LITERATURE REVIEW

Determining optimal measures of health-related quality of life, anxiety and depression for evaluating progress in the psychosocial care of cancer patients in New South Wales

January 2011

The NSW Government agency dedicated to the control and cure of cancer through prevention, detection, innovation, research and information.

cancer institute NSW

Cancer Institute NSW
Australian Technology Park
Level 9, 8 Central Avenue
Australian Technology Park
Eveleigh NSW 2015
Telephone (02) 8374 5600
Facsimile (02) 8374 3600
Email information@cancerinstitute.org.au
Website www.cancerinstitute.org.au/publications

Acknowledgments

Psycho-Oncology Co-operative Research Group
The Cancer Institute NSW commissioned the Psycho-Oncology Co-operative Research Group (PoCoG) at the University of Sydney to undertake this review and would like to thank in particular Tim Luckett, Phyllis Butow, Melanie Price, Nicole Rankin, Madeleine King and Mayumi Oguchi.

Cancer Institute NSW
Nadine Hackl and Gaynor Heading

Contents

List of tables

Abbreviations

BDI (-PC)	Beck Depression Inventory (- Primary Care)
BSI	Brief Symptom Inventory
CAT	Computer Adaptive Testing
CES-D	Center for Epidemiologic Studies Depression Scale
C-SOSI/ SOSI	(Calgary) Symptoms of Stress Inventory
DASS	Depression, Anxiety and Stress Scales
DASS	Depression Anxiety Stress Scales
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revised
DT	Distress Thermometer
EF	Emotional functioning scale of the EORTC QLQ-C30
EORTC QLQ-C30	European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire - Core
ES	Effect size
EWB	Emotional wellbeing scale of the FACT-G
FACIT	Functional Assessment of Chronic Illness and Therapy
FACT-G	Functional Assessment of Cancer and Therapy - General
FWB	Functional wellbeing scale of the FACT-G
GHQ	General Health Questionnaire
HADS	Hospital Anxiety and Depression Scale
HADS-A	Anxiety scale from the Hospital Anxiety and Depression Scale
HADS-D	Depression scale from the Hospital Anxiety and Depression Scale
HADS-T	Overall score from the Hospital Anxiety and Depression Scale
HRQoL	Health-related quality of life
IES-R	Impact of Event Scale - Revised
IRT	Item response theory
LASA	Linear Analogue Self Assessment
LOTE	Languages other than English
MCS	Mental component summary from the SF-12 or SF-36
MHI	Mental Health Inventory
MID	Minimally important difference
NCCN	National Comprehensive Cancer Network
NIH	National Institute of Health
NSW	New South Wales
PCL-C	Post-traumatic Stress Disorder Checklist - Civilian Version
PCS	Physical component summary from the SF-12 or SF-36
PF	Physical functioning scale of the EORTC QLQ-C30
PoCoG	Psycho-oncology Co-operative Research Group
POMS	Profile of Mood States
PROM	Patient reported outcome measure
PROMIS	Patient-reported Outcome Measurement Information System
PTSD	Post traumatic stress disorder
PWB	Physical wellbeing scale of the FACT-G
QLI-CV	Quality of Life Index - Cancer Version

QOL-CA	Quality of Life - Cancer
RCT	Randomised controlled trial
RF	Role functioning scale of the EORTC QLQ-C30
RQ	Research question
SCL-90-R	Symptom Checklist – 90 - Revised
SD	Standard deviation
SF	Social functioning scale of the EORTC QLQ-C30
SF-12	Medical Outcomes Study Short Form Health Survey-12
SF-36	Medical Outcomes Survey Short Form Health Survey-36
SPHERE	Somatic and Psychological Health Report
STAI	State-Trait Anxiety Inventory
SWB	Social/family wellbeing scale of the FACT-G
TMDS	Total mood disturbance score on the Profile of Mood States
UK	United Kingdom
USA	United States of America
VAS	Visual analogue scale

Executive summary

Background

During the period from June to August 2009, the Psycho-Oncology Co-operative Research Group (PoCoG) was commissioned by the Cancer Institute New South Wales (NSW) to design and undertake a critical review aimed at answering the following research questions.

Research questions

RQ1. Which patient-reported outcome measures (PROMs) have been used to measure outcomes of anxiety, depression, general distress and health related quality of life (HRQoL) in randomised controlled trials (RCTs) of psychosocial interventions carried out with English-speaking cancer patients and published within the past 10 years?

RQ2. Which of these candidate PROMs will be optimal for use by the Cancer Institute NSW in evaluating new interventions as judged against the following criteria?

i. Suitability for use by people undergoing active treatment for cancer of any type and stage;
ii. Efficiency, relating to number of items and length of time taken to assess each construct;
iii. Ease of administration via touch-screen computer;
iv. Demonstrated reliability and validity in English-speaking people with cancer;
v. Demonstrated ability to identify treatment effects in RCTs of psychosocial interventions carried out with English-speaking cancer patients and published within the past 10 years;
vi. Ease of interpretability regarding clinical cut-offs for anxiety, depression or distress and/or the importance of differences between mean scores or change in mean scores over time;
vii. Availability of comparison data from cancer and general populations;
viii. Practical issues relating to availability in languages other than English (LOTE) and cost.

RQ3. Are there PROMs that have a proven track record in evaluating outcomes from a wide variety of psychosocial interventions, or will different (or a combination of) PROMs be needed to evaluate certain intervention types?

Methods

The review was conducted via the following methodology:

Step 1. **Identification of** all anxiety, depression, distress and HRQoL **questionnaires** used to measure outcomes from RCTs of psychosocial interventions carried out with English-speaking cancer patients and published in the past 10 years, **thus answering research question 1**;

Step 2. **First filter of candidate questionnaires** aimed at selecting those showing sufficient promise to warrant further investigation. Criteria included: 1) a time to administer of 10 minutes or less for scales assessing anxiety, depression, distress or HRQoL; 2) amenability to administration via touch-screen computer; 3) suitability for use by people undergoing active treatment for cancer of any type and stage; and 4) availability of evidence for reliability and validity in an English-speaking cancer population;

Step 3. **Description of selected candidate PROMs** via collation of information about the authors, content, number of items, time to administer, response options, recall period, scoring, licensing requirements and costs of selected candidate PROMs;

Step 4. Detailed **review of the evidence for reliability and validity** of each PROM in the cancer setting;

Step 5. **Review of the evidence for the capacity** of each selected questionnaire **to detect important effects of** treatment in RCTs of **psychosocial interventions** carried out with English-speaking cancer patients and published in the past 10 years;

Step 6. **Synthesis** of evidence collected at Steps 1 through 5 **aimed at developing recommendations** for optimal PROM(s) for use by the Cancer Institute NSW (**research question 2**) **and identifying PROMs** especially well **suited to measuring outcomes from particular types of intervention** (**research question 3**).

Results and discussion

Research question 1

Altogether, **174 RCTs of psychosocial interventions** were identified that had been conducted with English-speaking cancer patients and published since 1999. Of these, **133 anxiety, depression and/or distress (using 32 PROMs), while 102 measured quality of life (using 18 PROMs)**; **62 studies measured both** psychological and quality of life outcomes.

The **5 most commonly used PROMs for measuring anxiety, depression and/or distress** were:

- Profile of Mood States (POMS; various versions);
- State-Trait Anxiety Inventory (STAI);
- Centre for Epidemiologic Studies Depression Scale (CES-D);

- Hospital Anxiety and Depression Scale (HADS); and
- Impact of Event Scale – Revised (IES-R).

The **5 most commonly used HRQoL measures** were:

- the core measure from the Functional Assessment of Chronic Illness and Therapy (FACIT) suite, the FACT-G;
- the core measure from the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ) suite, the EORTC QLQ-C30;
- Medical Outcomes Survey Short Form Health Survey-36 (SF-36v2);
- Cancer Rehabilitation and Evaluation Systems Short Form (CARES-SF); and
- Rotterdam Symptom Checklist (RSCL).

The FACT-G and QLQ-C30 together accounted for nearly half the instances of HRQoL measurement.

Research question 2

Anxiety, Depression and Distress

With regard to measures of anxiety, depression and distress, **the HADS** (Hospital Anxiety and Depression Scale) **scored highest** (weighted score = 72.5/100) due to its supreme efficiency in assessing all three constructs within 14 items and the wealth of evidence for its reliability, validity and ability to identify treatment effects from psychosocial interventions in cancer. However, there is a concern that the HADS-D's emphasis on anhedonia may reduce sensitivity to depressive disorders other than major depression.

The POMS-37 (Profile of Mood States) – an **unofficial short-form** of the POMS-65 - **rated next highest** (weighted score = 60). Unlike the HADS, the POMS-37 is also free to use. However, while evidence for convergent validity suggests the POMS assesses constructs closely related to clinical anxiety and depression, criterion validity has not been evaluated against the gold standard of a diagnostic interview. Like the HADS, the POMS emphasises anhedonia rather than offering broad coverage of depressive symptoms.

The **CES-D** (Center for Epidemiological Studies Depression Scale) **performed well as a measure of depression** (weighted score = 55). However, the fact it uses 20 items to assess a single construct makes it **inefficient**; a 15-item short form shows promise but has not been extensively validated.

The IES-R (Impact of Event Scale - Revised) and **PCL-C** (Post Traumatic Stress Disorder Checklist – Civilian version) **provide information about PTSD**, a psychological construct that is **substantially different from** that assessed by **general distress** measures. Of these two measures, the **PCL-C scored highest** (weighted score = 37.5) due to greater evidence for its validity in cancer, including criterion validity against a diagnostic interview for PTSD. The PCL-C could be added to the battery to provide supplementary information about PTSD should this be considered useful.

HRQoL

Overall, the **FACT-G scored highest** (weighted score = 90) among PROMs designed to evaluate HRQoL. The **EORTC QLQ-C30** was rated a **close second** (weighted score = 80).

Both PROMs are free to use, assess similar domains of physical, emotional, social and functional/role wellbeing and functioning and can be supplemented with 'modules' specific to various cancer types and treatments.

The relative suitability of the FACT-G versus QLQ-C30 for a given application rests principally on their different approaches to scaling. The FACT-G combines all of its items into its core domains and enables these to be summed to produce an overall score, whereas the QLQ-C30 provides scores for a number of symptoms in 15 separate scales and produces an overall score from summation of just two items. **Advantages of the FACT-G's approach include ease of analysis and reduced problems of multiple hypothesis testing** together with **increased precision when using the overall score**. The **QLQ-C30**, on the other hand, **offers greater flexibility** where outcomes of interest are likely to vary across studies and **include specific symptoms** as well as core HRQoL domains.

Future directions

Computer adaptive testing (CAT) is predicted to replace assessment of PROs by standardised, static questionnaires in years to come. CAT offers **increased precision for a given number of items, reduced floor and ceiling effects, avoidance of uninformative and clinically irrelevant questions** that otherwise unnecessarily burden patients and researchers **and the ability to adapt measures to each study without compromising comparability**. Both the EORTC and US National Institute of Health (NIH) have embarked on programs to develop, calibrate and validate item banks covering anxiety, depression and many HRQoL domains for use in CAT. The NIH initiative, entitled the Patient Reported Outcomes Measurement Information System (PROMIS), is developing generic as well as cancer-specific item banks. Neither EORTC nor PROMIS item banks are ready for the Cancer Institute NSW to begin using CAT immediately. However, it is likely that CAT will become an attractive option in the future. Fortunately, likely calibration of item banks against popular PROMs like the HADS, FACT-G and CES-D raise the potential for the Cancer Institute NSW to transition to CAT at a later time while retaining the facility to compare new with previous results.

Research question 3

Both the **FACT-G and HADS have proven track records in identifying treatment effects across a variety of psychosocial interventions**, suggesting they will likely be sufficient for evaluating outcomes regardless of the intervention chosen by the Cancer Institute NSW.

Recommendations

Results from the critical review can be used to make the following recommendations:

Recommendation 1. That the Cancer Institute NSW include in its battery of PROMs the **HADS** as an overall measure of anxiety, depression and distress. Where cost is a concern,

the HADS could be substituted with the **free-to-use POMS-37** unofficial short-form as a measure of anxiety and depression. However, its total mood disturbance score (TMDS) is generated through summation of too various scales to be recommended as a measure of distress across the full spectrum of clinical contexts;

Recommendation 2. That the Cancer Institute NSW include in its battery of PROMs the **CES-D** where depression is an outcome of specific interest;

Recommendation 3. That the Cancer Institute NSW include in its battery of PROMs the **FACT-G** as a measure of HRQoL.

Conclusion

The findings of the report address the research questions posed above and provide information to assist the Cancer Institute NSW in selecting optimal patient-reported outcome measures for evaluation of future system-based interventions aimed at improving the wellbeing of people treated for cancer in NSW.

1. Background and rationale for the review

In May 2009, the Cancer Institute New South Wales (NSW) was developing plans to evaluate new system-based interventions aimed at improving physical and psychosocial morbidity in people treated for cancer in NSW. A robust rationale was needed for the Institute's choice of optimal patient-reported outcome measures (PROMs). The Cancer Institute NSW commissioned the Psycho-oncology Co-operative Research Group (PoCoG) to design and conduct a critical review and recommend optimal measures based on appraisal of relevant information in the public domain.

After preliminary discussion, it was decided to limit the outcome measures of interest to those assessing anxiety, depression, distress and health-related quality of life (HRQoL). In the psycho-oncology literature, the terms anxiety and depression are generally used to refer to clinical diagnoses (see Appendix A for DSM-IV-TR (1)criteria for generalised anxiety disorder and major depressive disorder). However, they are sometimes used in a more lay sense, especially when assessed by direct questioning in some PROMs (e.g. "have you been anxious/depressed?") or used to describe mood states. Distress is a more general construct that is defined by the National Comprehensive Cancer Network (NCCN) as:

"a multifactorial unpleasant emotional experience of a psychological (cognitive, behavioural, emotional), social, and/or spiritual nature that may interfere with the ability to cope effectively with cancer, its physical symptoms and its treatment. Distress extends along a continuum, ranging from common normal feelings of vulnerability, sadness, and fears to problems that can become disabling, such as depression, anxiety, panic, social isolation, and existential and spiritual crisis" ((2) p.6).

Because we were interested in questionnaires' performance as outcome measures rather than in screening, the decision was made not to limit instruments a priori but rather to find out which measures were commonly used in clinical research and to compare them based on empirical evidence. While evidence for criterion validity of anxiety, depression and distress measures against the 'gold standard' of diagnostic interviews was considered important in establishing that PROMs assessed a construct of clinical importance, this was not prioritised over their ability to identify treatment effects from psychosocial interventions. In this sense, the current review was more data driven than a previous report on measures of distress commissioned by Australia's National Breast Cancer Centre and conducted by Anthony Love at Monash University (3). Love's review focused on the performance of instruments in identifying distress in women with breast cancer and made a number of a priori exclusions on the grounds that instruments were perceived not to measure clinical constructs. Because we intended to review a larger number of instruments than Love, the decision was also made to impose stricter criteria with regard to psychometric properties, namely that evidence should come primarily from research carried out with English-speaking cancer patients.

A number of definitions currently exist for HRQoL, none of which is universally accepted. However, there is general agreement that HRQoL is a multi- rather than uni-dimensional

construct (4). When operationalising HRQoL for the purposes of outcome measurement, the core components are most commonly defined as physical, emotional and social wellbeing and functioning (5).

These limits, agreed prior to the beginning of the review, are reflected in the Methods section below.

2. Research questions

RQ1. Which patient-reported outcome measures (PROMs) have been used to measure outcomes of anxiety, depression, distress and HRQoL in randomised controlled trials (RCTs) of psychosocial interventions carried out with English-speaking cancer patients and published within the past 10 years?

RQ2. Which of these candidate PROMs will be optimal for use by the Cancer Institute NSW in evaluating new interventions as judged against the following criteria?

ii. Suitability for use by people undergoing active treatment for cancer of any type and stage;
iii. Efficiency, relating to number of items and length of time taken to assess each construct;
iv. Ease of administration via touch-screen computer;
v. Demonstrated reliability and validity in English-speaking people with cancer;
vi. Demonstrated ability to identify treatment effects in RCTs of psychosocial interventions carried out with English-speaking cancer patients and published within the past 10 years;
vii. Ease of interpretability regarding clinical cut-offs for anxiety, depression and/or distress and/or the importance of differences between mean scores or change in mean scores over time;
viii. Availability of comparison data from cancer and general populations;
ix. Practical issues relating to availability in languages other than English (LOTE) and cost.

RQ3. Are there PROMs that have a proven track record in evaluating outcomes from a wide variety of psychosocial interventions, or will different (or a combination of) PROMs be needed to evaluate certain intervention types?

3. Method

3.1 Overview

During the period from June to August 2009, staff at PoCoG in collaboration with the Cancer Institute NSW designed and undertook a critical review, which was conducted via the following methodology.

Step 1. **Identification of** all anxiety, depression, distress and HRQoL **questionnaires** used to measure outcomes from RCTs of psychosocial interventions carried out with English-speaking cancer patients and published in the past 10 years, **thus answering research question 1**;

Step 2. **First filter of candidate questionnaires** aimed at selecting those showing sufficient promise to warrant further investigation. Criteria included: 1) a time to administer of 10 minutes or less for scales assessing anxiety, depression, distress or HRQoL; 2) amenability to administration via touch-screen computer; 3) suitability for use by people undergoing active treatment for cancer of any type and stage; and 4) availability of evidence for reliability and validity in an English-speaking cancer population;

Step 3. **Description of selected candidate PROMs** via collation of information about the authors, content, number of items, time to administer, response options, recall period, scoring, licensing requirements and costs of selected candidate PROMs;

Step 4. Detailed **review of the evidence for reliability and validity** of each PROM in the cancer setting;

Step 5. **Review of the evidence for** the capacity of **each selected questionnaire to detect important effects of treatment in RCTs of psychosocial interventions** carried out with English-speaking cancer patients and published in the past 10 years;

Step 6. **Synthesis** of evidence collected at Steps 1 through 5 aimed at developing recommendations for optimal PROM(s) for use by the Cancer Institute NSW (**research question 2**) and identifying PROMs especially well suited to measuring outcomes from particular types of intervention (**research question 3**).

3.2 Procedures

3.2.1 Step 1: *Identification of all anxiety, depression, distress and HRQoL questionnaires used to measure outcomes from RCTs of psychosocial interventions carried out with English-speaking cancer patients and published in the past 10 years*

RCTs were chosen as the primary means to identify candidate PROMs because they represent best practice in evaluative research and can be expected to provide information about the ability of PROMs to identify treatment effects that is relatively free from the limitations inherent in evidence from other designs and methodologies.

A systematic review of the literature was conducted to identify all reports of RCTs designed to evaluate psychosocial interventions carried out with English-speaking cancer patients and published in the past 10 years. Interest was limited to evidence from English-speaking countries to minimise problems in generalising results across languages and cultures; limiting results to RCTs conducted only in Australia was predicted to yield too few studies to be informative. Searches were limited to RCTs published between 1999 and 2009 to reduce the prominence of measures that were once popular but have become obsolete in recent years. This decision was informed by the twin assumptions that PROM methodology has developed and more measures have become available over time.

3.2.1.1 Systematic review strategy

Database search

Resources searched

MEDLINE, PSYCHINFO, EMBASE, AMED, CENTRAL and CINAHL were searched in May, 2009.

Search terms

Psychosocial interventions were searched for using descriptions collated by Jacobsen and Jim in 2008 (6) from guidelines for the management of distress in people with cancer published by: a) the NCCN (7) and b) the National Breast Cancer Centre and National Cancer Control Initiative of Australia (8). This list was supplemented where necessary with search terms used by other systematic reviews (6, 9-16), most notably those relating to mindfulness and acceptance commitment therapy (ACT). See Table 1, (adapted from Jacobsen and Jim (2008) (6) (p. 217) for the resulting list of psychosocial interventions and descriptions.

Appendix B provides lists of terms used to search each database. Where available, medical subject headings (MeSH) or similar were used to identify further, narrower, related and alternative terms used to categorise articles in each database. Relevant terms were 'exploded' to include subordinate categories wherever possible. Depending on a database's architecture, each super-ordinate category may have several layers of subordinate categories. Exploded terms are reported in Appendix B at the first layer only. Categories were not searched again once these had been identified and included within a previous search from which the results were to be combined. Regardless of the availability of MeSH terms, databases were generally searched using a keyword search comprised of terms derived from Table 1. EMBASE was exempt from this rule because the relevant, superordinate EMTREE facet was considered sufficiently broad to negate the need for further keyword searches.

Table 1: Descriptions of common psychosocial interventions

Behavioural therapy	A type of psychotherapy that focuses on identifying problematic behaviours and replacing them with more

	adaptive behaviours.
Complementary and alternative medicine	Including qigong, tai chi, yoga, aromatherapy and reike.
Cognitive therapy	A type of psychotherapy that focuses on recognising and changing maladaptive thought patterns to reduce negative emotions and facilitate psychological adjustment. Therapeutic techniques include mindfulness.
Cognitive-behavioural therapy	A type of psychotherapy that focuses on recognising and changing maladaptive thoughts and behaviours to reduce negative emotions and facilitate psychological adjustment.
Communication skills training	A set of techniques used to modify verbal and nonverbal interactions with the goals of reducing interpersonal conflict and increasing the accuracy of information exchanged.
Counselling (including support)	Generic term used to refer to psychosocial care provided by a qualified professional.
Education/psycho-education	Provision of information through print, audiovisual, or interpersonal channels designed to increase knowledge of a subject area and reduce uncertainty.
Exercise	When targeting anxiety, depression, distress or HRQoL
Expressive therapy (music therapy/art/writing/dance)	Use of music / art / drama / writing as an outlet for emotional expression.
Family therapy/counselling	A type of psychotherapy that focuses on modifying problematic interactions within a family through conjoint sessions with family members.
Guided imagery/meditation/hypnosis	Structured meditative activity using mental imagery to facilitate relaxation.
Problem-solving/coping therapy	A type of psychotherapy that focuses on generating, applying, and evaluating solutions to identified problems.
Psychotherapy	Generic term used to refer to psychosocial care provided by a qualified professional.
Relaxation training	Techniques for releasing physical or mental tension that may involve meditative activities, progressive tensing and relaxing of muscle groups, or use of guided imagery.
Screening/referral	Screening and referral for anxiety, depression, distress or reduced HRQoL, including clinical monitoring.
Stress management training	Techniques for managing stress that may include relaxation training, breathing exercises, or use of internal monologues
Support group	Meetings that may/may not be facilitated by a professional at which individuals discuss issues of common concern
Supportive-expressive group therapy	A type of psychotherapy that focuses on expression of emotions in a supportive group environment to reduce negative emotions and promote psychological adjustment

Manual searches

Further, relevant RCTs were sought via the reference lists of recent reviews of psychosocial interventions and patient-reported outcome assessment in oncology (6, 9-14, 16-28).

3.2.1.2 Study selection criteria and procedure

Articles returned by the above searches were searched manually to identify those complying with the following criteria.

Inclusion criteria

In order to be included, reports had to relate to:

1. an English-language RCT designed to evaluate a psychosocial intervention for people living with cancer. Where reports did not describe the language(s) and location(s) in which the research was carried out, reference was made to the institutions to which authors were affiliated. Affiliations to institutions in Australia, New Zealand, USA, UK, Ireland or Canada were assumed to be suggestive of the involvement of English-speaking participants. International studies were included provided they included a sample from at least one English-speaking country. Psychosocial interventions were defined according to the classification presented in Table 1;

2. results from a PROM reportedly used to assess anxiety, depression, distress or quality of life. To qualify as a measure of anxiety or depression, a questionnaire had only to include one or more scales ostensibly used to assess these constructs; content analysis was delayed until Step 3 and evaluation of measurement properties until Step 4. Because distress is an umbrella term commonly used to describe any unpleasant emotional experience (2), outcomes described in terms of mood, emotion/affect or stress were also included. Single item measures described as assessing the constructs of anxiety, depression or distress were also included. PROMs measuring quality of life were required to enable separate scores for physical, psychological/emotional and social functioning/wellbeing as a minimum (5). This rule was intended to exclude cases where 'quality of life' was used by authors to refer to only one or two limited dimensions of wellbeing (e.g., symptom status or sexual functioning).

Exclusion criteria

Specific exclusions related to reports of:

1. interventions aimed at improving outcomes in persons caring for someone with cancer;
2. interventions aimed exclusively at improving outcomes in persons cured of cancer or living with cancer more than 1 year post diagnosis and not currently on active treatment;
3. interventions aimed at improving outcomes in samples including people both with and without cancer, unless results were reported for those with cancer separately; and
4. outcomes relating to clearly defined psychological constructs other than anxiety, depression or distress such as coping, adjustment or self-esteem.

3.2.1.3 Data extraction

Information on the type of psychosocial intervention and PROM(s) used to measure anxiety, depression, distress and/or HRQoL were extracted from articles meeting the inclusion criteria. Where results were reported from the same RCT in more than one publication, articles were reviewed in combination as necessary. Where intervention types were administered in combination, intervention type was categorised according to the predominant content.

3.2.2 Step 2: *Appraisal of each candidate questionnaire aimed at selecting those showing sufficient promise to warrant further investigation.*

Samples of candidate PROMs identified in Step 1 were obtained together with corresponding manuals and information on relevant websites. Manuals and websites were searched for details of current, alternative versions (e.g., short-forms) that could also be included in the review. Where a PROM had been replaced by a revised version, the most recent version formed the primary focus of the review.

Each PROM was evaluated against the following criteria. Reviewing of any given PROM ceased once failure to meet a single criterion was confirmed.

3.2.2.1 Criterion 1: Time to administer

A time to administer of 10 minutes or less for scales assessing anxiety, depression, distress or HRQoL as defined above. In the brief given to PoCoG, the Cancer Institute NSW indicated that the whole battery, including measures of all constructs of interest, should take no more than 20 minutes to administer. A maximum of 10 minutes per construct was chosen to allow for flexibility in how the 20 minute maximum could be apportioned in the final battery. Times to administer whole measures were, wherever possible, taken from PROM manuals or official websites. Where no information about time to administer could be found, it was estimated as the longer of two times taken in independent 'practice runs' by separate reviewers. Times to administer scales of interest were then estimated according to their proportion of the whole instrument.

3.2.2.2 Criterion 2: Suitability for touch-screen administration

Amenability to administration via touch-screen computer. Decisions were made based on PROM item structure and response options. Linear or visual analogue scales (LASAs/VASs) were regarded as incompatible with computer administration because no report could be found of these having been successfully administered by computer, the programming requirements for this type of response option are relatively complex and there were concerns that the validity of VASs and LASAs might be compromised by changing their mode of administration.

3.2.2.3 Criterion 3: Suitability for use in cancer patients

Suitability for use by people undergoing treatment for cancer of any type and stage. For instruments not designed to be cancer-specific, this was evaluated by a review of the contents of each questionnaire rather than through reference to the populations in which the PROM had been previously used. PROMs suitable for use only by long-term survivors, people with cancers of specific types and/or stages or people undergoing certain treatments were excluded.

Items used to assess anxiety, depression or distress that related to somatic symptoms (tiredness, insomnia, tingling/numbness, nausea, loss of appetite, weight loss, trembling, difficult breathing, feeling hot, psychomotor retardation) or asked about memory, concentration, restlessness or interest in sex were considered problematic because responses might be confounded by symptoms from cancer or side-effects from treatment. Items asking about preoccupation with health and thoughts of dying (other than suicide) were also regarded as disadvantageous given that these are likely ubiquitous in certain cancer sub-groups. Such items were considered grounds for exclusion if they constituted one third or more of scale items.

3.2.2.4 Criterion 4: Availability of evidence for reliability and validity

Availability of evidence for reliability and validity of an English language version of the questionnaire in a cancer population was required because evidence for validity or reliability cannot be generalised beyond the particular populations and clinical contexts in which these properties have been tested.

Relevant articles reporting psychometrics were identified via reference to manuals and websites and through further searches of MEDLINE and PSYCHINFO. Databases were searched using the name and acronym of each candidate PROM as keywords combined with the medical subject heading (MeSH) terms ‘neoplasms’ and ‘psychometrics’ and the key words ‘neoplasm$’ and ‘psychometric$ OR valid$ OR reliab$’. Further articles were identified via the reference lists of relevant articles and reviews.

Research reports of many kinds have potential to provide insights into PROM construct validity or responsiveness. However, in order to make optimal use of time and resources, the decision was made to focus exclusively on reports giving information about two or more psychometric properties.

Evidence for reliability was defined as data and analyses relating to internal consistency, test-retest reliability or inter-rater reliability between patients and proxies. Types of validity considered of interest were content, internal, convergent/divergent, discriminant, criterion and predictive. In the case of measures of anxiety and/or depression, comparison with the ‘gold standard’ of a diagnostic interview was prioritised as evidence that scores from a given PROM are of clinical value. While evidence of this type is generally reported within the context of screening, it also has important implications for interpretation of data when a questionnaire is used as an outcome measure.

As in Step 1, interest was confined to evidence relating to the English-language version of each PROM.

3.2.3 Step 3: *Collection of information about candidate PROMs.*

For each PROM meeting the 4 criteria in Step 2, information about the authors, content, number of items, time to administer, response options, recall period, scoring, translation availability, licensing requirements and costs of each was extracted from relevant articles, websites, publications and manuals. The content of PROMs was compared with DSM-IV-TR criteria (1) for generalised anxiety disorder and major depressive disorder (see Appendix A) as a method complementary to psychometric evidence reviewed at Step 4 for delineating measures of general distress from those of anxiety and depression. A thesaurus was used to map between terms with similar meanings.

3.2.4 Step 4: *Review of evidence for reliability and validity.*

Evidence for validity or reliability in English-speaking cancer populations identified by means of the above searches was evaluated using a checklist adapted from the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (29) and another checklist developed for the Dementia Outcomes Measurement Suite (DOMS) project (30) (see Appendix C).

Evidence for each property was independently evaluated by two reviewers until inter-rater reliability (Kappa > 0.60) was achieved on at least 25 pairs of ratings. Any disagreements were resolved through discussion. Once inter-rater agreement had been established, the property in question was assessed by one reviewer only but other properties continued to be evaluated by both reviewers until Kappa > 0.60 was achieved.

Where pooled or meta-analyses were identified, these were reviewed instead of the cited, original articles themselves. Articles reporting on the reliability and validity of non-English versions of PROMs were reviewed as a secondary source of evidence. These were not rated using the checklist but were reviewed by a single researcher to identify findings likely to be important across language versions.

3.2.5 Step 5: *Review of evidence for the capacity to detect important effects of treatment in RCTs of psychosocial interventions.*

Articles returned in the systematic review conducted for Step 1 were reviewed to identify all RCTs of psychosocial interventions published since 1999 in which English-language versions of questionnaires meeting Step 2 criteria had been used.

3.2.5.1 RCT selection

Reported RCTs were subject to further exclusion criteria aimed at reducing limitations due to flawed design and methodology. Studies were excluded if they:

i. had an overall sample size of less than 10, which were considered too small for meaningful quantitative analysis; or

ii. had an attrition of more than 15% of the baseline sample (31), except where the authors provided evidence that the drop-out rate was of a similar magnitude and for similar reasons across arms and therefore unlikely to bias estimates of the effect of the intervention. Where attrition varied at different follow-up assessments, only those time-points where attrition was less than 15% were considered. Attrition was deemed to be of less concern in RCTs that used more than one candidate PROM

because these were considered to offer important information about relative performance under the same conditions, however compromised these may have been.

Reports were reviewed to identify instances where a candidate questionnaire had demonstrated an important treatment effect, defined as an effect size (ES) of at least 0.2. We used ES because it provides a standard unit for comparison across studies and, unlike statistical significance, it is independent of sample size. Conventionally, an ES of 0.2 is considered small, 0.5 moderate, and 0.8 large (32). A small ES was chosen based on previous work in which we found recommendations of 0.5 SD to be too conservative, at least regarding longitudinal change (33-35). Wherever possible, ESs were calculated as the difference between the mean change (from baseline) in the treatment group and comparison group, divided by the standard deviation at baseline (pooled across both groups). Where means and SDs were reported only at follow-up, ESs were calculated as the difference between follow-up means divided by the pooled follow-up SD. In these cases, however, evidence was required that the groups did not vary significantly in key characteristics at baseline.

In general, studies where no ES $\geq$0.2 was identified were treated as uninformative. This is because failure to identify a treatment effect may arise not only from the inadequacy of an outcome measure to detect a real difference but also:

i. inadequacy of the intervention to induce a real difference; or
ii. inadequacy of the RCT design and/or methodology to induce and/or detect a real difference, which could be due to a range of flaws including lack of power.

Inadequacy of a given PROM was inferred only where an ES of $\geq$0.2 was observed on another outcome measure assessing a similar or related construct. For this reason, RCTs where more than one candidate PROM had been used were considered especially informative.

3.2.5.2 Data extraction

Variables for each RCT and intervention were extracted from each report and tabulated as described in Table 2.

Table 2: Variables extracted from each report in Step 5

Parameter	Variables extracted
Sample	Type(s) of cancer
	Stage(s) of cancer
	Gender (% female)
	Mean/median age
	Country or countries
Intervention	Type, as defined in Table 1
Context of intervention	Service delivery context (e.g., out-patient)
	Treatment context (e.g. during active treatment, post-treatment)
Sample size	Numbers of participants at baseline
	Numbers of participants at each time-point
Outcomes	PROMs and any other, non patient-reported outcome measures reported
Results	Effect sizes for each candidate PROM or, where these were not reported, information needed to calculate these (e.g. pre- and post-intervention means and baseline standard deviations)

3.2.6 Step 6: *Synthesis of findings and discussion.*

Candidate PROMs from each category were systematically compared against the criteria outlined for **research question 2** above. The initial filtering process removed PROMs not suited for use by people undergoing active treatment for cancer of all types and stages, taking 10 minutes or longer to administer, not suitable for administration via touchscreen computer and lacking any evidence for reliability and validity in English-speaking cancer patients. Remaining PROMs were now rated on their efficiency (number of constructs assessed and number of items needed in each case), their estimated ease of comprehension when administered via touch-screen, the amount and consistency of evidence for reliability and validity and their track record in identifying treatment effects in RCTs of psychosocial interventions. When reviewing evidence for validity and performance as outcome measures, special prominence was given to results from studies that allowed two or more PROMs to be directly compared. For measures of anxiety, depression or distress, ease of interpretability was evaluated against information on cut-offs relative to a clinical diagnosis; information about minimally important differences (MIDs) derived from anchor-based or distribution-based approaches were prioritised for HRQoL instruments. The availability of comparison data was also considered to support interpretability regardless of PROM type. A summary score for each candidate PROM was generated through discussion by two reviewers using the values presented in Tables 3 and 4. Weightings for each property were developed in consultation with staff at the Cancer Institute NSW to ensure overall scores reflected the needs of their specific application. Ratings were intended to rank the PROMs rather than place them on an interval scale.

Table 3. Criteria and weights for use in generating overall scores for measures of anxiety, depression or distress

Evaluation Criteria	**Scoring system**	*Score*	*Weight*	*Weighted Score*
Number of psychological constructs assessed	0 = assesses anxiety OR depression OR distress 5 = assesses anxiety and distress OR depression and distress OR anxiety and depression 10 = assesses anxiety AND depression AND distress		1	
Length of each scale of interest	0 = 20+ items 5 = 11-19 items 10 = 10 items or less		1	
Complexity of computer administration and cognitive burden to patients	0 = demanding to understand or computer administer 5 = some difficulties to understand or computer administer 10 = easy to understand and computer administer		1	
Reliability in English-speaking cancer populations	0 = no evidence or reliability poor 5 = evidence for reliability inconsistent or from 1 or 2 studies only 10 = generally consistent evidence for reliability from several studies		1.5	
Validity in English-speaking cancer populations	0 = no evidence or validity poor 5 = evidence for validity inconsistent or from 1 or 2 studies only 10 = generally consistent evidence for validity from several studies		1.5	
Performance as a screening instrument as judged against comparison with the gold-standard of a diagnostic interview	0 = no evidence or screening performance unsatisfactory 5 = evidence for screening performance inconsistent or from 1 study only 10 = generally consistent evidence for screening performance from more than 1 study		1.5	
Proven ability to identify treatment effects in RCTs of psychosocial interventions in English-speaking cancer populations	0 = no evidence 5 = ES $\geq$ 0.2 identified in 1 to 3 RCTs 10 = ES $\geq$ 0.2 identified in 4 + RCTs		1.5	
Availability of comparison data	0 = no or minimal comparison data available 5 = substantial comparison data available from large-scale studies 10 = Comparison data available for subgroups (e.g., age and gender) in cancer and general populations		1	
OVERALL SCORE		**/80**	**N/A**	**/100**

Table 4. Criteria and weights for use in generating overall scores for measures of HRQoL

Evaluation Criteria	**Scoring system**	*Score*	*Weight*	*Weighted Score*
HRQoL domains assessed beyond core physical, psychological and social functioning/wellbeing	0 = includes scales for physical, emotional and social functioning/wellbeing only 5 = includes scales for a range of other domains/symptoms OR global QOL 10 = includes scales for range of other domains/symptoms AND global QOL		1	
Length of core domains combined	0 = 40+ items 5 = 21 – 39 items 10 = 20 items or less		1	
Complexity of computer administration and cognitive burden to patients	0 = demanding to understand or computer administer 5 = some difficulties to understand or computer administer 10 = easy to understand and computer administer		1	
Reliability in English-speaking cancer populations	0 = no evidence or reliability poor 5 = evidence for reliability inconsistent or from 1 or 2 studies only 10 = generally consistent evidence for reliability from several studies		1.5	
Validity in English-speaking cancer populations	0 = no evidence or validity poor 5 = evidence for validity inconsistent or from 1 or 2 studies only 10 = generally consistent evidence for validity from several studies		1.5	
Availability of evidence to aid interpretation of scores in terms of minimally important difference (MID)	0 = no/very limited evidence available to aid interpretation 5 = some evidence to aid interpretation 10 = substantial evidence to aid interpretation in a range of cancer populations		1.5	
Proven ability to identify treatment effects in RCTs of psychosocial interventions in English-speaking cancer populations	0 = no evidence 5 = ES ≥ 0.2 identified in 1 to 3 RCTs 10 = ES ≥ 0.2 identified in 4 + RCTs		1.5	
Availability of comparison data	0 = no or minimal comparison data available 5 = substantial comparison data available from large-scale studies 10 = Comparison data available for subgroups (e.g., age/gender) in cancer and general pops.		1	
OVERALL SCORE		**/80**	**N/A**	**/100**

After an initial comparison across intervention types, the performance of candidate PROMs in RCTs of each type of intervention were compared in order to identify those especially suited to measuring outcomes from particular types of intervention (**research question 3**). Since we expected to identify only a relatively small number of RCTs for each type of intervention, a category system was devised wherein similar intervention types were clustered together to allow comparison across a reasonable sample. Three

clinical psychologists on the PoCoG team (Phyllis Butow, Nicole Rankin and Melanie Price) organised the types of interventions into categories as outlined in Table 5.

Table 5. Clusters of intervention types devised by three clinical psychologists

Cluster	Types of interventions included
SG/CP	Support group, counselling, psychotherapy and family therapy
Education	Education and psycho-education
CBT/P/S	Cognitive-behavioural therapy, cognitive therapy, problem-solving/coping therapy and stress management training
E/CAM-P	Exercise and physical types of complementary and alternative medicine such as yoga and tai chi
CAM-NP	Non-physical kinds of complementary and alternative medicine such as aromatherapy, expressive therapy and supportive-expressive group therapy
Screening	Screening and referral interventions

4. Results and discussion

4.1 Step 1: *Identification of all anxiety, depression, distress and HRQoL questionnaires used to measure outcomes from RCTs of psychosocial interventions carried out with English-speaking cancer patients and published in the past 10 years.*

Altogether, we identified **174 RCTs** of psychosocial interventions in English-speaking cancer patients published since 1999. Of these, **133 measured anxiety, depression and/or distress using a total of 32 PROMs while 102 measured quality of life by means of 18 PROMs**; **62 studies used both psychological and quality of life measures**. Tables 6 and 7 summarise PROMs used in RCTs to assess anxiety, depression, distress and quality of life.

PROMs excluded from Table 6 on the basis that they assess psychological constructs other than anxiety, depression or distress were:

- Courtauld Emotional Control Scale (CECS) (36);
- Differential Emotions Scale (DES) (37);
- Perceived Stress Scale (PSS) (38);
- Mental Adjustment to Cancer (MAC) (39);
- Mood Scale, Positive States of Mind (PSOM) (40);
- Measure of Current Status (MOCS) (41).

Table 6. PROMs used to assess anxiety, depression and/or distress in English-language RCTs of psychosocial interventions published since 1999

Name	Acronym	Scales	Ref.	N of RCTs
State-Trait Anxiety Inventory	STAI	Anx	(42)	30
Profile of Mood States (original)	POMS-65	Anx, Dep, Dis	(43)	25
Center for Epidemiologic Studies Depression Scale	CES-D	Dep	(44)	21
Hospital Anxiety and Depression Scale	HADS	Anx, Dep, Dis	(45)	20
Impact of Event Scale - Revised	IES-R	Dis	(46)	16
Beck Depression Inventory - II	BDI-II	Dep	(47)	10
Visual Analogue Scale for anxiety	VAS Anxiety	Anx	Ad hoc	10
Profile of Mood States - 30	POMS-30	Anx, Dep, Dis	(48)	6
Positive and Negative Affect Schedule	PANAS	Dis	(49)	5
Symptom Checklist – 90 - Revised	SCL-90-R	Anx, Dep, Dis	(50)	5
Mental Health Inventory - 18	MHI-18	Anx, Dep	(51)	3
Brief Symptom Inventory	BSI	Anx, Dep, Dis	(52)	2
Derogatis Affects Balance Scale	DABS	Anx, Dep	(53)	2
Distress Thermometer	DT	Dis	(54)	2
Hamilton Depression Rating Scale	HDRS	Dep	(55)	2
Patient Health Questionnaire - 9	PHQ-9	Dep	(56)	2
Post-traumatic Stress Disorder Checklist-Civilian Version	PCL-C	Dis	(57)	2
Profile of Mood States - 11	POMS-11	Dis	(58)	2
(Calgary) Symptoms of Stress Inventory	C-SOSI/ SOSI	Anx, Dep	(59, 60)	2
(Bradburn) Affect Balance Scale	ABS	Anx, Dep, Dis	(61)	1
Anxiety composite (PANAS, SF-12, Index of Clinical Distress)	N/A	Anx	Ad hoc	1
Beck Anxiety Inventory	BAI	Anx	(62)	1
Depression Anxiety Stress Scales	DASS	Anx, Dep, Dis	(63)	1
General Health Questionnaire-28	GHQ-28	Anx, Dep	(64)	1
Geriatric Depression Scale (short form)	GDS	Dep	(65)	1
Mental Health Inventory - 5	MHI-5	Dis	(66)	1
Profile of Mood States – 10	POMS-10	Dis	(67)	1
Profile of Mood States - 14	POMS-14	Dis	(68)	1
Profile of Mood States – 37	POMS-37	Anx, Dep	(69)	1
Profile of Mood States Bi-Polar Form	POMS-Bi	Anx, Dep, Dis	(70)	1
Somatic and Psychological Health Report	SPHERE	Dis	(71)	1
Visual Analogue Scale for distress	VAS Distress	Dis	Ad hoc	1

Anx = anxiety; Dep = depression; Dis = distress

Table 7. PROMs used to assess quality of life in English-language RCTs of psychosocial interventions published since 1999

Name	Acronym	Ref	N of RCTs
Functional Assessment of Cancer Therapy - General with/without FACIT modules	FACT-G + modules	(72)	33
European Organization for Research and Treatment of Cancer Quality of Life Questionnaire – Core with/without EORTC QLQ modules	EORTC QLQ-C30 + modules	(73)	17
Medical Outcomes Survey Short Form Health Survey-36	SF-36	(74)	16
Cancer Rehabilitation and Evaluation Systems, Short Form	CARES-SF	(75)	5
Rotterdam Symptom Checklist	RSCL	(76)	5
Medical Outcomes Study Short Form Health Survey-12	SF-12	(77)	4
Quality of Life Instrument (Breast Cancer Version)	QOL-Breast	(78)	3
Cancer Rehabilitation and Evaluation Systems	CARES	(79)	2
Functional Living Index - Cancer	FLIC	(80)	2
McGill Quality of Life Questionnaire	MQOL	(81)	2
Measure Yourself Medical Outcome Profile	MYMOP	(82)	2
Quality of Life Index - Colostomy	QLI-CP	(83)	2
Satisfaction With Life Scale	SWLS	(84)	2
Atkinson Life Happiness Rating Scale	N/A	(85)	1
Hospice Quality of Life Index-Revised	HQOLI-R	(86)	1
Linear Analogue Self Assessment	LASA	(87)	1
Quality of Life - Cancer	QOL-CA	(88)	1
Quality of Life Index - Cancer Version	QLI-CV	(89)	1

4.2 Step 2: *Appraisal of each candidate questionnaire aimed at selecting those showing sufficient promise to warrant further investigation.*

Searches of manuals and websites identified a number of alternative versions that were reviewed alongside the 51 PROMs identified in Step 1. Decisions to include or exclude each candidate version are summarised below.

4.2.1 Excluded PROMs

4.2.1.1 Initial exclusions

PROMs assessing distress

A number of multi-dimensional PROMs yield total scores that are used as distress measures, including the **SCLR-90-R, BSI-53/18, MHI-38, GHQ-28 and POMS-65/Bi-Polar/30** as well as the POMS unofficial short-forms **POMS-37.** These **total scores were discounted** because they are generated through summation of scales that assess a range of distinct psychological constructs, many of which include a majority of somatic items.

PROMs assessing HRQoL

PROMs listed in Table 7 that were ostensibly used to assess quality of life but were immediately excluded because they failed to provide separate scales for physical, emotional and social wellbeing were:

- Atkinson Life Happiness Rating Scale (85);
- Cancer Rehabilitation and Evaluation Systems and short-form (CARES/CARES-SF) (75, 79);
- Measure Yourself Medical Outcome Profile (MYMOP) (82);
- Rotterdam Symptom Checklist (RSCL) (76);
- Satisfaction with Life Scale (SWLS) (84).

4.2.1.2 Evaluation against 4 criteria

Remaining PROMs were then evaluated against the 4 criteria of:

> 1) time to administer of 10 minutes or less for scales assessing anxiety, depression, distress or HRQoL;
> 2) amenability to administration via computer;
> 3) suitability for use by all people undergoing active treatment for cancer; and
> 4) availability of evidence for reliability and validity in an English-speaking cancer population.

Table 8 gives information about PROMs excluded against at least one of these criteria. It is important to note that some PROMs may have failed to meet more than one criterion if given the opportunity; reviewing ceased when failure to meet one criterion had been confirmed.

All PROMs excluded on the basis of difficulty with computer administration used linear or visual analogue scales (LASA/VAS). A number of measures showing promise for cancer clinical research were excluded due to lack of evidence for reliability and validity in an English-speaking oncology sample. Most surprising was the **lack of evidence for the STAI,** which has been widely used in cancer both as an outcome measure and as a comparison measure for establishing convergent validity for new PROMs. A further PROM **excluded due to lack of cancer-related validity data was the DASS**, the only Australian-developed PROM identified in Step 1. Kate Neilson and colleagues at Peter MacCallum Cancer Centre are currently comparing the DASS and HADS with a diagnostic interview in patients with head and neck cancer; results are expected by the end of 2009 (personal correspondence 31/07/09).

Several instruments were excluded against criterion 3 on the basis that they were designed for specific groups (GDS; QLI-CP; QOL–Breast/Ovarian/Cancer Survivors) and/or clinical contexts (HQOLI-R; MQOL). Table 9 gives more information about the anxiety and depression scales excluded against criterion 3 on the basis of comprising one third or more somatic or other items that were considered likely to comprise their performance in at least some clinical contexts.

Table 8. PROMs excluded after failing to meet one of four criteria in Step 2

Measure	Criteria			
	1	2	3	4
ABS				✓
BAI			✓	
BDI-II			✓	
BDI-SF/PC				✓
DABS				✓
DASS				✓
FLIC		✓		
GDS			✓	
GHQ-30/60				✓
HDRS			✓	
HQOLI-R			✓	
LASA		✓		
LSQ				✓
MHI-18				✓
MQOL			✓	
PANAS				✓
POMS-30/10/Bi Polar				✓
PHQ-9			✓	
PHQ-2				✓
QLI-CP			✓	
QOL-CA		✓		
QOL–Breast/Ovarian/Cancer Survivors			✓	
SF-8				✓
STAI				✓
VAS for assessing anxiety, depression or distress		✓		

Table 9. Anxiety and depression scales excluded due to problematic items

PROM	Scale	Problematic items	Problematic/ total items
BAI	Anx	Feeling hot, wobbly, hands trembling, tingly, unsteady, dizzy, choking, breathing, fear of dying	10/21
BDI-II	Dep	Energy, tiredness, sleep, eating, concentration, restlessness, interest in sex	7/21
HDRS	Dep	Insomnia early, insomnia middle, insomnia late, psychomotor retardation, psychomotor agitation, anxiety somatic, appetite, gastrointestinal, somatic symptoms general, hypochondriasis, weight loss	11/17
PHQ-9	Dep	Sleep, fatigue, appetite, concentration, restlessness	5/9

4.2.2 Included PROMs

Table 10 summarises the time taken to administer relevant scales from PROMs meeting all 4 criteria.

Two PROMs, the Impact of Event Scale – Revised (**IES-R**) and Post Traumatic Stress Disorder Checklist – Civilian version (**PCL-C**), are worthy of note in that they have been developed to mirror the diagnostic criteria for post traumatic stress disorder (PTSD) rather than to assess distress more generally. That said, the IES-R and its forerunner the Impact of Event Scale (IES) are often described in the literature as assessing 'cancer-related distress'. Since there is ongoing debate about how PTSD differs from other psychological constructs in the cancer setting, the decision was made to include the IES-R and PCL-C in the review alongside other distress measures.

Table 10. PROMs meeting Step 2 criteria: Minutes to administer scales

Measure	Anxiety	Depression	Distress	HRQoL[#]
BSI-53/18	< 2*	< 2*		
CES-D-20		5		
CES-D-15		< 4*		
C-SOSI		< 1**		
DT			<1*	
EORTC QLQ-C30				6
FACT-G				5-10
GHQ-28	< 1*	< 1*		
GHQ-12			< 1*	
HADS	< 3	< 3	2-5***	
IES-R			2**	
MHI-5			< 1**	
MHI-38	< 2**	< 2**		
PCL-C			< 3**	
POMS-65	< 2*	< 2*		
POMS-37	< 2*	< 2*		
POMS-11			< 1*	
POMS-14			< 1*	
QLI-CV				8
SCL-90-R	< 2*	< 2*		
SF-12				2
SF-36				5-10
SPHERE			< 1**	

* Extrapolated from time to administer cited in manual; ** Estimated as the longer of two times taken to self-administer by independent reviewers; *** The distress scale on the HADS is generated by summing scales for anxiety and depression; [#] times to administer HRQoL measures are for the whole PROM.

4.3 Step 3: *Collection of information about candidate PROMs.*

Appendix D summarises information about the authors, domains, number of items, time to administer, response options, recall period, scoring, availability of translations, licensing requirements and costs of PROMs meeting all four criteria at Step 2.

Appendix E compares content of measures of anxiety, depression and distress with terms used in the DSM-IV-TR criteria (1). It is evident that there are both considerable overlaps and considerable differences in the symptoms assessed by different instruments.

Anxiety scales from the POMS-65 shared most descriptors with the DSM criteria. This is surprising because the POMS-65 was not developed to assess clinical anxiety but rather tension-anxiety as a mood state. The POMS is also the only instrument that uses the term 'anxious'; the MHI-38 includes this item in its distress scale only.

Conversely, 'depressed' is absent from the depression scale of the POMS-65 even though DSM criteria make repeated reference to 'mood'. The depression scale most resembling the DSM criteria was the CES-D and SCL-90-R.

Of the distress scales, only the GHQ-12 bears much resemblance to DSM criteria. POMS-65 items similar to the diagnostic criteria have not been retained in the POMS-14 or POMS-11.

The degree to which respondents distinguish between terms such as 'nervous' and 'restless' or 'blue' and 'sad' as separate items in the same questionnaires is not clear but can be inferred to at least some degree by results relating to internal consistency and structure where these were identified in Step 4.

As mentioned earlier, content of the IES-R and PCL-C are organised into similar scales and have both been designed to mirror the diagnostic criteria for PTSD.

The content of HRQoL PROMs was considered too different to match on an item-by-item basis. Instead, readers are recommended to review online samples of the FACT-G, EORTC QLQ-C30, QLI-CV and SF-36/12 at the websites identified in Appendix D.

4.4 Step 4: *Review of evidence for reliability and validity.*

In all, some 250 articles were identified that reported on the psychometric properties of candidate PROMs meeting all four criteria at Step 2. Of these, **85 articles** reported studies evaluating English-language versions and **were amenable to review using the checklist** in Appendix C.

Inter-rater reliability of Kappa > 0.60 between two reviewers was achieved for ratings of the following properties: internal consistency (Kappa = 0.71); criterion validity (Kappa = 0.82); discriminant validity (Kappa = 0.63); convergent validity (Kappa = 0.67). Other properties were reported too infrequently for inter-rater reliability to be properly assessed. In these cases, both reviewers continued to rate each report, and consensus was reached via discussion. Ratings of the 85 articles are summarised in Appendix F.

Several further articles were identified that were not amenable to rating by the checklist but included relevant information in the form of direct comparisons between candidate PROMs or advice on interpreting scores. Key issues raised in these articles are discussed in the relevant sections below.

4.4.1 Measures of anxiety and depression

Note - when comparing evidence for criterion validity between studies, it is important to remember that screening performance will be influenced by the prevalence of the condition - in this case anxiety or depression - in the samples involved (90).

4.4.1.1 Calvary Symptoms of Stress Inventory (C-SOSI)

The C-SOSI is a brief, cancer-specific version of the Symptoms of Stress Inventory (SOSI) (59) which was developed for use in chronic disease more generally. Both versions assess a range of symptoms, including anxiety and depression in the case of the SOSI and anxiety in the case of the C-SOSI. The SOSI has reportedly been validated in US patients with melanoma and myocardial infarction (91) but relevant data do not appear to have been published. The C-SOSI has been validated in only one study, where the depression scale showed satisfactory internal consistency (alpha = 0.90) and strong correlations with the depression-dejection scale of the POMS-65 (0.87) and emotional functioning scale (EF) of the EORTC QLQ-C30 (- 0.76) in Canadian patients with mostly breast cancer (N = 344)(60). The C-SOSI also showed satisfactory factor structure in this sample.

4.4.1.2 Center for Epidemiological Studies – Depression (CES-D-20/15)

The CES-D has performed well when compared to a diagnostic interview in small samples of patients with mixed (92) and head and neck (93) cancer diagnoses. The CES-D has also been found to predict survival in 205 cancer patients with cancer of various types and stages (94).

Three short-forms of the CES-D have been developed, a 10-item, an 11-item and 15-item version. Neither the 10- nor 11-item versions have been validated in cancer patients, though the 11-item version performed satisfactorily in a sample of disease-free breast cancer survivors (95). The 15-item version was developed using factor analysis after the two interpersonal items from the long-form were found uninformative and three further items were found to have an unacceptable gender bias in a large sample of patients with heterogeneous cancer diagnoses and caregivers (96).

4.4.1.3 Hospital Anxiety and Depression Scale (HADS)

Of the depression and anxiety questionnaires reviewed here, the HADS has received by far the most attention with regard to evaluation of reliability and validity in cancer. Intuitively at least, it has an immediate advantage for use in oncology in that it has been developed to avoid somatic items that might be confounded by symptoms from disease or treatment. This aptitude has been broadly confirmed in breast cancer by a factor analysis that found all items to load more strongly onto a psychological than somatic factor (97). However, the HADS' performance as a screening measure has not always followed the pattern one

might expect. While it has demonstrated good screening performance in samples with poorer health status and those on active treatment compared with those off active treatment, it has also performed surprisingly well in those who were disease free and relatively poorly in those with metastatic or progressive disease (98-100). Optimal cut-offs for both anxiety and depression have also varied widely between studies, with those recommended by the instruments' developers performing poorly in some cases (101). Claims that the HADS items may function differently in samples drawn from the general population (102), however, have not been supported by research carried out in Australia (103).

Several studies suggest the HADS may be better at screening for anxiety than for depression (101, 104, 105). The main limitation of the HADS depression scale (HADS-D) has been said to relate to its over emphasis on anhedonia which, while appropriate when screening for major depression, may fail to identify minor depression or adjustment disorder with depressed mood (3). Indeed, one study found the overall HADS score (HADS-T) to be a more effective screening tool than the HADS-D for identifying predominantly minor depression in patients with head and neck cancer (93). However, results from a meta-analysis across language versions of the HADS presented by Alex Mitchell at the 2009 World Congress of the International Psycho-Oncology Society (IPOS) present a somewhat different picture. Mitchell's meta-analysis found the HADS-D to be superior to both the HADS-A and HADS-T in ruling out cases of mixed affective disorders (depression, anxiety, adjustment disorders combined); all three scores performed relatively poorly in ruling in cases (fraction correct scores = HADS-D 78.3%; HADS-A 65.9%; HADS-T 72.6%) (106). The three scores were equivalent in screening for depression only, which they were also better at ruling out than ruling in (fraction correct scores = HADS-D 81.4%; HADS-A 81.8%; HADS-T 83.4%). Results from this meta-analysis suggest poorer performance than did many English-language studies reviewed in the current report, raising the possibility that the HADS may perform differently across languages. Results from Mitchell's meta-analysis of head-to-head comparisons of the HADS with other screening instruments are expected to be published in 2010 (personal correspondence 4/8/2009).

4.4.1.4 Profile of Mood States (POMS)

As mentioned earlier, the POMS-65 (original) and its unofficial short-form, the POMS-37, differ from most other anxiety and depression PROMs reviewed here in that they have not been designed or used to screen for psychological disorders but rather to assess depression-dejection and tension-anxiety as they are manifested in mood. The POMS-37 was developed specifically for use with cancer patients who may be too unwell to complete all 65 items from the original (69). It closely correlated with and matched the performance of the POMS-65 in three samples (69, 107, 108). The depression-dejection and tension-anxiety scales from both versions have generally shown satisfactory internal consistency in cancer, although Cronbach alphas in excess of 0.90 in at least one subgroup in most studies suggest some item redundancy, especially with regard to the depression-rejection scale (69, 107-109). Both versions have demonstrated satisfactory convergent and discriminant validity (107-109), and the POMS-37 satisfactory factor structure (109).

A further, 19-item, 3-scale version of the POMS was recently developed using factor analysis in US women with breast cancer (110) but reported details were considered insufficient for inclusion in the current review.

4.4.1.5 Symptom Checklist-90-Revised (SCL-90-R) and Brief Symptom Index (BSI-53/18)

The continued popularity of the SCL-90-R despite the longstanding availability of two short-forms (BSI-53/18) indicates a belief on the part of some researchers that it retains properties not replaced by the newer versions. The SCL-90-R's anxiety and depression scales differ from those of the BSI-53 and 18 in being 10 versus 6 and 13 versus 6 items respectively. Because the anxiety and depression scales of the BSI-53 and BSI-18 are identical, the current review treated these instruments as synonymous. For the sake of accuracy, however, evidence is reported separately for each version.

Despite their widespread use, none of the three versions available has been satisfactorily validated in cancer, especially with regard to criterion validity against the gold standard of a diagnostic interview. Both the SCL-90-R and BSI-53 have successfully discriminated between cancer patients who met versus did not meet criteria for clinical levels of distress via diagnostic interview, but sensitivity and specificity were not reported (111, 112). In another study, the BSI-53 correlated closely with the Omega Screening Instrument and Inventory of Current Concerns and predicted 16 of 19 (82%) cases of future distress as defined by scores on the POMS and Psychosocial Adjustment to Illness Scale (PAIS) collected 12 months later (113). The anxiety and depression scales of the SCL-90-R have also demonstrated satisfactory internal consistency and test-retest reliability in Canadian cancer patients, though the hostility scale was problematic (114). Evidence for reliability and validity of the BSI-18 in cancer is limited to supportive data on internal consistency and factor structure in a large sample of people with heterogeneous cancer diagnoses (115) and adult survivors of childhood cancer (116). In the last of these studies, scales of the BSI-18 showed strong correlations (0.88 to 0.94) with corresponding scales of the SCL-90-R. A further study, with brain tumour patients, confirmed that scales assessing somatisation, obsessive-compulsive and psychotic disorder are likely to be confounded by symptoms of disease and treatment in some cancer groups (117).

4.4.2 Measures of distress

4.4.2.1 Distress Thermometer (DT)

The DT is a single-item measure that has been recommended for routine use in cancer clinics by the National Comprehensive Cancer Network (NCCN) either alone or in combination with a more comprehensive 'Problem List' (7). The Problem List asks respondents to report whether they have experienced each of 35 practical problems, family problems, spiritual/religious concerns, physical problems and emotional problems over the past week. It is intended for use in screening patients for referral but so far has been validated neither for this purpose nor as an outcome measure.

Proposed advantages of the single-item DT lie not only in its brevity but also in its visual appeal and avoidance of the terms ‘anxiety’ and ‘depression’ in favour of the less pathologising descriptor ‘distress’ (118). This last attribute is deceptive since evaluation of the DT’s psychometric properties has focused almost exclusively on its ability to identify clinical caseness.

While a large number of studies have evaluated the DT against clinical cut-offs on more established PROMs of anxiety and depression, evidence for criterion validity of the English-language form as judged against the gold standard of a diagnostic interview is lacking. Mitchell *et al.* (2007) concluded from their pooled analysis of 19 international DT studies published up until November 2006 that the DT was modest to poor at ruling in cases of anxiety or depression (however defined) and generally better at screening for depression than anxiety (119). Performance in ruling out cases was reported to be somewhat better.

Evidence for the validity of the DT beyond criterion validity is scant, with only one study reporting on the convergent and discriminant validity of the English-language version (118). The fact the DT demonstrated only low-to-moderate responsiveness in this study against changes observed on the HADS, GHQ-12 and BSI-18 cautions against dependence on it as an outcome measure.

4.4.2.2 Hospital Anxiety and Depression Scales (HADS)

Although the HADS manual advises against summing the HADS-A and HADS-D scores to generate an overall indicator of distress, the resulting HADS-T score has, in fact, been widely used as an outcome measure. Factor analysis in various cancer populations has yielded mixed support for the HADS-T (97, 100, 120-122) whereas Rasch analyses conducted both in cancer patients (123) and patients attending an out-patient musculoskeletal rehabilitation program (124) have been broadly supportive. Moreover, several studies in cancer have found the HADS-T to be superior to one or both scales in identifying clinical levels of distress (93, 98, 100, 123, 125).

4.4.2.3 Profile of Mood States (POMS)

Three further, unofficial POMS versions have been developed specifically for use in cancer and have simplified scaling that lends them to use as measures of overall distress. Two of these, the POMS-10 and 11, are summarised by means of a single total mood disturbance score (TMDS), while the POMS-14 yields scores in two scales, namely negative and positive affect, the former of which approximates to a measure of distress. Both the POMS-11 (58) and 14 (68) were constructed using factor analysis. The 14-item version was developed to avoid somatic items while the POMS-11 retains ‘weary’, as well as ‘bewildered’ and ‘muddled’. The last two items have the potential to be confounded by chemotherapy-induced cognitive impairment. The POMS-14 is the better validated of the two, having shown greater evidence of discriminant validity and convergent validity with Karnofsky Performance Status and Spitzer’s Quality of Life Index (126).

The POMS-10 was developed on an ad hoc basis for use in a breast cancer study (67) and has not been validated beyond reporting of satisfactory internal consistency.

4.4.2.4 Somatic and Psychological Health Report (SPHERE)

The SPHERE was developed from three instruments, including the GHQ-30, to screen for common mental disorders in Australian general practice (71). In addition to the RCT identified at Step 1 (127), it has been used in two Australian cancer clinical research studies both of which have focused on the relationship between distress and fatigue (128, 129). Recently, the SPHERE's 6 psychological items have shown satisfactory convergent validity (Kappa = 0.73) with the HADS-T in Australian cancer centre outpatients (130).

4.4.2.5 Measures of post-traumatic stress disorder (PTSD)

As indicated above, the IES-R and PCL-C differ from other distress measures reviewed here in that they have been developed to accord with the diagnostic criteria of a specific psychiatric disorder. The IES was developed prior to PTSD becoming officially recognised as a diagnosable disorder in DSM-III (131) and lacks 7 items of the IES-R that assess hyper-arousal. All three instruments have been used to assess distress arising from the experience of cancer and its treatment. In the current review, we focused primarily on the psychometric properties of the overall scores from these measures rather than the more specific constructs (intrusion/re-experiencing, avoidance and arousal) assessed by their scales.

Both the IES and PCL-C have been partially validated in a small sample (N=55) of women 6 to 60 months post-treatment for primary breast cancer (132). Here, the total scores from both showed high internal consistency (alpha = 0.93 [IES], 0.94 [PCL-C]) suggestive of some degree of item redundancy and correlated strongly (0.88). Both measures also demonstrated high inter-correlations between their own scales (0.68-0.93 [IES], 0.64-0.94 [PCL-C]).

Surprisingly, given their widespread use in oncology both as outcome measures and comparative measures to validate new PROMs, neither the IES nor the IES-R has been widely validated in this population. However, validation studies have been carried out with Australians at increased risk of hereditary breast cancer (IES-R) (133) and Greek (IES) (134), Taiwanese (IES) (135)and German (IES) (136) cancer patients, in whom the IES/-R have generally demonstrated satisfactory internal consistency, factor structure and convergent validity.

Less widely used in cancer research than the IES/IES-R, the PCL-C has been validated only in one further cancer study, again with women with breast cancer (N=82) (137). In this study, it showed good criterion validity against a diagnostic interview for PTSD with an optimal sensitivity of 100% and specificity of 83%.

4.4.3 HRQoL measures

4.4.3.1 EORTC QLQ-C30

The current review identified 19 articles reporting evidence for the reliability and validity of the QLQ-C30 in English-speaking or international cancer populations. Studies were conducted with patients from a large variety of cancer groups including samples with lung, head and neck, breast, prostate pleural mesothelioma, gastro-intestinal and mixed cancer diagnoses. Internal consistency has been variable in scales assessing role (RF) and cognitive (CF) functioning; those assessing physical (PF), emotional (EF) and social (SF) functioning and global quality of life have tended to fall within the alpha 0.70 - 0.90 range. RF has also been found problematic in at least two assessments of internal structure (73, 138), the majority of which have focused on item-scale and inter-scale correlations rather than factor structure. Interestingly, SF has often correlated at least as strongly with physical and/or functional as emotional scales both within the QLQ-C30 itself (139-141) and with other measures (142, 143) (see below for a comparison between the social scales of the QLQ-C30 and FACT-G).

Taken as a whole, there is persuasive evidence for the convergent validity of QLQ-C30 scales and the discriminant validity and responsiveness of PF, RF and global quality of life. A dearth of studies examining these last two properties within the context of psychosocial interventions means evidence for EF and SF is less compelling. Evidence collected across five studies of inter-rater reliability between patients and proxies suggest reliability varies depending on the domain and proxy concerned (144).

Analysis using item response theory (IRT) was carried out on data from a large sample (N=8,242) of palliative care patients and found the EF scale could be reduced to 2 rather than 4 items with little or no loss of measurement efficiency, raising the potential for a short-form of this scale (145).

There are numerous publications to assist with interpreting scores from the QLQ-C30. Table 11 gives information about articles of this type, citing reviews that summarise information across a number of studies wherever possible.

Table 11. Articles dedicated to interpretation of QLQ-C30 data

Type of information	Article	Description
Normative / reference data	(146)	Comparison of norms for Germany, Norway, Denmark and Sweden
	(147)	Using reference data for interpretation
	(148)	Official EORTC reference values for various cancer sites
Minimally important difference	(149)	Worked example of using norms to develop MIDs
	(33)	Clinically based MIDs derived from secondary analysis
	(150)	Interpreting scores by comparison with ratings of change
Overcoming missing data	(151)	Interpreting data in advanced gastrointestinal cancer

4.4.3.2 FACT-G

The current review identified 25 articles reporting on the reliability and validity of the FACT-G in English-speaking or international cancer populations. Validation samples have included patients with breast, brain, gastrointestinal, lung, melanoma, prostate, gynaecological, renal and mixed cancer diagnoses as well as those undergoing bone marrow transplantation, neurotoxic chemotherapy and immunotherapy.

Scaling of the FACT-G has been broadly confirmed by factor analysis, although the uni-dimensionality of the social and family wellbeing (SWB) and, to a lesser extent, emotional wellbeing (EWB) scales remains questionable (152-154). While these scales have sometimes shown low internal consistency (155-157), physical wellbeing (PWB), functional wellbeing (FWB) and even the overall FACT-G scale have sometimes shown internal consistencies in excess of alpha 0.90 (154, 158, 159), suggesting a degree of item redundancy. Secondary analysis of RCT data suggests that responsiveness might not be greatly affected if a number of items identified as poorly fitting by Rasch analysis were removed, thus raising the possibility of a future short-form (153).

Unusually for a HRQoL measure, the test-retest reliability of the FACT-G has been repeatedly evaluated, with Kappas found to be 0.70 or above in 6 out of 8 studies.

As in the case of the QLQ-C30, there is a range of information available to assist with interpreting scores from the FACT-G (see Table 12).

Table 12. Articles dedicated to interpretation of FACT-G data

Type of information	Article	Description
Normative / reference data	(160)	US norms and heterogeneous cancer reference values
	(161)	US norms for the FACT-GP
	(162)	Austrian norms and cancer survivor reference values
	(163)	Australian (QLD) norms for the FACT-GP
Minimally important difference	(164)	Reviews anchor and distribution based MIDs for the FACT-G, its scales and various modules
	(165)	Reviews anchor and distribution based MIDs for the FACT-G, its scales and various modules
	(35)	Meta-analysis of anchor based evidence for the FACT-G and its scales

4.4.3.3 Quality of Life Index – Cancer Version (QLI-CV)

As its name implies, the QLI-CV is a cancer-specific version of a HRQoL questionnaire originally designed for use in the general population (89). In modifying the original QLI, Ferrans (1990) relied on evidence from the literature rather than information from experts and patients to identify new domains relevant to oncology (166). The QLI-CV differs from the other HRQoL questionnaires reviewed here in that it asks respondents to rate both satisfaction with and importance of its various domains. Asking respondents to rate each construct twice makes the QLI-CV substantially longer than the QLQ-C30, FACT-G or SF instruments - 66 items in total. Modification of the original QLI complicates but does not exclude the potential to compare scores on the QLI-CV with those from the general population on the QLI or cardiac patients on the relevant version of the questionnaire.

Reports for the reliability and validity of the English version of the QLI-CV are limited to internal consistency, discriminant validity and comparison with an overall measure of satisfaction in the original validation article; it has been evaluated more rigorously in Chinese (167).

4.4.3.4 SF-36/12

Evidence of reliability and validity in cancer relate exclusively to the original versions of the SF-36 and 12, despite each having been replaced by a second version a decade or more ago. Changes made to the revised versions relate to questionnaire wording, layout and response options (168, 169).

The major potential advantage of generic PROMs like the SF-36 and 12 is their ability to enable comparison across different disease groups and the general population to assess relative burden. Reference data for a range of populations are available in their respective manuals.

The SF-36 has demonstrated at least some evidence for convergent validity and responsiveness in cancer patients (170, 171). All scales of the SF-36 have also shown good ability (ES > 0.70) to discriminate between depressed (n=24) and not depressed (n=9) patients with mixed cancer diagnoses (92). Two assessments of internal consistency have identified alphas for individual scales ranging from 0.65 to 0.93 (171, 172). The mental component (MCS) and physical component (PCS) summary scores of the SF-12 have demonstrated good internal consistency and test-retest reliability in a large sample (N=2,415) of men with prostate cancer (173). It is important to note that the MCS from both the SF-12 and SF-36 includes scores from both emotional and social scales.

4.5 Step 5: *Review of the evidence for the capacity of each selected questionnaire to detect important effects of treatment in RCTs of psychosocial interventions.*

A total of **124 RCTs** identified in Step 1 **used a candidate measure meeting all 4 criteria in Step 2.** Of these, 67 studies were excluded from ES calculation (see Figure 1 for details). Articles reporting the remaining **57 trials were reviewed for information relating to trial samples, the interventions they evaluated and ES measured by each PROM**. ES for one trial (174) could be calculated for between-group differences at follow-up only; this trial controlled for differences at baseline. Summary information is presented in Appendix G. Appendix H provides a summary of the performance of each PROM in identifying effects in RCTs of each type of intervention.

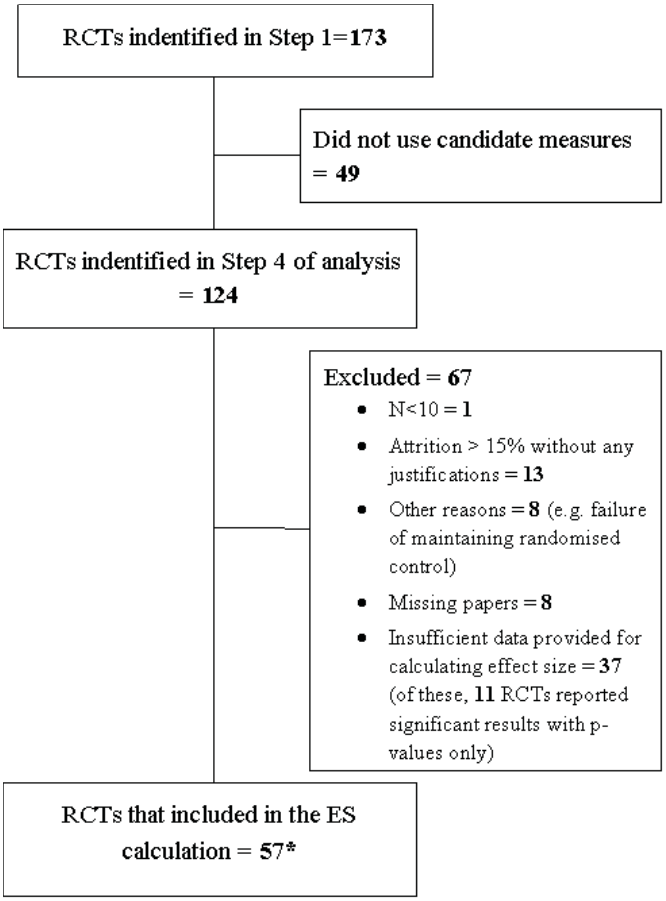

Figure 1. RCTs included and excluded at Step 5

4.6 Step 6: *Synthesis of findings and discussion.*

4.6.1 Psychometric comparisons between candidate measures based on review of evidence collected at Steps 4 and 5

Direct comparisons undertaken between candidate PROMs in studies identified at Step 2 have usually taken the form of correlations aimed at exploring the convergent validity of one or both measures in the absence of clear hypotheses or further comparisons to inform interpretation of results. Unsurprisingly, such comparisons generally show that scales from different PROMs ostensibly measuring the same or similar constructs correlate moderately or strongly. More useful are studies that compare the ability of different PROMs to discriminate between known groups, to screen for anxiety or depression, or to register changes in the construct over time. These studies are given priority in the following discussion.

4.6.1.1 Psychometric comparisons between measures of anxiety, depression and distress

Comparisons undertaken in studies of English-speaking cancer patients included in the current review have not been conclusive in recommending one particular PROM either in

terms of reliability and validity or performance as an outcome measure in RCTs of psychosocial interventions.

Studies of criterion validity are particularly difficult to compare because, as noted earlier, instrument performance will vary partly due to variation in the prevalence of anxiety and depression between different samples. Most useful are four studies that compared two or more candidate measures with a diagnostic interview, all of which included the HADS. Table 13 summarises evidence from these studies regarding the relative screening performance of instruments at their optimal cut-offs. The last of these studies (175) evaluated only a limited number of cut-offs for both the HADS and MHI-5, resulting in likely sub-optimal performance by both measures. The authors reported the optimal screening approach to be a two step screening process whereby patients were administered the HADS if they scored 11 or above on the MHI-5.

The relatively poor performance of both the HADS and GHQ-28 in patients with progressive disease (99) adds to other evidence that distress items may function differently in people with advanced cancer. In one palliative care sample, the single question ‘are you depressed?’ was found to have greater sensitivity and specificity in screening for depression than a number of standardised questionnaires, including the HADS (100). In that study, as in the palliative care sample referred to in Table 13 above (98), the HADS-T outperformed the HADS-D in screening for depression.

Table 13. PROM screening performance in studies comparing two psychological measures with a diagnostic interview

Ref.	Sample	Diagnosis	Measures compared	n	SE %	SP %	PPV %
(99)	Heterogeneous cancer	Anxiety or depression	*HADS-T*	284	80	76	41
			GHQ-28	ID	<80	<70	ID
	Disease-free		*HADS-T*	88	92	95	72
			GHQ-28	95	75	92	69
	Stable disease		*HADS-T*	113	83	78	42
			GHQ-28	102	<80	<70	ID
	Progressive disease		*HADS-T*	ID	<80	<70	ID
			GHQ-28	ID	<80	<70	ID
	On treatment		*HADS-T*	165	85	77	47
			GHQ-28	ID	<80	<70	ID
	Off treatment		HADS-T	ID	<80	<70	ID
			GHQ-28	133	88	79	41
(98)	Palliative care in-patients	Depression	*HADS-T*	79	77	94	72
			GHQ-12	79	ID	ID	ID
		Any ICD-10	*HADS-A*	79	83	77	59
			GHQ-12	79	ID	ID	ID
(93)	Head and neck cancer	Depression	*HADS-T*	60	100	95	86
			CES-D	60	100	85	63
(175)	Chemotherapy out-patients	Clinical ‘case’	HADS-T	172	60	85	55
			MHI-5	172	70	72	43

SE = sensitivity; SP = specificity; PPV = positive predictive value; ID = insufficient detail. Italics have been used to identify superior performing measures where comparisons were conclusive.

The **CES-D has performed well as a screening instrument for depression**. However, it has been evaluated in only two, small-scale studies with patients with heterogeneous (N=33) (92) and head and neck cancer (N=60) (93) diagnoses. In the latter of these, it was outperformed by the HADS-T but not HADS-D. Both studies focused on major depression only.

As stated above, while the C-SOSI and POMS-65 and 37 include scales assessing anxiety and/or depression, they were not designed nor have been used to screen for psychiatric disorders. Interestingly, the depression/depression-dejection scales from the C-SOSI and POMS-65 have been found highly correlated (0.87) in a sample of Canadian patients with mostly breast cancer (60). Moreover, the depression-dejection scale of the POMS-37 has correlated highly with the CES-D (0.63 (109); 0.80 (176)), while negative and positive cases identified by the GSI of the BSI-53 showed agreement with the POMS-65 TMDS (Kappa = 0.66) 12 months later (113). These findings, together with the overlap in contents between its tension-anxiety scale and DSM-IV-TR criteria for anxiety identified at Step 3, suggest that the **POMS may have clinical utility in cancer that has yet to be explored.**

Head-to-head comparisons between psychological measures enabled by RCTs identified in Step 1 were limited to 6 studies that used the IES-R or PCL-C in combination with the CES-D, POMS-65 or MHI-18. The combined use of PROMs intended to assess PTSD

with those assessing anxiety, depression and/or distress suggests the researchers involved believed that these different instruments might provide complementary results. Indeed, in all but one comparison, the ES identified by instruments from the two classes were substantially different, with PTSD measures identifying substantially larger effects than other psychological measures in two trials and vice versa in a further two. Table 14 summarises the relative performance of PROMs in these 6 trials.

Table 14. Relative effect sizes identified by PTSD and other psychological measures in the same RCTs

	CES-D	MHI-18	POMS-65
IES/-R	Dep=Ø (177) Dep=Ø (178) **-0.24** (177) Ø (178)	**Dep= -0.40** (179) Dis= Ø (179)	**Anx = 0.49** (180, 181) **Dep = 0.59** (180, 181) Ø (180, 181)
PCL-C	Dep= -0.51 (182) 0.44 (182)	- -	Anx = -0.41 (183)* Dep = -0.24 (183)* **1.47** (183)*

Ø = no effect above ≥ 0.2; Anx = anxiety scale; Dep = depression scale; bold font is used to identify ESs that are substantially larger (i.e. 'small' [0.2-0.49] versus 'moderate' [0.50-0.79] versus 'large' [0.80 +]) in each case. Note: Higher ES on all measures indicates greater morbidity.* The PCL-C indicates an opposite direction of effect to the POMS in this study. The MHI-18 was excluded as a candidate PROM at Step 2 but has been included in this table because it enables comparison with the IES.

4.6.1.2 Psychometric comparisons between HRQoL questionnaires

Cancer-specific versus generic

In general, disease-specific PROMs can be expected to have certain advantages and disadvantages when compared to their generic counterparts. Because disease-specific measures are targeted towards issues of particular importance to that disease, they are likely to be more sensitive to differences between clinically differentiable groups and responsive to clinically important changes. This hypothesis has been supported in a number of health conditions (184) although evidence is limited in cancer. In patients on active treatment or with advanced stage disease, generic PROMs may also demonstrate unacceptable floor effects that hamper their ability to register worsening of symptoms and, in severe cases, the effects of palliative treatment. On the other hand, generic PROMs enable comparison between disease groups and the general population, allowing a better understanding of the relative burden of different health conditions.

Two articles identified in the current review enable hypotheses testing in relation to relative responsiveness and distribution by offering head-to-head comparisons between the SF-36/12 (generic) and QLQ-C30 or FACT-G (disease-specific). One article in particular has used item response theory (IRT) to compare the distribution of scores on the SF-36 with those on both the QLQ-C30 and FACT-G in a single, large (N=1,163) sample (171). Surprisingly, the **SF-36 showed greater precision (lower standard error of the mean) in measuring the HRQoL not only of the highest but also the lowest scorers** in this study. However, these results appear to be based on total scores for each instrument; it is not clear what score was used for the SF-36.

A second, small study (N=34) of patients undergoing primary surgery for oral cancer compared the SF-36 and QLQ-C30 (170). Baseline HRQoL of patients in this study was high and remained relatively high to one year follow-up. As expected, the QLQ-C30 demonstrated more ceiling effects than the SF-36 in this group. However, unexpectedly, the physical scale of the SF-36 showed better responsiveness to deterioration between baseline and 3-months post-surgery (ES = - 0.50 vs - 0.27), while the reverse was true for relative responsiveness to improvement in emotional functioning (QLQ-C30) versus mental health (SF-36) over the same period (ES = 0.24 vs 0.58). Findings from this study are consistent with those from evaluative studies in gynaecological cancer reviewed previously (34), which similarly failed to show a clear advantage for the QLQ-C30 or FACT-G versus the SF-36/12 in terms of relative responsiveness. Interestingly, studies carried out in other health conditions have also found the SF-36 to perform surprisingly well in relative responsiveness compared to disease-specific questionnaires (185, 186). More research on this question in cancer is needed, especially with regard to potential complementary use of generic and cancer-specific HRQoL measures.

Other comparisons between the SF-36/12 and QLQ-C30 or FACT-G identified by the current review are limited to correlations between their various scales. Evidence of this type is difficult to interpret because no consensus exists regarding how similar generic and disease-specific versions of the same domains should be and the fact that constructs may be correlated without being similar. Studies conducted in a variety of cancer groups and languages have yielded **high correlations (≥ 0.50) between respective scales of the SF-36/12 and QLQ-C30** (170, 187, 188), while a further study with British head and neck cancer patients found an individualised HRQoL measure, the Patient Generated Index (PGI), overlapped similarly with the cognitive, emotional and mental functioning scales of both the SF-12 and QLQ-C30 (189). **Evidence for similarity or relatedness between the FACT-G and SF- instruments has been more mixed**, with most articles reporting generally high correlations with the SF-12 or SF-36 (152, 190, 191) but at least one reporting low to moderate correlations (0.17, 0.20 and 0.36) between social, emotional and physical scales on the FACT-G and social functioning, mental component summary (MCS) and physical component summary (PCS) on the SF-36 (192). In this study, carried out with US cancer patients aged 65 years and over, the FACT-G performed best in discriminating between clinically differentiated groups.

4.6.1.3 Psychometric comparisons between the EORTC QLQ-C30 and FACT-G

Across languages, **psychometric comparisons between the QLQ-C30 and FACT-G have failed to demonstrate superiority**. However there are important differences that may be useful in determining choice on a case-by-case basis. **While correlations between respective QLQ-C30 and FACT-G physical, emotional, role/functional scales and overall scales have tended to be high, those between the social scales have been moderate at best** (0.01 – 0.47) (140, 193-195) (159, 196). These lower correlations between social scales of the two measures make intuitive sense when one compares the content of constituent items; those in the FACT-G focus on impact on perceived social support and relationships whereas those in the QLQ-C30 assess impacts on social activities and family life (193). The distinction between social and family 'wellbeing' on the FACT-G and social 'functioning' as measured by the QLQ-C30 is further supported by the proven responsiveness of the former to RCTs of psychosocial interventions (183, 197-204) and the better sensitivity and specificity to performance status demonstrated by the latter (140). In other samples, unexpectedly low correlations between the emotional scales in German-speaking patients undergoing bone-marrow transplantation suggest that items from these respective scales may function differently in this population (195, 205, 206).

Other differences between the QLQ-C30 and FACT-G relate to the distribution of scores on each measure within different samples and therefore the extent to which each offers appropriate coverage of each domain or is at risk of floor and ceiling effects. In the Chang and Cella (1997) study referred to above, the authors **found the QLQ-C30 to show superior item efficiency (defined as person separation divided by the square root of the number of items) while the FACT-G showed better precision (lower standard error of the mean) in measuring the HRQoL of participants who were in the middle of the HRQoL range** (171). Data on floor and ceiling effects have been reported too infrequently to allow comparison of the QLQ-C30 and FACT-G in this respect, but means and standard deviations generally suggest that ceiling rather than floor effects are likely to be of more concern on both instruments and to a similar extent.

Perhaps **the most important difference between the QLQ-C30 and the FACT-G is their approach to scaling** and, in particular, the ways in which their overall quality of life scores are reached. In addition to the physical, emotional, social and functional/role scales contained in both measures, **the QLQ-C30 offers brief scales for cognitive functioning, financial impact and a range of symptoms** either not assessed by the FACT-G or else subsumed within its wellbeing scales. While the QLQ-C30's approach enables scores to be generated for outcomes that may be of independent interest to more general aspects of functioning, it generates 15 scores compared to the FACT-G's 5, which **complicates analysis and incurs problems of multiple hypothesis testing** (5). **In generating an overall score, the QLQ-C30 relies on responses to just two questions while the FACT-G allows summation of all 27 of its items**.

Classical test theory predicts that scales comprising a greater number of items should be more reliable and sensitive/responsive. A study by Cheung *et al.* (2005) (207)seems to confirm this by identifying evidence for **superior sensitivity of the FACT-G overall score relative to that of the QLQ-C30** in a sample of 452 Singaporean patients with heterogeneous cancer diagnoses. Based on cross-sectional data, Cheung *et al.* found the ES of the QLQ-C30's overall score in detecting a difference between patients with better (0 to 1) and worse (2 to 4) ECOG performance status to be 25% lower than that of the FACT-G, with important implications for sample size calculations. However, the only

other study included in the current review that offered an opportunity to test this hypothesis found overall scores from both measures to perform similarly in predicting performance status (140). It should also be noted that any superiority in sensitivity/responsiveness conferred by the FACT-G score's length will be offset where differential effects occur between composite scales.

A final difference between the QLQ-C30 and FACT-G **relates to their respective 'look and feel'**. With the exception of its emotional functioning items, the QLQ-C30 focuses on evaluation of functioning whereas the FACT-G encourages respondents to reflect on their feelings throughout. **Studies asking patients about** relative face validity, ease of comprehension and overall **preference have been inconclusive**, although the trend has generally favoured the QLQ-C30 (193, 194, 208-210). Two FACT-G items seem especially problematic both in terms of respondents' willingness to complete them and their relevance across cancer patients. The first of these relates to worry about dying and the second satisfaction with sex life; the second is optional and consistently leads to substantial missing data from the social scale, which can still be scored providing responses to 50% of the items are available.

4.6.1.4 Psychometric comparisons between psychological and HRQoL measures

Studies examining correlations between measures of anxiety, depression or distress and those of HRQoL generally show that the **emotional scales from HRQoL measures correlate strongly (≥ 0.50) or at least in the higher moderate range (0.40 – 0.50) with anxiety, depression and distress scales from dedicated measures** (60, 142, 156, 211-215). The lowest correlation (- 0.30) observed has been between the emotional scale of the QLQ-C30 and the IES (213), a result that is again consistent with the IES's purpose as a measure of a specific psychological disorder.

While lower, **correlations between psychological measures and physical and functional scales of HRQoL measures in these studies have still tended to be in the moderate range (0.30 – 0.50)**, suggesting that physical and psychological aspects of wellbeing are often inter-related. Conversely, too, the POMS includes scales of fatigue and vitality that correlate more strongly with physical and functional scales of HRQoL measures than emotional (156, 158, 211, 216), raising the possibility that its overall score may capture both emotional and physical dimensions in some clinical contexts.

Interestingly, there is evidence to suggest that the **emotional and role/functional scales of the QLQ-C30 and the FACT-G correlate closely with anxiety and depression respectively** (214, 215). One Austrian study found a combination of the EF and RF scales on the QLQ-C30 and a question asking about previous psychological/ psychotherapeutic treatment to be better than the HADS at predicting need for psychosocial treatment in women with breast cancer (N=105) (217).

Evidence for the SF-36 is less clear due to its different approach to scaling and the fact that the two summary scores (MCS and PCS) are reported more frequently than all 8 domains. The MCS and PCS are both generated from all 8 scales, each of which are weighted negatively or positively depending on the summary score. One Norwegian study found the HADS-A and HADS-D to explain similar amounts of MCS variance, 49% and 45% respectively; PCS variance explained was 5% and 10% (218). Constituent scales from the MCS correlated similarly to the HADS-A (0.54-0.75) and HADS-D (0.51-0.68),

with social functioning correlating identically with each (0.58) and only mental health showing a difference of any magnitude (0.75 and 0.68 respectively). At the same time, a recent US study found the SF-36's MCS to perform surprisingly well in screening for major depression in a small sample (N=34) of patients with mixed cancer diagnoses (92). In this study, optimal cut-offs on the MCS gave a positive predictive value (PPV) of 88% versus 92% for the CES-D. While PPV was not reported for individual scales, ESs between depressed and not-depressed patients were 0.7 or greater for all scales, with MCS scales demonstrating larger effects than those from the PCS.

Taken together, these findings raise the possibility that single or combined domains from the FACT-G, QLQ-C30 and possibly the SF-36 approximate closely enough to psychological measures to lessen the need to include the latter in the proposed battery of PROMs. Further testing of this idea is possible through head-to-head comparisons between HRQoL and psychological PROMs in 13 RCTs identified in Step 1. Table 15 summarises comparisons between the emotional scales of HRQoL measures and whichever anxiety, depression or distress scale identified the largest effect in the same study. Of 18 available comparisons, psychological measures identified substantially larger ESs in 6 (33%), emotional scales of HRQoL measures identified substantially larger effects in 1 (6%), and ESs were of comparable magnitude in 11 (61%).[1] Added to this is evidence from a further, non-randomised study identified in Step 4, in which the HADS-A showed better responsiveness than the FACT-G EWB in 20 Australian women with vulvar cancer post- versus pre-treatment (214). **These findings suggest that psychological and HRQoL measures have different capacities for identifying treatment effects in evaluations of psychosocial interventions and are worthwhile using in combination**.

Relative performance of HRQoL scales other than emotional compared to psychological measures was mixed. Of 11 comparisons, global quality of life scores identified substantially larger ESs than psychological measures in 3 (27%) cases (FACT-G (200); QLQ-C30 (219, 220), substantially smaller in 5 (46%) (180, 221-224) and similar in 3 (27%) (183, 225, 226). Interestingly, social scales performed rather more comparably to psychological measures: Of 10 available comparisons, social scales identified substantially larger ESs than psychological measures in 2 (20%) (SF-36 (227); QLQ-C30 (220)) cases, smaller ESs in 3 (30%) (180, 183, 223, 224) and similar ESs in 5 (50%) (225, 228-232). Even more surprising given the current review's focus on psychosocial interventions was the impressive performance by physical HRQoL scales. Of 15 available comparisons, physical scales identified substantially larger ESs than psychological measures in 3 (20%) cases (SF-36 (227); FACT-G (225); QLQ-C30 (220)), similar in 8 (53%) (200, 224, 228, 232-236) and substantially smaller in only 4 (27%) (180, 183, 222, 223, 230, 231). In all, physical scales of the QLQ-C30, FACT-G and SF-36/12 identified ESs of 0.2 or more in 1 (220), 6 (183, 200, 224, 225, 233, 237) and 4 (227, 235, 238, 239) trials respectively. While some of these trials evaluated exercise (240), education (238) or insomnia (200) interventions from which benefits to physical fitness or symptoms might be expected, others were concerned with psycho-educational, expressive or psycho-therapeutic interventions. **In many trials, physical scales performed similarly to emotional or social scales of the same instrument**, regardless of which HRQoL measure was used (200, 220, 224, 227, 232, 235, 237). While changes in disease or treatment status during the course of psychosocial intervention may have confounded measurement on physical

[1] Caution is needed in interpreting results from one of these trials, where the POMS-65 and FACT-G both identified moderate ESs but indicated opposite directions of effect.

scales in some of these studies, **the findings likely reflect the inter-relationship between physical and psychosocial domains of HRQoL and attest to the value of assessing physical wellbeing even when evaluating psychosocial interventions**.

Finally, given the evidence for a relationship between functional/role scales of the FACT-G and QLQ-C30 and depression discussed above, it is of interest to consider performance of these scales relative to psychological measures and other HRQoL scales in evaluating psychosocial interventions. Of 8 comparisons available, functional/role scales identified substantially larger effects than psychological measures in 3 (37%) cases (FACT-G (224, 225); QLQ-C30 (220)), substantially smaller in 3 (37%) (180, 223, 230, 231) and similar in 2 (20%) (200, 228). Of 8 RCTs where ESs could be calculated for both functional/role and emotional scales from the same HRQoL instrument, these differed substantially in 4 (50%) cases, identifying larger effects in 3 (FACT-G (200, 225); QLQ-C30 (220)) and smaller in 1 (FACT-G (183)). In one case, FWB on the FACT-G showed an ES closer to the CES-D than EWB (200), consistent with the hypothesis that FWB and depression are sometimes related. In other trials, however, functional or role scales performed substantially differently from PROMs assessing depression (180, 183, 220, 223-225, 230, 231).

4.6.1.5 Psychometric comparisons in relation to different subgroups

To find out whether there were PROMs with track records across a variety of interventions, we compared their performance across studies evaluating different intervention types (research question 3). Table 16 summarises the largest effects identified by each PROM in RCTs evaluating the various psychosocial intervention clusters. Based on the number of intervention types in which an ES of 0.2 or more was identified by each PROM, **there is most evidence for the broad-scale utility of the FACT-G and HADS**. However, the QLQ-C30's performance recommended its use over the FACT-G in evaluating education interventions, while the PCL-C excelled in evaluating complementary and alternative medicine interventions of a physical nature such as yoga or Tai Chi.

Table 15. Relative effect sizes identified by the emotional scales of HRQoL PROMs and psychological measures in the same RCTs

	QLQ-C30 (EF)	FACT-G (EWB)	SF-12*	SF-36*
CES-D	-0.61 (220)** Dep=-0.57 (220)**	Ø (228) Ø (200) Ø (241) Dep=Ø (228) **Dep= -0.24** (200) Dep=Ø (241)	Ø (242) Dep=Ø (242)	Ø (243) Ø (244) MH=0.38(245) Dep=0.36 (243) Dep=Ø (244) **Dep=-0.53** (245)
CES-D-15	- -	- -	- -	Ø (246) Dep=Ø (246)
HADS	0.27 (222) **Dis= -0.54** (222)	0.39(202) Dep = -0.67	- -	- -
IES-R	Ø (180, 181) Dis=Ø (180, 181)	0.46 (247) Dis= -0.23 (247)	- -	- -
PCL-C	- -	0.49 (183) **Dis=1.47** (183)	- -	- -
POMS-65	0.47(248) Ø (180, 181) Dep= -0.38 (248) **Dep=0.59** (180, 181)	**0.55** (249) Ø (197, 198) -0.49 (183) Anx= -0.22 (249) Dep=0.2 (197, 198) Anx= -0.41 (183)	- -	- -
SCL-90-R	- -	- -	- -	MCS=0.37 (250) Dep=-0.24 (250)

Ø = no effect size (ES) ≥0.2; - = comparison not available; Anx = anxiety scale; Dep = depression scale; Dis = distress scale; MCS= mental composite score; * for SF-12/36 , emotional scales were assumed to include role emotional, mental health or mental component summary (MCS); bold is used to identify ESs that are substantially larger (i.e. 'large' [0.80 +] versus 'moderate' [0.50-0.79] versus 'small' [0.2-0.49] versus 'no effect' [<0.2]) in each study. Note: higher ES for HRQoL indicates better HRQoL; higher ES for anxiety, depression or distress indicates greater morbidity; ** in this study, the QLQ-C30 and CES-D identified opposite effects, probably due to small sample size.

Table 16. Largest effects identified by each PROM in RCTs of psychosocial interventions

		Intervention types^					
PROMS	**Scales**	**SG/C/P**	**Education**	**CBT/P/S**	**E/CAM-P**	**CAM-NP**	**Screen.**
BSI-53/18	Anx	★★(233)	-	-	-	-	-
	Dep	★★★(233)	-	-	-	-	-
CES-D	Dep	★★★(251)	★★(220)	★(229)	Ø (177, 244)	★(221)	-
CES-D-15	Dep	-	Ø (220)	-	-	-	-
DT	Dis	-	★(252)	-	-	-	-
EORTC QLQ-C30	QOL	-	★★★(220)	-	★★(248)	Ø (180, 181, 221)	★(222)
	EF	-	★★(220)	-	★(248)	Ø (180, 181)	★(222)
	SF	-	★★★(220)	-	-	Ø (180, 181)	-
	PF	-	★★★(220)	-	-	Ø (180, 181)	Ø (222)
FACT-G	G	Ø (241, 249)	-	★★(200)	★(183)	★(253)	-
	EWB	★★(249)	Ø (228)	★(247)	★(183)	Ø (253)	-
	SWB	★(197, 198)	★(199)	★(202)	★★(203)	★★(253)	-
	PWB	★★★(233)	Ø (199, 254)	★★(202)	★(183)	★★(253)	-
HADS	Anx	★(255)	★★★(256)	★★(202)	-	★★★(257)	★(258)
	Dep	Ø (255)	★★★(256)	★★(202)	-	★(259)	★★(222)
	Dis	-	-	Ø (260)	-	-	★★(222)
IES-R	Dis	Ø (261)	Ø (178)	★(247)	★(177)	Ø (180, 181)	-
MHI-5	Dis	-	-	★(262)	-	-	-
PCL-C	Dis	★(182)	-	-	★★★(183)	-	-
POMS-11	Dis	★(263)	-	-	-	-	★(264)
POMS-14	Dis!	-	-	-	-	-	-
POMS-37	Anx	-	-	★(265)	-	-	-
	Dep	-	-	★(265)	-	-	-
POMS-65	Anx	★(249)	-	★★(91)	★(183)	★(180, 181)	-
	Dep	★(197, 198)	-	Ø (91)	★★(266)	★★(180, 181)	-
QLI-CV	QOL	-	Ø (267)	-	-	-	-
SCL-20**	Anx	-	★(268)	-	-	-	-
	Dep	-	★(268)	-	-	-	-
SCL-90-R	Anx	-	-	-	-	★(250)	-
	Dep	-	-	-	-	★(250)	-
SF-12	EF	-	★(269)	-	-	-	-
	SF	-	-	★★(270)	-	-	-
	PF	-	★★(269)	★★★(270)	-	-	-

PROMS	Scales	Intervention types^					
		SG/C/P	Education	CBT/P/S	E/CAM-P	CAM-NP	Screen.
SF-36	EF	-	Ø	Ø (245)	★(271)	★(272)	-
	SF	-	★(243)	-	★(244)	-	-
	PF	-	★(243)	-	★(244)	★(250)	-

Ø = no ES 0.20 or above; ★ = ES ≥0.20; ★★ = ES ≥0.50; ★★★ = ES ≥0.80; **** = SCL-20 derived from original version of SCL-90; Dis! = Distress scale for POMS-14 was assumed to be the negative affect scale; - = ES not available for this PROM from RCTs of this type of intervention. Intervention types: SG/C/P = support group, counselling, psychotherapy and family therapy; Education = education/psycho-education; CBT/P/S = cognitive-behavioural therapy, cognitive therapy, problem-solving/coping therapy and stress management training; E/CAM-P = Exercise and physical kinds of complementary and alternative medicine such as yoga and Tai Chi; CAM-NP = complementary and alternative medicine non-physical including aromatherapy, expressive therapy and supportive-expressive group therapy; Screen. = screening/referral. Note: Reported ES have been adjusted for the fact that higher ES for HRQoL indicates better HRQoL whereas higher ES for anxiety, depression or distress indicates greater morbidity.

4.6.2 Overall comparisons between candidate measures

In addition to psychometric properties, the current review compared PROMs with regard to the remaining criteria outlined in research question 1: efficiency, ease of administration via touch-screen computer, ease of interpretation and availability of comparison data. Tables 17 and 18 provide summary scores for psychological and HRQoL PROMs as rated against these properties using the weighted checklist in Appendix C. Information about the availability of translated forms and cost per use was considered supplementary by the Cancer Institute NSW and was presented earlier, in Appendix D.

4.6.2.1 Psychological measures

The **HADS scored highest overall** (weighted score = 72.5), partly because of the volume of evidence for its psychometric properties and partly because it provides scores for anxiety, depression and distress in only 14 items. The HADS has also been validated for administration via touch-screen (273).

When recommending a psychological measure for use by the Cancer Institute NSW, it is pertinent to consider Love's (2004) recommendations for choosing a measure to identify distress in women with breast cancer (3) (see Table 19). As was highlighted at the beginning of this document, Love's review differed from the current not only in its primary focus, which was on instruments' performance in screening rather than as outcome measures, but also in it's a priori exclusion of instruments on the grounds they did not measure clinical constructs and the greater weight it attached to performance in languages other than English (LOTE) and non-cancer populations.

Evidence for inferior performance of HADS-D in screening for depression was used by Love to caution against its use beyond a measure of anxiety. However, a recent meta-analysis suggests that the HADS-D may compete with the HADS-A and HADS-T in ruling out not only cases of depression but also anxiety and adjustment disorders (106). Moreover, the HADS-D performed comparably to the HADS-A in identifying intervention effects in RCTs identified in Step 1. This is important because Love highlighted the predominance of

anhedonia in HADS-D and HADS-T scaling as a potential barrier to their evaluation of psychosocial interventions.

Of the other psychological measures meeting criteria in Step 2, only the CES-D has consistent evidence of performance both in screening and as an outcome measure of RCTs (weighted score = 55). The CES-D was notably absent from Love's recommendations because he considered its assessment of symptom frequency rather than severity to be "idiosyncratic". The CES-D's less-than-intuitive response options led us to allocate it only a mid-range rating with regard to ease of administration. The other feature that counted against the CES-D was the relatively large number of items it requires to assess only a single construct. However, **the CES-D remains a cost-free alternative to the HADS in research where depression is the primary outcome.**

The **POMS-65** (weighted score = 55) and its unofficial short-form, the **POMS-37** (weighted score = 60), **scored highest after the HADS** for measures evaluating both anxiety and depression due to substantial evidence for their validity and ability to identify treatment effects in RCTs. Unlike the HADS and POMS-65, **the POMS-37 is also free to use**. Unfortunately, however, it provides an overall index of distress only through summation of all its scales, which include anger-hostility, vigour-activity, fatigue-inertia and confusion-bewilderment as well as tension-anxiety and depression-dejection. All versions of the POMS were excluded by Love on the basis that they assess anxiety and depression as mood states rather than as clinical constructs. While the current review was concerned with performance in measuring outcomes rather than in screening, evidence of criterion validity is important both to ensure that the constructs being assessed are clinically useful and to assist with interpretation of scores. The POMS' content and its performance in studies suggest that it may assess constructs closely related to clinical anxiety and depression. However, like the HADS, it emphasises anhedonia in assessing depression.

By contrast, the **IES-R and PCL-C**, which were also excluded from Love's review, appear to **measure a construct substantially different from anxiety, depression or general distress** and are likely to add supplementary information to a battery comprising the HADS and FACT-G should this be considered worth the extra administration time. **Of the IES-R and PCL-C, the latter scored highest in overall ratings** (weighted score = 37.5) because of the greater evidence for its validity, including criterion validity against a diagnosis of PTSD.

Table 17. Summary ratings for PROMs assessing anxiety (A), depression (D), distress (Dis) and post-traumatic stress disorder (PTSD)

	Constructs assessed	Validity	Reliability	Ability to identify treatment effects in RCTs	Performance in screening	Availability of comparison data	No. of psychological constructs	Length	Ease of administration and cognitive burden	Raw score	Weighted score
Weight		***1.5***	***1.5***	***1.5***	***1.5***	***1***	***1***	***1***	***1***		
HADS	A, D, Dis	7.5	7.5	15	7.5	10	10	10	10	**65**	**77.5**
POMS-37	A, D	15	7.5	7.5	0	5	5	10	10	**50**	**60**
CES-D	D	7.5	7.5	15	15	5	0	0	5	**40**	**55**
POMS-65	A, D	7.5	7.5	15	0	5	5	5	10	**45**	**55**
BSI-18/53	A, D	7.5	0	7.5	0	10	5	10	10	**45**	**50**
SCL-90-R	A, D	7.5	7.5	7.5	0	5	5	5	10	**40**	**47.5**
DT	Dis	7.5	0	7.5	0	5	0	10	10	**35**	**40**
GHQ-28	A, D	0	0	0	7.5	5	5	10	10	**35**	**37.5**
MHI-5	Dis	0	0	7.5	0	5	0	10	10	**30**	**32.5**
SPHERE	Dis	5	0	0	0	5	0	10	10	**30**	**32.5**
CES-D-15	D	7.5	7.5	0	0	5	0	5	5	**25**	**30**
MHI-38	A, D	0	0	0	0	5	10	10	5	**30**	**30**
POMS-11	Dis	7.5	0	7.5	0	0	0	5	10	**25**	**30**
C-SOSI	D	0	7.5	0	0	0	0	10	10	**25**	**27.5**
GHQ-12	Dis	0	0	0	7.5	5	0	5	10	**25**	**27.5**
POMS-14	Dis	7.5	0	0	0	0	0	10	10	**25**	**27.5**
PCL-C	PTSD	0	5	10	0	7.5	7.5	7.5	0	**30**	**37.5**
IES-R	PTSD	0	0	10	0	0	0	7.5	5	**20**	**22.5**

Table 18. Summary ratings for PROMs assessing HRQoL

	Domains/ global QOL	Length	Computer administration	Reliability	Validity	Evidence to aid interpretation	Ability to identify treatment effects in RCTs	Availability of comparison data	Raw score	Weighted score
Weight	***1***	***1***	***1***	***1.5***	***1.5***	***1.5***	***1.5***	***1***		
FACT-G	5	5	10	15	15	15	15	10	**70**	**90**
QLQ-C30	10	5	10	7.5	15	15	7.5	10	**65**	**80**
SF-36	0	5	5	7.5	7.5	7.5	15	10	**45**	**57.5**
SF-12	0	10	5	7.5	0	7.5	7.5	10	**40**	**47.5**
QLI-CV	5	0	10	0	7.5	0	0	5	**25**	**27.5**

Table 19. Recommendations made by Love (2004) (3) for measures identifying distress in breast cancer

Application	Recommended measure(s)
To detect both anxiety and depression, or distress, in a clinical setting.	BSI-18 & GHQ-12
To detect depression in a clinical setting.	BDI-SF
To detect anxiety in a clinical setting.	HADS
To provide information about psychological distress in a research project.	GHQ-30 & BSI-53
To form part of a research project concerning distress in women with breast cancer.	BDI-PC & GHQ-12

As a side issue, it is also worth noting that two measures recommended by Love were excluded by the current review because they have yet to be properly validated in oncology. Of these, the **Beck Depression Inventory – Primary Care (BDI-PC) shows** most **promise** because it omits the somatic content of the BDI-II and has been shown to out-perform the HADS-D in screening for depression in a sample of 50 US medical in-patients, 6% of whom had cancer (274). As Love notes, the second measure, the BDI short-form (BDI-SF), includes 3 somatic items out of 13 and so may be compromised in some cancer settings.

4.6.2.2 HRQoL measures

Given that the EORTC QLQ-C30 and FACT-G are the most widely used cancer-specific PROMS worldwide, it is of little surprise to find them performing similarly in terms of overall ratings summarised in Table 18. Historically, the QLQ-C30 has tended to have been used more in Europe and Canada and the FACT-G in the USA; both have been widely used in Australia. Both EORTC and FACIT 'suites' follow a similar approach in that they offer a range of disease-, treatment- and symptom-specific 'modules' that can be added to the 'core' QLQ-C30 and FACT-G as required. Questionnaires from both suites have been validated for administration via touch-screen computer, are available in a range of languages and can be used free of charge (although FACIT charges for its manual and questionnaire translations). Developers of both suites are also undertaking similar innovations in terms of the calibration of items for computer adaptive testing and development of 'weights' for use in cost utility analysis.

In the current review, **the FACT-G** (weighted score = 90) **scored marginally higher than the QLQ-C30** (weighted score = 80) thanks to more extensive evidence for its test-retest reliability and performance as an outcome measure in RCTs of psychosocial interventions. Australian normative data are also available for the FACT-G only (162). But probably the most important factor in choosing between the suites concerns their different approach to scaling (194). **The more numerous, shorter scales preferred by the EORTC offer greater flexibility** regarding outcomes of interest and may provide useful, unanticipated information about symptom relevance that may be missed using the FACIT system. While they come at the cost of more complex data for analysis and problems of multiple hypothesis testing, these advantages may be sufficient to recommend the QLQ-C30 where a standard questionnaire is wanted for application across a number of studies, some with outcomes relating to symptoms rather than the core HRQoL domains. On the other hand, **if core HRQoL domains and global quality of life are likely to be of exclusive interest, the scaling approach used in the FACT-G will simplify analysis and may offer greater precision with increased power** to detect treatment effects (275). This latter profile was considered advantageous to the Cancer Institute NSW's application.

4.6.3 Future directions in patient-reported outcome measurement

An important issue to keep in mind is the **progressive move towards computer adaptive testing (CAT)** in both distress (276) and HRQoL (277) measurement. In the future, 'static', standardised PROMs of the kind reviewed in this report may become obsolete, replaced either by CAT measures or related static short-forms.

CAT is a recent innovation that enables questionnaire content to be adapted to each individual respondent. It does this by using responses to previously-asked items to select the most informative next item from a large number available in a stored item 'bank'. Administration of items proceeds until a predefined level of precision has been reached and/or a predefined number of items have been asked. If the respondent reports no problems on a given item, the next item will describe a less severe manifestation to gain more information at the less severe end of the spectrum, and vice versa.

CAT has several advantages compared to traditional measurement by static questionnaires**, including increased precision for a given number of items, reduced floor and ceiling effects, avoidance of uninformative and clinically irrelevant questions** that otherwise unnecessarily burden patients and researchers, **and the ability to adapt measures to each study**. Calibration by means of item response theory (IRT) analysis locates items from each bank on an underlying 'latent trait', ensuring that scores for each domain are directly comparable across patients and studies even though the scores are based on different subsets of items.

Both the EORTC and the USA National Institute of Health (NIH) have embarked on programmes to develop, calibrate and validate item banks for use in CAT. Phase 1 of the NIH-led initiative, entitled 'Patient-reported Outcome Measurement Information System (PROMIS)', completed in 2008, focused largely on developing necessary collaborations and infrastructure as well as selecting items for use in the most important generic item banks. Phase 2 will further validate these item banks in a range of populations and languages and is expected to be completed by 2013. So far, attention has focused on item banks for emotional distress-anger, emotional distress-anxiety, emotional distress-depression, fatigue, pain-behavior, pain-impact, physical function, satisfaction with discretionary social activities, satisfaction with social roles, sleep disturbance, wake disturbance and global health. Further item banks for sexual functioning, perceived cognitive function and illness impact are under development. Items for the banks were drawn from a large number of existing PROMs including the HADS, FACIT measures and the SF-36v2. Item banks and the software required to administer them in CAT are available free of charge from the PROMIS website (http://www.nihpromis.org/). Item banks take an average of 1 to 2 minutes per domain to administer, and respondents are usually asked to recall their experience over the past week and respond using 5-point scales.

Published reports to date have largely focused on plans for the initiative (278, 279), documenting the processes involved in developing a domain framework (280) and establishing a qualitative item review process (281). Initial IRT modelling of the physical function item bank was carried out with data from 7 samples (N=17,726) of people from the general population and with chronic health conditions, including a medical outpatient sample that included cancer patients (282). Content validity of the social item banks was conducted by 25 patients, one of whom had cancer (283). Work is currently underway to evaluate item banks assessing physical function, fatigue,

pain, emotional distress, and social role participation using data from a number of samples including 1,000 cancer patients (284). T-score distributions on these domains for cancer and other populations are available from the PROMIS website. According to the website, a study has also recently been completed that will provide information about minimally important differences (MIDs) on cancer-specific versions of item banks assessing pain-impact, fatigue, physical function, anxiety and depression.

When completed, EORTC item banks will also enable cancer-specific CAT (285). Calibration of emotional, physical and social functioning is nearing completion while that for symptom domains such as fatigue and pain is underway. However, item banks are not yet available on the EORTC's website, and we could find no publications reporting validity or reliability.

Choice of PROMIS item banks for use in the planned Cancer Institute NSW study would have at least two advantages over and above those listed for CAT above. Software to administer PROMIS is free; while the HADS, FACT-G and QLQ-C30 have all been routinely administered using touch-screen computer, further programming would be needed to administer the full PROM battery via touchscreen technology. Also, unlike QLQ-C30 and FACT-G reference data, PROMIS reference data will be available for a large range of health conditions, enabling comparisons of disease impact and some forms of health economic evaluation (e.g., high level, allocative resource decisions). On the downside, there are no known plans to enable item banks to generate an overall score. Finally, given that the Cancer Institute NSW expects to accrue a relatively large sample (N > 300) for its evaluative study, the greater precision offered by CAT may be less important. Greater precision would, however, provide power for subgroup analyses.

Thus, **while it may be premature for the Cancer Institute NSW to begin using CAT at the present time, it is likely that this will become an attractive option in the future**. Fortunately, it is highly likely that PROMIS item banks will be calibrated against the HADS, FACT-G and CES-D in the future (personal correspondence 19/3/2010). This means that the Cancer Institute NSW could transition from the recommended PROMs to CAT at a later time and still retain the ability to compare new with previous results.

5. Limitations of the current review

The most important limitations of the current review concern the potential for selection bias in identifying and reviewing evidence for each PROM. This potential arose from the somewhat conservative limits we imposed on evidence for psychometric properties and PROM somatic content. These limits were used in order to make optimal use of time and resources and maximise confidence in generalising evidence to the Cancer Institute NSW's application of PROMs in research with predominantly English-speaking Australians with heterogeneous cancer diagnoses. It is important to emphasise, therefore, that a review aimed at informing a different application might have generated somewhat different recommendations.

Commonly used PROMs excluded because of lack of evidence for psychometric properties in English-speaking cancer samples were the STAI, DASS and BDI-PC. All three PROMs may show potential in the future and, indeed, the STAI is already widely used in cancer clinical research. Evidence that might be available from clinical studies regarding responsiveness and discriminant or convergent validity was discounted with the aim of narrowing the field early on to PROMs for which a greater diversity of evidence was available.

With regard to our emphasis on the evidence for English-language versions, it seems likely that only the IES/IES-R and QLQ-C30 were seriously affected. The QLQ-C30 in particular is widely used in non-English speaking European countries whereas its main competitor, the FACT-G, tends to be used more in the USA. Of 75 RCTs excluded at Step 1 because they used LOTE PROMs, 26 had used the QLQ-C30 while only 6 had used the FACT-G. The trend for excluded LOTE-version validation articles was similar though less marked, with 36 QLQ-C30 articles excluded compared to 27 for the FACT-G.

Also, while measures were excluded on the basis of somatic content, it should be noted that these have sometimes performed well in cancer studies (92). The rationale for excluding such instruments a priori was that evidence for sound performance by somatic items in one clinical context could not be generalised to the full range expected in the Cancer Institute NSW application.

Finally, it is important to note that the amount of evidence available for PROM performance is at least partly a function of how commonly it has been used. PROMs may become popular for a range of reasons not always related to superior performance. Once established, it is likely that PROMs are often used by researchers simply because they are a 'safe' choice and offer potential for comparing results with previous studies. This phenomenon seems likely to have introduced further, unavoidable bias to the review process. On the other hand, publication bias, whereby positive results are more likely to be published, should not have adversely affected our results inasmuch as poorly performing PROMs are likely to have presented a lower profile, consistent with the aims of the review.

6. Recommendations

The following recommendations are informed by the results from the critical review.

Recommendation 1. *That the Cancer Institute NSW include within its battery of PROMs* ***the HADS*** *as an overall measure of anxiety, depression and distress.*

The HADS (weighted score = 72.5) is by far the most efficient measure of anxiety, depression and distress available, enabling scores to be generated for all three constructs from just 14 items. There is substantial evidence for its reliability, validity and performance as an outcome measure in cancer clinical research. It has been validated for administration via touch-screen computer. In addition to the English-language version, it is available in four of the five most commonly spoken NSW community languages (Arabic, Cantonese, Italian, Greek).

Where cost is a concern, the HADS could be substituted with the **free-to-use POMS-37** (weighted score = 60) unofficial short-form as a measure of anxiety and depression. However, the total mood disturbance score (TMDS) of the POMS-37 is generated through summation of too various scales to be recommended as a measure of distress across the full spectrum of clinical contexts.

Recommendation 2. *That the Cancer Institute NSW include within its battery of PROMs* ***the CES-D*** *where depression is an outcome of specific interest.*

The CES-D (weighted score = 55) is free to use and has demonstrated good criterion validity in both studies where it has been compared with a diagnostic interview for depression. There is also substantial evidence for its reliability, validity and performance as an outcome measure in cancer clinical research. The CES-D has been administered via touch-screen and is available in four of the five most commonly spoken NSW community languages (Arabic, Cantonese, Italian, Greek). However, its unusual response options and the fact it uses 20 items to assess a single construct weigh against its routine inclusion in the battery except where depression is an outcome of specific interest. A 15-item short form shows promise but has not been extensively validated.

Recommendation 3. *That the Cancer Institute NSW include within its battery of PROMs* ***the FACT-G*** *as a measure of HRQoL.*

The FACT-G (weighted score = 90) is among the most widely used cancer-specific HRQoL measures. There is substantial evidence for its reliability, validity and performance as an outcome measure. It is available free of charge and can be supplemented as necessary with modules assessing issues of specific concern to people with particular types of cancer, undergoing certain treatments or experiencing specific symptoms. It simplifies analysis and reduces multiple hypothesis-testing by summarising symptoms and functioning in only 4 scales and enables a precise overall quality of life score to be generated from the sum of all items. It has been validated for administration via touch-screen and is available in all five most commonly spoken NSW community languages (Arabic, Cantonese, Italian, Greek, Vietnamese). A

wealth of published evidence is available for assisting the interpretation of scores, including data from the Australian general population.

7. Conclusion

The findings of the report address the research questions posed above and provide information to assist the Cancer Institute NSW in selecting optimal patient-reported outcome measures for evaluation of future system-based interventions aimed at improving the wellbeing of people treated for cancer in NSW.

8. References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth edition, text revised (DSM-IV-TR). Washington DC: American Psychiatric Publishing, Inc; 2000.
2. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology: Distress Management. 2007; Available from: http://www.nccn.org/professionals/physician_gls/PDF/distress.pdf.
3. Love A. The identification of psychological distress in women with breast cancer. Sydney; available online at http://www.nbocc.org.au/resources/documents/IPD_The_identification_of_psychological_distress_in_women_with_breast_cancer.pdf: National Breast Cancer Centre2004.
4. Ferrans CE. Definitions and conceptual models of quality of life. In: J. Lipscomb CCG, & C. Snyder editor. Outcomes Research in Cancer: Measures, Methods, and Applications. Cambridge, UK: Cambridge University Press; 2005. p. 14-30.
5. Fayers PM, Machin D. Quality of life: The assessment, analysis and interpretation of patient-reported outcomes. 2nd ed. Chichester, UK: John Wiley and Sons, Ltd; 2007.
6. Jacobsen PB, Jim HS. Psychosocial interventions for anxiety and depression in adult cancer patients: achievements and challenges. CA Cancer J Clin. 2008 Jul-Aug;58(4):214-30.
7. National Comprehensive Cancer Network (NCCN). NCCN practice guidelines for the management of psychosocial distress. Oncology. 1999;13:113-47.
8. National Breast Cancer Centre & National Cancer Control Initiative (NBCC & NCCI). Clinical Practice Guidelines for the Psychosocial Care of Adults with Cancer. Sydney, Australia: National Breast Cancer Centre; 2003.
9. Rodin G, Lloyd N, Katz M, Green E, Mackay JA, Wong RKS, et al. The treatment of depression in cancer patients: a systematic review. Supportive Care in Cancer. 2007 Feb;15(2):123-36.
10. Williams S, Dale J. The effectiveness of treatment for depression/depressive symptoms in adults with cancer: a systematic review. Br J Cancer. 2006 Feb 13;94(3):372-90.
11. Osborn RL, Demoncada AC, Feuerstein M. Psychosocial interventions for depression, anxiety, and quality of life in cancer survivors: meta-analyses. Int J Psychiatry Med. 2006;36(1):13-34.
12. Uitterhoeve RJ, Vernooy M, Litjens M, Potting K, Bensing J, De Mulder P, et al. Psychosocial interventions for patients with advanced cancer - a systematic review of the literature. Br J Cancer. 2004 Sep 13;91(6):1050-62.
13. Newell SA, Sanson-Fisher RW, Savolainen NJ. Systematic review of psychological therapies for cancer patients: overview and recommendations for future research. Journal of the National Cancer Institute. 2002;94(8):558-84.
14. Love A. The identification of psychological distress in women with breast cancer: The National Breast Cancer Centre; 2004.
15. Smith JE, Richardson J, Hoffman C, Pilkington K. Mindfulness-based stress reduction as supportive therapy in cancer care: Systematic review. JAN Journal of Advanced Nursing. 2005;52(3):315-27.

16. Hersch J, Juraskova I, Price M, Mullen B. Psychosocial interventions and quality of life in gynaecological cancer patients: a systematic review. Psychooncology. 2008;17:epub ahead of print.
17. Coyne JC, Lepore SJ, Palmer SC. Efficacy of psychosocial interventions in cancer care: evidence is weaker than it first looks.[see comment][comment]. Annals of Behavioral Medicine. 2006 Oct;32(2):104-10.
18. Gottlied BH, Wachala ED. Cancer support groups: A critical review of empirical studies. Psycho-Oncology. 2007;16(5):379-400.
19. Graves KD. Social cognitive theory and cancer patients' quality of life: a meta-analysis of psychosocial intervention components. Health Psychology. 2003 Mar;22(2):210-9.
20. Lepore SJ, Coyne JC. Psychological interventions for distress in cancer patients: a review of reviews.[see comment]. Annals of Behavioral Medicine. 2006 Oct;32(2):85-92.
21. Luebbert K, Dahme B, Hasenbring M. The effectiveness of relaxation training in reducing treatment-related symptoms and improving emotional adjustment in acute non-surgical cancer treatment: A meta-analytical review. Psycho-Oncology. 2001;10(6):490-502.
22. Owen JE, Klapow JC, Hicken B, Tucker DC. Psychosocial interventions for cancer: review and analysis using a three-tiered outcomes model. Psycho-Oncology. 2001 May-Jun;10(3):218-30.
23. Rajasekaran M, Edmonds P, Higginson I. Systematic review of hypnotherapy for treating symptoms in terminally ill adult cancer patients. Palliative Medicine. 2005 Jul;19(5):418-26.
24. Redd WH, Montgomery GH, DuHamel KN. Behavioral intervention for cancer treatment side effects. Journal of the National Cancer Institute. 2001 Jun 6;93(11):810-23.
25. Rehse B, Pukrop R. Effects of psychosocial interventions on quality of life in adult cancer patients: meta analysis of 37 published controlled outcome studies. Patient Education & Counseling. 2003 Jun;50(2):179-86.
26. Roffe L, Schmidt K, Ernst E, Roffe L, Schmidt K, Ernst E. A systematic review of guided imagery as an adjuvant cancer therapy. Psycho-Oncology. 2005 Aug;14(8):607-17.
27. Semple CJ, Sullivan K, Dunwoody L, Kernohan WG. Psychosocial interventions for patients with head and neck cancer: past, present, and future. Cancer Nursing. 2004 Nov-Dec;27(6):434-41.
28. Van Kuiken D. A meta-analysis of the effect of guided imagery practice on outcomes. Journal of Holistic Nursing. 2004 Jun;22(2):164-79.
29. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6(2):doi: 10.1186/471-2288-6-2.
30. Sansoni J, Marosszeky N, Jeon Y-H, Chenoweth L, Hawthorne G, King M, et al. Final Report: Dementia Outcomes Measurement Suite Project. Wollongong: University of Wollongong, Development CfHS;2007.
31. Physiotherapy Evidence Database (PEDro). The PEDro Scale. 1999 [13th June 2009]; Available from: http://www.pedro.org.au/scale_item.html.
32. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.

33. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of Life Research. 1996 Dec;5(6):555-67.
34. Luckett T, King MT, Butow PN, Friedlander M, Paris T. Assessing health-related quality of life in gynecological oncology: A systematic review of questionnaires and their ability to detect clinically important differences and change. Int J Gynecol Cancer. in press;accepted 30th April, 2009.
35. King MT, Stockler MS, Cella D, Osoba D, Eton D, Thompson J, et al. Meta-analysis provides evidence-based effect sizes for a cancer-specific quality of life questionnaire, the FACT-G. Journal of Clinical Epidemiology accepted 12/5/09.
36. Watson M, Greer S. Development of a questionnaire measure of emotional control. J Psychosom Res. 1983;27(4):299-305.
37. Izard CE, Dougherty, Bloxom B, Kotsch NE. The differential emotions scale: A method of measuring the subjective experience of discrete emotions. Unpublished manuscript 1974.
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983 Dec;24(4):385-96.
39. Watson M, Greer S, Young J, Inayat Q, et al. Development of a questionnaire measure of adjustment to cancer: The MAC scale. Psychological Medicine Vol 18(1) Feb 1988, 203-209. 1988.
40. Horrowitz M, Adler N, Kegeles S. A scale for measuring the occurrence of positive states of mind: A preliminary report. Psychosom Med. 1988;50:477-83.
41. Carver CS. Measure of Current Status. 2006 [cited 2009 13th August]; Available from: http://www.psy.miami.edu/faculty/ccarver/sclMOCS.html.
42. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA.: Consulting Psychologists Press; 1983.
43. McNair DM, Lorr M, Droppleman LF. Manual for the profile of Mood States (POMS): Revised. San Diego, CA: Educational and Industrial Testing Service; 1992.
44. Center for Epidemiologic Studies. Center for Epidemiologic Studies Depression Scale (CES-D). Rockville, MD: National Institute of Mental Health; 1971.
45. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983 Jun;67(6):361-70.
46. Weiss D, Marmar C. The Impact of Event Scale -Revised. In: Wilson J, Keane T, editors. Assessing psychological trauma and PTSD. New York: Guildford; 1997.
47. Beck AT, Steer RA, Brown GK. BDI®-II Manual. San Antonio, Texas, USA: The Psychological Corporation; 1996.
48. McNair DM, Heuchert, J.W.P. Profile of Mood States Technical Update. North Tonawanda, NY: MHS; 2003.
49. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scale. Journal of Personality and Social Psychology. 1988;54:1063-70.
50. Derogatis LR, et al. SCL-90: an outpatient psychiatric rating scale. preliminary report. Psychopharmacology Bulletin. 1974;9:13-28.
51. Veit CT, Ware JE. The Structure of Psychological Distress and Well-Being in General Populations Journal of Consulting and Clinical Psychology. 1983;51(5):730-42.
52. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: An introductory report. Psychological Medicine Vol 13(3) Aug 1983, 595-605. 1983.

53. Derogatis LR, Rutigliano PJ. Derogatis Affects Balance Scale (DABS). In: Spilker B, editor. Quality of Life and Pharmacoeconomics In Clinical Trials. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1996.
54. Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma: a pilot study. Cancer. 1998 May 15;82(10):1904-8.
55. Hamilton M. Development of a rating scale for primary depressive illness. British Journal of Social and Clinical Psychology. 1967;6:278-96.
56. Spitzer RL, Kroenke K, Williams JB. Patient Health Questionnaire primary care study group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Journal of the American Medical Association. 1999;282:1737-44.
57. Weathers FW, Litz BT, Herman DS, Huska JA, Keane T. The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Meeting of the International Society for Traumatic Stress Studies; San Antonio, TX1993.
58. Cella DF, Jacobsen PB, Orav EJ, Holland JC, Silberfarb PM, Rafla S. A brief POMS measure of distress for cancer patients. J Chronic Dis. 1987;40(10):939-42.
59. Leckie MS, Thompson E. Symptoms of Stress Inventory. Serial (Book,Monograph), Seattle, WA.: University of Washington; 1979.
60. Carlson LE, Thomas BC. Development of the Calgary Symptoms of Stress Inventory (C-SOSI). International Journal of Behavioral Medicine. 2007;14(4):249-56.
61. Bradburn N. The Structure of Psychological Well-Being. Chicago: Aldine; 1969.
62. Beck AT, Steer RA. Manual for the Beck Anxiety Inventory. San Antonio, Texas: Psychological Corporation; 1990.
63. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd ed. Sydney: Psychology Foundation; 1993.
64. Goldberg D, Williams P. A user's guide to the General Health Questionnaire (GHQ). London: GL Assessment; 1988.
65. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res. 1982;17:37- 49.
66. Stewart AL, Hays RD, Ware JE. The MOS short-form General Health Survey: reliability and validity in a patient population. Med Care. 1988;26:724.
67. Carver CS, Pozo C, Harris SD, Noriega V, Scheier MF, Robinson DS, et al. How coping mediates the effect of optimism on distress: A study of women with early stage breast cancer. Journal of Personality and Social Psychology Vol 65(2) Aug 1993, 375-390. 1993.
68. Guadagnoli E, Mor V. Measuring cancer patients' affect: Revision and psychometric properties of the Profile of Mood States (POMS). Psychological Assessment: A Journal of Consulting and Clinical Psychology. 1989;1(2):150-4.
69. Shacham S. A shortened version of the Profile of Mood States. Journal of Personality Assessment Vol 47(3) Jun 1983, 305-306. 1983.
70. Lorr M, McNair DM, Heuchert JWP. Profile of Mood States Bipolar Supplement: POMS-Bi. North Tonawanda, NY: MHS; 2003.
71. Hickie IB, Davenport TA, Hadzi-Pavlovic D, Koschera A, Naismith SL, Scott EM, et al. Development of a simple screening tool for common mental disorders in general practice.[see comment]. Med J Aust. 2001 Jul 16;175 Suppl:S10-7.

72. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. Journal of Clinical Oncology. 1993;11(570-579).
73. 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;85:365-76.
74. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30(6):473-83.
75. Schag CA, Ganz PA, Heinrich RL. CAncer Rehabilitation Evaluation System--short form (CARES-SF). A cancer specific rehabilitation and quality of life instrument. Cancer. 1991 Sep 15;68(6):1406-13.
76. De Haes JCJM, van Knippenberg FCE, Neijt JP. Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. Br J Cancer. 1990;62:1034-8.
77. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33.
78. Ferrell BR, Grant M, Funk B, Garcia N, Otis-Green S, Schaffner ML. Quality of life in breast cancer. Cancer Practice. 1996 Nov-Dec;4(6):331-40.
79. Schag CA, Heinrich RL. Development of a comprehensive quality of life measurement tool: CARES. Oncology (Williston). 1990 May;4(5):135-8;discussion 47.
80. 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):472-83.
81. Cohen SR, Mount BM, Strobel MG, Bui F. The McGill Quality of Life Questionnaire: a measure of quality of life appropriate for people with advanced disease. A preliminary study of validity and acceptability. Palliat Med. 1995 Jul;9(3):207-19.
82. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey.[see comment]. BMJ. 1996 Apr 20;312(7037):1016-20.
83. Padilla GV, Grant MM. Quality of life as a cancer nursing outcome variable. ANS Adv Nurs Sci. 1985 Oct;8(1):45-60.
84. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess. 1985;49:71-5.
85. Atkinson T. Is Satisfaction a Good Measure of the Perceived Quality of Life? Annual Meeting of the American Statistical Association; August 15-18; Chicago, Illinois1977.
86. McMillan SC, Mahon M. Measuring Quality of Life in Hospice Patients Using a Newly Developed Hospice Quality of Life Index. Quality of Life Research. 1994;3(6):437-47.
87. Selby PJ, Chapman JA, Etazadi-Amoli J, Dalley D, Boyd NF. The development of a method for assessing the quality of life of cancer patients. Br J Cancer. 1984;50(1):13-22.
88. Padilla GV, Presant C, Grant MM, Metter G, Lipsett J, Heide F. Quality of life index for patients with cancer. Res Nurs Health. 1983 Sep;6(3):117-26.

89. Ferrans CE, Powers MJ. Quality of life index: development and psychometric properties. ANS Adv Nurs Sci. 1985 Oct;8(1):15-24.
90. Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technology Assessment. 2004;8(25):1-234.
91. Speca M, Carlson LE, Goodey E, Angen M. A randomized, wait-list controlled clinical trial: the effect of a mindfulness meditation-based stress reduction program on mood and symptoms of stress in cancer outpatients. Psychosomatic Medicine. 2000 Sep-Oct;62(5):613-22.
92. Hopko DR, Bell JL, Armento MEA, Robertson SMC, Hunt MK, Wolf NJ, et al. The phenomenology and screening of clinical depression in cancer patients. Journal of Psychosocial Oncology. 2008;26(1):31-51.
93. Katz MR, Kopek N, Waldron J, Devins GM, Tomlinson G. Screening for depression in head and neck cancer. Psychooncology. 2004 Apr;13(4):269-80.
94. Brown KW, Levy AR, Rosberger Z, Edgar L. Psychological distress and cancer survival: A follow-up 10 years after diagnosis. Psychosomatic Medicine Vol 65(4) Jul-Aug 2003, 636-643. 2003.
95. Carpenter JS, Andrykowski MA, Wilson J, Hall LA, Rayens MK, Sachs B, et al. Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. Issues Ment Health Nurs. 1998 Sep-Oct;19(5):481-94.
96. Stommel M, Given BA, Given CW, Kalaian HA, Schulz R, McCorkle R. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). Psychiatry Res. 1993 Dec;49(3):239-50.
97. Johnston M, Pollard B, Hennessey P. Construct validation of the hospital anxiety and depression scale with clinical populations. J Psychosom Res. 2000 Jun;48(6):579-84.
98. Le Fevre P, Devereux J, Smith S, Lawrie SM, Cornbleet M. Screening for psychiatric illness in the palliative care inpatient setting: a comparison between the Hospital Anxiety and Depression Scale and the General Health Questionnaire-12. Palliat Med. 1999 Sep;13(5):399-407.
99. Ibbotson T, Maguire P, Selby P, Priestman T, Wallace L. Screening for anxiety and depression in cancer patients: the effects of disease and treatment. Eur J Cancer. 1994;30A(1):37-40.
100. Lloyd-Williams M, Friedman T, Rudd N. An analysis of the validity of the Hospital Anxiety and Depression scale as a screening tool in patients with advanced metastatic cancer. J Pain Symptom Manage. 2001 Dec;22(6):990-6.
101. Hall A, A'Hern R, Fallowfield L. Are we using appropriate self-report questionnaires for detecting anxiety and depression in women with early breast cancer? Eur J Cancer. 1999 Jan;35(1):79-85.
102. Groenvold M, Fayers PM, Sprangers MA, Bjorner JB, Klee MC, Aaronson NK, et al. Anxiety and depression in breast cancer patients at low risk of recurrence compared with the general population: a valid comparison? Journal of Clinical Epidemiology. 1999 Jun;52(6):523-30.
103. Osborne RH, Elsworth GR, Sprangers MAG, Oort FJ, Hopper JL. The value of the Hospital Anxiety and Depression Scale (HADS) for comparing women with early onset breast cancer with population-based reference women. Quality of Life Research. 2004 Feb;13(1):191-206.
104. Hopwood P, Howell A, Maguire P. Screening for psychiatric morbidity in patients with advanced breast cancer: validation of two self-report questionnaires. Br J Cancer. 1991 Aug;64(2):353-6.

105. Ibbotson T, Maguire P, Selby P, Priestman T, Wallace L. Screening for anxiety and depression in cancer patients: The effects of disease and treatment. Eur J Cancer. 1994;30(1):37-40.
106. Mitchell AJ, editor. How Accurate is the HADS as a Screen for Emotional Complications of Cancer? A Metaanalysis Against Robust Psychiatric Interviews. International Psycho-Oncology Society (IPOS) 11th World Congress; 2009; Vienna, Austria: Psycho-Oncology.
107. DiLorenzo TA, Bovbjerg DH, Montgomery GH, Valdimarsdottir H, Jacobsen PB. The application of a shortened version of the profile of mood states in a sample of breast cancer chemotherapy patients. British Journal of Health Psychology. 1999;4:315-25.
108. Curran SL, Andrykowski MA, Studts JL. Short Form of the Profile of Mood States (POMS-SF): Psychometric information. Psychological Assessment. 1995;7(1):80-3.
109. Baker F, Denniston M, Zabora J, Polland A, Dudley WN. A POMS short form for cancer patients: psychometric and structural evaluation. Psychooncology. 2002 Jul-Aug;11(4):273-81.
110. Moadel AB, Shah C, Wylie-Rosett J, Harris MS, Patel SR, Hall CB, et al. Randomized controlled trial of yoga among a multiethnic sample of breast cancer patients: effects on quality of life.[see comment]. Journal of Clinical Oncology. 2007 Oct 1;25(28):4387-95.
111. Derogatis LR, Morrow GR, Fetting J, Penman D, Piasetsky S, Schmale AM, et al. The prevalence of psychiatric disorders among cancer patients. Jama. 1983 Feb 11;249(6):751-7.
112. Payne DK, Hoffman RG, Theodoulou M, Dosik M, Massie MJ. Screening for anxiety and depression in women with breast cancer. Psychiatry and medical oncology gear up for managed care. Psychosomatics. 1999 Jan-Feb;40(1):64-9.
113. Zabora JR, Smith-Wilson R, Fetting JH, Enterline JP. An efficient method for psychosocial screening of cancer patients. Psychosomatics. 1990;31(2):192-6.
114. Fitch MI, Osoba D, Iscoe N, Szalai JP. Predicting psychological distress in patients with cancer: conceptual basis and reliability evaluation of a self-report questionnaire. Anticancer Res. 1995 Jul-Aug;15(4):1533-42.
115. Zabora J, BrintzenhofeSzoc K, Jacobsen P, Curbow B, Piantadosi S, Hooker C, et al. A new psychosocial screening instrument for use with cancer patients. Psychosomatics: Journal of Consultation Liaison Psychiatry Vol 42(3) May-Jun 2001, 241-246. 2001.
116. Recklitis CJ, Parsons SK, Shih M-C, Mertens A, Robison LL, Zeltzer L. Factor structure of the brief symptom inventory--18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. Psychological Assessment. 2006 Mar;18(1):22-32.
117. Kaplan CP, Miner ME, Mervis L, Newton H, McGregor JM, Goodman JH. Interpretive risks: the use of the Hopkins Symptom Checklist 90-Revised (SCL 90-R) with brain tumour patients. Brain Inj. 1998;12(3):199-205.
118. Gessler S, Low J, Daniells E, Williams R, Brough V, Tookman A, et al. Screening for distress in cancer patients: is the distress thermometer a valid measure in the UK and does it measure change over time? A prospective validation study. Psychooncology. 2008 Jun;17(6):538-47.
119. Mitchell AJ. Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders.[see comment]. Journal of Clinical Oncology. 2007 Oct 10;25(29):4670-81.

120. Moorey S, Greer S, Watson M, Gorman C, Rowden L, Tunmore R, et al. The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer.[see comment]. Br J Psychiatry. 1991 Feb;158:255-9.
121. Rodgers J, Martin CR, Morse RC, Kendell K, Verrill M. An investigation into the psychometric properties of the Hospital Anxiety and Depression Scale in patients with breast cancer. Health Qual Life Outcomes. 2005;3:41.
122. Smith AB, Selby PJ, Velikova G, Stark D, Wright EP, Gould A, et al. Factor analysis of the Hospital Anxiety and Depression Scale from a large cancer population. Psychol Psychother. 2002 Jun;75(Pt 2):165-76.
123. Smith AB, Wright EP, Rush R, Stark DP, Velikova G, Selby PJ. Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. Psychooncology. 2006 Sep;15(9):817-27.
124. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). British Journal of Clinical Psychology. 2007 Mar;46(1):1-18.
125. Chaturvedi S. Clinical irrelevance of HAD factor structure: Comment. British Journal of Psychiatry Vol 159 Aug 1991, 298. 1991.
126. Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. J Chronic Dis. 1981;34:585-97.
127. Graham PH, Browne L, Cox H, Graham J. Inhalation aromatherapy during radiotherapy: results of a placebo-controlled double-blind randomized trial. J Clin Oncol. 2003 Jun 15;21(12):2372-6.
128. Bennett B, Goldstein D, Lloyd A, Davenport T, Hickie I. Fatigue and psychological distress--exploring the relationship in women treated for breast cancer. Eur J Cancer. 2004 Jul;40(11):1689-95.
129. Goldstein D, Bennett B, Friedlander M, Davenport T, Hickie I, Lloyd A. Fatigue states after cancer treatment occur both in association with, and independent of, mood disorder: a longitudinal study. BMC Cancer. 2006;6:240.
130. Clover K, Carter GL, Adams C, Hickie I, Davenport T. Concurrent validity of the PSYCH-6, a very short scale for detecting anxiety and depression, among oncology outpatients. Australian and New Zealand Journal of Psychiatry. 2009;43(7):682-8.
131. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington, DC: American Psychiatric Association; 1980.
132. Cordova MJ, Andrykowski MA, Kenady DE, McGrath PC, Sloan DA, Redd WH. Frequency and correlates of posttraumatic-stress-disorder-like symptoms after treatment for breast cancer. J Consult Clin Psychol. 1995 Dec;63(6):981-6.
133. Thewes B, Meiser B, Hickie IB. Psychometric properties of the Impact of Event Scale amongst women at increased risk for hereditary breast cancer. Psycho-Oncology Vol 10(6) Nov-Dec 2001, 459-468. 2001.
134. Mystakidou K, Tsilika E, Parpa E, Galanos A, Vlahos L. Psychometric properties of the Impact of Event Scale in Greek cancer patients. J Pain Symptom Manage. 2007 Apr
Apr;33(4):454-61.
135. Chen S-C, Lai Y-H, Liao C-T, Lin C-C. Psychometric testing of the Impact of Event Scale-Chinese Version (IES-C) in oral cancer patients in Taiwan. Supportive Care in Cancer. 2005 Jul
Jul;13(7):485-92.
136. Ferring D, Filipp S-H. Statistical testing of the Impact of Event Scale: Reliability and stability. Diagnostica. 1994;40(4):344-62.

137. Andrykowski MA, Cordova MJ, Studts JL, Miller TW. Posttraumatic stress disorder after treatment for breast cancer: prevalence of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C) as a screening instrument. J Consult Clin Psychol. 1998 Jun;66(3):586-90.
138. Osoba D, Zee B, Pater J, 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 Oct;3(5):353-64.
139. 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 Jun;8(4):311-7.
140. Sharp LK, Knight S, Nadler R, Albers M, Moran E, Kuzel T, et al. Quality of life in low-income patients with metastatic prostate cancer: Divergent and convergent validity of three instruments. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation. 1999 Aug;8(5):461-70.
141. Blazeby JM, Conroy T, Hammerlid E, Fayers P, Sezer O, Koller M, et al. Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. Eur J Cancer. 2003 Jul;39(10):1384-94.
142. Niezgoda H, Pater J. A validation study of the domains of the core EORTC Quality of Life Questionnaire. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation. 1993 Oct;2(5):319-25.
143. King MT, Dobson AJ, Harnett PR. A comparison of two quality-of-life questionnaires for cancer clinical trials: the functional living index--cancer (FLIC) and the quality of life questionnaire core module (QLQ-C30). Journal of Clinical Epidemiology. 1996 Jan;49(1):21-9.
144. Sneeuw KCA, Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. Journal of Clinical Epidemiology. 2002 Nov;55(11):1130-43.
145. Bjorner JB, Petersen MA, Groenvold M, Aaronson N, Ahlner-Elmqvist M, Arraras JI, et al. Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. Quality of Life Research. 2004 Dec;13(10):1683-97.
146. Fayers PM. Interpreting quality of life data: population-based reference data for the EORTC QLQ-C30.[comment]. Eur J Cancer. 2001 Jul;37(11):1331-4.
147. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Using reference data on quality of life--the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3). Eur J Cancer. 1998 Aug;34(9):1381-9.
148. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. EORTC QLQ-C30 reference values. Brussels: European Organisation for Research and Treatment of Cancer; 2008.
149. Gulbrandsen N, Hjermstad MJ, Wisloff F, Nordic Myeloma Study G. Interpretation of quality of life scores in multiple myeloma by comparison with a reference population and assessment of the clinical importance of score differences. Eur J Haematol. 2004 Mar;72(3):172-80.
150. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of Clinical Oncology. 1998 Jan;16(1):139-44.

151. Nordin K, Steel J, Hoffman K, Glimelius B. Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. Br J Cancer. 2001 Nov 2;85(9):1265-72.
152. Ashing-Giwa KT, Kim J, Tejero JS. Measuring quality of life among cervical cancer survivors: preliminary assessment of instrumentation validity in a cross-cultural study. Quality of Life Research. 2008 Feb;17(1):147-57.
153. Smith AB, Wright P, Selby PJ, Velikova G. A Rasch and factor analysis of the Functional Assessment of Cancer Therapy-General (FACT-G). Health Qual Life Outcomes. 2007;5:19.
154. Winstead-Fry P, Schultz A. Psychometric assessment of the Functional Assessment of Cancer Therapy-General (FACT-G) scale in a rural sample. Cancer. 1997;79(12):2446-52.
155. List MA, D'Antonio LL, Cella DF, Siston A, Mumby P, Haraf D, et al. The Performance Status Scale for head and neck cancer patients and the Functional Assessment of Cancer Therapy-Head and Neck (FACT-H&N) scale: A study of utility and validity. Cancer. 1996;77(11):2294-301.
156. Ward WL, Hahn EA, Mo F, Hernandez L, Tulsky DS, Cella D. Reliability and validity of the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) quality of life instrument. Quality of Life Research. 1999;8(3):181-95.
157. Cella DF, Bonomi AE, S.R L, Tulsky DS, Kaplan E, Bonomi P. Reliability and validity of the Functional Assessment of Cancer Therapy - Lung (FACT-L) quality of life instrument. Lung Cancer. 1995;12(2):199-220.
158. Heffernan N, Cella D, Webster K, Odom L, Marone M, Passik S, et al. Measuring health-related quality of life in patients with hepatobiliary cancers: The Functional Assessment of Cancer Therapy - Hepatobiliary (FACT-Hep) questionnaire. Journal of Clinical Oncology. 2002;20(9):2229-39.
159. Darling G, Eton DT, Sulman J, Casson AG, Celia D. Validation of the functional assessment of cancer therapy esophageal cancer scale. Cancer. 2006 Aug 15;107(4):854-63.
160. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patients norms for The Functional Assessment if Cancer Therapy-General (FACT-G). Evaluation & the Health Professions. 2005 Jun;28(2):192-211.
161. Cella D, Zagari MJ, Vandoros C, Gagnon DD, Hurtz H-J, Nortier JWR. Epoetin alfa treatment results in clinically significant improvements in quality of life in anemic cancer patients when referenced to the general population. Journal of Clinical Oncology. 2003 Jan 15;21(2):366-73.
162. Holzner B, Kemmler G, Cella D, De Paoli C, Meraner V, Kopp M, et al. Normative data for functional assessment of cancer therapy--general scale and its use for the interpretation of quality of life scores in cancer survivors. Acta Oncologica. 2004;43(2):153-60.
163. Janda M, DiSipio T, Hurst C, Cella D, Newman B. The Queensland cancer risk study: general population norms for the Functional Assessment of Cancer Therapy-General (FACT-G). Psychooncology. 2009;18:606-14.
164. Webster K, Cella D, Yost KJ. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. . Health Qual Life Outcomes. 2003:1-79; http://www.hqlo.com/content/1/1/.
165. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. Evaluation & the Health Professions. 2005 Jun;28(2):172-91.

166. Ferrans CE. Development of a quality of life index for patients with cancer. Oncology Nursing Forum. 1990 May-Jun;17(3 Suppl):15-9; discussion 20-1.
167. Luo J, Sun Y. Preliminary results of the Chinese version of the Quality of Life Index (QLI) administered in Chinese cancer patients. Chinese Mental Health Journal. 1999;13(1):4-7.
168. Ware JE, Kosinski MA, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. User's manual for the SF-36v2 health survey. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2007.
169. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to score Version 2 of the SF-12 Health Survey. Lincoln, Rhode Island: Quality Metric Incorporated; 2004.
170. Rogers SN, Lowe D, Brown JS, Vaughan ED. A comparison between the University of Washington Head and Neck Disease-Specific measure and the Medical Short Form 36, EORTC QOQ-C33 and EORTC Head and Neck 35. Oral Oncology. 1998 Sep;34(5):361-72.
171. Chang CH, Cella D. Equating health-related quality of life instruments in applied oncology settings. Physical Medicine and Rehabilitation: State of the Art Reviews. 1997;11(2):397-406.
172. Forjaz MJ, Guarnaccia CA. Hematological cancer patients' quality of life: Self versus intimate or non-intimate confidant reports. Psychooncology. 1999 Nov-Dec;8(6):546-52.
173. Lebeau T, Perrotte P, Valiquette L, Benard F, McCormack M, Saad F, et al. Validation of prostate cancer index and SF-12 short forms. Can J Urol. 2005 Dec;12(6):2873-9.
174. Helgeson VS, Cohen S, Schulz R, Yasko J. Education and peer discussion group interventions and adjustment to breast cancer. Arch Gen Psychiatry. 1999 Apr;56(4):340-7.
175. Cull A, Gould A, House A, Smith A, Strong V, Velikova G, et al. Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. Br J Cancer. 2001 Dec 14;85(12):1842-9.
176. Conerly RC, Baker F, Dye J, Douglas CY, Zabora J. Measuring depression in African American cancer survivors: The reliability and validity of the Center for Epidemiologic Study-Depression (CES-D) Scale. J Health Psychol. 2002 Jan;7(1):107-14.
177. Cohen L, Warneke C, Fouladi RT, Rodriguez MA, Chaoul-Reich A. Psychological adjustment and sleep quality in a randomized trial of the effects of a Tibetan yoga intervention in patients with lymphoma. Cancer. 2004 May 15;100(10):2253-60.
178. Stanton AL, Ganz PA, Kwan L, Meyerowitz BE, Bower JE, Krupnick JL, et al. Outcomes from the Moving Beyond Cancer psychoeducational, randomized, controlled trial with breast cancer patients. J Clin Oncol. 2005 Sep 1;23(25):6009-18.
179. Manne SL, Ostroff JS, Winkel G, Fox K, Grana G, Miller E, et al. Couple-focused group intervention for women with early stage breast cancer. J Consult Clin Psychol. 2005 Aug;73(4):634-46.
180. Lemieux J, Beaton, D. E., Hogg-Johnson, S., Bordeleau, L, J., Hunter, J., Goodwin, P. J. Responsiveness to change to change due to supportive-expressive group therapy, improvement in mood and disease progression in women with metastatic breast cancer. Qual Life Res. 2007;16:1007-17.
181. 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 May;60(5):448-55.
182. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer. 2003 Mar 1;97(5):1164-73.
183. Levine EG, Eckhardt J, Targ E. Change in post-traumatic stress symptoms following psychosocial treatment for breast cancer. Psychooncology. 2005 Aug;14(8):618-35.
184. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. Journal of Clinical Epidemiology. 2003 Jan;56(1):52-60.
185. Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol. 2008;8:26.
186. Hagen KB, Smedstad LM, Uhlig T, Kvien TK. The responsiveness of health status measures in patients with rheumatoid arthritis: comparison of disease-specific and generic instruments. J Rheumatol. 1999 Jul Jul;26(7):1474-80.
187. Apolone G, Filiberti A, Cifani S, Ruggiata R, Mosconi P. Evaluation of the EORTC QLQ-C30 questionnaire: a comparison with SF-36 Health Survey in a cohort of Italian long-survival cancer patients. Ann Oncol. 1998 May;9(5):549-57.
188. Kuenstner S, Langelotz C, Budach V, Possinger K, Krause B, Sezer O. The comparability of quality of life scores. a multitrait multimethod analysis of the EORTC QLQ-C30, SF-36 and FLIC questionnaires. Eur J Cancer. 2002 Feb;38(3):339-48.
189. Llewellyn CD, McGurk M, Weinman J. Head and neck cancer: to what extent can psychological factors explain differences between health-related quality of life and individual quality of life? Br J Oral Maxillofac Surg. 2006 Oct;44(5):351-7.
190. Gil KM, Gibbons HE, Jenison EL, Hopkins MP, von Gruenigen VE. Baseline characteristics influencing quality of life in women undergoing gynecologic oncology surgery. Health Qual Life Outcomes. 2007;5:25.
191. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology. 2000 Dec 20;56(6):899-905.
192. Overcash J, Extermann M, Parr J, Perry J, Balducci L. Validity and reliability of the FACT-G scale for use in the older person with cancer. American Journal of Clinical Oncology. 2001;24(6):591-6.
193. Conroy T, Mercier M, Schraub S, al e. Comparison of three quality of life instruments for cancer patients: FACT-G, EORTC QLQ-C30 and FLIC Quality of Life Research. 2000;9:274.
194. Blazeby JM, Kavadas V, Vickery CW, Greenwood R, Berrisford RG, Alderson D. A prospective comparison of quality of life measures for patients with esophageal cancer. Quality of Life Research. 2005 Mar;14(2):387-93.
195. Holzner B, Kemmler G, Sperner-Unterweger B, Kopp M, Dunser M, Margreiter R, et al. Quality of life measurement in oncology--a matter of the assessment instrument? Eur J Cancer. 2001 Dec;37(18):2349-56.
196. Holzner B, Bode RK, Hahn EA, Cella D, Kopp M, Sperner-Unterweger B, et al. Equating EORTC QLQ-C30 and FACT-G scores and its use in oncological research. Eur J Cancer. 2006 Dec;42(18):3169-77.

197. Sandgren AK, McCaul KD. Short-term effects of telephone therapy for breast cancer patients. Health Psychology. 2003 May;22(3):310-5.
198. Sandgren AK, McCaul KD, Sandgren AK, McCaul KD. Long-term telephone therapy outcomes for breast cancer patients. Psychooncology. 2007 Jan;16(1):38-47.
199. Dwight-Johnson M, Ell K, Lee PJ, Dwight-Johnson M, Ell K, Lee P-J. Can collaborative care address the needs of low-income Latinas with comorbid depression and cancer? Results from a randomized pilot study. Psychosomatics. 2005 May-Jun;46(3):224-32.
200. Dirksen SR, Epstein DR, Dirksen SR, Epstein DR. Efficacy of an insomnia intervention on fatigue, mood and quality of life in breast cancer survivors. Journal of Advanced Nursing. 2008 Mar;61(6):664-75.
201. Graves KD, Carter CL, Anderson ES, Winett RA. Quality of life pilot intervention for breast cancer patients: use of social cognitive theory. Palliative & Supportive Care [serial on the Internet]. 2003; (2): Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/037/CN-00556037/frame.html.
202. Espie CA, Fleming L, Cassidy J, Samuel L, Taylor LM, White CA, et al. Randomized controlled clinical effectiveness trial of cognitive behavior therapy compared with treatment as usual for persistent insomnia in patients with cancer. J Clin Oncol. 2008 Oct 1;26(28):4651-8.
203. Moadel AB, Shah C, Wylie-Rosett J, Harris MS, Patel SR, Hall CB, et al. Randomized controlled trial of yoga among a multiethnic sample of breast cancer patients: effects on quality of life.[see comment]. J Clin Oncol. 2007 Oct 1;25(28):4387-95.
204. Mutrie N, Campbell A.M., Whyte F., McConnachie A., Emslie C., Lee L., Kearney N., Walker A. and Ritchie D. Benefits of supervised group exercise programme for women being treated for early stage breast cancer: Pragmatic randomised controlled trial British Medical Journal 2007;334(7592):517-20
205. Kopp M, Schweigkofler H, Holzner B, Nachbaur D, Niederwieser D, Fleischhacker WW, et al. EORTC QLQ-C30 and FACT-BMT for the measurement of quality of life in bone marrow transplant recipients: a comparison. Eur J Haematol. 2000 Aug;65(2):97-103.
206. Kemmler G, Holzner B, Kopp M, Dunser M, Margreiter R, Greil R, et al. Comparison of two quality-of-life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30. Journal of Clinical Oncology. 1999 Sep;17(9):2932-40.
207. Cheung Y-B, Goh C, Thumboo J, Khoo K-S, Wee J. Variability and sample size requirements of quality-of-life measures: a randomized study of three major questionnaires.[see comment]. Journal of Clinical Oncology. 2005 Aug 1;23(22):4936-44.
208. Rodary C, Pezet-Langevin V, Garcia-Acosta S, Lesimple T, Lortholary A, Kaminsky MC, et al. Patient preference for either the EORTC QLQ-C30 or the FACIT Quality Of Life (QOL) measures: a study performed in patients suffering from carcinoma of an unknown primary site (CUP). Eur J Cancer. 2004 Mar;40(4):521-8.
209. Mehanna HM, Morton RP. Patients' views on the utility of quality of life questionnaires in head and neck cancer: a randomised trial. Clin Otolaryngol. 2006 Aug;31(4):310-6.
210. Snyder CF, Dy SM, Hendricks DE, Brahmer JR, Carducci MA, Wolff AC, et al. Asking the right questions: investigating needs assessments and health-related

quality-of-life questionnaires for use in oncology clinical practice. Support Care Cancer. 2007 15:1075-85.
211. McQuellon RP, Russell GB, Cella DF, Craven BL, Brady M, Bonomi A, et al. Quality of life measurement in bone marrow transplantation: development of the Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT) scale. Bone Marrow Transplant. 1997 Feb;19(4):357-68.
212. Paterson AG, Trask PC, Wagner LI, Esper P, Redman B. Validation of the FACT-BRM with interferon-alpha treated melanoma patients. Quality of Life Research. 2005 Feb;14(1):133-9.
213. Sherman AC, Simonton S, Adams DC, Vural E, Owens B, Hanna E. Assessing quality of life in patients with head and neck cancer: cross-validation of the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Head and Neck module (QLQ-H&N35). Arch Otolaryngol Head Neck Surg. 2000 Apr;126(4):459-67.
214. Janda M, Obermair A, Cella D, Perrin LC, Nicklin JL, Ward BG, et al. The Functional Assessment of Cancer-Vulvar: reliability and validity. Gynecol Oncol. 2005 May;97(2):568-75.
215. Skarstein J, Aass N, Fossa SD, Skovlund E, Dahl AA. Anxiety and depression in cancer patients: relation between the Hospital Anxiety and Depression Scale and the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire. J Psychosom Res. 2000 Jul;49(1):27-34.
216. Brady MJ, D.F C, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast (FACT-B) quality of life instrument. Journal of Clinical Oncology. 1997;15:974-86.
217. Meraner V, Giesinger J, Kemmler G, Taucher S, Hubalek M, Weber B, et al. Development of a screening tool for the identification of psychooncological treatment need in breast cancer patients. Psychooncology. 2009;18:DOI: 10.1002/pon.480.
218. Fossa S, Dahl AA. Short Form 36 and Hospital Anxiety and Depression Scale. A comparison based on patients with testicular cancer. J Psychosom Res. 2002;52(2):79-87.
219. Culos-Reed SN, Robinson JL, Lau H, O'Connor K, Keats MR. Benefits of a physical activity intervention for men with prostate cancer. J Sport Exerc Psychol. 2007 Feb;29(1):118-27.
220. Katz MR, Irish JC, Devins GM, Katz MR, Irish JC, Devins GM. Development and pilot testing of a psychoeducational intervention for oral cancer patients. Psychooncology. 2004 Sep;13(9):642-53.
221. Wilkinson SM, Love SB, Westcombe AM, Gambles MA, Burgess CC, Cargill A, et al. Effectiveness of aromatherapy massage in the management of anxiety and depression in patients with cancer: a multicenter randomized controlled trial. J Clin Oncol. 2007 Feb 10;25(5):532-9.
222. Kornblith AB, Dowell JM, Herndon JE, 2nd, Engelman BJ, Bauer-Wu S, Small EJ, et al. Telephone monitoring of distress in patients aged 65 years or older with advanced stage cancer: a cancer and leukemia group B study. Cancer. 2006 Dec 1;107(11):2706-14.
223. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ, Lemieux J, et al. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. Journal of Clinical Epidemiology. 2007 May;60(5):448-55.

224. Powell CB, Kneier A, Chen L-m, Rubin M, Kronewetter C, Levine E. A randomized study of the effectiveness of a brief psychosocial intervention for women attending a gynecologic cancer clinic. Gynecol Oncol. 2008 Oct;111(1):137-43.
225. Courneya KS, Friedenreich CM, Sela RA, Quinney HA, Rhodes RE, Handman M. The group psychotherapy and home-based physical exercise (group-hope) trial in cancer survivors: physical fitness and quality of life outcomes. Psychooncology. 2003 Jun;12(4):357-74.
226. Owen JE, Klapow JC, Roth DL, J.L S, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer Ann Behav Med. 2005;30(1):54-64.
227. Carmack Taylor CL, Demoor C, Smith MA, Dunn AL, Basen-Engquist K, Nielsen I, et al. Active for Life After Cancer: a randomized trial examining a lifestyle physical activity program for prostate cancer patients. Psychooncology. 2006 Oct;15(10):847-62.
228. Steel JL, Nadeau K, Olek M, Carr B, et al. Preliminary results of an individually tailored psychosocial intervention for patients with advanced hepatobiliary carcinoma. J Psychosoc Oncol. 2007;25(3):19-42.
229. Downe-Wamboldt BL, Butler LJ, Melanson PM, Coulter LA, Singleton JF, Keefe JM, et al. The effects and expense of augmenting usual cancer clinic care with telephone problem-solving counseling. Cancer Nurs. 2007 Nov-Dec;30(6):441-53.
230. Sandgren AK, McCaul KD. Short-term effects of telephone therapy for breast cancer patients. Health Psychol. 2003 May;22(3):310-5.
231. Sandgren AK, McCaul KD. Long-term telephone therapy outcomes for breast cancer patients. Psychooncology. 2007 Jan;16(1):38-47.
232. Giesler RB, Given B, Given CW, Rawl S, Monahan P, Burns D, et al. Improving the quality of life of patients with prostate carcinoma: a randomized trial testing the efficacy of a nurse-driven intervention. Cancer. 2005 Aug 15;104(4):752-62.
233. Cole BS. Spiritually-focused psychotherapy for people diagnosed with cancer: A pilot outcome study. Mental Health, Religion & Culture. 2005 Sep;8(3):217-26.
234. Lepore SJ, Helgeson VS, Eton DT, Schulz R. Improving quality of life in men with prostate cancer: a randomized controlled trial of group education interventions. Health Psychol. 2003 Sep;22(5):443-52.
235. Monti DA, Peterson C, Kunkel EJS, Hauck WW, Pequignot E, Rhodes L, et al. A randomized, controlled trial of mindfulness-based art therapy (MBAT) for women with cancer. Psychooncology. 2006 May;15(5):363-73.
236. McCorkle R, Dowd M, Ercolano E, Schulman-Green D, Williams AL, Siefert ML, et al. Effects of a nursing intervention on quality of life outcomes in post-surgical women with gynecological cancers. Psychooncology. 2009;18(1):62-70.
237. Hanser SB, Bauer-Wu S, Kubicek L, Healey M, Manola J, Hernandez M, et al. Effects of a music therapy intervention on quality of life and distress in women with metastatic breast cancer. J Soc Integr Oncol. 2006;4(3):116-24.
238. Anderson KO, Mendoza TR, Payne R, Valero V, Palos GR, Nazario A, et al. Pain education for underserved minority cancer patients: a randomized controlled trial. Journal of Clinical Oncology. 2004 Dec 15;22(24):4918-25.
239. Daubenmier JJ, Weidner G, Marlin R, Crutchfield L, Dunn-Emke S, Chi C, et al. Lifestyle and health-related quality of life of men with prostate cancer managed with active surveillance. Urology. 2006 Jan;67(1):125-30.

240. Courneya KS, Friedenreich CM, Quinney HA, Fields ALA, Jones LW, Fairey AS. A randomized trial of exercise and quality of life in colorectal cancer survivors. European Journal of Cancer Care. 2003 Dec;12(4):347-57.
241. Courneya KS, Friedenreich CM, Sela RA, Quinney HA, Rhodes RE, Handman M, et al. The group psychotherapy and home-based physical exercise (group-hope) trial in cancer survivors: physical fitness and quality of life outcomes. Psychooncology. 2003 Jun;12(4):357-74.
242. McCorkle R, Dowd M, Ercolano E, Schulman-Green D, Williams AL, Siefert ML, et al. Effects of a nursing intervention on quality of life outcomes in post-surgical women with gynecological cancers. Psychooncology. 2009 Jan;18(1):62-70.
243. Giesler RB, Given B, Given CW, Rawl S, Monahan P, Burns D, et al. Improving the quality of life of patients with prostate carcinoma: a randomized trial testing the efficacy of a nurse-driven intervention. Cancer. 2005 Aug 15;104(4):752-62.
244. Carmack Taylor CL, Demoor C, Smith MA, Dunn AL, Basen-Engquist K, Nielsen I, et al. Active for life after cancer: A randomized trial examining a lifestyle physical activity program for prostate cancer patients. Psychooncology. 2006;15(10):847-62.
245. Krischer MM, Xu P, Meade CD, Jacobsen PB, Krischer MM, Xu P, et al. Self-administered stress management training in patients undergoing radiotherapy. J Clin Oncol. 2007 Oct 10;25(29):4657-62.
246. Lepore SJ, Helgeson VS, Eton DT, Schulz R. Improving Quality of Life in Men With Prostate Cancer: A Randomized Controlled Trial of Group Education Interventions. 2003.
247. Owen JE, Klapow JC, Roth DL, Shuster JL, Jr., Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Annals of Behavioral Medicine. 2005 Aug;30(1):54-64.
248. Culos-Reed SN, Carlson LE, Daroux LM, Hately-Aldous S. A pilot study of yoga for breast cancer survivors: physical and psychological benefits. Psychooncology. 2006 Oct;15(10):891-7.
249. Powell CB, Kneier A, Chen LM, Rubin M, Kronewetter C, Levine E, et al. A randomized study of the effectiveness of a brief psychosocial intervention for women attending a gynecologic cancer clinic. Gynecologic Oncology. 2008 Oct;111(1):137-43.
250. Monti DA, Peterson C, Kunkel EJ, Hauck WW, Pequignot E, Rhodes L, et al. A randomized, controlled trial of mindfulness-based art therapy (MBAT) for women with cancer. Psychooncology. 2006 May;15(5):363-73.
251. Badger T, Segrin C, Dorros SM, Meek P, Lopez AM. Depression and anxiety in women with breast cancer and their partners. Nursing Research [serial on the Internet]. 2007; (1): Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/611/CN-00574611/frame.html.
252. Orringer JS, Fendrick AM, Trask PC, Bichakjian CK, Schwartz JL, Wang TS, et al. The effects of a professionally produced videotape on education and anxiety/distress levels for patients with newly diagnosed melanoma: a randomized, prospective clinical trial. Journal of the American Academy of Dermatology. 2005 Aug;53(2):224-9.
253. Hanser SB, Bauer-Wu S, Kubicek L, Healey M, Manola J, Hernandez M, et al. Effects of a music therapy intervention on quality of life and distress in women with metastatic breast cancer. J Soc Integr Oncol. 2006;4(3):116-24.

254. Steel JL, Nadeau K, Olek M, Carr BI. Preliminary results of an individually tailored psychosocial intervention for patients with advanced hepatobiliary carcinoma. J Psychosoc Oncol. 2007;25(3):19-42.
255. Kissane DW, Bloch S, Smith GC, Miach P, Clarke DM, Ikin J, et al. Cognitive-existential group psychotherapy for women with primary breast cancer: a randomised controlled trial. Psychooncology. 2003 Sep;12(6):532-46.
256. Sharpe M, Strong V, Allen K, Rush R, Postma K, Tulloh A, et al. Major depression in outpatients attending a regional cancer centre: screening and unmet treatment needs. British Journal of Cancer. 2004 Jan 26;90(2):314-20.
257. Schofield P, Payne S. A pilot study into the use of a multisensory environment (Snoezelen) within a palliative day-care setting. International journal of palliative nursing [serial on the Internet]. 2003; (3): Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/239/CN-00423239/frame.html.
258. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl). 2006 May;15(2):163-71.
259. Clark M, Isaacks-Downton G, Wells N, Redlin-Frazier S, Eck C, Hepworth JT, et al. Use of preferred music to reduce emotional distress and symptom activity during radiation therapy. Journal of Music Therapy. 2006;43(3):247-65.
260. Pitceathly C, Maguire P, Fletcher I, Parle M, Tomenson B, Creed F. Can a brief psychological intervention prevent anxiety or depressive disorders in cancer patients? A randomised controlled trial. Ann Oncol. 2009;20(5):928-34.
261. Manne SL, Ostroff JS, Winkel G, Fox K, Grana G, Miller E, et al. Couple-focused group intervention for women with early stage breast cancer. J Consult Clin Psychol. 2005 Aug;73(4):634-46.
262. Allen SM, Shah AC, Nezu AM, Nezu CM, Ciambrone D, Hogan J, et al. A problem-solving approach to stress reduction among younger women with breast carcinoma: a randomized controlled trial. Cancer. 2002 Jun 15;94(12):3089-100.
263. Edmonds CVI, Lockwood GA, Cunningham AJ. Psychological response to long term group therapy: A randomized trial with metastatic breast cancer patients. Psychooncology. 1999;8(1):74-91.
264. Rosenbloom SK, David E. Victorson, D. E., Hahn, E. A., Peterman, A. H., Cella, D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. Psychooncology. 2007;16 1069–79.
265. Bailey DE, Mishel MH, Belyea M, Stewart JL, Mohler J, Bailey DE, et al. Uncertainty intervention for watchful waiting in prostate cancer. Cancer Nurs. 2004 Sep-Oct;27(5):339-46.
266. Pinto BM, Clark MM, Maruyama NC, Feder SI, Pinto BM, Clark MM, et al. Psychological and fitness changes associated with exercise participation among women with breast cancer. Psychooncology. 2003 Mar;12(2):118-26.
267. Ward S, Donovan H, Gunnarsdottir S, Serlin RC, Shapiro GR, Hughes S. A Randomized Trial of a Representational Intervention to Decrease Cancer Pain (RIDcancerPain). Health Psychology. 2008;27(1):59-67.
268. Strong V, Waters R, Hibberd C, Murray G, Wall L, Walker J, et al. Management of depression for people with cancer (SMaRT oncology 1): a randomised trial.[see comment]. Lancet. 2008 Jul 5;372(9632):40-8.

269. Andersen BL, Farrar WB, Golden-Kreutz DM, Glaser R, Emery CF, Crespin TR, et al. Psychological, behavioral, and immune changes after a psychological intervention: a clinical trial. J Clin Oncol. 2004 Sep 1;22(17):3570-80.
270. Given B, Given CW, McCorkle R, Kozachik S, Cimprich B, Rahbar MH, et al. Pain and fatigue management: results of a nursing randomized clinical trial. Oncology Nursing Forum Online. 2002 Jul;29(6):949-56.
271. Cook CA, Guerrerio JF, Slater VE, Cook CAL, Guerrerio JF, Slater VE. Healing touch and quality of life in women receiving radiation treatment for cancer: a randomized controlled trial. Alternative Therapies in Health & Medicine. 2004 May-Jun;10(3):34-41.
272. Sandel SL, Judge JO, Landry N, Faria L, Ouellette R, Majczak M, et al. Dance and movement program improves quality-of-life measures in breast cancer survivors. Cancer Nurs. 2005 Jul-Aug;28(4):301-9.
273. Velikova G, Wright EP, Smith AB, Cull A, Gould A, Forman D, et al. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. Journal of Clinical Oncology. 1999 Mar;17(3):998-1007.
274. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. Behav Res Ther. 1997 Aug;35(8):785-91.
275. Cheung Y-B, Lim C, Goh C, Thumboo J, Wee J. Order effects: a randomised study of three major cancer-specific quality of life instruments. Health Qual Life Outcomes. 2005;3:37.
276. Smith AB, Rush R, Velikova G, Wall L, Wright EP, Stark D, et al. The initial development of an item bank to assess and screen for psychological distress in cancer patients.[erratum appears in Psychooncology. 2007 Aug;16(8):724]. Psychooncology. 2007 Aug;16(8):724-32.
277. Cook KF, Teal CR, Bjorner JB, Cella D, Chang C-H, Crane PK, et al. IRT health outcomes data analysis project: an overview and summary. Quality of Life Research. 2007;16 Suppl 1:121-32.
278. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol. 2005 Sep-Oct;23(5 Suppl 39):S53-7.
279. Ader DNP. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007 May;45(5):S1-S2.
280. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years.[see comment]. Med Care. 2007 May;45(5 Suppl 1):S3-S11.
281. DeWalt DAMDMPH, Rothrock NP, Yount SP, Stone AAP, on behalf of the PCG. Evaluation of Item Candidates: The PROMIS Qualitative Item Review. Med Care. 2007 May;45(5):S12-S21.
282. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS).[see comment]. Journal of Clinical Epidemiology. 2008 Jan;61(1):17-33.
283. Castel LD, Williams KA, Bosworth HB, Eisen SV, Hahn EA, Irwin DE, et al. Content validity in the PROMIS social-health domain: a qualitative analysis of focus-group data. Quality of Life Research. 2008 Jun;17(5):737-49.

284. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007 May;45(5 Suppl 1):S22-31.
285. Bottomley A, Vachalec S, Bjordal K, Blazeby J, Flechtner H, Ruyskart P. The development and utilisation of the European Organisation for research and treatment of cancer quality of life group item bank. Eur J Cancer. 2002 Aug;38(12):1611-4.
286. Hann D, Winter K, Jacobsen P. Measurement of depressive symptoms in cancer patients: evaluation of the Center for Epidemiological Studies Depression Scale (CES-D). J Psychosom Res. 1999 May;46(5):437-43.
287. Stockler MR, Osoba D, Goodwin P, Corey P, Tannock IF. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. European Organization for Research and Treatment of Cancer. Journal of Clinical Epidemiology. 1998 Feb;51(2):137-45.
288. Stockler MR, Osoba D, Corey P, Goodwin PJ, Tannock IF. Convergent discriminitive, and predictive validity of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) assessment and comparison with analogous scales from the EORTC QLQ-C30 and a trial-specific module. European Organisation for Research and Treatment of Cancer. Core Quality of Life Questionnaire. Journal of Clinical Epidemiology. 1999 Jul;52(7):653-66.
289. Osoba D, Aaronson N, Zee B, Sprangers M, 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;6:103-8.
290. Nowak AK, Stockler MR, Byrne MJ. Assessing quality of life during chemotherapy for pleural mesothelioma: feasibility, validity, and results of using the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire and Lung Cancer Module. Journal of Clinical Oncology. 2004 Aug 1;22(15):3172-80.
291. Lagergren P, Fayers P, Conroy T, Stein HJ, Sezer O, Hardwick R, et al. Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-OG25, to assess health-related quality of life in patients with cancer of the oesophagus, the oesophago-gastric junction and the stomach. Eur J Cancer. 2007 Sep;43(14):2066-73.
292. Blazeby JM, Fayers P, Conroy T, Sezer O, Ramage J, Rees M, et al. Validation of the European Organization for Research and Treatment of Cancer QLQ-LMC21 questionnaire for assessment of patient-reported outcomes during treatment of colorectal liver metastases. Br J Surg. 2009 Mar;96(3):291-8.
293. Blazeby JM, Conroy T, Bottomley A, Vickery C, Arraras J, Sezer O, et al. Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-STO 22, to assess quality of life in patients with gastric cancer. Eur J Cancer. 2004 Oct;40(15):2260-8.
294. Webster K, Odom L, Peterman A, Lent L, Cella D. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: Validation of version 4 of the core questionnaire. Quality of Life Research. 1999;8(7):604.
295. Weitzner MA, Meyers CA, Gelke C, Byrne K, Cella DF, Levin VA. The Functional Assessment of Cancer Therapy (FACT) scale: Development of a brain scale and revalidation of the general version (FACT-G) in patients with primary brain tumors. Cancer. 1995;75(5):1151-61.

296. Cormier JN, Davidson L, Xing Y, Webster K, Cella D. Measuring quality of life in patients with melanoma: development of the FACT-melanoma scale.[see comment]. J Support Oncol. 2005 Mar-Apr;3(2):139-45.
297. Basen-Engquist K, Bodurka-Bevers D, Fitzgerald MA, Webster K, Cella D, Hu S, et al. Reliability and Validity of the Functional Assessment of Cancer Therapy-Ovarian. Journal of Clinical Oncology. 2001;19(6):1809-17.
298. Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology. 1997 Dec;50(6):920-8.
299. Bacik J, Mazumdar M, Murphy BA, Fairclough DL, Eremenco S, Mariani T, et al. The functional assessment of cancer therapy-BRM (FACT-BRM): a new tool for the assessment of quality of life in patients treated with biologic response modifiers. Quality of Life Research. 2004 Feb;13(1):137-54.
300. Calhoun EA, Welshman EE, Chang CH, Lurain JR, Fishman DA, Hunt TL, et al. Psychometric evaluation of the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (Fact/GOG-Ntx) questionnaire for patients receiving systemic chemotherapy. Int J Gynecol Cancer. 2003 Nov-Dec;13(6):741-8.
301. 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 & Treatment. 2001 Aug;68(3):273-82.
302. Lyons KD, Bakitas M, Hegel MT, Hanscom B, Hull J, Ahles TA. Reliability and Validity of the Functional Assessment of Chronic Illness and Therapy - Palliative Care (FACIT-Pal) Scale Journal of Pain and Symptom Management. 2008.
303. Wagner LI, Beaumont JL, Ding B, Malin J, Peterman A, Calhoun E, et al. Measuring health-related quality of life and neutropenia-specific concerns among older adults undergoing chemotherapy: validation of the Functional Assessment of Cancer Therapy-Neutropenia (FACT-N). Supportive Care in Cancer. 2008 Jan;16(1):47-56.
304. Yellen SB, Cella DF, Webster KA, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. Journal of Pain and Symptom Management. 1997;13:63-74.
305. Cella D, Peterman A, Hudgens S, Webster K, Socinski MA. Measuring the side effects of taxane therapy in oncology: the functional assesment of cancer therapy-taxane (FACT-taxane). Cancer. 2003 Aug 15;98(4):822-31.
306. Carroll BT, Kathol RG, Noyes R, Jr., Wald TG, Clamon GH. Screening for depression and anxiety in cancer patients using the Hospital Anxiety and Depression Scale. Gen Hosp Psychiatry. 1993 Mar;15(2):69-74.
307. Sellick SM, Edwardson AD. Screening new cancer patients for psychological distress using the Hospital Anxiety and Depression Scale. Psycho-Oncology Vol 16(6) Jun 2007, 534-542. 2007.
308. Walker J, Postma K, McHugh GS, Rush R, Coyle B, Strong V, et al. Performance of the Hospital Anxiety and Depression Scale as a screening tool for major depressive disorder in cancer patients. J Psychosom Res. 2007 Jul;63(1):83-91.
309. Love AW, Kissane DW, Bloch S, Clarke D. Diagnostic efficiency of the Hospital Anxiety and Depression Scale in women with early stage breast cancer. Aust N Z J Psychiatry. 2002 Apr;36(2):246-50.
310. Love AW, Grabsch B, Clarke DM, Bloch S, Kissane DW. Screening for depression in women with metastatic breast cancer: a comparison of the Beck

Depression Inventory Short Form and the Hospital Anxiety and Depression Scale. Aust N Z J Psychiatry. 2004 Jul;38(7):526-31.
311. Manne S, Schnoll R. Measuring cancer patients' psychological distress and well-being: a factor analytic assessment of the Mental Health Inventory. Psychological Assessment. 2001 Mar;13(1):99-109.
312. Cordova MJ, Studts JL, Hann DM, Jacobsen PB, Andrykowski MA. Symptom structure of PTSD following breast cancer. J Trauma Stress. 2000 Apr;13(2):301-19.
313. Badger T, Segrin C, Meek P, Lopez AM, Bonham E, Sieger A, et al. Telephone interpersonal counseling with women with breast cancer: symptom management and quality of life. Oncology Nursing Forum Online. 2005 Mar;32(2):273-9.
314. Penedo FJ, Dahn JR, Molton I, Gonzalez JS, Kinsinger D, Roos BA, et al. Cognitive-behavioral stress management improves stress-management skills and quality of life in men recovering from treatment of prostate carcinoma. Cancer. 2004 Jan 1;100(1):192-200.
315. Penedo FJ, Molton I, Dahn JR, Shen BJ, Kinsinger D, Traeger L, et al. A randomized clinical trial of group-based cognitive-behavioral stress management in localized prostate cancer: development of stress management skills improves quality of life and benefit finding. Annals of Behavioral Medicine. 2006 Jun;31(3):261-70.
316. Savard J, Simard S, Giguere I, Ivers H, Morin CM, Maunsell E, et al. Randomized clinical trial on cognitive therapy for depression in women with metastatic breast cancer: psychological and immunological effects. Palliative & Supportive Care. 2006 Sep;4(3):219-37.
317. Daubenmier JJ, Weidner G, Marlin R, Crutchfield L, Dunn-Emke S, Chi C, et al. Lifestyle and health-related quality of life of men with prostate cancer managed with active surveillance. Urology. 2006 Jan;67(1):125-30.
318. Anderson KO, Mendoza TR, Payne R, Valero V, Palos GR, Nazario A, et al. Pain education for underserved minority cancer patients: a randomized controlled trial. J Clin Oncol. 2004 Dec 15;22(24):4918-25.
319. Doorenbos A, Given B, Given C, Verbitsky N, Cimprich B, McCorkle R, et al. Reducing symptom limitations: a cognitive behavioral intervention randomized trial. Psychooncology. 2005 Jul;14(7):574-84.

APPENDICES

Appendix A: DSM-IV-TR criteria for generalised anxiety disorder and major depressive disorder (1)

Generalised Anxiety Disorder

A. At least 6 months of "excessive anxiety and worry" about a variety of events and situations. Generally, "excessive" can be interpreted as more than would be expected for a particular situation or event. Most people become anxious over certain things, but the intensity of the anxiety typically corresponds to the situation.

B. There is significant difficulty in controlling the anxiety and worry. If someone has a very difficult struggle to regain control, relax, or cope with the anxiety and worry, then this requirement is met.

C. The presence for most days over the previous six months of 3 or more (only 1 for children) of the following symptoms:

1. Feeling wound-up, tense, or restless
2. Easily becoming fatigued or worn-out
3. Concentration problems
4. Irritability
5. Significant tension in muscles
6. Difficulty with sleep

D. The symptoms are not part of another mental disorder.

E. The symptoms cause "clinically significant distress" or problems functioning in daily life. "Clinically significant" is the part that relies on the perspective of the treatment provider. Some people can have many of the aforementioned symptoms and cope with them well enough to maintain a high level of functioning.

F. The condition is not due to a substance or medical issue

Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
Note: Do note include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.

(1) depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). **Note:** In children and adolescents, can be irritable mood.

(2) markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)

(3) significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. **Note:** In children, consider failure to make expected weight gains.

(4) insomnia or hypersomnia nearly every day

(5) psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)

(6) fatigue or loss of energy nearly every day

(7) feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)

(8) diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)

(9) recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms do not meet criteria for a Mixed Episode.

C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).

E. The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

Major Depressive Disorder

Single Episode

A. Presence of a single Major Depressive Episode

B. The Major Depressive Episode is not better accounted for by Schizoaffective Disorder and is not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.

C. There has never been a Manic Episode, a Mixed Episode, or a Hypomanic Episode. **Note**: This exclusion does not apply if all the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects of a general medical condition.

Recurrent

A. Presence of two or more Major Depressive Episodes. **Note**: To be considered separate episodes, there must be an interval of at least 2 consecutive months in which criteria are not met for a Major Depressive Episode.

B. The Major Depressive Episodes are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.

C. There has never been a Manic Episode, a Mixed Episode, or a Hypomanic Episode. **Note**: This exclusion does not apply if all the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects or a general medical condition.

Appendix B: Lists of terms used to search each database

Database(s)	Search terms
MEDLINE and AMED	1. exp behavior therapy/ or exp "biofeedback (psychology)"/ or exp cognitive therapy/ or exp desensitization, psychologic/ or exp relaxation therapy/ 2. exp counseling/ or exp directive counseling/ or exp pastoral care/ 3. exp psychotherapy, group/ or exp family therapy/ 4. exp mind-body therapies/ or exp aromatherapy/ or exp breathing exercises/ or exp hypnosis/ or exp "imagery (psychotherapy)"/ or exp laughter therapy/ or exp meditation/ or exp mental healing/ or exp "mind-body relations (metaphysics)"/ or exp psychophysiology/ or exp tai ji/ or exp therapeutic touch/ or exp yoga/ 5. exp sensory art therapies/ or exp art therapy/ or exp color therapy/ or exp dance therapy/ or exp music therapy/ or exp play therapy/ or exp psychodrama/ 6. exp psychotherapy/ or exp autogenic training/ or exp bibliotherapy/ or exp crisis intervention/ or exp gestalt therapy/ or exp nondirective therapy/ or exp psychoanalytic therapy/ or exp psychotherapeutic processes/ or exp psychotherapy, brief/ or exp psychotherapy, multiple/ or exp psychotherapy, rational-emotive/ or exp reality therapy/ or exp socioenvironmental therapy/ 7. (psycho$ intervention or behaviour$ therapy or cognitive therapy or cognitive-behaviour$ therapy or communication skills training or counsel$ or psycho-education or family therapy or guided imagery or music therapy or problem-solving therapy or psychotherapy or relax$ or stress management or support group or supportive-expressive group therapy or hypno$ or meditat$ or desensiti$ or mindful$ or acceptance commitment therapy).mp. [mp=title, original title, abstract, name of substance word, subject heading word] 8. 1 or 2 or 3 or 4 or 5 or 6 or 7 9. neoplasm.mp. or exp Neoplasms/ 10. anxiety.mp. or exp Anxiety/ or exp Anxiety Disorders/ 11. exp Depression/ or depression.mp. 12. distress.mp. 13. quality of life.mp. or exp "Quality of Life"/ 14. 10 or 11 or 12 or 13 15. 8 and 9 and 14 16. limit 15 to (english language and yr="1999 -Current" and randomized controlled trial)
PSYCHINFO	1. exp behavior therapy/ or exp behavior modification/ or exp aversion therapy/ or exp conversion therapy/ or exp exposure therapy/ or exp implosive therapy/ or exp reciprocal inhibition therapy/ or exp "response cost"/ or exp systematic desensitization therapy/ or exp anger control/ or exp anxiety management/ or exp behavior/ or exp behavior contracting/ or exp cognitive behavior therapy/ or exp counterconditioning/ or exp eye movement

desensitization therapy/ or exp paradoxical techniques/
2. exp cognitive therapy/ or exp cognitive techniques/ or exp cognitive restructuring/ or exp rational emotive behavior therapy/ or exp self instructional training/ or exp self management/
3. exp communication skills training/ or exp training/ or exp assertiveness training/ or exp communication skills/ or exp human relations training/ or exp social skills training/
4. exp counseling/ or exp group counseling/ or exp pastoral counseling/ or exp peer counseling/ or exp psychotherapeutic counseling/ or exp rehabilitation counseling/ or exp counseling psychology/ or exp counselors/ or exp health care services/
5. exp psychoeducation/ or exp client education/ or exp health education/
6. exp family therapy/ or exp conjoint therapy/ or exp family intervention/ or exp family life education/
7. exp guided imagery/ or exp psychotherapeutic techniques/ or exp hypnotherapy/ or exp imagery/
8. exp music therapy/ or exp creative arts therapy/ or exp movement therapy/ or exp recreation therapy/
9. exp problem solving/
10. exp psychotherapy/ or exp adlerian psychotherapy/ or exp adolescent psychotherapy/ or exp analytical psychotherapy/ or exp autogenic training/ or exp brief psychotherapy/ or exp child psychotherapy/ or exp client centered therapy/ or exp conversion therapy/ or exp eclectic psychotherapy/ or exp emotion focused therapy/ or exp existential therapy/ or exp experiential psychotherapy/ or exp expressive psychotherapy/ or exp eye movement desensitization therapy/ or exp geriatric psychotherapy/ or exp gestalt therapy/ or exp group psychotherapy/ or exp individual psychotherapy/ or exp insight therapy/ or exp integrative psychotherapy/ or exp interpersonal psychotherapy/ or exp logotherapy/ or exp narrative therapy/ or exp persuasion therapy/ or exp primal therapy/ or exp psychoanalysis/ or exp psychodrama/ or exp psychodynamic psychotherapy/ or exp psychotherapeutic counseling/ or exp rational emotive behavior therapy/ or exp reality therapy/ or exp relationship therapy/ or exp solution focused therapy/ or exp supportive psychotherapy/
11. exp relaxation therapy/ or exp progressive relaxation therapy/ or exp autogenic training/ or exp meditation/ or exp muscle relaxation/ or exp posthypnotic suggestions/ or exp psychotherapeutic techniques/ or exp systematic desensitization therapy/
12. exp stress management/ or exp cognitive techniques/ or exp stress/
13. support groups/ or exp twelve step programs/ or exp self help techniques/ or exp social support/
14. (psycho$ intervention or behaviour$ therapy or cognitive therapy or cognitive-behaviour$ therapy or communication skills training or counsel$ or psycho-education or family therapy or guided imagery or music therapy or problem-solving therapy or psychotherapy or relax$ or stress management or support group or

	supportive-expressive group therapy or hypno$ or meditat$ or desensiti$ or mindful$ or acceptance commitment).mp. [mp=title, abstract, heading word, table of contents, key concepts] 15. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 16. exp Anxiety Disorders/ or exp Anxiety/ or anxiety.mp. 17. exp "Depression (Emotion)"/ or depression.mp. or exp Major Depression/ 18. quality of life.mp. or exp "Quality of Life"/ 19.exp Distress/ or distress.mp. 20. 16 or 17 or 18 or 19 21. neoplasm$.mp. or exp Neoplasms/ 22. 15 and 20 and 21 23. limit 22 to (english language and "treatment outcome/randomized clinical trial" and yr="1999 -Current")
EMBASE	1. 'psychological and psychiatric procedures, techniques and concepts'/exp 2. 'quality of life'/exp 3. 'anxiety'/exp 4. 'depression'/exp 5. ‘distress’/exp OR 'distress' 6. 2 OR 3 OR 4 OR 5 7. 1 AND 6 AND [randomized controlled trial]/lim AND [cancer]/lim AND [english]/lim AND [embase]/lim AND [1999-2009]
CENTRAL	1. MeSH descriptor Quality of Life explode all trees 2. MeSH descriptor Anxiety explode all trees 3. MeSH descriptor Depression explode all trees 4. (anxiety):kw or (depression):kw or (quality of life):kw or (distress):kw 5. MeSH descriptor Relaxation Therapy explode all trees 6. MeSH descriptor Music Therapy explode all trees 7. MeSH descriptor Psychotherapy explode all trees 8. MeSH descriptor Counseling explode all trees 9. MeSH descriptor Behavior Therapy explode all trees 10. (psycho$ intervention OR behaviour$ therapy OR cognitive therapy OR cognitive-behaviour$ therapy OR communication skills training OR counsel$ OR psycho-education OR family therapy OR guided imagery OR music therapy OR problem-solving therapy OR psychotherapy OR relax$ OR stress management OR support group OR supportive-expressive group therapy OR hypno$ OR meditat$ OR desensiti$ OR mindful$ OR acceptance commitment):kw 11. (neoplasms):kw 12. (#1 OR #2 OR #3 OR #4) 13. (#5 OR #6 OR #7 OR #8 OR #9 OR #10) 14. (#11 AND #12 AND #13) from 1999 to 2009 in Clinical Trials
CINAHL	S1 (MM "Behavior Therapy+") or (MM "Cognitive Therapy") Search modes - Boolean/Phrase

S2	(MM "Communication Skills Training") or (MM "Communication Skills")	Search modes - Boolean/Phrase
S3	(MH "Counseling+")	Search modes - Boolean/Phrase
S4	(MM "Psychoeducation")	Search modes - Boolean/Phrase
S5	(MM "Family Therapy")	Search modes - Boolean/Phrase
S6	(MM "Guided Imagery")	Search modes - Boolean/Phrase
S7	(MM "Music Therapy")	Search modes - Boolean/Phrase
S8	(MH "Psychotherapy+") or (MM "Psychotherapy, Brief") or (MM "Psychotherapy, Group")	Search modes - Boolean/Phrase
S9	(MH "Relaxation Techniques+")	Search modes - Boolean/Phrase
S10	(MM "Stress Management")	Search modes - Boolean/Phrase
S11	(MH "Support Groups+")	Search modes - Boolean/Phrase
S12	psycho* intervention OR behaviour* therapy OR cognitive therapy OR cognitive-behaviour* therapy OR communication skills training OR counsel* OR psycho-education OR family therapy OR guided imagery OR music therapy OR problem-solving therapy OR psychotherapy	Search modes - Boolean/Phrase
S13	relax* OR stress management OR support group OR supportive-expressive group therapy OR hypno* OR meditat* OR desensiti* OR mindful* OR acceptance commitment	Search modes - Boolean/Phrase
S14	(MH "Anxiety+")	Search modes - Boolean/Phrase
S15	(MH "Depression+")	Search modes - Boolean/Phrase
S16	(MH "Quality of Life+")	Search modes - Boolean/Phrase
S17	depression or anxiety or quality of life or distress	Search modes - Boolean/Phrase
S18	(MH "Neoplasms+")	Search modes - Boolean/Phrase
S19	neoplasm*	Search modes - Boolean/Phrase
S20	S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13	Search modes - Boolean/Phrase
S21	S14 or S15 or S16 or S17	Search modes - Boolean/Phrase
S22	S18 or S19	Search modes - Boolean/Phrase

S23	S20 and S21 and S22	Limiters - Published Date from: 199901-2009 English Language; Research Article; Exclude MEDLINE records; Publication Type: Clinica Trial

Appendix C: Checklist for evaluating evidence for PROM reliability, validity and interpretability and responsiveness in cancer [2]

RELIABILITY	Ratings
Internal consistency The extent to which items in a scale are inter-correlated in accordance with hypothesised internal structure; a measure of the homogeneity of items within a scale.	+ Cronbach's alpha of 0.70 to 0.90 for all scales in accordance with evidence-based dimensionality as currently or previously reported. ± Alpha of 0.70 to 0.90 reported for at least half scales in accordance with dimensionality or for all scales but no evidence for dimensionality currently or previously reported. - Alpha <0.70 or > 0.90 for more than half the scales. Blank = No information found on content validity
Test – retest The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change has occurred that would be expected to confer a change in scores.	+ Intraclass correlation coefficient (ICC) ≥ 0.70 and satisfactory time intervals reported. ± ICC ≥ 0.70 reported but method doubtful (e.g. time intervals unsuitable or not reported) - ICC <0.70 Blank = No information found on content validity
Inter – rater The extent to which ratings given by a proxy (e.g., family member, clinician) agree with those of the patient him/herself.	+ Adequate agreement reported (Kappa or ICC ≥ 0.60) based on sound methodology. ± Marginal agreement reported (Kappa or ICC 0.50 – 0.59) or ≥ 0.60 but methods doubtful. - Kappa or ICC <0.50. Blank = No information found on content validity

[2] adapted from checklists developed as part of the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) and Dementia Outcome Measurement Suite (DOMS) Project

VALIDITY	Ratings
Content The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+ Patients and experts were involved during item selection and/or item reduction or are asked to thoroughly appraise content validity ± Patients were consulted for acceptability, reading and/or comprehension only - No patient involvement Blank = No information found on content validity
Convergent/divergent The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.	+ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used ± Method doubtful or limited/inadequate evidence reported (e.g. inadequate choice of comparison measure; no hypotheses made and the scales compared were not developed to assess identical constructs) - Results not in agreement with hypotheses. Blank = No information provided
Internal Structure Assesses whether factor analysis, item response theory (IRT) or multi-trait scaling was applied in order to provide empirical support for the dimensionality of the questionnaire.	+ Substantial evidence provided to support internal structure via factor analysis or IRT ± Some evidence provided to support internal structure. Limitations may relate to design, sample size, analysis (e.g., only multi-trait scaling was used) or results - Failed a test of dimensionality Blank = No evidence provided

VALIDITY	Ratings
Discriminant validity The scale differentiates between relevant categories of respondent (e.g. clinically depressed versus not, on treatment versus off treatment)	+ Scale differentiates between relevant categories of respondents in accordance with hypotheses ± Method doubtful or scale differentiated between categories of respondent but no clear hypotheses made and expected difference was not obvious - Scale fails to differentiate between relevant categories of respondents Blank = No information provided
Criterion validity Information on the relationship of scores to long-form measures or clinical diagnosis is provided	+ Satisfactory comparison made to criterion measures (correlations, area under the curve, sensitivity/ specificity) ± Comparison made using doubtful methods or criterion measure unsatisfactory (e.g., another PROM was substituted for diagnostic interview) - Results unfavourable Blank = No evidence provided
Predictive validity The degree of agreement between scales and an independent event (e.g., number of hospital stays)	+ Results were acceptable in accordance with the hypotheses and an adequate event was chosen ± Limited/inadequate evidence reported (e.g., inadequate choice of event; no hypotheses made and the comparison measure would not obviously have been expected to correlate) - Results not supportive of hypotheses. Blank = No evidence provided

RESPONSIVENESS	Ratings
Floor and ceiling effects The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved Authors should provide descriptive statistics of the distribution of scores	+ Descriptive statistics of the distribution of scores were presented and less than 5% of respondents achieved the highest or lowest possible score ± Descriptive statistics of the distribution of scores were presented and between 6% and 15% of respondents achieved the highest or lowest possible score - Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score Blank = No evidence provided
Responsiveness to change The ability to detect important change over time in the concept being measured	+ Hypotheses were formulated and results were in agreement, using an adequate metric (effect size, standardised response mean, comparison with external standard) ± Method doubtful (e.g., no hypotheses made and expected direction of change not obvious) - Measure found inadequately responsive when an important change was expected Blank = No evidence provided

INTERPETABILITY	**Ratings**
Reports to assist with interpretability. The degree to which one can assign qualitative meaning to quantitative scores Do authors provide the following? 1. Presentation of means and SD of scores before and after clinically important event (e.g., change in treatment, disease or performance status) 2. Comparative data on the distribution of scores in relevant subgroups 3. Information on the relationship of scores to well-known measures (including widely used PROMs) or clinical diagnoses 4. Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced	+ Authors provide 2 or more types of information to enable interpretability ± Authors provide one type of information to assist with interpretability Blank = No information provided or sample sizes too small to be useful.

Appendix D: Characteristics of candidate PROMS that met all four criteria in Step 2

PROM	Author(s)	Domains	N of items	TTA	Response options	Recall period	Scoring	LOTE***	Licensing requirements	Costs
BSI-18	Derogatis LR	Anxiety, Depression, Somatisation	18	4	5-point scale	Past 7 days	Global severity index, scale	None	Registration by a medical practitioner required - https://www.pearsonpsychcorp.com.au/	Manual $132; Forms $3.96 pp (hand-scored forms also available at $1.98 pp); Q Local Software required for computer-based scoring - annual license fee for network version $250
BSI-53		Somatisation, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, Psychoticism	53	8-12			Global severity index, positive symptom total, positive symptom distress index, scale			
CES-D	Centre for Epidemiologic Studies from the National Institute of Mental Health	Depression	20	5	4-point scale	Past week	Total only	Arabic, Cantonese, Greek, Italian	Contact NIMHinfo@nih.gov	Free for non-commercial use
CES-D-15	Stommel M, Given BA, Given CW et al		15	4*				Need to adapt from CES-D	Contact manfred.stommel@hc.msu.edu	Free for non-commercial use
C-SOSI	Carlson LE & Thomas BC	Depression, Anger, Muscle Tension,	56	<1**	5-point scale	Past week	Total, scale	None	Contact tombejoy@cancerb	Free for non-commercial use

PROM	Author(s)	Domains	N of items	TTA	Response options	Recall period	Scoring	LOTE***	Licensing requirements	Costs
		Cardiopulmonary Arousal, Sympathetic Arousal, Neurological/GI, Cognitive Disorganization, Upper Respiratory Symptoms							oard.ab.ca	
DT	Roth AJ, Kornblith AB, Batel-Copel L, et al	Distress	1	< 1**	10-point scale	Past week	Total only	Unknown	None – available from http://www.nccn.org/professionals/physician_gls/PDF/distress.pdf	Free for non-commercial use
EORTC QLQ-C30	EORTC Quality of Life Group	Physical Functioning, Emotional Functioning, Social Functioning, Role Functioning, Cognitive Functioning, Financial Impact, Pain, Fatigue, Nausea and Vomiting, Dyspnea, Insomnia, Appetite loss, Constipation, Diarrhea, Global Health Status /	30	11	4-point and 7-point scales	Past week	Total (sum of 2 questions); scale	Arabic, Cantonese, Greek, Italian, Vietnamese	Complete a user agreement form at http://groups.eortc.be/qol/questionnaires_qlqc30.htm	Free for non-commercial use

PROM	Author(s)	Domains	N of items	TTA	Response options	Recall period	Scoring	LOTE***	Licensing requirements	Costs
		QOL								
FACT-G	Cella D	Physical Well-being, Emotional Wellbeing, Social/Family Wellbeing, Functional Well-being	27	5-10	5-point scale	Past 7 days	Total (sum of all scales); scale	Arabic, Cantonese, Greek, Italian, Vietnamese	Complete a user agreement form at http://www.facit.org/registration/registration_landing.aspx	Questionnaire is free but scoring manual and translations cost
GHQ-28	Goldberg D	Somatic Symptoms, Anxiety and Insomnia, Severe Depression, Social Dysfunction	28	3-4	4-point scale	Past few weeks	Total; domain	Arabic, Cantonese, Greek	Complete a user agreement form at www.gl-assessment.co.uk	Manual 75 GBP; Forms 75 pence pp
GHQ-12		Distress	12	<1*			Total only	Arabic, Cantonese, Greek, Italian		Forms 50 pence pp
HADS	Snaith RP & Zigmond AS	Anxiety, Depression	14	2-5	4-point scale	Past 7 days	Total (unofficial), scale	Arabic, Cantonese, Greek, Italian	Complete a user agreement form at www.gl-assessment.co.uk	Manual 25 GBP; Forms 55 pence per patient
IES-R	Weiss DS & Marmar CR	Intrusion, Avoidance, Hyperarousal	22	2**	4-point scale	Past 7 days	Total, scale	Cantonese, Italian	Contact daniel.weiss@ucsf.edu	Free for non-commercial use
MHI-38	Veit CT, Ware JE	Psychological Distress (Anxiety, Depression, Loss of Behavioural Emotional Control), Psychological Wellbeing (General Positive	38	8**	6-point scale	Past month	Total, scale	Arabic, Cantonese, Greek, Italian, Vietnamese (as Australian community languages)	None – available from http://www.rand.org/health/surveys_tools/mos/mos_mentalhealth.html	Free for non-commercial use

PROM	Author(s)	Domains	N of items	TTA	Response options	Recall period	Scoring	LOTE***	Licensing requirements	Costs
MHI-5	Ware, JE	Affect, Emotional Ties) Psychological Distress	5	< 1**			Total only	Need to adapt from MHI-38	None – need to adapt from MHI-38	Free for non-commercial use
PCL-C	Weathers FW, Litz BT, Herman DS et al	Re-experiencing, Avoidance or Numbing, Hyper-arousal	17	< 3**	5-point scale	Past month	Total, scale		Register at - http://www.ncptsd.va.gov/ncmain/assessment/assessmt_request_form.html	Free for non-commercial use
POMS-65	Lorr M, McNair DM & Heuchert JW	Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigour, Fatigue-Inertia, Confusion-Bewilderment.	65	10	5-point scale	Past week	Total Mood Disturbance Score (TMDS), scales	Cantonese, Greek, Italian, Vietnamese	Purchase required - see http://www.mhs.com	Manual US$27; Forms US$1.32 pp
POMS-37	Shacham S		37	5*				Need to adapt from POMS-65	None – available in Shacham *et al.* (1983) (69)	Free for non-commercial use
POMS-11	Cella DF, Jacobsen PB, Orav EJ et al	Total Mood Disturbance	11	<1*			TMDS only	Need to adapt from POMS-65	None – available in Cella *et al.* (1987) (58)	Free for non-commercial use
POMS-14	Guadagnoli E & Mor V	Negative and Positive Affect	14	<1*			Scale only	Need to adapt from POMS-65	None – available in Guadagnoli *et al.* (1989) (68)	Free for non-commercial use
QLI-CV	Ferrans CE	Health and Functioning, Socioeconomic, Psychological / Spiritual, Family	66	10	6-point scale	Now	Total, scale	Arabic, Italian	None – available from http://www.uic.edu/orgs/qli/questionaires/questionnairehome.htm	Free for non-commercial use
SCL-90-R	Derogatis, LR	Somatization, Obsessive-	90	?*	5-point scale	Past 7 days	Global severity	None	Registration by a medical practitioner	Manual US$132; Forms US$3.96 pp

PROM	Author(s)	Domains	N of items	TTA	Response options	Recall period	Scoring	LOTE***	Licensing requirements	Costs
		Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, Psychoticism					index, scale		required - https://www.pearsonpsychcorp.com.au/	(hand-scored forms also available at US$1.98 pp); Q Local Software required for computer-based scoring - annual license fee for network version US$250
SF-36v2 SF-12v2	Ware J	Physical Functioning, Role-physical, Bodily pain, General health, Vitality, Social Functioning, Role-emotional, Mental Health	36 12	10 2-5	3-point and 5-point scales	In general, on a typical day, during the past 4 weeks	Physical Component and Mental Component Summaries, scales	Arabic, Cantonese, Greek, Italian, Vietnamese Cantonese, Greek, Italian	Purchase required - see http://www.qualitymetric.com/DefaultPermissions/RequestInformation/tabid/233/Default.aspx/	Obtain a quote via website
SPHERE	Hickie IB, Davenport TA, Hadzi-Pavlovic D et al	Psychological distress, somatic distress	12	2*	3-point scale	Past few weeks	Total score only	None	None – available from ianh@med.usyd.edu.au	Free for non-commercial use

TTA = time to administer scales of interest; LOTE = languages other than English; pp = per participant. * Extrapolated from time to administer cited in manual; ** Estimated as the longer of two times taken to self-administer by independent reviewers; *** NSW top 5 languages other than English = Arabic, Cantonese, Italian, Greek, Vietnamese. None of the questionnaires are available in Australian Aboriginal or Torres Strait Islander languages.

Appendix E: Content comparison between DSM-IV-TR criteria and PROMs assessing anxiety, depression or distress

DSM-IV-TR Diagnostic Criteria		Anxiety							Depression									Distress					
Generalised anxiety disorder	**Major depressive disorder**	**BSI-18/53 (6)**	**GHQ-28 (7)**	**HADS (7)**	**MHI-38 (6)**	**POMS-65 (9)**	**POMS-37 (6)**	**SCL-90-R (10)**	**BSI-18/53 (6)**	**C-SOSI (8)**	**CES-D-20 (20)**	**CES-D-15 (15)**	**GHQ-28 (7)**	**HADS (7)**	**MHI-38 (13)**	**POMS-65 (15)**	**POMS-37 (8)**	**SCL-90-R (13)**	**DT**	**GHQ-12**	**MHI-5**	**POMS-11 (11)**	**POMS-14 (7)***	**SPHERE (6)$**
Anxiety, worry, tense, tension in muscles, wound up, struggle to relax		2	2	3	6	7	5	6		2			1		1			1		1	2	1		1
Distressed																			1					
Struggle to regain control																								
Restless	Restless	1		1	1	1	1	1																
Difficulty with sleep	Insomnia or hypersomnia		2								1	1								1				
Easily fatigued or worn out	Fatigue or loss of energy																	1				1		
Irritability	Irritability (in children)		1																					
Concentration problems	Diminished ability to concentrate										1	1								1				
	Diminished ability to think																					2		
	Depressed, sad								1	2	5	5		2	6	8	6	2		2	3	6	1	1
	Loss of interest	1												1				2						
	Loss of pleasure										1	1		3						1				
	Feels empty																	1						
	Tearful									1	1				1			1						
	Weight gain or loss																							
	Appetite increase or decrease										1	1												
	Feeling slowed down													1										
	Feeling worthless								1		1	1	1			2	1	1		1				1

DSM-IV-TR Diagnostic Criteria		Anxiety							Depression										Distress					
Generalised anxiety disorder	**Major depressive disorder**	**BSI-18/53 (6)**	**GHQ-28 (7)**	**HADS (7)**	**MHI-38 (6)**	**POMS-65 (9)**	**POMS-37 (6)**	**SCL-90-R (10)**	**BSI-18/53 (6)**	**C-SOSI (8)**	**CES-D-20 (20)**	**CES-D-15 (15)**	**GHQ-28 (7)**	**HADS (7)**	**MHI-38 (13)**	**POMS-65 (15)**	**POMS-37 (8)**	**SCL-90-R (13)**	**DT**	**GHQ-12**	**MHI-5**	**POMS-11 (11)**	**POMS-14 (7)***	**SPHERE (6)$**
	Feeling guilty															2								
	Indecisiveness																			1				
	Thoughts of death									1			3		1									
	Suicidal ideation								1				2		1			1						
Number of items assessing DSM-IV-TR criteria for anxiety		**3**	**5**	**4**	**7**	**8**	**6**	**7**	**0**	**2**	**2**	**2**	**1**	**0**	**1**	**0**	**0**	**2**	**1**	**3**	**2**	**2**	**0**	**1**
Number of items assessing DSM-IV-TR criteria for depression		**2**	**3**	**1**	**1**	**1**	**1**	**1**	**3**	**4**	**11**	**10**	**6**	**7**	**9**	**12**	**7**	**10**	**0**	**7**	**3**	**9**	**1**	**2**

Note – Numbers in brackets indicate number of items in each scale. Numbers in cells indicate the number of items in each scale containing content similar to each DSM-IV-TR criterion. Some items are positively worded so that a higher rating indicates lower anxiety, depression or distress. * Only the negative affect scale of the POMS-14 was considered. $ Only the 6-item psychological distress scale was considered.

Appendix F: Evidence for reliability, validity, responsiveness and interpretability of each candidate PROM in English-speaking cancer populations

Symptom Checklist-90-Revised (SCL-90-R) and Brief Symptom Index (BSI-53/18)

Version	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
SCL-90-R	US Cancer centre new admissions with mixed diagnoses (N=215)	(111)							+					Some
SCL-90-R	Canadian patients with mixed cancer diagnoses	(114)	±	±										
BSI-53	US patients with mixed cancer diagnoses (N=30)	(113)					±				±			
BSI-53	US women with breast cancer attending ambulatory cancer clinics (N=275)	(112)					+		+					
BSI-18	US patients with mixed cancer diagnoses (N=1,543)	(115)						+		±				Some

Calvary Symptoms of Stress Inventory (C-SOSI)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
Canadian patients with mostly breast cancer registered for a stress-management program (N=344)	(60)	+				+	+						

Center for Epidemiologic Studies Depression Scale (CES-D)

CES-D-20/15	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
CES-D-20	US breast cancer patients undergoing radiotherapy, chemotherapy or bone marrow transplantation (n=117) and healthy controls (n=62)	(286)	+	-			+		+				+	Substantial
CES-D-20	Canadian head and neck cancer patients (N=60)	(93)								+				Some
CES-D-20	Depressed and not depressed US patients with mixed cancer diagnoses (N=33)	(92)								+				Some
CES-D-20 CES-D-15	US patients with mixed cancer diagnoses (n=708) and caregivers of chronically ill patients (n=504)	(96)	+ +					± +						Some

Distress Thermometer (DT)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
Analysis of pooled data on criterion validity drawn from 19 studies	(119)								±				Some
British patients with mixed cancer diagnoses (N=171)	(118)					±			±			-	Some
British women with breast cancer (N=321)								+	±				Some

EORTC QLQ-C30

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
European patients with lung cancer (N=305)	(73)	±			±		±	+				+	Substantial
Canadian men with advanced hormone resistant prostate cancer receiving chemotherapy (N=161)	(287)											+	
	(288)					+		+		±			
US mostly advanced head and neck cancer patients (N=120)	(213)	±				+	±	+				±	
British oral cancer patients undergoing surgery (N=29)	(170)					±					-	±	
Canadian and Dutch patients with mixed cancer diagnoses (N=1,310)	(289)	+											

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
Canadian and European patients with mostly breast, ovarian and lung cancer on chemotherapy enrolled in an anti-emetic trial	(138)	±					±	+				±	Some
Canadian patients with mixed cancer diagnoses (N=96)	(142)					+							
Canadian metastatic breast cancer patients (N=150)	(139)					±	±	+					Some
Review of 5 studies comparing self- and proxy-report in various cancer groups	(144)			±									
Australian metastatic breast, ovarian and colon cancer patients (N=98)	(143)					±	±	±				±	Some
US patients with cancer (88%), HIV, (8%) or cancer and HIV (4%) (N=1,163)	(171)	±				±							Some
US low income, mostly Afro-Americans, with metastatic prostate cancer (N=110)	(140)	±				+		+					Some
Australian patients with stages 1, 2 or 4 pleural mesothelioma on chemotherapy (N=53)	(290)	ID						±		±		±	

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
European patients with mostly early stage oesophagus, oesophago–gastric junction or stomach cancer mostly undergoing treatment (N=300)	(291)							+					Some
British, French and German patients being treated for liver metastases from colorectal cancer (N=356)	(292)											±	Some
European and Australian gastric cancer patients on treatment with curative or palliative intent (N=219)	(293)											±	Some
European and Australian oesophageal cancer patients on treatment with curative or palliative intent (N=419)	(141)						±					±	Substantial
Canadian patients with oesophageal cancer undergoing surgery or scheduled for neoadjuvant chemoradiation before surgery (N=83)	(159)	-				+							Some

FACT-G

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
US patients with mixed cancer diagnoses (N=545)	(72)	±	+		+	+	±	+					Substantial
US patients with mixed cancer diagnoses (N=99)	(294)	ID			ID	ID						ID	
US rural patients with mixed cancer diagnoses (N=344)	(154)	ID				+	±						
British patients with mixed cancer diagnoses (N=436)	(153)	+					±						
US patients aged 65 years or older with mixed cancer diagnoses (n=85) and community sample (n=27)	(192)	-				±		+					Substantial
US patients with cancer (88%), HIV, (8%) or cancer and HIV (4%) (N=1,163)	(171)	+				±							Some
US women with breast cancer (N=342)	(216)	±				+	±	+					Substantial
US patients with brain tumours (N=101)	(295)	±	±			±		±					Substantial
Us patients with advanced colorectal cancer (N=60) Us patients with colorectal cancer at various stages (N=63)	(156)	+ ±				+ +		+ +					Substantial

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
Canadian patients with oesophageal cancer undergoing surgery or scheduled for neoadjuvant chemoradiation before surgery (N=83)	(159)					+		+				+	Substantial
US patients with head and neck cancer (N=151)	(155)	±						+					Some
US patients with hepatobiliary cancers, 57% of whom had no evidence of disease (N=51)	(158)	±	+			±		+					Some
US patients with lung cancer (N=116)	(157)	±				+		±					Some
US patients with melanoma (N=273)	(296)	+	+			+		+					Some
US women with ovarian cancer (N=232)	(297)	+	+			+		+					Substantial
US men with prostate cancer, mostly early stage (N=130)	(298)	-						+					Some
Australian women with vulvar cancer (N=97)	(214)	+				+		+				±	Some
US patients receiving bone marrow transplantation for various cancers (N=187)	(211)	-				+		+				±	Substantial
US patients with advanced renal cell carcinoma (N=191)	(299)	±						±				+	Some
US melanoma patients treated with interferon-*a* (N=21)	(212)	±	±					±				±	

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
US women with ovarian cancer undergoing neurotoxic chemotherapy (N=99)	(300)	-						±				±	
British women with breast cancer participating in a trial of sentinel node guided axillary therapy (n=279) and attending a lymphoedema clinic (n=29).	(301)	+	+					±				±	Some
US cancer patients with anaemia (n=375) and US general population (n=1078)	(161)	+					±				±		Substantial
US patients with advanced cancer (N=275)	(302)	+				+	±						Some
US patients aged 65 years or older undergoing chemotherapy for lung, breast or ovarian cancer or non-Hodgkin's lymphoma (N=852)	(303)	±				+		±					Some
US patients with mixed cancer diagnoses (N=49)	(304)	±	+			+		+					
US patients with mixed cancer diagnoses undergoing taxane therapy (N=230)	(305)	+						+				+	Substantial

General Health Questionnaire (GHQ)

GHQ-12/28	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
GHQ-12	British palliative care patients (N=79)	(98)								±				
GHQ-28	British patients who were disease-free at assessment (N=95) British patients with mixed cancer diagnoses who were off treatment at assessment (N=133)	(99)								± +				Some

Hospital Anxiety and Depression Scale (HADS)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
British patients with mixed cancer diagnoses (N=568)	(120)	±					+						
British breast cancer patients on chemotherapy and/or hormone therapy (N=110)	(121)	+					+						Some
US in- and out-patients with mixed cancer diagnoses (N=809)	(306)							±					
British patients with advanced cancer of various types (N=100)	(100)	+					-		-				Some
Canadian patients with mixed cancer diagnoses (N=3,035)	(307)							±					
British patients with mixed cancer diagnoses (N=1,474)	(122)	+						±	-				Some
British out-patients with mixed cancer diagnoses (N=635)	(308)								+				Some
British palliative care patients with mixed cancer diagnoses (N=79)	(98)								±				Some

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
British chemotherapy patients with mixed cancer diagnoses (N=172)	(175)					+			-				Substantial
British early breast cancer patients (N=266)	(101)								±				Some
British women with breast cancer (n=258), myocardial infarction (N=108) or following stroke (n=68)	(97)	±					±						Some
British patients with mixed cancer diagnoses (N=284) British patients who were disease-free at time of assessment (N=88) British patients with mixed cancer diagnoses who were stable at time of assessment (N=113) British patients with mixed cancer diagnoses who were on treatment at time of assessment (N=165)	(99)								± + ± ±				Some

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
British patients with mixed cancer diagnoses (N=1,855) of whom a subset completed a diagnostic interview (n=381)	(123)						±		±				
Australian early stage breast cancer patients enrolled in an RCT of cognitive-existential group therapy (N=303)	(309)						±						Some
Australian women with stage 4 breast cancer being screened for depression only (N=227)	(310)						±						
Canadian head and neck cancer patients being screened for depression only (N=60)	(93)						+						Some
Australian women with vulvar cancer (N=97)	(214)	+										+	
US patients aged 65 years or older undergoing chemotherapy for lung, breast or ovarian cancer or non- Hodgkin's lymphoma (N=852)	(303)	+											Some

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
US women with breast cancer attending ambulatory cancer clinics (N=275)	(112)					+		+					

Impact of Event Scale (IES)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
US women 6 to 60 months post-completion of primary breast cancer therapy (N=55)	(132)	-				+	±						Some

Mental health Inventory (MHI)

Version	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
MHI-5	British patients with mixed cancer diagnoses undergoing chemotherapy (N=172)	(175)					+			±				Some
MHI-38	US patients with mixed cancer diagnoses (N=433)	(311)					+	±						Some

Post Traumatic Stress Disorder Checklist – Civilian Version (PCL-C)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
US women 6 to 60 months post-completion of breast cancer therapy (N=55)	(132)	±				+	±						Some
US women 2 to 72 months post-completion of breast cancer therapy (N=142)	(312)						±						
US women with breast cancer (N=82)	(137)								+				Some

Profile of Mood States (POMS)

POMS version	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
POMS-37* POMS-65*	US cancer patients (unspecified) enrolled in a pain treatment study (N=83)	(69)	± ±				+							
POMS-14	US newly diagnosed lung, breast, colorectal cancer patients (N=428) US mixed cancer diagnoses on chemotherapy (N=225)	(68)					+ +	±	±					

POMS version	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
	Both above samples combined								+					
POMS-37*	US women with early breast cancer scheduled to receive intravenous adjuvant chemotherapy post mastectomy (N=162) plus healthy controls (N=55)	(107)	+						+				±	Substantial
POMS-65*			+						+				±	
POMS-11	US mixed cancer (other than lung) patients post diagnosis but pre-treatment (N=619)	(58)	-					+						
	US lung cancer patients post diagnosis but pre-treatment (N=296)		-											
	US pancreatic (N=119) versus gastric cancer (N=128) patients								+					Some

POMS version	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
POMS-37 POMS-65	US patients with nonmetastatic breast cancer (n=47) or undergoing bone marrow transplantation for heterogeneous cancers (n=85) and healthy controls (n=76)	(108)	± ±				+							
POMS-37	US patients with mixed cancer diagnoses awaiting bone marrow Transplantation (N=428)	(109)	+				+	+	+					Substantial

* Ratings relate only to the depression-dejection and tension-anxiety scales

Quality of Life Index – Cancer Version (QLI-CV)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	Conv V	IS	DV	CriV	PV	FCE	Resp	
US women with breast cancer at a mean of 7.93 years since diagnosis, 27% of whom were undergoing active treatment (N=371)	(166)	-				+		+					

SF-36/12

SF-36/12	Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
			IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
SF-36	British oral cancer patients undergoing surgery (N=29)	(170)					±					±	±	
SF-36	US patients with cancer (88%), HIV, (8%) or cancer and HIV (4%) (N=1,163)	(171)	±				±							Some
SF-36	Depressed and not depressed US patients with mixed cancer diagnoses (N=33)	(92)							±	±				
SF-36	Ethnic minority US patients with haematological cancers (N=45)	(172)	±		±									
SF-12	Canadian men with prostate cancer (N=2,415)	(173)	+	+										

Somatic and Psychological Health Report (SPHERE)

Sample	Ref.	Reliability			Validity						Responsiveness		Information to assist interpretation
		IC	TR	IR	CV	ConvV	IS	DV	CriV	PV	FCE	Resp	
Australian cancer centre outpatients with mixed diagnoses (N=340)	(130)					+							Some

ID = insufficient detail available to enable rating; IC = internal consistency; TR = test-retest reliability; IR = inter-rater reliability (self and proxy); CV = content validity; Conv V = convergent validity; IS = internal structure; DV = discriminant validity; Cri V = criterion validity; PV = predictive validity; FCE = floor and ceiling effects; Resp = responsiveness

Appendix G: Total 57 studies included for effect size (ES) calculation in Step 4

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(313)	Women with breast cancer (Stage I-III) receiving chemo-, radio- or hormone-therapy.	48	100	53.9	US	C	CES-D†	Dep = **-0.37**
(251)	Stage I-III breast cancer patients receiving adjuvant treatment, and their partners	75	100	54.1	US	C	CES-D†	Dep = **-1.04**
(249)	Gynaecologic cancer patients first time at the clinic, 16% of whom were cancer free.	64	100	49.6	US	C	FACT-G POMS-65†	G = 0.06 EWB= **0.55** SWB= -0.12 PWB= **0.48** FWB = **0.71** Anx = **-0.22** Dep = -0.04
(197, 198)	Breast cancer patients (stages I-III), recruited 1-3 months after diagnosis and receiving adjuvant chemo-, radio-, hormone- or combined therapy.	222	100	54.5	US	C	POMS-65† FACT-B	Anx = -0.17 Dep = **0.20** G = N/A EWB= 0.03 SWB= **-0.28** PWB= -0.06 FWB = **0.20**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(177)	Patients with lymphoma who were undergoing treatment or who had concluded treatment within the past 12 months.	39	63	51	US	C-p	IES-15† CES-D†	Dis= **-0.24** Dep = -0.13
(271)	Gynaecological and breast cancer patients undergoing radiation therapy	62	100	50.7	US	C-p	SF-36	VT= **0.22** MH= 0.18 SF= **0.26** RE= **0.21** PF= 0.03 RP = **0.20** BP= 0.15 GH = **0.34**
(248)	Cancer survivors (average 56 months after diagnosis)	38	95	50	Canada	C-p	POMS-65† QLQ-C30	Anx = -0.06 Dep = **-0.38** Global QOL= **0.64** EF= **0.47**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(183)	Breast cancer patients, within 18 months of initial diagnosis.	26	100	48	US	C-p	POMS-65† PCL-C† FACT-G	Anx = **-0.41** Dep = **-0.24** Dis = **1.47** G = **-0.48** EWB = **-0.49** SWB = -0.07 PWB = **-0.24** FWB = -0.02
(203)	Multiethnic (42% African American, 31% Hispanic) oncology outpatients with breast cancer (stages I to III) diagnosed within previous 5 years and able to speak/read English or Spanish.	71	100	54.81	US	C-p	FACT-G	G= **0.42** EWB = **0.43** SWB = **0.59** PWB = 0.16 FWB = **0.26**
(257)	Palliative day care patients with heterogeneous cancers.	26	33.3	68.2	US	C-np	HADS†	Anx= **-2.12** Dep = N/A

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(221)	Heterogeneous cancer patients with an estimated life expectancy of more than 3 months.	288	87	52.1	UK	C-np	CES-D† QLQ-C30	Dep = **-0.24** Global QOL = 0.06
(200)	Breast cancer at least 3 months post completion of primary treatment and without current evidence of disease.	72	100	58	US	CBT	CES-D† FACT-B	Dep = **- 0.24** G= **0.51** EWB= 0.17 SWB= **0.36** PWB= **0.33** FWB= **0.27**
(202)	Breast, prostate, bowel, or gynaecological cancer patients, who satisfied diagnostic criteria for chronic insomnia.	150	69	59.3	UK	CBT	HADS† FACT-G	Anx = **-0.57*** Dep = **-0.67*** EWB = **0.39*** SWB = **0.42*** PWB = **0.74*** FWB = **1.17***
(314)	Men who had undergone radical prostatectomy or radiotherapy for stage I or II prostate carcinoma in the past 18 months.	92	0	63.1	US	CBT	FACT-G	G = **0.29**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(315)	Men who had undergone radical prostatectomy or radiation therapy for stage I or II prostate cancer.	191	0	65.1	US	CBT	FACT-G	G = **0.20**
(260)	Heterogeneous cancer patients without depression or anxiety who were commencing chemotherapy or radiotherapy for the first time.	465	69	51.4	UK	CBT	HADS†	Dis = -0.18
(201)	Women diagnosed with breast cancer (any stage) within the past 5 years.	32	100	55.7	US	CT	FACT-B	EWB = -0.08 SWB = **0.24** PWB = **-0.35** FWB = 0.03
(316)	Women with metastatic breast cancer.	37	100	51.6	US	CT	HADS†	Anx = **-0.51** Dep = **-0.50**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(244)	Men with prostate cancer receiving continuous androgen-ablation therapy with expectation to continue therapy for 1 year or more.	134	0	69.2	US	E	CES-D† SF-36	Dep = -0.17 PCS = **-0.28** MCS = -0.13 VT = **-0.36** SF = **-0.35** RE = -0.15 PF = **-0.43** RP = **-0.30** BP = -0.14 GH = -0.13
(317)	Men with prostate cancer undergoing active surveillance.	82	0	65.6	US	E	SF-36	MCS= -0.12 PCS = **0.23**
(204)	Women with early stage breast cancer receiving chemotherapy, radiotherapy or both.	201	100	51.6	UK	E	FACT-G	G = 0.00 EWB = 0.12 SWB = **0.22** PWB = -0.04 FWB = -0.02
(266)	Women with stage 0–II breast cancer receiving post-operative chemotherapy and/or radiation.	24	100	52.5	US	E	POMS-65	Anx = **-0.40** Dep = **-0.60**
(318)	African-American and Hispanic outpatients with cancer-related pain with a socioeconomic disadvantage.	97	63	56.5	US	Ed	SF-12	MCS = **-0.73** PCS = **0.39**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(199)	Low income Latina women at least 3 months past initial diagnosis of cervical or breast cancer.	55	100	N/A	US	Ed	FACT-G	SWB= **0.30** PWB = 0.00
(243)	Stage T1a–T2c prostate carcinoma patients scheduled to undergo or having undergone surgery, external beam radiation, or brachytherapy.	99	0	63.8	US	Ed	CES-D† SF-36	Dep = **0.36*** VT= 0.18* MH = 0.17* SF = **0.21*** RE = **0.29*** PF = 0.14* RP = **0.32*** BP = **0.25*** GH = 0.09*
(220)	Newly diagnosed oral cancer patients (mixed stages) undergoing curative surgical treatment.	19	37	56.7	Canada	Ed	CES-D† QLQ-C30	Dep = **-0.57** PF = **0.80** RF = **0.80** CF = **-0.97** EF = **-0.61** SF = **-0.80** Global QOL = **-0.90**
(246)	Men who were recently treated for prostate cancer (stages I-III).	250	0	65.1	US	Ed	CES-D (15) † SF-36	Dep = -0.17 MCS= 0.09 PCS = 0.15

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(242)	Women with ovarian cancer due to start post-operative chemotherapy.	123	100	60.3	US	Ed	CES-D† SF-12	Dep = -0.09 MCS = 0.14 PCS = 0.19
(252)	Newly diagnosed melanoma patients (stages I-III).	217	N/A	N/A	US	Ed	DT†	Dis = **0.32**
(256)	Outpatients with major depressive disorder attending breast, gynaecological, bladder, prostate, testicular and colorectal cancer clinics.	60	57	28	UK	Ed	HADS	Anx = **-1.49** Dep = **-0.97**
(178)	Women with breast cancer within 6 weeks after surgery.	418	100	87.2	US	Ed	CES-D† IES-R† SF-36	Dep = 0.11 Dis = 0.11 VT = **0.25**
(254)	Patients diagnosed with hepatobiliary carcinoma or chronic liver disease (80% of hepatobiliary patients were in the advanced stage).	28	36	67	US	Ed	CES-D† FACT-Hep	Dep = 0.19* G = 0.04* EWB = 0.08* SWB = 0.05* PWB = 0.00* FWB = 0.002*

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(268)	Outpatients with major depressive disorder and a life expectancy of greater than 6 months.	200	70.5	56.6	UK	Ed	SCL-90† QLQ C-30	Anx = -0.10* Dep = -0.17* PF= 0.06*
(267)	Metastatic cancer patients who had experienced cancer-related pain during the past two weeks.	176	57.4	N/A	US	Ed	QLI-CV	QOL = -0.16
(174)	Women with stage I-III breast cancer.	312	100	48.25	US	Ed	SF-36	PCS= 0.18! MCS= 0.11!
(259)	Patients undergoing curative radiotherapy for heterogeneous cancers.	63	38	55.79	US	Ex	HADS†	Anx = **0.38** Dep = **-0.31**
(253)	Women with metastatic breast cancer with a life expectancy greater than 6 months.	42	100	52	US	Ex	HADS† FACT-G	Anx = 0.15 Dep = -0.1 Dis = 0.16 G = **0.39** EWB = -0.17 SWB = **0.56** PWB = **0.6** FWB = **0.32**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(250)	Women with any cancer, 67% undergoing active treatment during the study period.	111	100	53.6	US	Ex	SF-36 SCL-90-R	MCS = **0.37** PCS = **0.21** Anx = **-0.21**† Dep = **-0.24**†
(272)	Women who underwent a lumpectomy or more extensive surgery for breast cancer at least 1 month earlier.	38	100	59.6	US	Ex	SF-36	MCS = **0.48** PCS = -0.17
(261)	Women with early stage breast cancer who had undergone surgery within the last 6 months and were married or cohabiting, and their significant others.	238	100	49.5	US	F	IES-15† (MHI-18†)	Dis = -0.17 Anx = **-0.38** Dep = **-0.40**
(233)	Individuals diagnosed or re-diagnosed with heterogeneous cancers 2 months – 2yrs previously.	16	81	52.9	US	P	BSI-53† FACT-G (PWB)	Anx = **-0.67** Dep = **-1.85** PWB = **1.15**
(241)	Heterogeneous cancers (41% breast), spread evenly across early and advanced stage disease.	108	84.4	51.55	Canada	P	CES-D† FACT-G	Dep = -0.14 G = N/A EWB = -0.05 SWB = 0.10 PWB = **0.29** FWB = **0.38**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(262)	Younger women with stage I-III breast cancer.	164	100	N/A	US	PS	MHI-5†	Dis = 0.23
(319)	Newly diagnosed solid tumour cancer patients who were within the first two cycles of chemotherapy.	237	100	59.55	US	PS	CES-D†	Dep = -0.15
(229)	Breast, lung, and prostate cancer patients who were receiving treatments other than chemotherapy.	175	60.5	62.3	Canada	PS	CES-D†	Dep = **-0.31**
(270)	Heterogeneous cancer patients who reported pain and fatigue at baseline.	113	72	58	US	PS	SF-36	SF = **0.52** RP = **0.86**
(247)	Early stage breast cancer patients receiving chemotherapy or radiation therapy.	62	100	51.9	US	PS	IES-22† FACT-G	Dis = **-0.23** G = **0.41** EWB = **0.46**
(258)	Patients with heterogeneous cancers receiving various treatments.	80	59.8	N/A	Australia	SC	HADS†	Anx = **-0.30** Dep = **-0.46**

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(222)	Older patients who had advanced stage breast, colon or prostate cancer.	129	47.5	73.5	US	SC	HADS† QLQ-C30	Anx = **-0.23** Dep = **-0.51** Dis = **-0.54** Global QOL = **0.34** EF = **0.27** PF = -0.13
(264)	Patients with metastatic breast, lung or colorectal cancer.	213	67	59	US	SC	POMS-11†	Dis = **0.36**
(180, 181)	Women with metastatic breast cancer receiving chemo-, radiation or hormonal therapy.	218	100	50	Canada	SE	POMS-65† IES-15† QLQ-C30	Anx = **0.49*** Dep = **0.59*** Dis = 0.04* Global QOL = 0.13* EF = -0.1* SF = 0.01* PF = 0.06* RF = 0.00*
(263)	Metastatic breast cancer patients.	66	100	50.7	Canada	SG	POMS-11†	Dis = **-0.23**
(255)	Women with primary breast cancer.	303	100	46.4	Australia	SG	HADS†	Anx = **0.23** Dep = -0.10

Ref.	Samples	N	% Female	Mean age	Country	Int.	PROMS	Effect size^
(182)	Women with early stage breast cancer.	72	100	49.5	US	SG	CES-D† PCL-C54†	Dep = **-0.51** Dis = **-0.44**
(265)	Older men with localised prostate cancer.	39	0	75.4	US	SM	POMS-37†	Anx = **0.25** Dep = **-0.47**
(245)	Patients undergoing radiotherapy for heterogeneous cancers.	101	71	61	US	SM	CES-D SF-36	Dep = **-0.53** MH = **0.38** MCS = **0.34**
(91)	Heterogeneous cancer patients.	90	36.5	51.9	Canada	SM	POMS-65† SOSI†	Anx = **-0.63** Dep = -0.08 Anx = -0.18 Dep = **-0.28**

Results are given for all scales reported in each article. N= total number of patients at baseline included in the analysis by the authors; * = effect size (ES) taken directly from the article; N/A = information not available (or scales not used); † = negative values indicates improvement in the construct being measured; ^Effect size – ESs for all reported scales are given except POMS-65 total mood disturbance score and EORTC QLQ-C30 cognitive functioning, financial impact and symptoms scales. Figures in bold indicate an ES 0.2 or above; ! = ES for between group difference at follow-up; Int. = Intervention category as follows: support group (SG), counselling (C), psychotherapy (P), family therapy (F), education/psycho-education (E), cognitive-behavioural therapy (CBT), cognitive therapy (CT), problem-solving/coping therapy (PS), stress management training (SM), exercise (E), CAM-physical (C-p), screening (SC), CAM-non physical (C-np), expressive therapy (Ex) and supportive-expressive group therapy (SE); FACIT measures: G= total score, EWB = emotional wellbeing; SWB = social/family wellbeing; PWB = physical wellbeing; FWB = functional wellbeing; EORTC-QLQ-C30: PF = physical functioning; RF = role functioning; EF = emotional functioning; SF = social functioning; Global QOL = global quality of life score; SF-36/12: PF = physical functioning; RP = role functioning; VT= vitality; RE = role emotional; MH = mental health; GH = general health; BP = bodily pain. PCS = physical component summary; MCS = mental component summary; Anx = anxiety scale; Dep = depression scale; Dis = distress score. Note: effect sizes reported are the largest in each trial

regardless of time-points or groups being compared. (MHI-18) indicates that while the MHI-18 was excluded at Step 2, outcome data from this PROM were included because they enable direct comparison with the IES-15.

Appendix H: Largest effect sizes identified in RCTs of each intervention type

Intervention cluster: Support group (SG), counselling (C), psychotherapy (P) and family therapy (F) (Number of RCTs: 11)

	Study detail			Emotional			HRQoL^			
PROM	**Ref.**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
BSI-53	(233)	P	16	-0.67	-1.85	-	-	-	-	-
CES-D	(241)	P	108	-	Ø	-	-	-	-	-
CES-D	(313)	C	48	-	-0.37	-	-	-	-	-
CES-D	(182)	SG	72	-	-0.51	-	-	-	-	-
CES-D	(251)	C	75	-	1.04	-	-	-	-	-
FACT-G	(233)	P	16	-	-	-	ID	ID	ID	1.15
FACT-G	(241)	P	108	-	-	-	Ø	Ø	Ø	0.29
FACT-G	(249)	C	64	-	-	-	Ø	0.55	Ø	0.48
FACT-G	(197, 198)	C	222	-	-	-	ID	Ø	-0.28	Ø
HADS	(255)	SG	303	0.23	Ø	ID	-	-	-	-
IES-R	(261)	F	238	-	-	Ø	-	-	-	-
PCL-C	(182)	SG	72	-	-	-0.44	-	-	-	-
POMS-11	(263)	SG	66	-	-	0.23	-	-	-	-
POMS-65	(249)	C	72	-0.22	Ø	-	-	-	-	-
POMS-65	(197, 198)	C	222	Ø	0.20	-	-	-	-	-

Intervention cluster: Education/psycho-education (E) (Number of RCTs: 12)

	Study detail			Emotional			HRQoL^			
PROM	**Ref.**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
CES-D	(220)	E	19	-	-0.57	-	-	-	-	-
CES-D	(178)	E	418	-	Ø	-	-	-	-	-
CES-D	(243)	E	99	-	0.36	-	-	-	-	-
CES-D	(242)	E	123	-	Ø	-	-	-	-	-
CES-D	(254)	E	28	-	Ø	-	-	-	-	-
CES-D-15	(246)	E	250	-	Ø	-	-	-	-	-
DT	(252)	E	217	-	-	0.32	-	-	-	-
EQLQ-C-30	(220)	E	19	-	-	-	-0.90	-0.61	-0.80	0.80
EQLQ-C-30	(268)	E	200	-	-	-	ID	ID	ID	Ø
FACT-G	(199)	E	55	-	-	-	ID	ID	0.30	Ø
FACT-G	(254)	E	28	-	-	-	ID	Ø	Ø	Ø
HADS	(256)	E	60	-1.49	-0.97	ID	-	-	-	-
IES-R	(178)	E	418	-	-	Ø	-	-	-	-
QLI-CV	(267)	E	176	-	-	-	Ø	-	-	-
SCL-20**	(268)	E	200	Ø	Ø	-	-	-	-	-
SF-12	(269)	E	97	-	-	-	-	0.39	ID	-0.73
SF-12	(242)	E	123	-	-	-	-	Ø	ID	Ø
SF-36	(243)	E	99	-	-	-	-	Ø	0.21	0.32
SF-36	(246)	E	250	-	-	-	-	Ø	ID	Ø
SF-36	(174)	E	312	-	-	-	-	Ø*	ID	Ø*

Intervention cluster: Cognitive-behavioural therapy (CBT), Cognitive therapy (CT), Problem-solving/coping therapy (PS) and Stress management training (SM) (Number of RCTs: 15)

	Study detail			Emotional			HRQoL^			
PROM	**Ref.**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
CES-D	(319)	PS	237	-	Ø	-	-	-	-	-
CES-D	(229)	CT	175	-	-0.31	-	-	-	-	-
CES-D	(200)	CBT	72	-	-0.24	-	-	-	-	-
CES-D	(245)	SM	101	-	-0.53	-	-	-	-	-
FACT-G	(314)	CBT	92	-	-	-	0.29	ID	ID	ID
FACT-G	(247)	PS	62	-	-	-	0.41	0.46	ID	ID
FACT-G	(315)	CBT	191	-	-	-	0.20	ID	ID	ID
FACT-G	(200)	CBT	72	-	-	-	0.51	Ø	0.36	0.33
FACT-G	(201)	CT	32	-	-	-	ID	Ø	0.24	-0.35
FACT-G	(202)	CBT	150	-	-	-	ID	0.39	0.42	0.74
HADS	(202)	CBT	150	-0.57	-0.67	ID	-	-	-	-
HADS	(316)	SM	37	-0.51	-0.50	ID	-	-	-	-
HADS	(260)	CBT	465	ID	ID	Ø	-	-	-	-
IES-R	(247)	PS	62	-	-	-0.23	-	-	-	-
MHI-5	(262)	PS	164	-	-	0.23	-	-	-	-
POMS-37	(265)	SM	39	0.25	-0.47	-	-	-	-	-
POMS-65	(91)	SM	90	-0.63	Ø	-	-	-	-	-
SF-12	(270)	PS	113	-	-	-	-	ID	0.52	0.86
SF-36	(245)	SM	101	-	-	-	-	0.38	ID	ID
SOSI	(91)	SM	90	Ø	-0.28	-	-	-	-	-

Intervention cluster: Exercise (E) and CAM-physical (C-p) (Number of RCTs: 9)

	Study detail			Emotional			HRQoL^			
PROM	**Ref**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
CES-D	(244)	E	134	-	Ø	-	-	-	-	-
CES-D	(177)	C-p	39	-	Ø	-	-	-	-	-
EQLQ-C30	(248)	C-p	38	-	-	-	0.64	0.47	ID	ID
FACT-G	(203)	C-p	71	-	-	-	0.42	0.43	0.59	Ø
FACT-G	(204)	E	201	-	-	-	Ø	Ø	0.22	Ø
FACT-G	(183)	C-p	26	-	-	-	-0.48	-0.49	Ø	-0.24
IES-R	(177)	C-p	39	-	-	-0.24	-	-	-	-
PCL-C	(183)	C-p	26	-	-	1.47	-	-	-	-
POMS-65	(183)	C-p	26	-0.41	-0.24	-	-	-	-	-
POMS-65	(248)	C-p	38	Ø	-0.38	-	-	-	-	-
POMS-65	(266)	E	24	-0.40	-0.60	-	-	-	-	-
SF-36	(244)	E	134	-	-	-	-	Ø	-0.35	-0.43
SF-36	(271)	C-p	62	-	-	-	-	0.21	0.26	0.20
SF-36	(317)	E	82	-	-	-	-	Ø	ID	0.23

Intervention cluster: CAM-non physical-NP (C-np), Expressive therapy (Ex) and Supportive-expressive group therapy (SE) (Number of RCTs: 9)

	Study detail			Emotional			HRQoL^			
PROM	**Ref.**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
CES-D	(221)	C-np	288	-	0.24	-	-	-	-	-
EQLQ-C30	(221)	C-np	288	-	-	-	Ø	ID	ID	ID
EQLQ-C30	(180, 181)	SE	133	-	-	-	Ø	Ø	Ø	Ø
FACT-G	(253)	Ex	42	-	-	-	0.39	Ø	0.56	0.60
HADS	(259)	Ex	63	0.38	0.31	ID	-	-	-	-
HADS	(257)	C-np	26	-2.12	ID	ID	-	-	-	-
HADS	(253)	Ex	42	Ø	Ø	Ø	-	-	-	-
IES-R	(180, 181)	SE	133	-	-	Ø	-	-	-	-
POMS-65	(180, 181)	SE	133	0.49	0.59	-	-	-	-	-
SCL-90-R	(250)	Ex	111	-0.21	-0.24	-	-	-	-	-
SF-36	(272)	Ex	38	-	-	-	-	0.48	ID	Ø
SF-36	(250)	Ex	111	-	-	-	-	0.37	ID	0.21

Intervention cluster: Screening (SC) (Number of RCTs: 3)

	Study detail			Emotional			HRQoL^			
PROM	**Ref.**	**Int.**	**N**	**Anxiety**	**Depression**	**Distress**	**QOL**	**Emotional**	**Social**	**Physical**
EQLQ-C30	(222)	SC	129	-	-	-	0.34	0.27	ID	Ø
HADS	(222)	SC	129	-0.23	-0.51	-0.54	-	-	-	-
HADS	(258)	SC	80	-0.30	-0.46	ID	-	-	-	-
POMS-11	(264)	SC	213	-	-	0.36	-	-	-	-

Ref. = Reference; Int. = Intervention category; N = total sample at baseline or the final sample size used for analysis in a study; *= effect size for between group difference at follow-up; - = not measured; ID = insufficient detail or information not given; Ø = no effect size 0.2 or above. Note: higher effect size for HRQoL indicates better HRQoL; higher effect size for anxiety, depression or distress indicates greater morbidity.

^*HRQoL scores*

- 'QOL' represents the effect size for FACT-G and QLI-CV total, and EORTC-QLQ-C30 global QOL scale scores
- 'Emotional' represents the effect size for FACT emotional wellbeing and EORTC-QLQ-C30 emotional functioning scales and for SF-12/36 the highest ES of role emotional, mental health or mental component summary (MCS)
- 'Social' represents the effect size for FACT social wellbeing, EORTC-QLQ-C30 social functioning and SF-12/36 social functioning scale scores
- 'Physical' represents the effect size for FACT physical wellbeing, EORTC-QLQ-C30 physical functioning and for SF-12/36, the highest ES of role functioning, role physical or physical component summary.