validity

is supported via item generation methodology. Items were generated using semi-structured interview input from cancer patients and oncology specialists. Patients first completed other QoL questionnaires in order to provide them with insight into potential QoL issues of relevance to them whilst the specialists reviewed these instruments and endorsed any items they felt were important as well as highlighting any QoL issues they felt were not covered in these instruments. Pilot testing and data reduction proceeded (Cella et al., 1993).

Functional Assessment of Cancer Therapy – Breast (FACT-B)

The FACT-B is a 10 item breast cancer specific module that supplements the Functional Assessment of Cancer Therapy – General (FACT-G). The 10 items are added to the 27 core items of the FACT-G, becoming a 37 item questionnaire. The FACT-B consists of 5 subscales: physical wellbeing (PW, 7 items), social and family wellbeing (SW, 7 items), emotional wellbeing (EW, 6 items), functional wellbeing (FW, 7 items) and the Breast Cancer Subscale (BCS, 10 items). Scores can be produced through three different calculations; a combined total of all domains (FACT-B total), the Breast Cancer Score (BCS) and a Treatment Outcome Index (TOI) can be calculated by summing the FACT-G physical and functional domains and the BCS. Answers are provided on a scale of Not at all; A little; Somewhat; Quite a bit; Very Much, in the context of 'during the past seven days'. Items are scored from 0-4, with negatively-phrased items requiring reverse response scores. Higher scores represent better well-being on each of the dimensions or better global QoL when combined.

Functional Living Index – Cancer (FLIC)

Also known as the Manitoba Functional Living Questionnaire, the FLIC (Schipper et al., 1984) is a 22-item functionally oriented QoL instrument that is scored with a total score and includes five subscales: physical well-being and ability, emotional state, sociability, family situation, and nausea. Questions are answered using a 7-point Likert-type linear analogue scale with the respondent marking the answer on the continuum. The response is then measured to the nearest tenth and recorded in the positive direction, as needed, with 1=worst and 7=best rating.

The FLIC was developed initially by a team of patients and their spouses, and health-care and oncologist specialists (Clinch, 1996) before being validated longitudinally on 837 cancer patients (Schipper et al., 1984). This was followed by a pilot study with lung cancer patients aimed at testing the patient-oriented nature of the instrument (Finkelstein et al., 1988).

CANCER-SPECIFIC INSTRUMENTS DOMAINS AND SCORING

Table 7: Domains and Scoring

Instrument name (total no. items)	Domains (no. items)	Response options	Scoring	Mode of administration Completion time	Licensing information
European Organization for Research and Treatment of Cancer Quality of Life core Questionnaire, EORTC QLQ-C30 (30)	Global health status/QoL (2) *Functional scales:* Physical functioning (5) Role functioning (2) Emotional functioning (4) Cognitive functioning (2) Social functioning (2) *Symptom scales:* Fatigue (3) Nausea & vomiting (2) Pain (2) Dyspnoea (1) Insomnia (1) Appetite loss (1) Constipation (1) Diarrhoea (1) Financial difficulties (1)	4-point Likert scales where 1=not at all (best), 4=very much (worst). 7-point Likert scales for global health and overall QoL questions, where 1=very poor, 7=excellent Except for PF, responses are based on the past week	Subscale scores transformed into 0-100 scores using an algorithm. Higher scores on functional scales and global items indicate better functioning; higher scales on symptom scales indicate worse symptomatology. Aggregation of subscale scores not recommended by developers.	Interview, telephone, or self-administration. Electronic versions of the QLQ-C30 under development. 10-15 minutes	No charge for use in academic settings, but written consent required for each study. Royalty fee, based on no. of patients, payable for commercial studies.

Table 8: Summary of cancer-specific instruments: health status domains (after Fitzpatrick et al., 1998)

Instrument	Instrument domains								
	Physical function	**Symptoms**	**Global judgement of health**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**	**Satisfaction with care**
EORTC QLQ-C30	x	x	x	x	x	x	x	x	
FACT-C	x	x	x	x	x		x	x	

* crosses in bold indicate content of the colorectal cancer-specific module; the remainder indicate domains covered by the core

APPENDIX E: LICENSING & CONTACT DETAILS

European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
EORTC Headquarters
Quality of Life Department
Ave. E. Mounier 83, B.11
1200 Brussels Belgium
Fax: +32 (0)2 779 45 68
Tel: +32 (0)2 774 1678
E mail: ken.cornelissen@eortc.be
http://groups.eortc.be/qol/questionnaires_qlqc30.htm
http://groups.eortc.be/qol/questionnaires_modules.htm

Functional Assessment of Cancer Therapy – General Version (FACT-G)
FACIT.org
381 South Cottage Hill Ave
Elmhurst, IL 60126 USA
Tel: (+1) 877 828 3228
Fax: (+1) 630 279 9465
E mail: information@facit.org
http://www.facit.org/qview/qlist.aspx

REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ *et al*. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 1993; 85(5):365-376.

Alfano CM, McGregor BA, Kuniyuki A, Reeve BB, Bowen DJ, Baumgartner KB *et al*. Psychometric properties of a tool for measuring hormone-related symptoms in breast cancer survivors. *Psycho-Oncology* 2006a; 15(11):985-1000.

Alfano CM, McGregor BA, Kuniyuki A, Reeve BB, Bowen DJ, Smith AW *et al*. Psychometric evaluation of the Brief Cancer Impact Assessment among breast cancer survivors. *Oncology (Basel)* 2006a; 70(3):190-202.

Allard NC. Day surgery for breast cancer: effects of a psychoeducational telephone intervention on functional status and emotional distress. *Oncology Nursing Forum* 2007; 34(1):E133-141.

Andrykowski MA, Schmidt JE, Salsman JM, Beacham AO, Jacobsen PB. Use of a case definition approach to identify cancer-related fatigue in women undergoing adjuvant therapy for breast cancer. *Journal of Clinical Oncology* 2005; 23(27):6613-22.

Aranda S, Schofield P, Weih L, Yates P, Milne D, Faulkner R *et al*. Mapping the quality of life and unmet needs of urban women with metastatic breast cancer. *European Journal of Cancer Care* 2005; 14(3):211-22.

Avis N, 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 survivors. Health and Quality of Life Outcomes. 2006 Dec 1;4:92

Bardwell WA, Major JM, Rock CL, Newman VA, Thomson CA, Chilton JA *et al*. Health-related quality of life in women previously treated for early-stage breast cancer. *Psycho-Oncology* 2004; 13(9):595-604.

Bardwell WA, Natarajan L, Dimsdale JE, Rock CL, Mortimer JE, Hollenback K *et al*. Objective cancer-related variables are not associated with depressive symptoms in women treated for early-stage breast cancer. *Journal of Clinical Oncology* 2006; 24(16):2420-2427.

Bergner M, Bobbitt RA, Kressel S, et al. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Service* 1976; 6(3):393-415.

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

Bower JE, Ganz PA, Desmond KA, Bernaards C, Rowland JH, Meyerowitz BE *et al*. Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. *Cancer* 2006; 106(4):751-758.

Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR *et al*. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *Journal of Clinical Oncology* 1997; 15(3):974-86.

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

Byar KL, Berger AM, Bakken SL *et al*. Impact of adjuvant breast cancer chemotherapy on fatigue, other symptoms, and quality of life. *Oncology Nursing Forum* 2006; 33(1):E18-26.

Campbell A, Mutrie N, White F, McGuire F, Kearney N. A pilot study of a supervised group exercise programme as a rehabilitation treatment for women with breast cancer receiving adjuvant treatment. *European Journal of Oncology Nursing* 2005; 9(1):56-63.

Carver CS, Pozo-Kaderman C, Price AA, Noriega V, Harris SD, Derhagopian RP *et al*. Concern about aspects of body image and adjustment to early stage breast cancer. *Psychosomatic Medicine* 1998; 60(2):168-74.

Casso D, Buist DS, Taplin S. Quality of life of 5-10 year breast cancer survivors diagnosed between age 40 and 49. *Health and Quality of Life Outcomes* 2004; **2**:25.

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi AE *et al*. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *Journal of Clinical Oncology* 1993; 11(3 Suppl.2):570-9.

Cella DF, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Quality of Life Research* 2002; 11(3):207-221.

Claus EB, Petruzella S, Carter D, Kasl S. Quality of life for women diagnosed with breast carcinoma in situ. *Journal of Clinical Oncology* 2006; 24(30):4875-4881.

Clinch JJ. The Functional Living Index-Cancer: ten years later. In: Spilker B, editor. Quality of Life and Pharmacoecoeconomics in Clinical Trials. Philadelphia, USA: Lippincott-Raven, 1996: 215-225.

Coates AS, Gebski V, Signorini D, Murray P, McNeil D, Byrne M *et al*. Prognostic value of quality of life scores during chemotherapy for advanced breast cancer. Australian New Zealand Breast Cancer Trials Group. *Journal of Clinical Oncology* 1992; 10(12):1833-8.

Conner-Spady B, Cumming C, Nabholtz JM, Jacobs P, Stewart D. Responsiveness of the EuroQol in breast cancer patients undergoing high dose chemotherapy. *Quality of Life Research* 2001; 10(6):479-86.

Coster S, Poole K, Fallowfield LJ. The validation of a quality of life scale to assess the impact of arm morbidity in breast cancer patients post-operatively. *Breast Cancer Research and Treatment* 2001; 68(3):273-82.

Crane PK, Gibbons LE, Narasimhalu K, Lai JS, Cella D. Rapid detection of differential item functioning in assessments of health-related quality of life: The Functional Assessment of Cancer Therapy. *Quality of Life Research* 2007; 16(1):101-114.

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

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

Dausch BM, Compas BE, Beckjord E, Luecken L, Anderson-Hanley C, Sherman M *et al*. Rates and correlates of DSM-IV diagnoses in women newly diagnosed with breast cancer. *Journal of Clinical Psychology in Medical Settings* 2004; 11(3):159-69.

Davies NJ, Kinman G, Thomas RJ, Bailey TA. Information satisfaction in breast and prostate cancer patients: implications for quality of life. *Psycho-Oncology* 2008a; 17(10):1048-1052.

Davies NJ, Kinman G, Thomas RJ, Bailey, TA. Health baseline comparison theory: predicting quality of life in breast and prostate cancer. *Health Psychology Update* 2008b; 17(3):3-11.

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

Department of Health, London, December 2007. NHS Cancer Reform Strategy.

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

Doorenbos A, Given B, Given C, Verbitsky N. Physical functioning: effect of behavioral intervention for symptoms among individuals with cancer. *Nursing Research* 2006; 55(3):161-171.

Elliott BA, Renier CM, Haller IV, Elliott TE. Health-related quality of life (HRQoL) in patients with cancer and other concurrent illnesses. *Quality of Life Research* 2004; 13(2):457-62.

Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *Journal of Clinical Epidemiology* 2004; 57(9):898-910.

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

Ferrell BR, Dow KH, Leigh S, Ly J, Gulasekaram P. Quality of life in long-term cancer survivors. *Oncology Nursing Forum* 1995; 22(6):915-22.

Figueiredo MI, Cullen J, Hwang Y, Rowland JH, Mandelblatt JS. Breast cancer treatment in older women: does getting what you want improve your long-term body image and mental health? *Journal of Clinical Oncology* 2004; 22(19):4002-9.

Finkelstein DM, Cassileth BR, Bonomi PD, Ruckdeschel JC, Ezdinli EZ, Wolter JM. A pilot study of the Functional Living Index-Cancer (FLIC) scale for the assessment of quality of life for metastatic lung cancer patients. *American Journal of Clinical Oncology* 1988; 11(6):630-633.

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

Fitzpatrick R, Bowling A, Gibbons E, Haywood K, Jenkinson C, Mackintosh A, Peters M. A Structured Review of Patient-Reported Measures in Relation to Selected Chronic Conditions, Perceptions of Quality of Care, and Carer Impact. Oxford: National Centre for Health Outcomes Development, November 2006.

Ganz PA, Schag CAC, Cheng HL. Assessing the quality of life: a study in newly-diagnosed breast cancer patients. *Journal of Clinical Epidemiology* 1990; 43(1):75-86.

Garland MR, Lavelle E, Doherty D, Golden-Mason L, Fitzpatrick P, Hill A *et al*. Cortisol does not mediate the suppressive effects of psychiatric morbidity on natural killer cell activity: a cross-sectional study of patients with early breast cancer. *Psychological Medicine* 2004; 34(3):481-90.

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

Glick HA, Shpall EJ, LeMaistre CF, McCarron TJ, Lu ZJ, Erder MH *et al*. Empirical criteria for the selection of quality-of-life instruments for the evaluation of peripheral blood progenitor cell transplantation. *International Journal of Technology Assessment in Health-Care* 1998; 14(3):419-30.

Golden-Kreutz DM, Thornton LM, Wells DG, Frierson GM, Jim HS, Carpenter KM *et al*. Traumatic stress, perceived global stress, and life events: prospectively predicting quality of life in breast cancer patients. *Health Psychology* 2005; 24(3):288-96.

Gordon LG, Scuffham P, Battistutta D, Graves N, Tweeddale M, Newman B. A cost-effectiveness analysis of two rehabilitation support services for women with breast cancer. *Breast Cancer Research and Treatment* 2005; 94(2):123-133.

Gotay C, Muraoka M, Quality of life in long term survivors of adult onset cancers. *Journal of the National Cancer Institute.* 1998; 90: 656-67.

Grabsch B, Clarke DM, Love A, McKenzie DP, Snyder RD, Bloch S *et al*. Psychological morbidity and quality of life in women with advanced breast cancer: a cross-sectional survey. *Palliative & Supportive Care* 2006; 4(1):47-56.

Griggs JJ, Sorbero ME, Mallinger JB, Quinn M, Waterman M, Brooks B *et al*. Vitality, mental health, and satisfaction with information after breast cancer. *Patient Education & Counseling* 2007; 66(1):58-66.

Grunfeld E, Levine MN, Julian JA, Coyle D, Szechtman B, Mirsky D *et al*. Randomized trial of long-term follow-up for early-stage breast cancer: a comparison of family physician versus specialist care. *Journal of Clinical Oncology* 2006; 24(6):848-855.

Hack TF, Pickles T, Bultz BD, Reuther JD, Weir LM, Degner LF *et al*. Impact of providing audiotapes of primary adjuvant treatment consultations to women with breast cancer: a multisite, randomized, controlled trial. *Journal of Clinical Oncology* 2003; 21(22):4138-44.

Hack TF, Degner LF, Watson P, Sinha L. Do patients benefit from participating in medical decision making? Longitudinal follow-up of women with breast cancer. *Psycho-Oncology* 2006; 15(1):9-19.

Hegel MT, Moore CP, Collins ED, Kearing S, Gillock KL, Riggs RL *et al*. Distress, psychiatric syndromes, and impairment of function in women with newly diagnosed breast cancer. *Cancer* 2006; 107(12):2924-2931.

Helgeson VS, Tomich PL. Surviving cancer: a comparison of 5-year disease-free breast cancer survivors with healthy women. *Psycho-Oncology* 2005; 14(4):307-17.

Hopwood P, Haviland J, Mills J, Sumo G, Bliss JM. The impact of age and clinical factors on quality of life in early breast cancer: an analysis of 2208 women recruited to the UK START Trial (Standardisation of Breast Radiotherapy Trial). *Breast* 2007; 16(3):241–251.

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

Jeruss JS, Hunt KK, Xing Y, Krishnamurthy S, Meric-Bernstam F, Cantor SB *et al*. Is intraoperative touch imprint cytology of sentinel lymph nodes in patients with breast cancer cost effective? *Cancer* 2006; 107(10):2328-2336.

Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *Journal of Clinical Epidemiology* 2007; 60(5):448-455.

Lev EL, Owen SV. Counseling women with breast cancer using principles developed by Albert Bandura - corrected. *Perspectives in Psychiatric Care* 2000; 36(4):131-8.

Levine EG,.Targ E. Spiritual correlates of functional well-being in women with breast cancer. *Integrative Cancer Therapies* 2002; 1(2):166-74.

Lovrics PJ, Cornacchi SC, Barnabi F, Whelan T, Goldsmith CH. The feasibility and responsiveness of the Health Utilities Index in patients with early-stage breast cancer: a prospective longitudinal study. *Quality of Life Research* 2008; 17(2):333-345.

Manning-Walsh J. Spiritual struggle: effect on quality of life and life satisfaction in women with breast cancer. *Journal of Holistic Nursing* 2005a; 23(2):120-40.

Manning-Walsh J. Social support as a mediator between symptom distress and quality of life in women with breast cancer. *JOGNN: Journal of Obstetric Gynecologic and Neonatal Nursing* 2005b; 34(4):482-93.

Mansel RE, Fallowfield L, Kissin M, Goyal A, Newcombe RG, Dixon JM et al. Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. *Journal of the National Cancer Institute* 2006; 98(9):599-609.

Mar Fan HG, Houede-Tchen N, Yi QL, Chemerynsky I, Downie FP, Sabate K *et al*. Fatigue, menopausal symptoms, and cognitive function in women after adjuvant chemotherapy for breast cancer: 1- and 2-year follow-up of a prospective controlled study. *Journal of Clinical Oncology* 2005; 23(31):8025-32.

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

McLachlan SA, Devins GM, Goodwin PJ. Validation of the European Organization for Research and Treatment of Cancer quality of life questionnaire (QLQ-C30) as a measure of psychosocial function in breast cancer patients. *European Journal of Cancer* 1998; 34(4):510-517.

McLachlan SA, Devins GM, Goodwin PJ. Factor analysis of the psychosocial items of the EORTC QLQ-C30 in metastatic breast cancer patients participating in a psychosocial intervention study. *Quality of Life Research* 1999; 8(4):311-317.

Meyerowitz BE, Watkins IK, Sparks FC. Quality of life for breast cancer patients receiving adjuvant chemotherapy. *American Journal of Nursing* 1983; 2(232):235.

Mondry TE, Riffenburgh RH, Johnstone PAS. Prospective trial of complete decongestive therapy for upper extremity lymphedema after breast cancer therapy. *Cancer Journal* 2004; 10(1):42-8.

Muss HB, Tu D, Ingle JN, Martino S, Robert NJ, Pater JL et al. Efficacy, toxicity, and quality of life in older women with early-stage breast cancer treated with letrozole or placebo after five years of tamoxifen: NCIC CTG intergroup trial MA.17. *Journal of Clinical Oncology* 2008; 26(12):1956-1964.

National Institute for Clinical Excellence, London, August 2002. Guidance on Cancer Services: Improving Outcomes for Breast Cancer - Manual Update.

Office for National Statistics (2007). Annual Update: Cancer incidence and mortality in the United Kingdom and constituent countries, 2002–04. *Health Statistics Quarterly* **35**:78–83.

Ohira T, Schmitz KH, Ahmed RL, Yee D. Effects of weight training on quality of life in recent breast cancer survivors: the Weight Training for Breast Cancer Survivors (WTBS) study. *Cancer* 2006; 106(9):2076-2083.

Osoba D, Zee B, Pater JL, Warr D, Kaizer L, Latreille J. Psychometric properties and responsiveness of the EORTC Quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian, and lung cancer. *Quality of Life Research* 1994; 3(5):353-64.

Osoba D, Aaronson NK, Zee B, Sprangers MAG, Te Velde A. Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. *Quality of Life Research* 1997; 6(2):103-108.

Osoba D, Rodrigues G, Myles J, Zee B, Pater JL. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology* 1998; 16(1):139-144.

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

Pearce NJ, Sanson-Fisher R, Campbell HS. (2008) Measuring quality of life in cancer survivors: a methodological review of existing scales. Psychooncology. 17(7):629-40.

Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health and Quality of Life Outcomes* 2007; **5**:70.

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

Potter S, Thomson HJ, Greenwood RJ, Hopwood P, Winters ZE. Health-related quality of life assessment after breast reconstruction. *Br J Surg* 2009; 96(6):613-620.

Purushotham AD, Upponi S, Klevesath MB, Bobrow L, Millar K, Myles JP *et al*. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. *Journal of Clinical Oncology* 2005; 23(19):4312-21.

Romero C, Friedman LC, Kalidas M, Elledge R, Chang J, Liscum KR. Self-forgiveness, spirituality, and psychological adjustment in women with breast cancer. *Journal of Behavioral Medicine* 2006; 29(1):29-36.

Sandel SL, Judge JO, Landry N, Faria L, Ouellette R, Majczak M. Dance and movement program improves quality-of-life measures in breast cancer survivors. *Cancer Nursing* 2005; 28(4):301-9.

Schag CAC, Heinrich RL. Development of a comprehensive quality of life measurement tool: CARES. *Oncology (Huntington)* 1990; 4(5):135-138.

Schag CAC, Ganz PA, Heinrich RL. CAncer Rehabilitation Evaluation System-Short Form (CARES-SF): a cancer specific rehabilitation and quality of life instrument. *Cancer* 1991; 68(6):1406-1413.

Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the Functional Living Index-Cancer: development and validation. *Journal of Clinical Oncology* 1984; 2(5):472-483.

Shelby RA, Lamdan RM, Siegel JE, Hrywna M, Taylor KL. Standardized versus open-ended assessment of psychosocial and medical concerns among African American breast cancer patients. *Psycho-Oncology* 2006; 15(5):382-397.

Shilling V, Jenkins V. Self-reported cognitive problems in women receiving adjuvant therapy for breast cancer. *European Journal of Oncology Nursing* 2007; 11(1):6-15.

Smith AB, Wright P, Selby PJ, Velikova G. A Rasch and factor analysis of the Functional Assessment of Cancer Therapy-General (FACT-G). *Health and Quality of Life Outcomes* 2007; **5**:19.

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

Snyder CF, Garrett-Mayer E, Brahmer JR, Carducci MA, Pili R, Stearns V et al. Symptoms, supportive care needs, and function in cancer patients: how are they related? *Quality of Life Research* 2008; 17(5):665-677.

Taenzer PA, Speca M, Atkinson MJ, Bultz BD, Page SA, Harasym P et al. Computerized quality of life screening in an oncology clinic. *Cancer Practice* 1997; 5(3):168-175.

Targ EF, Levine EG. The efficacy of a mind-body-spirit group for women with breast cancer: a randomized controlled trial. *General Hospital Psychiatry* 2002; 24(4):238-48.

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

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

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

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

Ware JE, Kosinski M, Dewey JE *et al*. How to score and interpret single-item health status measures: a manual for users of the SF-8 Health Survey. Lincoln, RI, USA: Quality Metric Incorporated, 2001.

Wilburn O, Wilburn P, Rockson SG. A pilot, prospective evaluation of a novel alternative for maintenance therapy of breast cancer-associated lymphedema. *BMC Cancer* 2006; **6**:84.

Wilson RW, Hutson LM, VanStry D. Comparison of 2 quality-of-life questionnaires in women treated for breast cancer: the RAND 36-Item Health Survey and the Functional Living Index-Cancer. *Physical Therapy* 2005; 85(9):851-60.

Wyatt GK, Donze LF, Beckrow KC. Efficacy of an in-home nursing intervention following short-stay breast cancer surgery. *Research in Nursing and Health* 2004; 27(5):322-31.

Yost KJ, Yount SE, Eton DT, Silberman C, Broughton HA, Cella D. Validation of the Functional Assessment of Cancer Therapy-Breast Symptom Index (FBSI). *Breast Cancer Research and Treatment* 2005; 90(3):295-8

Zebrack BJ, Ganz PA, 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.

Zhou X, Cella D, Cameron D, Amonkar MM, Segreti A, Stein S et al. Lapatinib plus capecitabine versus capecitabine alone for HER2+ (ErbB2+) metastatic breast cancer: quality-of-life assessment. *Breast Cancer Res Treat* 2009; 117(3):577-589.

Zick SM, Sen A, Feng Y, Green J, Olatunde S, Boon H. Trial of essiac to ascertain its effect in women with breast cancer (TEA-BC). *Journal of Alternative and Complementary Medicine* 2006; 12(10):971-980.