Patient-reported depression measures in cancer: a meta-review

Wakefield, C.E.; Butow, P.N.; Aaronson, N.K.; Hack, T.F.; Hulbert-Williams, N.J.; Jacobsen, P.B.; International Psycho-Oncology Society Research Committee

Link to publication

Patient-report depression measures in cancer: A meta-review

Claire E Wakefield., Phyllis N Butow, Neil Aaronson, Thomas F Hack, Nicholas Hulbert-Williams, Paul B Jacobsen, on behalf of the International Psycho-Oncology Society Research Committee.

Affiliations:

Discipline of Paediatrics, School of Women's and Children's Health, UNSW Medicine, University of New South Wales, NSW, Australia (C E Wakefield, PhD);

Kids Cancer Centre, Sydney Children's Hospital, High Street, Randwick, NSW, Australia (C E Wakefield, PhD);

Centre for Medical Psychology and Evidence-based Decision-making (CeMPED) and the Psycho-Oncology Co-operative Research Group (PoCoG), School of Psychology, University of Sydney, Sydney, New South Wales, Australia (P N Butow, PhD);

Department of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands (N Aaronson, PhD);

Department of Nursing, Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada (T F Hack, PhD)

Chester Research Unit for the Psychology of Health, Department of Psychology, University of Chester, Chester, UK (N Hulbert-Williams, PhD);

Department of Health Outcomes and Behavior, Moffitt Cancer Center and Research Institute, Tampa, FL, USA (P B Jacobsen, PhD).

Running head: Assessing depression in cancer: A meta-review

Correspondence to: Claire E. Wakefield. Email: c.wakefield@unsw.edu.au. Full postal address: Behavioural Sciences Unit, Kids Cancer Centre, Sydney Children's Hospital, Level 1, High St, Randwick, NSW, AUSTRALIA, 2031. Ph: +612 9382 3113. Fax: +612 9382 1789.

Summary

Depression measurement in cancer care is complex and inconsistent. It is difficult for investigators to select the best–performing tool. We conducted a meta-review to integrate the findings of reviews of patient-report depression measures used as screeners or case-finders in oncology. We searched Medline, PsycINFO, EMBASE and grey literature from 1999-2014. We identified 19 reviews representing 372 primary studies assessing more than 50 depression measures. We used 11 high-quality reviews to guide our analysis, which was organized by measurement goal and target population. The Hospital Anxiety Depression Scale was the most recommended, and criticized, depression screener. Few reviews evaluated case-finding performance or measure responsiveness, or measure suitability for particular populations. This meta-review demonstrates that the available measure selection advice is conflicting. By being fully cognizant of the benefits and limitations of depression measurement, investigators can improve the accuracy of their data and achieve more sophisticated interpretations of their findings.

Introduction

Psycho-oncology has seen an exponential rise in research documenting the prevalence, measurement and experience of depression in cancer.[3,4] MedLine records that 'depression' and 'cancer' were addressed together in an average of 192 citations/year in the 1980's, rising to an average of >1000 citations/year between 2006-2015.[5,6] Mirroring this rise, clinicians and researchers have utilized numerous patient-reported outcome measures to assess depression in individuals affected by cancer.[4,7,8,9]

Unrecognized and untreated depression can have deleterious implications for long term quality of life,[11,12] treatment adherence[13] health service use,[14,15] requests for death,[11] and mortality.[16-18] Opinion leaders therefore often recommend that all patients be evaluated for depressive symptoms at regular intervals across the trajectory of cancer care.[12] Accurate and timely measurement of depression can ensure that the prevalence of depression across populations and stages is neither under- nor over-estimated.[19,20] This data is needed to inform clinical practice and the allocation of appropriate resources to psychosocial services.[3,19]

Available depression measures, however, yield differing data,[21] with one meta-analysis of 211 studies, using only 4 different depression measures, reporting a range of 8% to 24% cancer patients affected by depression.[22] It is not possible to elucidate whether this variability is due to actual differences in depression prevalence across cancer types or stages, is an artefact of the instrument used in each study,[22] or is a function of each study's characteristics (eg, sample size and representativeness).[6,23] The use of a wide range of depression measures in clinical practice and research across studies has prevented simple cross-population and cross-cultural comparisons.[7,24] Also lost has been the ability to pool data,[8] and to compare outcomes across cancer types,[22] across time,[8] and across disease stages.[24]

For measures to assess depression see http://www.scalesandmeasures.net/search.php

Numerous reviews of depression measures, as well as evaluation tools to assess measure quality, are available. Available reviews however differ in focus (eg, providing a generic summary, or systematically appraising evidence) and methods (eg, their search, appraisal, synthesis, and analysis).[25] The IPOS Research Committee therefore conducted a meta-review (an 'overview of reviews' or

3

'umbrella review'[25]) of depression outcome measures used as screeners or case-finders in adults with or recovering from cancer. Meta-reviews present a unique approach to knowledge integration, enabling the aggregation and synthesis of multiple reviews into a single document.[26] They are particularly useful for exploring consistency of findings across reviews and revealing consensus.[25,27] The aims of this meta-review were:

1. To identify and critically appraise, using a gold-standard checklist, the available reviews of depression measures for use with adults with cancer;[28]
2. To aggregate the results of the captured reviews into one accessible report;
3. To identify consensus between reviews; and
4. To identify a set of 'candidate measures' for further detailed consideration of their appropriateness for measuring depression in adults affected by cancer.

Methods

Search strategy and selection criteria

This meta-review was undertaken following the steps recommended by Cooper and Koenka.[27] After formulating the problem, the steps included (1) searching the literature, (2) gathering information from articles/reports, (3) evaluating the quality of the evidence, (4) analyzing and integrating the outcomes of research, (5) interpreting the evidence, and (6) presenting the results.

Review selection (Step 1: Searching the literature[27])

We searched three databases of peer-reviewed journals, two grey literature databases, Google Scholar and reference lists of eligible reviews, for studies published between 1999 and 2014 (detailed in Panel A). The 15-year timeframe aligns with other reviews,[5,29] and ensured that depression measures developed prior to 1999, but used (and reviewed) in the past 15 years, were included.

Inclusion/exclusion criteria

Domains assessed: The meta-review aimed to capture reviews of patient-reported outcome measures of depression in adults with, or recovering from, cancer using a standardized paper or online questionnaire. Reviews of measures that included one or more subscales assessing depression (eg, quality of life outcome measures with a depression subscale) were eligible if the subscale's psychometric properties were

reported separately from the performance of the complete measure. Reviews assessing outcome measures for specific cancer diagnoses (eg, breast cancer-specific measures) were also eligible, as were reviews assessing 'ultra-short' (1-4 items) and 'short' (5-20 items) instruments.[4]

Types of reviews: We included published systematic (as defined by the PRISMA Statement; with or without meta-analyses)[28] and narrative reviews summarizing data collected from adults (aged 18+ years) diagnosed with any type of cancer, at any stage of the cancer experience (including palliative care and survivorship). Given the evidence that 'grey literature' plays an important role in guiding policy and practice,[30,31] we also included reviews published in reports, discussion papers, briefings, and practice guidelines.[30,31] Reviews of measures used for screening (ie, to 'rule-out' patients without depression with minimal missed cases [false negatives])[7], case-finding (ie, to 'rule in' those who have depression with minimal false positives)[7] and assessment (ie, to determine the extent of depressive symptoms) were eligible.[32]

Exclusions: We excluded primary studies and reviews of non-questionnaire measures, such as face-to-face or telephone-delivered clinical interviews. There is a lack of depression research[33-36] in non-English speaking populations, however we restricted the meta-review to those published in English because expert review was not possible in other languages and translation was beyond the scope of the project. We excluded other related domains, such as sadness, grief, suicidal ideation, melancholy, hopelessness, demoralization, adjustment disorder, and quality of life. We excluded measures of generalized 'distress' due to their lack of specificity in terms of psychological morbidity, unless they were specifically evaluated as depression screeners or case-finders.[37] We also excluded reviews on individuals without a cancer diagnosis (eg, those at increased risk of cancer, partners, caregivers, and family members). When multiple reviews published by the same first author were captured, we utilized the article with the highest quality (defined by the PRISMA statement), unless the reviews addressed substantively different research questions. For example, the 2010 Luckett, Butow [5] review was included rather than their 2012 review[38] due to substantial overlap in research questions, methodology and findings.

Data extraction and classification (Step 2: Gathering the information[27])

CEW and EGR reviewed all abstracts and full-text articles. Consensus regarding inclusion or exclusion of articles was achieved by discussion, and for remaining disagreements, by consultation with all authors. Captured articles were categorized by their primary purpose as assessing depression measures as i) screening tools, ii) case-finding tools, or iii) on their capacity to detect change (table 1). We extracted the following data for screening tools: sensitivity and specificity (pooled or weighted), screening utility index, recommended 'cut-point' scores, and the review's recommendations. For articles that did not report summary sensitivity and specificity scores, we calculated medians and ranges of scores where possible. Data collected from reviews assessing case-finding capacity included: case-finding area under the curve (AUC) and positive utility index (UI+). Data gathered from reviews assessing capacity to detect change included: weighted score for responsiveness and effect sizes detected.

Critical appraisal (Step 3: 'Evaluating the quality of the evidence[27])

CEW and EGR independently appraised the captured reviews using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement criteria,[28] supplemented by the PRISMA Explanation and Elaboration document.[39] When the investigators disagreed on any assessment, the issue was resolved through discussion with each other or with all authors. We decided, a priori, to focus our analysis on reviews that met 20 or more of the 27 PRISMA criteria because a recent assessment of the quality of PRISMA reporting in 236 reviews showed that approximately 70% of reviews meet 20 of the 27 PRISMA criteria.[40] Only measures recommended by at least one high-scoring review were considered as possible candidate measures suitable for detailed assessment. Narrative reviews were not critically appraised because their purpose and methods differ from systematic reviews.[41]

Results (Step 4: Analyzing the outcomes[27])

We identified 19 eligible reviews with good inter-rater reliability, including 12 systematic and 7 narrative reviews (figure 1). The captured reviews represented 372 original studies and assessed more than 50 depression measures. The Medline-

EMBASE-PsychInfo search was most effective, yielding 78·9% sensitivity (15/19 eligible reviews were captured with these searches) and 10·3% specificity (15 eligible articles were captured out of 145 abstracts). Reviews originated from the United States (n=8), the United Kingdom (n=6), Australia (n=3), Canada (n=1), and the Netherlands (n=1), and focused on mixed cancer diagnoses, or on a specific cancer population (eg, older patients[11,42]). The goals of each review varied. Nine reviews assessed the suitability of depression measures as screening tools,[1,4,5,7,8,19,22,43,44] while three assessed their suitability for case-finding.[7,43,44] Several reviews examined appropriate cut-points of specific measures,[8,19,45,46] while others examined their usefulness in particular populations (eg, geriatric patients[11,42]). One review assessed the responsiveness of depression measures in detecting the effect of psychological interventions.[5] One review assessed the performance of measures across five stages in the cancer trajectory,[8] while others focused on mixed diagnoses and treatment stages.[3,5-7,19,20,22,43-47]

Critical appraisal

Table 2 summarizes the critical appraisal of each systematic review. All systematic reviews provided a sound rationale, a structured summary of findings, a description of their objectives and some discussion of findings (n=12). The PRISMA criteria least likely to be met were assessing the risk of bias within and across studies (four reviews assessed bias within studies[4,7,19,22] and four assessed bias across studies[4,43-45]). Five reviews failed to acknowledge their limitations and no review provided review protocol/registration details, suggesting that protocol registration for reviews is not yet common practice.[48] Eleven of the 12 systematic reviews met at least 20 of the PRISMA criteria. The findings of high scoring reviews are summarized in Tables 3-5. Tables 6 and 7 present the findings of lower scoring systematic reviews and the narrative reviews.

Aggregation of meta-review results (summary of high scoring reviews' recommendations)

Screening

The Hospital Anxiety and Depression Scale (HADS) was the most widely evaluated measure, with nine reviews assessing the HADS against other measures,[1,4,5,7,22,42,43] or alone.[19,44] Positive features reported included its popularity (enabling cross-study

comparisons),[6,8] and its ability to perform adequately across different stages of the cancer trajectory.[8] The HADS was described as performing well in identifying major depression within pre-treatment (with a cut-point of 7 for the HADS-Depression subscale [HADS-D]) and post-treatment populations (cut-point between 9 and 11 for HADS-D).[8] The most commonly-used threshold to determine depression prevalence during active treatment was a subscale score of 8 or above,[8,19,22] although this cut-point was poorly supported in one review.[19] Each of the HADS subscales received moderate screening utility index scores for depression in one review (ranging from 0·65-0·71),[7] although these figures vary substantially across reviews.[7,44]

Several reviews converged on the limitations of the HADS, highlighting the differing performance between the HADS-Total scale (HADS-T), the HADS-Anxiety subscale (HADS-A), and the HADS-D,[7] and the variability in recommended cut-points (ranging from 4 to 11).[4,6,8,15,19] Mitchell and colleagues also suggested that the HADS-T or HADS-A (rather than the HADS-D) could be used as the first choice for a depression screening measure.[44] HADS-A may also perform as well as HADS-T in identifying depression in palliative care,[1] although four reviews argued that the HADS was least suited for advanced cancer patients and for those receiving palliative care.[1,4,7,8]

Several reviews assessed the screening performance of the Beck Depression Inventory (BDI), and/or its variations (the BDI-II and the BDI-short form, BDI-SF).[1,4,5,7,8,15,22,42,43] In each case, the BDI's performance was considered in comparison with other measures. Reviews assessed the BDI favourably, highlighting its generalizability across cancer types and disease stages,[4] its adequate screening performance[7] and its potential usefulness in older patients.[42] One review described the BDI as 'excellent' for a long measure, due to its good reliability and validity.[4] Several reviews noted that the BDI has appropriate sensitivity and specificity,[1,4,7,8] although some evidence suggests it has poorer specificity before and after cancer treatment.[8] The BDI, however, is limited somewhat by its length (21 items), reducing its acceptability.[4,43] It also has a longer recall period (two weeks), potentially limiting its usefulness in some contexts.[42] The BDI has also been criticized for including items with a somatic emphasis.[5] The BDI-SF, with only 13

items may address some of these limitations; however, it may not perform as well psychometrically.[4]

Several reviews evaluated ultra-short depression screeners, such as the Distress Thermometer (DT).[7,38,45,47] Despite not specifically targeting depression, the DT showed good sensitivity and specificity as a depression screener and had high clinical acceptability in one review,[7] and good sensitivity to change in another.[4] Its performance also appeared comparable to the Brief Symptom Inventory-18 items (BSI-18) and General Health Questionnaire-12 items (GHQ-12) in palliative care.[8] However, given the potential high rates of false-negatives when using the DT, one review recommended the DT (and other ultra-short tools) not be used in isolation for depression screening.[7]

Three reviews highlighted positive features of the Zung Self-Rating Depression Scale (ZSDS) (including being able to assess mood variation and having predictive validity data available).[8,42,43] However, one review reported that it had good specificity, but poor sensitivity, at the time of cancer diagnosis.[8] The Patient Health Questionnaire-9 (PHQ-9) was also evaluated by three reviews.[4,5,42] Vodermeier and colleagues[4] recognized its strong psychometric properties in medical populations, however rated it poorly due to low reliability and validity in cancer patients. Nelson and colleagues argued that the recall length of two weeks makes it less useful, for geriatrics in particular.[42] Some PHQ-9 questions were highlighted as less appropriate for those undergoing active treatment (ie, in regards to sleep, fatigue, appetite, concentration, and restlessness).[5]

Three reviews provided a positive appraisal of the Edinburgh Depression Scale (EDS) or the Brief Edinburgh Depression Scale (BEDS).[1,4,8] In palliative care, one review recommended the EDS due to the absence of somatic items,[1] with another arguing that the EDS can perform better than the HADS in this population.[4] The Center for Epidemiologic Studies Depression Scale (CES-D) was evaluated by two reviews.[4,42] It was the highest ranked short tool in one review (particularly the negative affect subscale).[4] Although comprehensively evaluated, Nelson and colleagues argued that it was less suitable for geriatric patients because it includes only two of the most common seven depression symptoms in geriatric patients.[42]

One review described the General Health Questionnaire-28 (GHQ-28) as 'excellent' due to its high sensitivity and specificity, however the authors also expressed concerns about its length.[4] Despite its good screening performance, one review expressed caution regarding its use during active treatment because of a lack of validation studies.[8] The 12-item version of the GHQ (GHQ-12) was reviewed as 'good' by Thekkumpurath and colleagues,[1] although they reported it had inferior psychometric properties to the HADS in advanced cancer. Few studies have reported GHQ-12 parameters, making it difficult to compare with other tools.[1] The screening performance of the remaining scales was assessed in too few reviews to draw conclusions regarding their usefulness.

Case-finding

Evidence for the case-finding performance of depression measures was rare and less convincing, with only three reviews evaluating case-finding performance. One review assessed the HADS alone,[44] while two compared the performance of multiple measures.[7,43] Mitchell and colleagues (2010) reported that all HADS subscales had poor case-finding utility indices (ranging from 0·27-0·29).[44] The remaining reviews reported moderate case-finding indices for the HADS.[7,43] However, while the HADS indices were higher than the DT indices in both reviews,[7,43] they were lower than that calculated for the BDI-II, which was graded as having mid-level evidence for case-finding capacity.[7] Few data were available regarding the case-finding performance of depression measures for different populations or at different treatment stages, however there was some evidence that the BDI-18 did not perform as an effective case-finder in palliative care.[7,43]

Responsiveness

One review assessed the capacity of depression measures to detect treatment effects after participation in randomized controlled trials of psychological interventions.[5] The CES-D, the HADS, and the POMS-SF received the highest weighted scores (out the 16 measures assessed). The CES-D was most highly endorsed for studies in which depression was the sole focus, with an average detected effect size of -0·27 (ranging from -0·36 to -1·04). The length of the CES-D (20 items) was highlighted as a limitation. The HADS was the best supported measure if the study had a broader

focus, such as for studies also assessing anxiety, mixed affective disorders or general distress. In these studies, the detected depression effect sizes averaged 0·45 (range 0·1 to 0·97). HADS was however, considered less suitable for detecting minor depression.[5] The POMS-37 received a good score (ranked third) due to its good validity and reliability, although the reviewers highlighted that both the HADS and POMS-SF may be less suitable than the CES-D because they rely more heavily on anhedonia.[5]

Consensus and discordance between reviews (Step 5: Interpreting the evidence[27])
Table 8 summarizes the positive and negative features of the measures considered in this meta-review. Figures 2 and 3 present specific recommendations and cautions against each reviewed measure (organized by goal of measurement and by target population). No candidate measure was recommended by all reviews, nor did reviews unanimously recommend any one measure for a particular population or a specific time-point. Instead, the reviews agreed that every measure had important positive features, balanced by significant limitations. Many of the highlighted positive features and disadvantages of each measure were common across reviews. For example, multiple reviews highlighted the benefits associated with the popularity of the HADS (for cross comparisons),[6,8] and agreed that the variability of recommended HADS cut-points was a limitation.[4,8,19,22,46] Several reviews also agreed that the HADS was less well suited for palliative care and advanced cancer patients.[1,8,44] There were insufficient independent reviews of the case-finding ability and responsiveness of depression measures to identify consensus between reviews.

There were several disagreements between reviews regarding screening. When considering the HADS for example, it was unclear which scale or subscale (HADS-T, HADS-D or HADS-A) was most suitable, particularly since different reviews assessed one, or each of the scales/subscales.[7] Regarding the BDI and the CES-D, each was described as potentially useful for older, terminally ill patients,[11] yet were also highlighted as less suitable for this population.[42] The remaining measures were assessed by too few reviews to enable the identification of any disagreements regarding their suitability for screening, case finding, and detecting change.

Discussion

11

In conducting this meta-review, the IPOS Research Committee hoped to ensure that investigators are utilizing the best performing measures, encourage consistency across data collection and reporting, and create future opportunities for cross-study comparisons. However, depression measures that were commonly recommended by some reviews were criticized by others.[4,20] No review provided unqualified support for the adoption of any measure, and there were no measures identified as ideal for all types of patients and time points. The HADS was most often considered an acceptable depression screener and was most thoroughly evaluated,[1,4,7,8,44] although it remains unclear whether the HADS-D, HADS-T, or even the HADS-A, is the best choice.[44] Some captured reviews also recommended the CES-D, both for screening,[4,5] and for detecting change.[5] Reviews assessing case-finding capacity were rare, with the BDI-II being recommended if no verbal assessment is possible.[7] Panel B places these findings into context.

This meta-review identified several measures as having particular potential for specific time points/populations. In palliative settings, the EDS (or BEDS) and the DT appear promising.[1,8] The HADS appears less useful in this population, due to its focus on anhedonia,[5] although generally the HADS is appraised positively because of its reduced reliance on somatic symptoms.[20] Interestingly, there is some evidence to suggest that omitting somatic symptoms from depression screeners does not significantly improve the performance of common depression measures.[7,49]

Despite many published studies attributing disappointing findings to floor or ceiling effects,[50] the captured reviews did not address the risk of floor/ceiling effects in each measure. This is a significant limitation of many fixed length tools that might be overcome by computer adaptive testing (CAT). Given their recent development however, this meta-review failed to capture reviews of CAT measures of depression.[51] CAT utilizes computer-aided logic, such that participants' responses to previous items are used to select the most informative next item from a pool of possible items. CAT can reduce floor/ceiling effects because individuals at the top or bottom of a scale receive items that are targeted toward their level, improving precision and clinical relevance in comparison to the fixed length tools considered here.[50,52] Given the rapid progress of the patient-reported outcome measurement information system (PROMIS) initiative,[53] and the European Organisation for

Research and Treatment of Cancer (EORTC) project,[54] it is possible that these approaches may complement, or even replace, the fixed length outcome measures assessed in this meta-review.[7]

Despite identifying some evidence supporting the use of several depression measures in cancer, this meta-review also revealed major challenges faced by the screening and case-finding field. Patient-reported measures of depression can be useful to facilitate early detection,[55] can cost less than structured clinical interviews[5] and can be used across centers and countries.[4] However, without effective after care, even the best performing measures will be of minimal benefit to patients.[56] Indeed, many argue that screening should only be conducted in settings where there are appropriate resources available for in-depth assessment and treatment of those identified as in need.[56-58] Only a minority of cancer centers has implemented routine screening into practice,[4] and there is a dearth of implementation studies that demonstrate direct benefits to patients from screening.[7,43,44] It is also possible that currently available screeners are least suitable for those who may have the greatest need (eg, non-English speaking, and socially and economically disadvantaged, patients). It is important too, to recognize that all measures miss positive cases and falsely identify negative cases, both of which can have deleterious impacts.

The potential role of alternative approaches to identifying patient needs for help with depression therefore warrants further consideration. It may be simpler in many cases to directly ask patients what they need, face-to-face.[58,59] It is also possible that different measures or approaches may be appropriate for research versus clinical care.[4] More work exploring measure acceptability from clinicians' perspectives would be useful in this regard, as there is some evidence that clinicians find depression measures less acceptable than patients.[53,54] Indeed, many continue to rely on their clinical judgement[43,44] given that no patient-report measure has 100% concordance with clinical interview[15,60] and that most guidelines recommend screening should be used in conjunction with thorough clinical assessment anyway.[1,6,32,44]

Finally, given the often significant overlap between depression symptoms and other outcomes (such as distress, anxiety, fatigue, and pain), multidimensional assessment

might also be more useful than focusing on the single construct of depression.[4,8,15,61,62] In practice, clinicians are rarely solely interested in depression, potentially limiting the relevance of this meta-review to clinical practice.[56] However, assessment of multiple dimensions within one measure can create statistical and clinical challenges in interpreting the data,[63] suggesting that there still is a role for depression-specific measurement when it is the primary focus of a research project.

Future research

Further studies providing head-to-head (within sample) comparisons between measures and in comparison to clinical interviews are needed. There is a dearth of research assessing depression measures' case-finding ability and responsiveness. Further work is needed to investigate the practical usefulness and clinical acceptability of depression measures. For many patients, depression may not be their primary concern, with practical needs, such as support with financial difficulties, and family and social concerns, often being more highly endorsed as core concerns for patients than depression.[7,34,59] Recent research also suggests that that anxiety may be of greater importance to patients than depression, especially for cancer survivors.[64,65] The most useful screeners therefore might have a broader focus, beyond depression, and may also benefit from an assessment of patients' desire for help with the concerns they endorse.[59]

Additional population-specific work is also warranted. Some groups are more likely to under-report depression (eg, older patients, patients from different cultures, and those whose primary language is not English), meaning that measure performance may vary across groups.[11,42] One review reported that two-thirds of their reviewed studies did not report the race/ethnicity of their sample, or specifically targeted a homogenous English-speaking sample, highlighting the literature dearth in this area.[34] Furthermore, recorded depression prevalence varies widely among patients with different tumor types, with patients with some cancers (eg, breast, head and neck, and malignant melanoma) reporting high depression levels, despite often greater needs in patients with other diseases (eg, lung cancer).[66] Indeed, there may be little conceptual or statistical equivalence between the same measures used in different populations,[34] putting some groups at risk of receiving inappropriate services if screening programs are used to determine access.[6,34]

Meta-review limitations

Meta-reviews can overlook newly published papers.[27] The fact that CAT measures were not captured demonstrates that there may be newer, but well-validated, measures that have not yet generated enough research to have been captured in a review.[27,67] The PHQ-9 is another good example, because while it was reviewed less favourably in this meta-review, recent research suggests it may be a promising screener[12] and case-finder,[68] and may be quite responsive.[69] The captured reviews also provided different data and used different methods, making it difficult to draw firm conclusions, however future meta-reviews could utilize more sophisticated tools (such as Rasch measurement) to allow a common metric to be generated for different depression measures.[70] There were also overlaps in some of the captured reviews, which may have meant that the views of some, more prolific, research groups were over-represented.[27] Failure to capture non-English reviews was also a limitation.

Conclusion

Patient-report outcome measures can play an important role in cancer care, and when used appropriately, can serve as a cost-effective, equitable means of identifying patients in need of clinical assessment and treatment.[4,6] Given that current clinician and researcher choice of depression measures can be arbitrary (based on popularity, familiarity, personal preference, or perceived relevance),[8] there is a need for more consensus on optimal depression measures in cancer.[5,20] It is not likely, however, that a single tool exists, or will be developed, that meets the needs of every clinical and research purpose.[7] Despite this, there are still clear benefits to improving consistency in measurement across the field where possible.[7] This meta-review highlights commonly recommended depression measures in cancer while giving due consideration to the limitations of the depression measurement field and to the limitations of individual measures.

Conflicts of interest

The authors declared no conflicts of interest.

Contributors

CW, PB, NA, NH, and PJ contributed to the design, and supervision, of the review. CW and ER analysed and interpreted the data with the support of PB. All authors drafted the paper and read, commented on, and approved the final draft.

References

1. Thekkumpurath P, Venkateswaran C, Kumar M, Bennett MI. Screening for Psychological Distress in Palliative Care: A Systematic Review. *J Pain Symptom Manage* 2008; **36**(5): 520-8.

2. Harzing A-W. A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics* 2012; **98**(1): 1-11. doi: 10.1007/s11192-013-0975-y

3. Carlson LE, Bultz BD. Cancer distress screening: Needs, models, and methods. *J Psychosom Res* 2003; **55**(5): 403-9.

4. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. *J Natl Cancer Inst* 2009; **101**(21): 1464-88.

5. Luckett T, Butow PN, King MT, et al. A review and recommendations for optimal outcome measures of anxiety, depression and general distress in studies evaluating psychosocial interventions for English-speaking adults with heterogeneous cancer diagnoses. *Supportive Care in Cancer* 2010; **18**(10): 1241-62.

6. Trask PC. Assessment of depression in cancer patients. *JNCI Monographs* 2004; **2004**(32): 80-92.

7. Mitchell AJ, Meader N, Davies E, et al. Meta-analysis of screening and case finding tools for depression in cancer: evidence based recommendations for clinical practice on behalf of the Depression in Cancer Care consensus group. *Journal of Affective Disorders* 2012; **140**(2): 149-60. doi: 10.1016/j.jad.2011.12.043

8. Ziegler L, Hill K, Neilly L, et al. Identifying psychological distress at key stages of the cancer illness trajectory: A systematic review of validated self-report measures. *J Pain Symptom Manage* 2011; **41**(3): 619-36.

9. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB Journal* 2008; **22**(2): 338-42.

10. Jacsó P. Google Scholar: the pros and the cons. *Online information review* 2005; **29**(2): 208-14.

11. King DA, Heisel MJ, Lyness JM. Assessment and psychological treatment of depression in older adults with terminal or life-threatening illness. *Clinical Psychology: Science and Practice* 2005; **12**(3): 339-53.

12. Andersen BL, DeRubeis RJ, Berman BS, et al. Screening, assessment, and care of anxiety and depressive symptoms in adults with cancer: an American Society of Clinical Oncology guideline adaptation. *J Clin Oncol* 2014: JCO. 2013.52. 4611.

13. McLoone JK, Wakefield CE, Evans NT, Cohn RJ. When the 'all clear' is not so clear: Families' coping with fear of childhood cancer recurrence. *Under review* 2013.

14. Simon GE, VonKorff M, Barlow W. Health care costs of primary care patients with recognized depression. *Arch Gen Psychiatry* 1995; **52**(10): 850.

15. Wilson KG, Chochinov H, De Faye B, Breitbart W. Diagnosis and management of depression in palliative care. *Handbook of psychiatry in palliative medicine* 2000: 25-49.

16. Giese-Davis J, Collie K, Rancourt KM, Neri E, Kraemer HC, Spiegel D. Decrease in depression symptoms is associated with longer survival in patients with metastatic breast cancer: a secondary analysis. *J Clin Oncol* 2011; **29**(4): 413-20.

17. Faller H, Bülzebruck H, Drings P, Lang H. Coping, distress, and survival among patients with lung cancer. *Arch Gen Psychiatry* 1999; **56**(8): 756-62.

18. Satin JR, Linden W, Phillips MJ. Depression as a predictor of disease progression and mortality in cancer patients. *Cancer* 2009; **115**(22): 5349-61.

19. Carey M, Noble N, Sanson-Fisher R, MacKenzie L. Identifying psychological morbidity among people with cancer using the Hospital Anxiety and Depression Scale: time to revisit first principles? *Psycho-Oncology* 2012; **21**(3): 229-38. doi: 10.1002/pon.2057

20. Pirl WF. Instruments in psycho-oncology. Psychooncology. 2nd ed. New York, NY: Oxford University Press; US; 2010: 119-30.

21. Massie MJ. Prevalence of depression in patients with cancer. *JNCI Monographs* 2004; **2004**(32): 57-71.

22.   Krebber A, Buffart L, Kleijn G, et al. Prevalence of depression in cancer patients: a meta-analysis of diagnostic interviews and self-report instruments. *Psycho-Oncology* 2014; **23**(2): 121-30.

23.   Janberidze E, Hjermstad MJ, Haugen DF, et al. How Are the Patient Populations Characterized in Studies Investigating Depression in Advanced Cancer? Results From a Systematic Literature Review. *J Pain Symptom Manage* 2014; **48**(4): 678-98. doi: 10.1016/j.jpainsymman.2013.11.013

24.   Mühlan H, Bullinger M, Power M, Schmidt S. Short forms of subjective quality of life assessments from cross-cultural studies for use in surveys with different populations. *Clinical Psychology & Psychotherapy* 2008; **15**(3): 142-53.

25.   Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 2009; **26**(2): 91-108. doi: 10.1111/j.1471-1842.2009.00848.x

26.   Becker LA, Oxman AD. Overviews of reviews. *Cochrane Handbook for Systematic Reviews of Interventions John Wiley & Sons: Chichester* 2008: 607-31.

27.   Cooper H, Koenka AC. The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist* 2012; **67**(6): 446-62. doi: 10.1037/a0027119

28.   Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151**(4): 264-9.

29.   Watanabe SM, Nekolaichuk CL, Beaumont C. The Edmonton Symptom Assessment System, a proposed tool for distress screening in cancer patients: development and refinement. *Psychooncology* 2012; **21**(9): 977-85.

30.   Lawrence A, Houghton J, Thomas J, Weldon PR. Where is the evidence? Realising the value of grey literature for public policy & practice: a discussion paper. 2014.

31.   Lawrence A, Houghton J, Thomas J, Weldon P. Where is the evidence: Realising the value of grey literature for public policy and practice

Melbourne, Australia: Swinburne Institute for Social Research, 2014.

32.   Howell D, Keller-Olaman S, Oliver T, et al. A pan-Canadian practice guideline: screening, assessment and care of psychosocial distress (depression, anxiety) in adults with cancer. *Toronto: Canadian Partnership Against Cancer (Cancer Journey Action Group) and the Canadian Association of Psychosocial Oncology* 2010.

33.   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. *Journal of Health Psychology* 2002; **7**(1): 107-14.

34.   Kayser K, Acquati C, Tran TV. No patients left behind: A systematic review of the cultural equivalence of distress screening instruments. *J Psychosocial Oncology* 2012; **30**(6): 679-93.

35.   Luckett T, Goldstein D, Butow PN, et al. Psychological morbidity and quality of life of ethnic minority patients with cancer: a systematic review and meta-analysis. *The lancet oncology* 2011; **12**(13): 1240-8.

36.   Hogan M. Depression Often Undiagnosed, Untreated in Low-Income, Minority Women with Cancer. *Oncology Times* 2005; **27**(12): 36-7.

37.   Hearn J, Higginson IJ. Outcome measures in palliative care for advanced cancer patients: a review. *J Public Health* 1997; **19**(2): 193-9.

38.   Luckett T, Butow, P., Price, M., Rankin, N., King, M., & Oguchi, M. 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. . Sydney, Australia: Psycho-Oncology Co-operative Research Group (PoCoG), 2011.

39.   Schreiber PJ. Women's career development patterns. *New directions for adult and continuing education* 2002; **1998**(80): 5-13.

40.   Willis BH, Quigley M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol* 2011; **11**(1): 163. doi: 10.1186/1471-2288-11-163

41.   Cronin P, Ryan, F., Coughlan, M. Undertaking a literature review: a step-by-step approach. *Br J Nurs* 2008; **17**(1): 38-43.

42.   Nelson CJ, Cho C, Berk AR, Holland J, Roth AJ. Are gold standard depression measures appropriate for use in geriatric cancer patients? A systematic evaluation of self-report depression instruments used with geriatric, cancer, and geriatric cancer samples. *J Clin Oncol* 2010; **28**(2): 348-56.

43.   Mitchell AJ. Short screening tools for cancer-related distress: a review and diagnostic validity meta-analysis. *J Natl Compr Canc Netw* 2010; **8**(4): 487-94.

44.   Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. *J Affect Disord* 2010; **126**(3): 335-48.

45.   Donovan KA, Grassi, L., McGinty, H.L., & Jacobsen, P.B. Validation of the Distress Thermometer worldwide: state of the science. *Psychooncology* 2014; **23**(3): 241-50.

46.   Morse R, Kendell K, Barton S. Screening for depression in people with cancer: the accuracy of the hospital anxiety and depression scale. *Clinical Effectiveness in Nursing* 2005; **9**(3): 188-96.

47.   Institute NC. Assessment and Diagnosis. August 28th 2014. http://cancer.gov/cancertopics/pdq/supportivecare/depression/HealthProfessional (accessed September 24th 2014).

48.   Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; **151**(4): W-65-W-94.

49.   Mitchell AJ. The value of somatic and non-somatic symptoms in diagnosing depression with comorbid physical disease: A systematic review and meta-analysis. *Psychooncology* 2014; **23**(Suppl. 3): 8.

50.   Fries J, Rose M, Krishnan E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *The Journal of rheumatology* 2011; **38**(8): 1759-64.

51.   Streiner DL, Norman GR. Health measurement scales. Oxford: Oxford University Press; 2008.

52.   Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010; **63**(11): 1179-94.

53.   Whelan CT. The role of income, life-style deprivation and financial strain in mediating the impact of unemployment on psychological distress: Evidence from the Republic of Ireland. *Journal of Occupational and Organizational Psychology* 2011; **65**(4): 331-44.

54.   Colagiuri B, Christensen S, Jensen AB, Price MA, Butow PN, Zachariae R. Prevalence and predictors of sleep difficulty in a national cohort of women with primary breast cancer three to four months postsurgery. *J Pain Symptom Manage* 2011; **42**(5): 710-20.

55.   Fischer D, Wedel B. Anxiety and depression disorders in cancer patients: incidence, diagnosis and therapy. *Memo-Magazine of European Medical Oncology* 2012; **5**(1): 52-4.

56.   Mitchell AJ, Vahabzadeh A, Magruder K. Screening for distress and depression in cancer settings: 10 lessons from 40 years of primary-care research. *Psycho‐Oncology* 2011; **20**(6): 572-84.

57.   Coyne JC, Thompson R, Palmer SC, Kagee A, Maunsell E. Should we screen for depression? Caveats and potential pitfalls. *Applied and Preventive Psychology* 2000; **9**(2): 101-21.

58.   Palmer SC, Coyne JC. Screening for depression in medical care: pitfalls, alternatives, and revised priorities. *Journal of Psychosomatic Research* 2003; **54**(4): 279-87. doi: 10.1016/S0022-3999(02)00640-2

59.   Garssen B, Van Der Lee M. Re: Screening for emotional distress in cancer patients: A systematic review of assessment instruments. *J Natl Cancer Inst* 2010; **102**(7): 507-8.

60.   Pirl WF. Evidence report on the occurrence, assessment, and treatment of depression in cancer patients. *JNCI Monographs* 2004; **2004**(32): 32-9.

61.   Mitchell AJ, Chan M, Bhatti H, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. *The lancet oncology* 2011; **12**(2): 160-74.

62.   Paice JA. Assessment of symptom clusters in people with cancer. *Journal of the National Cancer Institute Monographs*

2004; **2004**(32): 98-102. doi: 10.1093/jncimonographs/lgh009

63.   Kaasa S, Loge JH. Quality of life in palliative care: principles and practice. *Palliat Med* 2003; **17**(1): 11-20.

64.   Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. *The lancet oncology* 2013; **14**(8): 721-32.

65.   Linden W, Vodermaier A, MacKenzie R, Greig D. Anxiety and depression after cancer diagnosis: Prevalence rates by cancer type, gender, and age. *J Affect Disord* 2012; **141**(2): 343-51.

66.   Mehnert A, Brähler E, Faller H, et al. Four-Week Prevalence of Mental Disorders in Patients With Cancer Across Major Tumor Entities. *J Clin Oncol* 2014: JCO-2014. doi: 10.1200/JCO.2014.56.0086

67.   Limburg H, Shaw AK, McBride ML. Impact of childhood cancer on parental employment and sources of income: a Canadian pilot study. *Pediatr Blood Cancer* 2008; **51**(1): 93-8.

68.   Carlson RH. Plenary Study Shows Quick Method to Screen for Depression in Cancer Patients. *Oncology Times* 2014; **36**(1): 15-6. doi: 10.1097/01.cot.0000441978.88895.b2

69.   Johns SA, Kroenke K, Krebs EE, Theobald DE, Wu J, Tu W. Longitudinal comparison of three depression measures in adult cancer patients. *J Pain Symptom Manage* 2013; **45**(1): 71-82. doi: 10.1016/j.jpainsymman.2011.12.284

70.   Fischer HF, Tritt K, Klapp BF, Fliege H. How to compare scores from different depression scales: equating the Patient Health Questionnaire (PHQ) and the ICD-10-Symptom Rating (ISR) using Item Response Theory. *International journal of methods in psychiatric research* 2011; **20**(4): 203-14. doi: 10.1002/mpr.350

71.   Lloyd-Williams M. Screening for depression in palliative care patients: a review. *Eur J Cancer Care (Engl)* 2001; **10**(1): 31-5.

72.   Love A. The identification of psychological distress in women with breast cancer. Sydney, Australia: National Breast Cancer Centre

2004.

73.   Ziegler L, Hill K, Neilly L, et al. Identifying psychological distress at key stages of the cancer illness trajectory: a systematic review of validated self-report measures. *J Pain Symptom Manage* 2011; **41**(3): 619-36.

74.   Garavan TN, Carbery R. Explaining managers' participation in career-focused learning and development. 2010.

75.   King DA, Heisel MJ, Lyness JM. Assessment and Psychological Treatment of Depression in Older Adults with Terminal or Life-Threatening Illness. *Clinical Psychology: Science and Practice* 2005; **12**(3): 339-53.

Panel A:

<div style="border:1px solid black; padding:10px">

Search and selection criteria

We identified eligible reviews by searching three electronic databases of peer-reviewed journals (Medline, EMBASE, and PsychInfo), limited to human studies published between 1999 and 2014. We used the following search terms in Medline: [oncol$ OR neoplasm OR cancer OR tumor OR tumour OR leuk$ OR haematol$ OR hematol$ OR palliative] AND [review OR meta-anal$ OR 'literature review' OR 'systematic review' OR 'research synthesis' or 'narrative review'] AND [scale OR questionnaire OR measure OR survey] AND [reliability OR validity OR validation OR psychometric] AND [depression OR depressive]. The term 'palliative care' was included because the majority of palliative care patients are diagnosed with cancer.[1] This search structure was then adapted to suit EMBASE and PsychInfo and re-run. We also identified reviews by searching two grey literature databases (OpenGrey and Grey Literature Report). We also searched the reports of the following relevant organizations: the National Institute for Health and Clinical Excellence, Cancer Australia, U.S. National Comprehensive Cancer Network, U.S. National Cancer Institute, Cancer Research UK). Due to its expanding coverage[2] and growing role in academic work (particularly in facilitating article retrieval[9,10]), we also searched Google Scholar. The reference lists of identified reviews were also searched.

</div>

Panel B:

**Research in context**

Evidence before this study

Patient-report depression measures are increasingly used in cancer care. It is unclear which measures perform best. Investigators therefore choose measures based on advice in the literature (which is often conflicting), or worse, based on popularity, familiarity or personal preference.

Added value of this study

This meta-review integrates the findings of all reviews about depression measure selection in cancer, published in the last 15 years. It identifies consensus and disagreements across reviews and provides guidance about the most, and least, recommended measures, organized by study goal and target population.

Implications of all the available evidence

No currently available depression measure performs well enough to meet the needs of every clinical practice or research study. It is critical that if investigators choose to undertake depression screening or case-finding, they utilize the best available measure for their population and study design while being cognizant of their limitations. In doing so, depression measurement will become more accurate and consistent. Improved awareness of measure limitations will facilitate a more sophisticated interpretation of the data provided by patient-report depression tools.
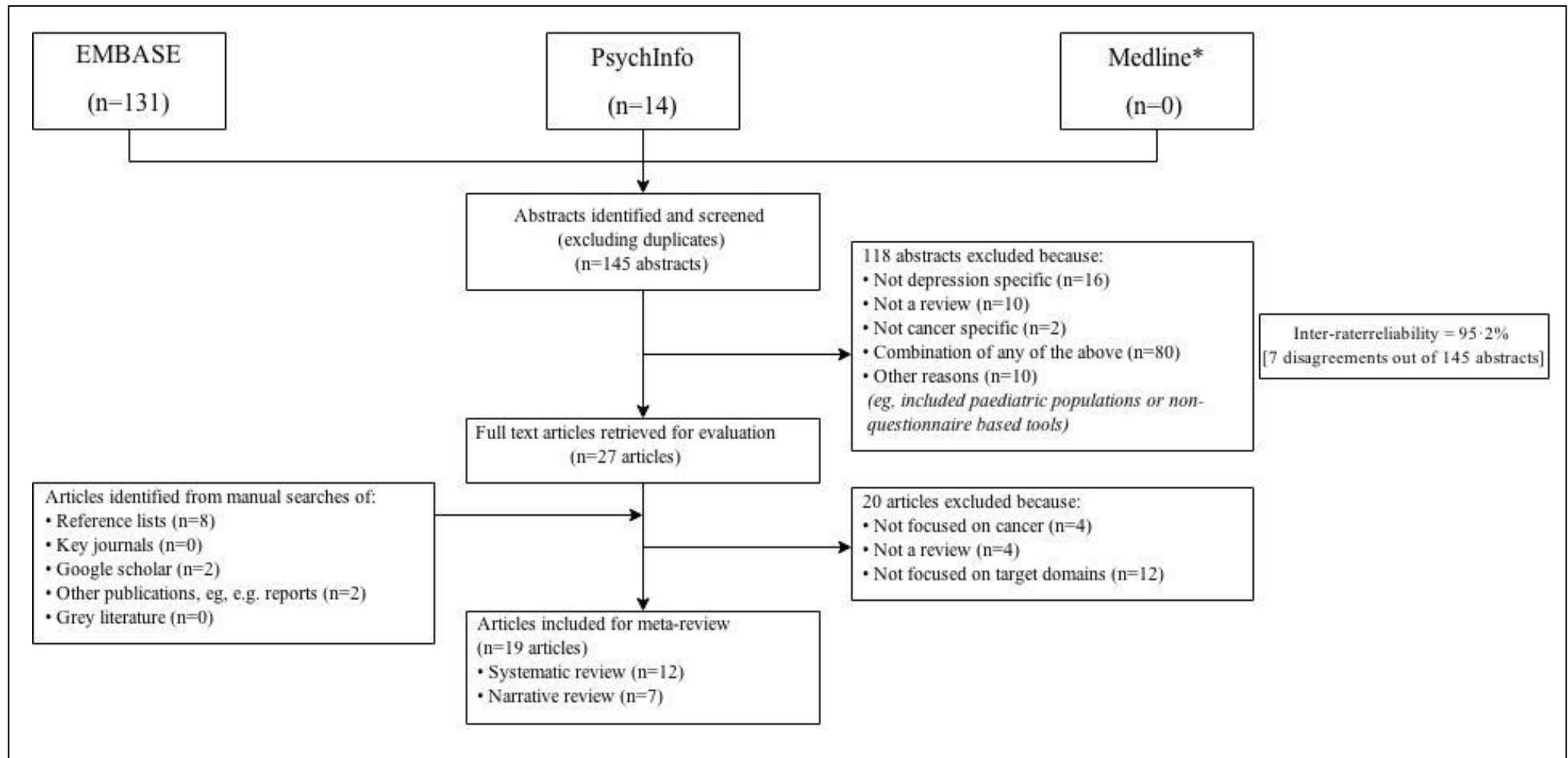
Figure 1: **Search process and articles captured for the meta-review (as at 7th November 2014)**
Note. Databases were searched in the following order: EMBASE, PsychInfo and Medline
(*Medline captured no additional abstracts that were not already captured by EMBASE and/or PsychInfo).

23
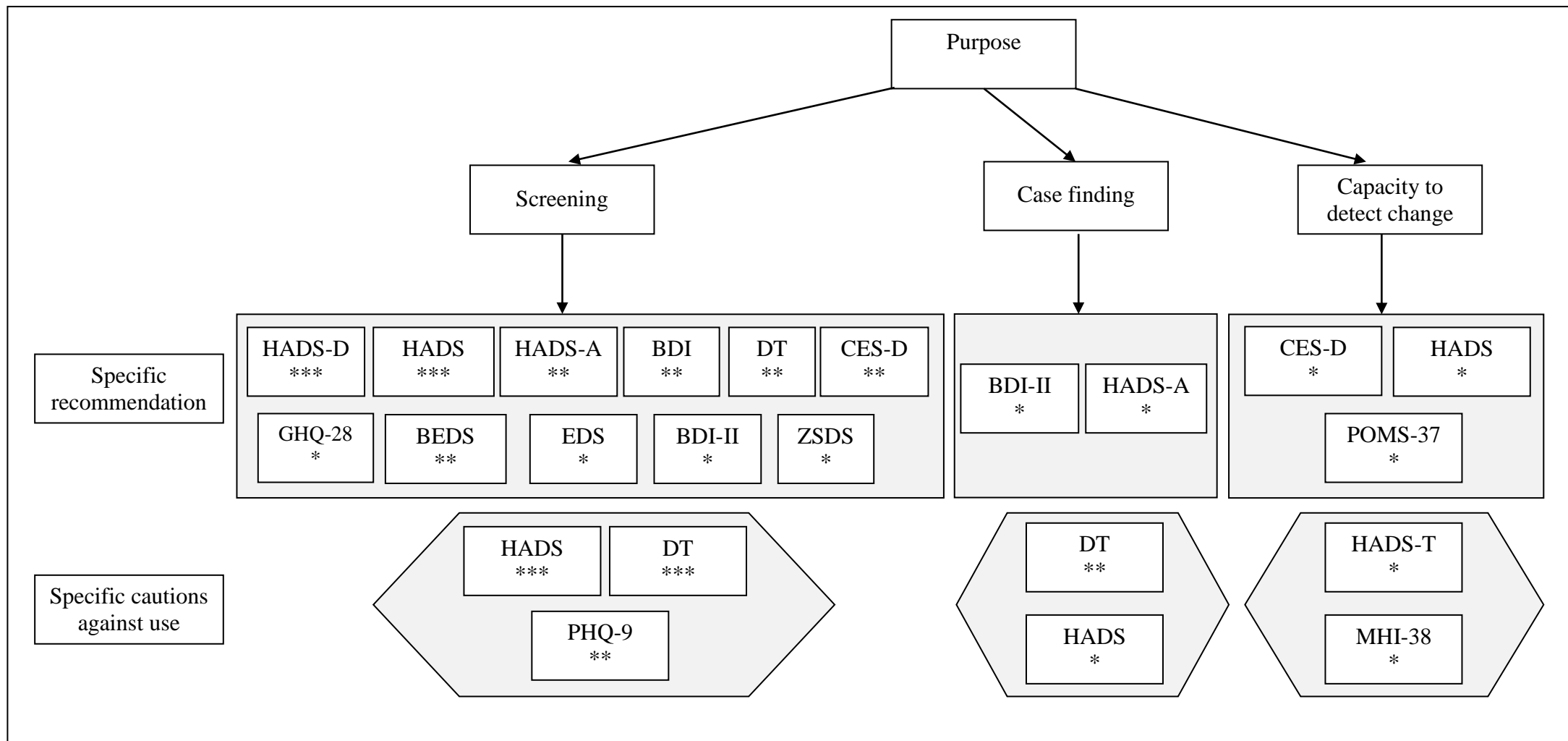
Figure 2: **Depression measures specifically recommended or cautioned against (organized by goal)**
*recommended/cautioned against by one review; **recommended/cautioned against by two reviews; ***recommended/cautioned against by three or more reviews.
HADS-D= Hospital Anxiety and Depression Scale-Depression subscale. HADS=Hospital Anxiety and Depression Scale. HADS-A= Hospital Anxiety and Depression Scale-Anxiety subscale. BDI=Beck Depression Inventory. DT=Distress Thermometer. CES-D=Center of Epidemiologic Studies-Depression. GHQ-28=General Health Questionnaire-28 items. BEDS=Brief Edinburgh Depression Scale. EDS=Edinburgh Depression Scale. BDI-II=Beck Depression Inventory-II. ZSDS=Zung Self-rating Depression Scale. POMS-37=Profile of Mood States-37 items. PHQ-9=Patient Health Questionnaire-9 items. MHI-38=Mental Health Inventoru-38 items.
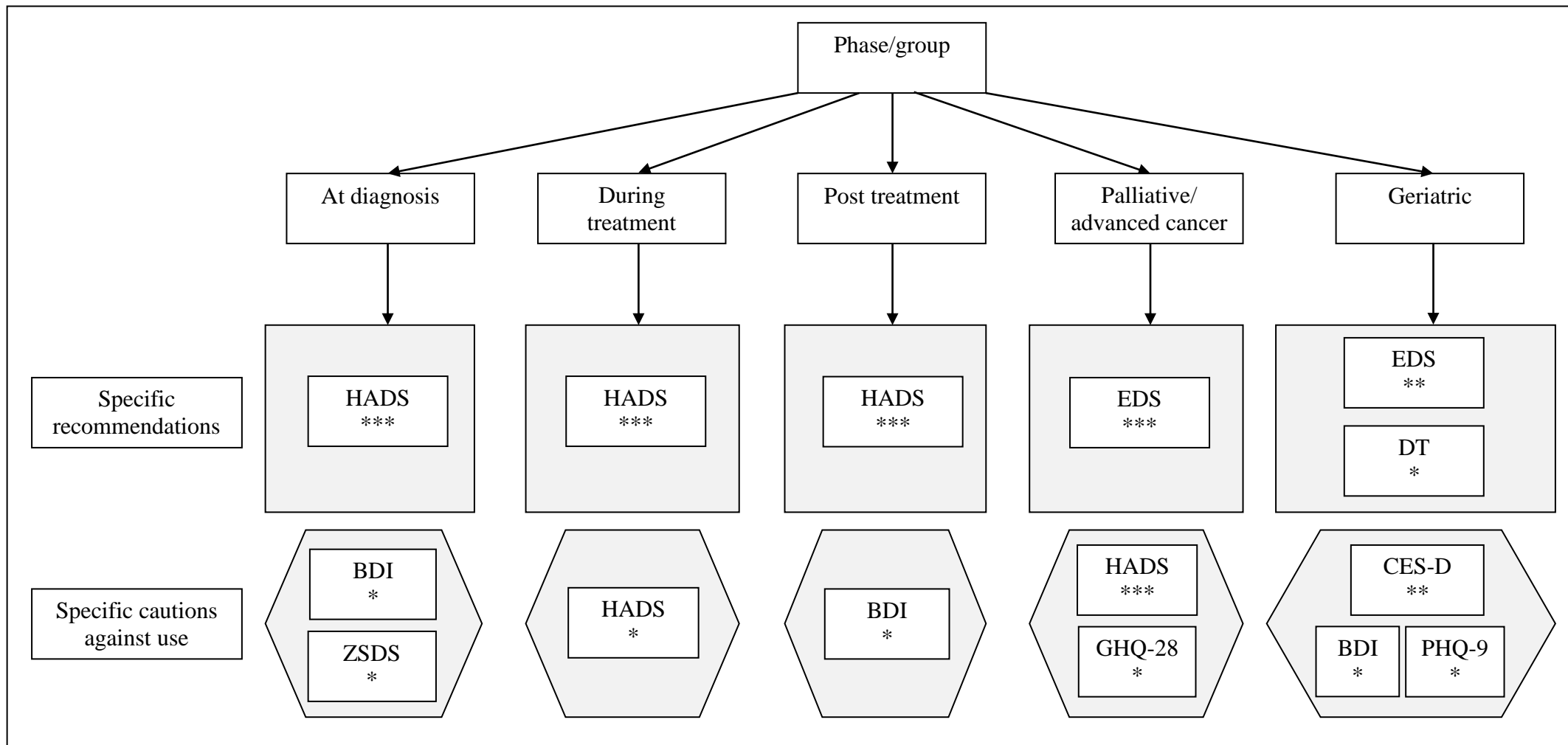
Figure 3: **Depression measures specifically recommended by, or cautioned against use (organized by target treatment phase/patient group)**
*recommended/cautioned against by one review; **recommended/cautioned against by two reviews; ***recommended/cautioned against by three or more reviews.
HADS=Hospital Anxiety and Depression Scale. EDS=Edinburgh Depression Scale. DT=Distress Thermometer. BDI=Beck Depression Inventory. ZSDS=Zung Self-rating Depression Scale. GHQ-28=General Health Questionnaire-28 items. CES-D=Center of Epidemiologic Studies-Depression. PHQ-9=Patient Health Questionnaire-9 items

First author (year), country	Type of review[+]	Population	Depression measures reviewed[+]	Main aims of review[§]	Description	PRISMA criteria[d]
Carey et al (2012)[19] Australia	Systematic	10 samples of cancer patients, mixed diagnoses, Europe and Asia	HADS	Screening	Describes the findings and quality of ten HADS validation studies against the SCID. Compares validated cut points with current cut points used in prevalence studies.	20
Carlson and Bultz (2003)[3] United States	Narrative	Narrative review regarding all cancer patients	BDI CES-D ZSDS	General recommendation	Provides general evaluation of commonly used measures, plus editorial comments. Highlights three depression specific measures. Other measures (eg, HADS) not evaluated for depression specifically.	-
Donovan et al (2014)[45] United States	Systematic	Mixed cancer diagnoses and treatment stages	DT	Screening	Summarizes the translations and validations of the DT in cancer patients worldwide. Assesses the translated versions' ability to detect clinically significant distress.	21
King et al (2005)[11] United States	Narrative	Narrative review of depression assessment in terminally ill older patients	BDI CES-D EDS GDS HADS MEQ	General recommendation	Suggests all six measures have 'some usefulness'. GDS highlighted as having fewer somatic items. No instrument is clearly preferred for use at end-of-life.	-
Krebber et al (2014)[22] Netherlands	Systematic	211 samples of cancer populations, mixed diagnoses and treatment stages	BDI BSI-18/53 CES-D HADS-D	Screening	Assesss prevalence of depression in cancer patients assessed by diagnostic interview and patient-report measures, and examines differences in prevalence between measures, type of cancer and treatment stage.	24
Lloyd-Williams (2001)[71] United Kingdom	Narrative	Narrative review of depression in palliative care	*Top 6 of 8* BDI EDS GHQ HADS MEQ ZSDS	General recommendation	Provides a review of frequently used screening tools and highlights their use in the detection of depression in palliative care. Suggests that BDI, ZSDS, HADS and EPDS are most appropriate as a screening in palliative care.	-
Love (2004)[72] Australia	Narrative	Women with breast cancer	*Top 6 of 36* BDI-SF BSI-18 BSI-5 GHQ-12 GHQ-30 HADS	General recommendation	Review 36 measures of depression or general distress. Recommends the BDI-SF as the most appropriate screening tool to detect depression in a clinical setting, as well as several other measures which are suitable for varying applications.	-

First author (year), country	Type of review [*]	Population	Depression measures reviewed [+]	Main aims of review [‡]	Description	PRISMA criteria[§]
Luckett et al (2010)[5] Australia	Systematic	42 samples of patients receiving treatment, mixed diagnoses	*Top 6 of 16* BSI-18/53 CES-D DT HADS POMS-37/65 SCL-90-R	Sensitivity to change	Reviews measures to assess distress (including depression) in RCTs of psycho-oncology interventions published between 1999-2009. Measures were filtered for initial suitability and then evaluated for reliability, validity and capacity to detect intervention effects.	20
Mitchell (2010)[43] United Kingdom	Systematic	Cancer and palliative populations	45 short tools, many depression focused, including DT, HADS and PDI	Screening Case-finding	Examines the merits of short and ultra-short measures (1-14 items) which have been validated against defined distress in cancer/palliative care, and compares single and multiple administrations.	22
Mitchell et al (2010)[44] United Kingdom	Systematic	24 samples of cancer and palliative populations, mixed diagnoses	24 studies testing HADS against clinical interview.	Screening Case-finding	Examines the diagnostic validity of the HADS-D, HADS-A and HADS-T in identifying depression, anxiety and distress in cancer. Highlights the sensitivity and specificity of subscales, real world comparison of methods and application of each subscale.	22
Mitchell et al (2012)[7] United Kingdom	Systematic	33 samples of cancer and palliative populations, mixed diagnoses	BDI-II DT EDS HADS-A HADS-D HADS-T	Screening Case-finding	Reviews the validity of screening and case-finding depression measures in cancer using a standardized rating system for validity and acceptability. All studies assessed for quality.	26
Morse et al (2005)[46] United Kingdom	Systematic	10 samples of cancer/palliative populations, mixed diagnoses/stages	HADS	Screening	Examines the sensitivity and specificity of the HADS-D when compared with a clinical interview, from 10 primary studies. Also compares cut-points and prevalence between studies.	13
Wilson et al (2000)[15] United States	Narrative	20 samples of advanced cancer/palliative cancer patients, mixed diagnoses.	*Top 6 of 7* BDI-SF DT EDS HADS RSCL VAS	General recommendation	Provides an overview of depression assessment (diagnostics interviews and patient-report measures) in palliative care in cancer, and the discusses challenges in the assessment of depression in palliative care.	-

(Continues on next page)

27

First author (year), country	Type of review [*]	Population	Depression measures reviewed [+]	Main aims of review [‡]	Description	PRISMA criteria[§]
Ziegler et al (2011)[73] United Kingdom	Systematic	22 samples of cancer populations across the trajectory (before, during and after treatment, palliative)	*Top 2 for each time-point* Pre treatment/ BDI, HADS During treatment HADS, MHI-5 Post treatment BDI, HADS Palliative care BDI, BEDS	Screening	Assesses the performance of measures across the cancer trajectory, and the feasibility of measures for use within palliative populations. 48 measures captured assessing psychological distress, with only 18 of these reporting trajectory stage.	21

Note: Only patient-reported outcome measures in survey/questionnaire format were included. 'HADS' has been used to refer to all subscales (HADS-D, HADS-A, HADS-T)

*Reviews were classified as either a narrative or systematic review. [+]For studies that reviewed more than six measures, the top six recommended measures were included in the table in alphabetical order. [‡]Main aims of systematic reviews were categorized as evaluating measures for screening, case-finding, and/or detecting change in depression. [§]See appendix for line-by-line critical appraisal of each systematic review. HADS=Hospital Anxiety and Depression Scale. SCID=Structured Clinical Interview for DSM-IV. BDI=Beck Depression Inventory. CES-D=Center of Epidemiologic Studies-Depression. ZSDS=Zung Self-rating Depression Scale. DT=Distress Thermometer. EDS=Edinburgh Depression Scale. GDS=Geriatric Depression Scale. MEQ=Mood Evaluation Questionnaire. BSI-18/53=Brief Symptom Inventory-18/53 items. HADS-D=Hospital Anxiety and Depression Scale-Depression Subscale. GHQ=General Health Questionnaire. BDI-SF=Beck Depression Inventory-Short Form. BSI-5=Brief Symptom Inventory. GHQ-12=General Health Questionnaire-12. GHQ-30=General Health Questionnaire-30. POMS-37/65=Profile of Mood States. SCL-90-R=Symptom-Checklist-90-Revised. PDI=Psychological Distress Inventory. HADS-A=Hospital Anxiety Depression Scale-Anxiety subscale. HADS-T=Hospital Anxiety Depression Scale-Total. BDI-II=Beck Depression Inventory-II. RSCL=Rotterdam Symptom Checklist. VAS=Visual Analogue Scale. MHI-5=Mental Health Inventory-5 items. BEDS=Brief Edinburgh Depression Scale.

Table 1. **Summary of captured reviews of depression measures in cancer**

Author	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Total
Carey (2012)[19]		x	x	x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x		x	x		x	x	22
Luckett (2010)[5]	x	x	x	x		x	x		x	x	x		x	x		x	x	x		x	x		x	x	x	x		20
Mitchell (2010)[43]	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x		x		22
Mitchell (2010)[44]	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x		x	x	23
Mitchell (2012)[7]	x	x	x	x		x	x	x	x	x	x	x	x	x		x	x	x		x	x		x	x	x	x	x	23
Nelson (2010)[42]	x	x	x	x		x	x	x	x	x			x	x		x	x	x		x			x	x	x	x	x	20
Thekkumpurath (2008)[1]	x	x	x	x		x	x	x	x	x	x		x	x		x	x	x		x	x		x	x		x		20
Vodermaier (2009)[4]	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	26
Ziegler (2011)[73]	x	x	x	x		x	x		x	x	x		x	x		x	x	x		x	x		x	x	x	x	x	21
Morse (2005)[46]		x	x	x		x	x		x							x	x			x	x		x	x		x		13
Donavon (2014)[45]		x	x	x		x	x	x	x	x	x		x	x	x	x	x	x		x	x		x	x	x	x		21
Krebber (2014)[22]	x	x	x	x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x	24

Note: The 27-item checklist comprises the following summarized checkpoints (*'x' if the captured review met this criterion*): 1: Title: Identify as systematic review, meta-analysis, or both. 2: Structured summary: Provide a structured summary. 3: Rationale: Describe the rationale for the review. 4: Objectives: Provide an explicit statement of questions being addressed. 5: Protocol and registration details. 6: Eligibility criteria. 7: Information sources: Describe all information sources in the search and date last searched. 8: Search: Present full electronic search strategy for at least one database. 9: Study selection: State the process for selecting studies. 10: Data collection process: Describe method of data extraction. 11: Data items: List and define all variables for which data were sought. 12: Risk of bias in individual studies: Describe methods used for assessing risk of bias. 13: Summary measures: State the principal summary measures. 14: Synthesis of results: Describe the methods of handling data and combining results of studies. 15: Risk of bias across studies: Specify any assessment of risk of bias that may affect the cumulative evidence. 16: Additional analyses: Describe methods of additional analyses. 17: Study selection: Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage. 18: Study characteristics: For each study, present characteristics for which data were extracted. 19: Risk of bias within studies: Present data on risk of bias of each study. 20: Results of individual studies: Provide: summary data for each intervention group, effect estimates and confidence intervals. 21: Synthesis of results: Present results of each meta-analysis done. 22: Risk of bias across studies: Present results of any assessment. 23: Additional analysis: Give results of additional analyses, if done. 24: Summary of evidence: Summarize the main findings; consider their relevance to key groups. 25: Limitations: Discuss limitations at study and outcome level and at review-level. 26: Conclusions: Provide interpretation of results in the context of other evidence, and implications for future research. 27: Funding: Describe sources of funding.

Table 2. **Assessment of captured reviews, using the PRISMA 27-item checklist**

29

First author (year), country	Criterion measure/ reference standard	Reported sensitivity (ranges) +	Reported specificity (ranges) +	Other data	Findings/ recommendations	Conclusion/ comments
Carey et al (2012)[19] Australia	SCID	HADS threshold for major depression: *Median (standard deviation)* HADS-A≥7: 0·82 (0·17) HADS-A≥9: 0·87 (0·70-0·95)* HADS-A≥11: 0·62 (0·11) HADS-D≥5: 0·88 (NR) HADS-D≥7: 0·82 (0·17) HADS-D≥9: 0·86 (0·10) HADS-D≥10: 0·82 (NR) HADS-T≥15: 0·87 (0·70-0·95)* HADS-T≥17: 0·71 (0·07) HADS-T≥19: 0·76 (0·08)	HADS threshold for major depression *Median (standard deviation)* HADS-A≥7: 0·83 (0·12) HADS-A≥9: 0·83 (0·78-0·86)* HADS-A≥11: 0.84 (0.11) HADS-D≥5: 0·60 (NR) HADS-D≥7: 0·73 (NR) HADS-D≥9: 0·76 (0·10) HADS-D≥10: 0·95 (NR) HADS-T≥15:0.85 (0.81-0.89)* HADS-T≥17: 0.8 (0.05) HADS-T≥19: 0.86 (0.15)	NR	None of the validation studies met all five criteria for selection of a screening measure. Showed that the most commonly used cut points are not well supported by validation studies, with studies recommending cut points as varied as 5 to 10 for the HADS-D.	Recommend caution in use of HADS due to variability of thresholds across studies.
Donovan et al (2014)[45] United States	HADS, BSI-53, Psychiatric DSM diagnosis	*Median (range; standard deviation)* DT: 0·83 (0·50-1·00; 0·11)	*Median (range; standard deviation)* DT: 0·68 (0·28-0·98; 0·16)	Area under curve (ROC analysis) *Median (range)* DT: 0·79 (0·47-0·90)	There is a lack of agreement regarding cut-point for clinically significant distress. Most commonly identified cut-point was 4 (52% of studies use this score).	While the DT has widespread acceptance as a brief screening measure, different language versions should be used with caution due inconsistent cut-points.
Krebber et al (2014)[22] Netherlands	Robust DSM or ICD psychiatric diagnosis	NR	NR	Mean prevalence of depression 8-24% in cancer patients HADS-D≥8: Pooled mean=0·18 (0·16-0·20) HADS-D≥11: Pooled mean=0·07 (0·06-0·08) CES-D≥16: Pooled mean=0·24 (0·21-0·26)	Prevalence of depression varies significantly by measure used, type of cancer and treatment stage. Prevalence of major depression was higher when patient-report instruments were used compared with diagnostic interview.	Patient-report measures may overestimate the presence of depression. Caution needs to be taken at different treatment stages and for different types of cancer.

First author (year), country	Criterion measure/ reference standard	Reported sensitivity (ranges) [+]	Reported specificity (ranges) [+]	Other data	Findings/ recommendations	Conclusion/ comments
Mitchell (2010)[43] United Kingdom	Interview defined distress	Not reported for depression alone. *Weighted specificity (95% CI):* DT: 0·79 (0·70-0·86) HADS: 0·70 (0·56-0·83) PDI: 0·68 (0·54-0·79)	Not reported for depression alone *Weighted specificity (95% CI):* DT: 0·67 (0·60-0·74) HADS: 0·81 (0·73-0·87) PDI: 0·90 (0·82-0·95)	*Screening utility index* DT: 0·682 HADS: 0·662 PDI: 0·667	Out of >45 measures, these depression focused tools were categorized as having diagnostic accuracy/validity: BDI-SF, EDS, HSF, HSI, BEDS, BCFD, PHQ-1, PHQ-2, HADS-D. The HADS, DT, PDI and DT+IT had comparable accuracy.	Given comparable performance, measure choice can be guided by acceptability or cost-effectiveness. Although short and ultra-short measures are useful, they should not be relied on in isolation.
Mitchell et al (2010)[44] United Kingdom	Psychiatric diagnosis with DSM or ICD	*Weighted specificity (95% CI):* HADS-A: 0·81 (0·74-0·86) HADS-D: 0·72 (0·55-0·85) HADS-T: 0·82 (0·74-0·89)	*Weighted specificity (95% CI):* HADS-A: 0·78 (0·66-0·88) HADS-D: 0·83 (0·74-0·90) HADS-T: 0·77 (0·63-0·89)	*Screening utility index:* HADS-A: 0·75 (good) HADS-D: 0·79 (good) HADS-T: 0·74 (good)	HADS-T, HADS-D and HADS-A performed adequately as screeners. Clinical utility index for syndromal depression screening (UI-) was 'good' for all three. HADS appeared to perform marginally better in non-palliative settings.	HADS-T and HADS-A performed better than HADS-D, however considering clinical utility, all three subscales were equally recommended. HADS acceptability limited by its length, reverse worded items and mix of anxiety and depression symptoms.
Mitchell et al (2012)[7] United Kingdom	Psychiatric diagnosis with DSM or ICD	*Pooled weighted sensitivity across cancer populations (95% CI):* BDI-II: 0·84 (0·65-0·96) DT: 0·80 (0·76-0·85) EPDS: 0·67 (0·52-0·80) HADS-A: 0·77 (0·69-0·84) HADS-D: 0·67 (0·55-0·78) HADS-T: 0·76 (0·70-0·82)	*Pooled weighted specificity across cancer populations (95% CI):* BDI-II: 0·87 (0·83-0·91) DT: 0·71 (0·64-0·78) EPDS: 0·85 (0·79-0·90) HADS-A: 0·84 (0·72-0·93) HADS-D: 0·83 (0·76-0·90) HADS-T: 0·79 (0·60-0·94)	*Screening utility index, recommendation grade:* BDI-II: 0·82, C DT: 0·71, C EPDS: 0·65, C HADS-A: 0·71, C HADS-D: 0·65, C HADS-T: 0·69, C	BDI-II performed adequately (but had lower acceptability). In advanced cancer, two-stem questions rated highest. Across groups, two verbal questions graded highest. HADS-A had moderate performance (above some depression-specific measures).	Optimal tool for screening appears to be two stem questions. BDI-II appeared to be best performing paper questionnaire. Omitting somatic symptoms did not appear to improve performance.

(Continues on next page)

31

First author (year), country	Criterion measure/ reference standard	Reported sensitivity (ranges) [+]	Reported specificity (ranges) [+]	Other data	Findings/ recommendations	Conclusion/ comments
Nelson et al (2010)[42] United States	NR	NR	NR	BDI-II, CES-D, ZSDS have been most extensively tested in geriatrics (although CES-D assesses less common symptoms in the elderly). PHQ-9 and BDI-II have longer recall periods (2 weeks), which may be too long for geriatric patients.	Many measures are validated in geriatric adults and cancer patients; however, no psychometric information was identified for geriatric cancer patients. Little validation data available on GDS-15, however GDS-30 is more extensively tested.	CES-D may be the most appropriate for elderly cancer patients. Further validation of depression measures for elderly needed.
Thekkump- urath et al (2008)[74] United Kingdom	Psychiatric diagnosis with DSM, ICD or WHO	Top 6 of 10 *Median (standard deviation)* BDI-SF: 0·79 (NR) BEDS: 0·72 (NR) EDS: 0·71 (0·01) GHQ: NR HADS: 0·77 (0·04) MEQ: NR	Top 6 of 10 *Median (standard deviation)* BDI-SF: 0·71 (NR) BEDS: 0·83 (NR) EDS: 0·08 (0·04) GHQ: NR HADS: 0·69 (0·03) MEQ: NR	VAS performed poorly compared with other measures. EDS fairs well, but needs further evaluation. BEDS is promising.	HADS total score recommended as an indicator of general distress, but not in palliative care.	Few studies have examined the validity of measures against structured interview in palliative care. Unidimensional tools are appealing due to reduced respondent burden in palliative care.
Vodermaier et al (2004)[4] Canada	Variable- interviews and questionna ires	Top 6 of 33 *Median (range; standard deviation)* BDI: 0·89 (0·52-0·92; 0·19) BSI-18:0·94 (0·91-0·97; 0·04) CDQ: 0·96 (0·68-1·00; 0·15) CES-D: 1·0 (1·0-1·0; 0) GHQ-28: 0·84 (0·75-0·93; 0·13) HADS: 0·73 (0·16-1·0; 0·17)	Top 6 of 33 *Median (range; standard deviation)* BDI: 0·93 (0·9-1·0; 0·05) BSI-18: 0·89 (0·85-0·93; 0·06) CDQ: 0·9 (0·68-0·98; 0·13) CES-D: 0·82 (0·79-0·85; 0·04) GHQ-28: 0·92 (0·92-0·92; 0) HADS: 0·77 (0·4-1·0; 0·14)	*Overall judgement:* BDI=Excellent BSI-18=Good CDQ=Excellent CES-D=Excellent GHQ=Excellent HADS=Good	*Ultra short*: CDQ performed best. *Short*: CES-D (particularly CES-D negative affect scale) judged 'excellent'. HADS was adequate, especially before and after treatment. HADS cuts offs variable. *Long*: BDI and GHQ-28 met all evaluation criteria.	No ultra short paper measures were recommended. CES-D recommended as best short measure, followed by HADS. HADS-T might be better than subscales in non-psychiatric patients. BDI and GHQ-28 were best long measures.

Ziegler et al (2011)[73] United Kingdom	SCID	See paper for detailed sensitivity and specificity data across all time points and measures. **T1:** HADS recommended (English, Japanese, Italian). Subscale cut-points ranged between 4-8, and 11-19 for total score. BDI-SF: 'very good' sensitivity, acceptable specificity. ZDSD: poor sensitivity, good specificity. **T2:** HADS plus MHI-5 recommended to screen for clinically significant distress. GHQ-28 demonstrated excellent sensitivity and specificity for psychosocial morbidity, however not recommended as a preferred option due to lack of recent validation studies. **T3:** HADS recommended with adequate sensitivity and specificity. BDI adequate sensitivity, lower specificity. **T4:** BEDS recommended (cut-point: 6). BDI (cut-point: 4): good sensitivity, low specificity. HADS did not perform adequately.	HADS was most often used. Recommended use of HADS-T if used in palliative population. No single measure supported for use throughout trajectory. Feasibility and respondent burden also important considerations. Cost of HADS may need to be considered.	HADS recommended for depression pre and post treatment, especially given its feasibility and acceptability (however appropriate cut-point varied). During treatment, HADS best if administered with MHI-5. During palliation/relapse, BEDS performed best in identifying depression.

Note: 'HADS' has been used to refer to 'HADS-A', 'HADS-D' and 'HADS-T'

*where the threshold was found in only 1 study, the range from the study was reported where possible. †For studies that reviewed more than six measures, the top six recommended measures were included in the table in alphabetical order. T1=Pretreatment/diagnosis.T2=Active treatment. T3=Post treatment. T4=Palliative care

SCID=Structured clinical interview for DSM-IV. HADS=Hospital Anxiety and Depression Scale. HADS-A=Hospital Anxiety and Depression Scale -Anxiety subscale. HADS-D=Hospital Anxiety and Depression Scale -Depression subscale. HADS-T=Hospital Anxiety and Depression Scale – Total. BSI-53= Brief Symptom Inventory-53 items. DSM=Diagnostic and Statistical Manual. DT=Distress Thermometer. ICD=International Classification of Disease. NR=Not reported, or only one study examined the measure so not data available. CES-D=Centre for Epidemiologic Studies-Depression Scale. PDI=Psychological Distress Inventory. BDI-SF=Beck Depression Inventory-Short From. EDS=Edinburgh Depression Scale. HSF=Hornheide Short Form. HIS=Hornheide Screening Instrument. BEDS=Brief Edinburgh Depression Scale. BCFD=Brief Case-find for Depression. PHQ-1/2=Patient Health Questionnaire-1/2 items. DT+IT=Distress Thermometer with Impact Thermometer. UI-=Negative Utility Index. BDI-II= Beck's Depression inventory-II. EPDS=Edinburgh Postnatal Depression Scale. ZSDS=Zung Self-Rating Depression Scale. PHQ-9. GDS-15/30=Geriatric Depression Scale-15/30 items. WHO=World Health Organization, GHQ=General Health Questionnaire. MEQ=Mood Evaluation Questionnaire. BDI=Beck Depression Inventory. BSI-18=Brief Symptom Inventory-18 items. CDQ=Combination Depression Question. GHQ-28=General Health Questionnaire-28 items. MHI-5= Mental health inventory-5 items.

Table 3. **Measures for depression screening: Reviews meeting at least 20 PRISMA criteria**

Author [type of review], country	Criterion measure/ reference standard	Case-finding AUC (UI+)	Case-finding Grade of recommendation	Other data	Findings/recommendations	Conclusion/comments
Mitchell (2010)[43] United Kingdom	Semi-structured interview	DT: 0·643 HADS-T: 0·70 PDI: 0·787	NR	*See table 2*	Data on short/ultra-short measures is too in complete to make recommendation	Accuracy of short and ultra-short tools is modest, but these measures should not be solely relied on. Comparison of HADS, PDI, GHQ-12 and DT suggested similar screening capacity for general distress, thus choice of tool should be based on cost and acceptability.
Mitchell et al (2010)[44] United Kingdom	Robust psychiatric diagnosis according to the DSM or ICD	HADS-A: 0·28 HADS-D: 0·27 HADS-T: 0·29	HADS-A: Poor HADS-D: Poor HADS-T: Poor	*See table 2*	HADS-T, HADS-D and HADS-A performed poorly in case-finding. Clinical utility index for case finding (UI+) was 'poor' for the scales,	HADS-T, HADS-D and HADS-A are likely to be suitable for screening for depression in cancer (not case-finding).
Mitchell et al (2012)[7] United Kingdom	Robust psychiatric diagnosis according to the DSM, ICD or WHO	BDI-II: 0·78 DT: 0·66 EDS: 0·73 HADS-D: 0·72	BDI-II: C DT: C EDS: C HADS-D: C	*See table 2*	BDI-II has level 2 evidence, but with low acceptability). In non-palliative settings, one question approach was graded highest for case-finding.	Optimal tool for case-finding appears to be one or two verbal questions. BDI-II appeared to be best performing paper questionnaire. Optimal tool for screening and case-finding appears to be two stem questions. BDI-II appeared to be best performing paper questionnaire.

DT=Distress Thermometer. HADS-T=Hospital Anxiety and Depression Scale-Total. PDI=Psychological Distress Inventory. NR=Not reported, or only one study examined the measure so not data available. GHQ-12=General Health Questionnaire-12 items. DSM=Diagnostic and Statistical Manual. ICD=International Classification of Disease. HADS-A=Hospital Anxiety and Depression Scale-Anxiety subscale. HADS-D=Hospital Anxiety and Depression Scale-Depression subscale. UI+=Positive utility index. WHO=World Health Organization. BDI-II=Beck Depression Inventory-II. EDS=Edinburgh Depression Scale.

Table 4. **Detailed summary of reviews of measures for depression <u>case-finding</u> meeting a minimum of 20 criteria outlined in the PRISMA Statement**

Author (year), country	Weighted score for assessing depression (max 100)[a]	Depression effect sizes Median (standard deviation)[a]	Findings/recommendations	Conclusion/comments
Luckett et al (2010)[5] Australia	*Top 6 of 16* BSI-18/53: 50 CES-D: 55 DT: 40 HADS:77.5 POMS-37/65: 60/55 SCL-90-R: 47·5	*Top 6 of 16* BSI-18/53: -1·85 (*1 study only*) CES-D: -0·21 (0·34) DT: 0·32 (*1 study only*) HADS: -0·5 (0·27) POMS-37/65: -0·24 (0·37) SCL-90-R: -0·205 (0·05)	CES-D and HADS have frequently been used to detect change for RCTs, 21 and 20 times respectively for this review. ABS, BAI, BDI-Primary Care, DASS, DABS, GDS, GHQ-30/60, HDRS, MHI-18, PHQ-2, PHQ-9, PANAS, POMS-10/30/Bipolar, STAI and VAS were excluded in initial filter due to lack of evidence for psychometric properties in English-speaking cancer populations or due to somatic emphasis. BSI-18, GHQ-28, MHI-38, SCL-90 received low scores, and were not recommended. Highest scoring measures were HADS, POMS-37/65 and CES-D. HADS-D and POMS-37 criticized for emphasis on anhedonia.	HADS-D may be less suitable for patients with advanced cancer and for detecting minor depression. Recommend using HADS-D, HADS-A and HADS-T where mixed affective disorders are the outcome of interest. CES-D recommended if depression is the sole focus.

[a] For studies that reviewed more than six measures, the top six recommended measures were included in the table in alphabetical order

BSI-18/53=Brief Symptom Inventory-18/53 items. CES-D=Center of Epidemiologic Studies-Depression. DT=Distress Thermometer. HADS=Hospital Anxiety and Depression Scale. POMS-37/65=Profile of Mood States-37-65 items. SCL-90-R=Symptom Checklist-90-Revised. RCT=Randomized Controlled Trial. ABS=Affect Balance Scale. BAI=Beck Anxiety Inventory. BDI-Primary Care=Beck Depression Inventory-Primary Care. DASS=Depression Anxiety Stress Scale. DABS=Derogatis Affect Balance Scale. GDS=Geriatric Depression Scale. GHQ-30/60=General Health Questionnaire-30/60 items. HDRS=Hamilton Depression Rating Scale. MHI-18=Mental Health Inventory-18. PHQ-2=Patient Health Questionnaire-2 items. PHQ-9=Patient Health Questionnaire-9 items. PANAS=Positive and Negative Affect Scale. POMS-10/30/Bipolar=Profile of Mood States-10/30/Bipolar items. STAI=State-Trait Anxiety Inventory. VAS=Visual Analogue Scale.GHQ-28=General Health Questionnaire-28 items. MHI-38=Mental Health Inventory-38 items. SCL-90=Symptom Checklist-90 items. POMS-37/65=Profile of Mood states-37/65 items. HADS-D=Hospital Anxiety and Depression Scale-Depression Subscale. HADS-A=Hospital Anxiety and Depression Scale-Anxiety Subscale. HADS-T=Hospital Anxiety and Depression Scale-Total.

Table 5. **Detailed summary of reviews of measures used to detect change in interventions meeting a minimum of 20 criteria outlined in the PRISMA Statement.**

Author (year), country	Findings/recommendations	Conclusion/comments
Morse et al (2005)[46] United Kingdom	Found that the HADS-D had a large degree of variability in sensitivity and specificity across a range of cancer populations. HADS-D shows high sensitivity but much at the expense of false positives. Cut-points for the HADS-D scale ranged from 5 to 11. Results of the review suggest a threshold of 8 to achieve high sensitivity, and lowering to 5 for some cancer populations.	Although HADS-D appears to be quite useful, it is suggested that no single tool can be used across treatment and disease pathway. Screening methods should be adapted according to location of assessment, stage of cancer (eg, use of HADS is not recommended in palliative care) and diagnosis.

HADS-D=Hospital Anxiety and Depression Scale-Depression subscale. HADS=Hospital Anxiety Depression Scale

Table 6. **General findings of recommendation reviews meeting fewer than 20 criteria outlined in the PRISMA Statement.**

Author [type of review], country	Population	Depression measures reviewed[a]	General description	Findings/recommendations	Conclusion/comments
Carlson et al (2003)[3] United States	All cancer patients	*Top 6 of 12* BDI BSI-18/53 CES-D GHQ-12/20/28/30/60 HADS ZSDS	Provides general evaluation of the psychometric properties of commonly in use, plus editorial comments. Highlighted three depression specific measures. Other measures (eg, HADS) not evaluated for capacity to assess depression specifically.	CES-D described as 'excellent' in providing population-based assessment of depression. BDI described as 'good', focused purely on depression, but not highly responsive to change. GHQ is losing favour in psycho-oncology settings.	CES-D received highest qualitative rating of depression specific measures. Review highlights potential for electronic administration of measures.
King et al (2005)[11] United States	Terminally ill older patients	BDI-SF CES-D EDS GDS HADS MEQ	HADS, EDS, GDS, MEQ, BDI, CES-D have 'some usefulness'. GDS highlighted as having fewer somatic items.	The listed patient-report measures have 'some usefulness' in the older patient population, however they should be used cautiously. Existing studies found to be unable to differentiate between measures based on psychometric properties.	Consider augmenting assessment with clinical interview. Possibly use measures to quantify severity of symptoms and change over time, rather than for case identification.
Love (2004)[72] Australia	Women with breast cancer	*Top 6 of 17* BDI BDI-SF BSI-18 BSI-53 GHQ-12 GHQ-30	Provides a summary of patient-report screening tools used for psychological distress in women with breast cancer. Evaluated the psychometric properties and provided pros and cons of each measure.	Found BDI-SF most appropriate tool to use to detect depression in a clinical setting. BSI-18 and GHQ-12 also recommended, although GHQ-12 has not widely been used in cancer groups. Although CES-D has been widely used in cancer populations, it has limitations as a suitable choice for a screening tool. Similarly, HADS might not be suitable for screening for depression.	The most suitable screening tool will depend on the intended use. BDI-SF was considered the most suitable screening tool to detect depression in a clinical setting.
Pirl (2010)[20] United States	Mixed diagnoses: all individuals with cancer	BDI BDI-II BSI-18 CES-D HADS PHQ-9	Provides a narrative guide to the use of commonly used outcome measures in psycho-oncology.	BDI described as widely used with cancer patients, although it contains some somatic items which may affect scores. BDI-II described as having 'good performance', easy to use, quick to complete and less emphasis on somatic symptoms. CES-D described as growing in popularity, with good validity/reliability in specific cancer patients (eg, breast). HADS has least focus on somatic symptoms.	BDI, CES-D and HADS may all be reasonable depression screening measures in ambulatory cancer patients.

Author [type of review], country	Population	Depression measures reviewed[a]	General description	Findings/recommendations	Conclusion/comments
Trask (2004)[6] United States	Cancer patients of mixed diagnoses	*Top 6 of 10* BDI CES-D GDS HADS RSCL ZSDS	Review aimed to provide a summary of the available means of assessing depression in cancer, including patient-report measures.	HADS is most used and was explicitly developed for the medically unwell. Appropriate cut-points for HADS-T and HADS-D vary by study. If HADS is to be used, it should be followed by in-depth assessment. Large variability in sensitivity, specificity, misclassification rate and positive predictive value of each of the reviewed measures. HADS performs well in those who are disease free or have stable disease, but not with progressive cancer.	Clinical interview still considered gold standard diagnostic tool. Depression measures for children and the elderly needed to be considered separately. Clear need for further research.
Wilson et al (2000)[15] United States	Palliative cancer patients, mixed diagnoses.	*Top 6 of 9* BDI-SF DT EDS HADS PHQ-9	Provides an overview of depression assessment in palliative care in cancer, including criterion-based diagnostic systems (eg, DSM-IV) and patient-report scales.	Although many depression measures in the literature, all those reviewed have been developed or tested with medical populations. None of the questionnaires provide perfect concordance with structured diagnostic interviews, with HADS showing twice the rates of "definite depression" than that reported in studies of diagnostic interviews. HADs however has outperformed GHQ in two studies. Optimal HADS cut-points vary between studies. Measures as VAS, such as the DT may be preferable in palliative care, although caution should be made as 'distress' does not indicate 'clinical depression'.	Given the necessary trade-off between sensitivity and specificity for all measures, practical considerations also important (eg, brevity). Outcomes measures can be useful when clinical interviews not feasible, however they are unlikely to match the diagnostic efficiency of a brief interview. Suggested that patient-report measures be used only as indices of general distress.

Note: 'HADS' has been used to refer to 'HADS-A', 'HADS-D' and 'HADS-T.

[a] For studies that reviewed more than six measures, the top six recommended measures were included in the table in alphabetical order

BDI=Beck Depression Inventory. BSI-18/53=Brief Symptom Inventory-18/53 items. CES-D=Center of Epidemiologic Studies-Depression. GHQ-12/20/28/30/60=General Health Questionnaire-12/20/28/30/60 items. HADS=Hospital Anxiety and Depression Scale. ZSDS=Zung Self-rating Depression Scale. GHQ=General Health Questionnaire. BDI-SF=Beck Depression Inventory-Short Form. EDS=Edinburgh Depression Scale. GDS=Geriatric Depression Scale. MEQ=Mood Evaluation Questionnaire. BDI-II=Beck Depression Inventory-II. PHQ-9=Patient Health Questionnaire-9 items. RSCL=Rotterdam Symptom Checklist. HADS-T=Hospital Anxiety and Depression Scale-Total. HADS-D=Hospital Anxiety and Depression Scale-Depression subscale. DT=Distress Thermometer.

Table 7. **General findings of narrative reviews**

Measure	Reviewed by	Positive features	Negative features
BEDS	[1,8]	**Screening:** • Recommended for palliative care by two systematic reviews.[1,8]	
BDI	[3-5,8,43]	**Screening:** • Recommended by two narrative[3,20] and two systematic[4,43] reviews • 'Easy to use' and quick to complete,[20] well tested in geriatric patients[42] • Recommended for older, terminally ill in one narrative review[75] • Recommended at diagnosis in one systematic review.[8] • Described as more accurate than the VAS in one narrative review[15] **Case finding:** • Adequate performance in one review[43]	• Described as not highly responsive to change in narrative review[3] • Not recommended by systematic review due to low property scores[5] • Had lower acceptability scores than other short screening tools[43] and is long[4] • Recall period may be too long for geriatric patients[42] • Contains some somatic symptoms[20] • Sensitivity appeared to decrease after treatment and during palliative care[8] • Symptoms assessed not as well suited to geriatric patients.[42]
CES-D	[3,4,42]	**Screening:** • Recommended for use in two narrative[3,20] and two systematic[4,5] reviews. • Recommended for older, terminally ill in one narrative review[75] and well tested in geriatric patients[42] • Freely available and quick to complete[20]	
EDS	[1,75]	**Screening:** • Described as having 'some usefulness' in narrative review for older, terminally ill[75] • Described as 'fairing well' in one systematic review in comparison to other tools in palliative care[1]	
GHQ-28	[4]	**Screening:** • Judged as 'excellent' in one systematic review.[4]	• Is longer than other measures.[4]
HADS	[3-8,19,75]	**Screening:** • One of the most widely used scales,[6,8] allowing cross-study comparison. • Highest scoring measure in one systematic review,[5] judged as 'good' by another.[4] • Highlighted as performing adequately as a screening tool (HADS-T, D and A)[7] • HADS-As performed moderately well in comparison to other short screening tools.[43] • Recommended at diagnosis, and post-treatment.[8] • Described as having 'some usefulness' in narrative review for older, terminally ill[75]	• HADS-T, A and D performed poorly in case-finding.[7] • Most commonly used cut-point (8) not well supported in validation studies.[4,6,8,15,19] • Criticized for emphasis on anhedonia,[5] although has less focus on somatic symptoms than other measures.[20] • Less suitable for advanced cancer,[4,5] palliative care[7,8] and for detecting minor depression.[5] • Performs less well during active treatment (unless combined with the MHI-5)[8] • Cost may be a barrier.[8] • Should be followed with clinical interview.[6]
PHQ-9	[4,5,42]		• Judged as 'poor' due to low reliability, low criterion measure and low validity.[4]

			• Length of recall (2-weeks) is quite burdensome for geriatric population.[42] • Multiple items problematic for those undergoing active treatment for cancer of any type and stage.[5]
POMS-37		**Screening:** • High scoring measure in one systematic review.[5]	
ZSDS	3,8,42	**Screening:** • Described as 'good' in narrative review.[3] • One of the most tested in geriatric patients.[42] • Judged as having good specificity at diagnosis in one study.[8]	• Judged as having poor sensitivity at diagnosis in one study.[8]

Brief Edinburgh Depression Scale (BEDS); Beck Depression Inventory (BDI); Centre of Epidemiologic Studies-Depression (CES-D); Edinburgh Depression Scale (EPS); General Health Questionnaire-28 items (GHQ-28); Hospital Anxiety and Depression Scale (HADS); Hospital Anxiety and Depression Scale-Anxiety subscale (HADS-A); Hospital Anxiety and Depression Scale-Depression subscale (HADS-D); Hospital Anxiety and Depression Scale-Total (HADS-T); Mental Health Inventory-5 items (MHI-5); Patient Health Questionnaire-9 items (PHQ-9); Profile of Mood States-37 items (POMS-37); Zung Self-rating Depression Scale (ZSDS)

Table 8: **Summary of positive and negative features of candidate measures**