Assessing the quality of diagnostic test accuracy meta-analyses

Methodological Quality of Meta-Analyses of the Diagnostic Accuracy of Depression Screening Tools Was Suboptimal Based on AMSTAR

Danielle B. Rice[a.b]; Ian Shrier[b,c]; Lorie A. Kloda[d]; Andrea Benedetti[c,e,f]; Brett D. Thombs*[a-c,e,g-h]

[a]Department of Psychiatry, McGill University, Montreal, Canada; [b]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Canada; [c]Department of Epidemiology, Biostatistics, and Occupational Health, Montreal, Canada; [d]Library, McGill University, Montreal, Canada; [e]Department of Medicine, McGill University, Montreal, Canada; [f]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Canada; [g]Department of Educational and Counselling Psychology, McGill University, Montreal, Canada; [h]Department of Psychology, McGill University, Montreal, Canada; and [i]School of Nursing, McGill University, Montreal, Quebec, Canada.

*Corresponding Author

Address for correspondence: Brett D. Thombs, PhD.; Jewish General Hospital; 4333 Ste. Catherine Road; Montreal, Quebec, Canada, H3T 1E4; Tel (514) 340-8222 ext. 5112; Fax (514) 340-8124; Email: brett.thombs@mcgill.ca

Funding Support: There was no specific funding for this study. Ms. Rice was supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship. Dr. Thombs was supported by an Investigator Salary Award from the Arthritis Society. None of the funding

sources had involvement in the study design, data collection, analysis, interpretation of data, writing of the report or the decision to submit the article for publication.

Conflicts of Interest: None of the authors have conflicts of interest to declare.

ABSTRACT

Objective. To evaluate the quality of meta-analyses of depression screening tools using an AMSTAR tool adapted for diagnostic test accuracy reviews.

Study Design. We searched MEDLINE and PsycINFO from January 1, 2005 through October 31, 2014 for recent meta-analyses in any language on the diagnostic accuracy of depression screening tools. Two reviewers independently assessed methodological quality using the adapted AMSTAR tool.

Results. We identified 16 eligible meta-analyses. The majority provided a list of included studies (100%), included a comprehensive literature search (94%) and assessed risk of bias of included studies (69%). Meta-analyses less consistently reported included study characteristics (44%), listed excluded studies (31%), included non-published evidence (25%) and assessed selective cutoff reporting (25%). Meta-analyses rarely reported that duplicate study selection or data extraction occurred (13%), incorporated risk of bias findings into conclusions (13%), mentioned 'a priori' protocols (6%), or reported on conflicts of interest (0%) or funding sources (0%) of primary studies. Only 2 of 16 included meta-analyses complied with at least 7 of 14 adapted AMSTAR items.

Conclusions. The methodological quality of most meta-analyses of the diagnostic test accuracy of depression screening tools is suboptimal. Improving quality will reduce the risk of inaccurate estimates of accuracy and inappropriate inferences.

Keywords: AMSTAR, depression, screening, diagnostic test accuracy, meta-analyses, quality

Running Title: Assessing the quality of diagnostic test accuracy meta-analyses

Word Count: 3,532

What is New?

Key findings

- Using a 14-item version of the AMSTAR tool adapted to evaluate methodological quality of meta-analyses of the diagnostic accuracy of depression screening tools, we found that only two of the 16 meta-analyses were rated *yes* for at least 50% of the adapted AMSTAR items.
- Few studies addressed key items related to bias, including selective cutoff reporting and inclusion of risk of bias assessments in conclusions.

What this adds to what is known?

- Concerns have been raised that primary studies of the diagnostic accuracy of depression screening tools may exaggerate accuracy estimates, but the quality of meta-analyses of the diagnostic accuracy of depression screening tools had not been reviewed. Our study found that most of these meta-analyses are of suboptimal quality and do not adequately address concerns raised about the quality of primary studies.

What is the implication and what should change now?

- Consumers of research should cautiously interpret results of many existing meta-analyses that were of low quality, did not adequately address issues related to bias, and may present overly enthusiastic accuracy estimates.
- Future meta-analyses of the diagnostic test accuracy of depression screening tools should attend to all adapted AMSTAR items, especially those regarding ‘a priori’ design protocols, duplicate study selection, selective cut-off reporting, and incorporating risk of bias findings into conclusions.

1.1 INTRODUCTION

Major depression is a disabling mood disorder that is present in 5-10% of primary care patients, including 10-20% of patients with chronic medical conditions [1, 2]. There are effective interventions to reduce the burden of depression, but most patients with depression do not receive adequate mental health care [3, 4]. Routine depression screening, which involves using self-report depression symptom questionnaires to attempt to identify patients who may have depression, has been proposed as a way to improve depression identification and management, but is controversial, and recommendations on screening are inconsistent [5].

The United States Preventative Services Task Force (USPSTF) recommends depression screening in primary care, but only when integrated systems for assessment, referral, and follow-up are available [1]. The UK National Screening Committee and the Canadian Task Force on Preventative Health Care (CTFPHC) recommend against depression screening due to the lack of evidence that depression screening would improve depressive symptoms or reduce the number of patients with depression [6, 7]. In its 2013 guideline, the CTFPHC expressed specific concern that published studies of the diagnostic accuracy of depression screening tools may exaggerate accuracy estimates [7]. Numerous specialty medical societies recommend depression screening in inpatient and outpatient settings (e.g., cancer, diabetes, heart disease, stroke) [8-14], but these recommendations are not based on systematic evidence reviews.

Concerns have been raised about the quality of existing primary studies on depression screening tool accuracy. Many primary studies have been conducted in samples too small to provide precise estimates. As a result, cutoff scores identified as optimal vary dramatically across studies [15, 16]. Many of these studies, however, selectively report accuracy results from a data-driven optimal cutoff and a small range of alternative cutoffs around it, and the cutoffs for

which data are reported are not consistent across studies [17-19]. Another concern relates to the inclusion of patients already diagnosed or being treated for depression in these studies, even though these patients would not be screened in clinical practice. One review found that more than 95% of almost 200 primary studies on the diagnostic accuracy of depression screening tools included already diagnosed or treated patients. Including these already diagnosed patients would overestimate the ability of a tool to identify previously unidentified patients who would be detected by a screening tool [20].

High-quality systematic reviews and meta-analyses can highlight shortcomings in primary studies. They can also provide guidance on how to improve research in order to address important health care questions. However, this will only occur to the degree that systematic reviews and meta-analyses are conducted rigorously, reflecting current standards for evidence synthesis [21, 22]. No studies, however, have evaluated the quality of existing systematic reviews or meta-analyses of the diagnostic accuracy of depression screening tools.

The Assessing the Methodological Quality of Systematic Reviews (AMSTAR) tool is an 11-item checklist that was developed to assess the scientific quality and rigor of systematic reviews for treatment effects from randomized trials [23]. We adapted AMSTAR to assess systematic reviews and meta-analyses of diagnostic test accuracy studies. The primary objective of our study was to evaluate the quality of meta-analyses of the diagnostic accuracy of depression screening tools published in journals indexed in the MEDLINE and PsycINFO databases, using the adapted AMSTAR tool.

1.2 MATERIALS AND METHODS

1.2.1 Identification of Meta-Analyses on the Diagnostic Accuracy of Depression Screening Tools

We searched MEDLINE and PsycINFO (both on the OvidSP platform) from January 1, 2005 through October 31, 2014 for meta-analyses in any language on the diagnostic accuracy of depression screening tools. We restricted the search to this period in order to identify relatively recent meta-analyses. We adapted a search strategy originally designed to identify primary studies on the diagnostic accuracy of depression screening tools, which was developed by a medical librarian and peer-reviewed by another medical librarian [17], by adding search terms designed to restrict the results to meta-analyses. The strategy was then adapted for PsycINFO. A medical librarian adapted the meta-analysis search strategies and conducted the search. The complete search strategies used for MEDLINE and PsycINFO can be found in Appendix A.

We included publications of meta-analyses, but not systematic reviews without meta-analyses, in order to focus on commonly used depression screening tools, which are more likely to be evaluated in systematic reviews with meta-analyses. Eligible publications had to include one or more meta-analyses that: (1) included a documented systematic review of the literature using at least one electronic database; (2) statistically combined results from ≥ 2 primary studies; and (3) reported measures of diagnostic accuracy (e.g., sensitivity, specificity, diagnostic odds ratio) of one or more depression screening tools compared to a reference standard diagnosis of depression based on a clinical interview or validated diagnostic interview (e.g., Composite International Diagnostic Interview). Meta-analyses that only compared scores on one self-report screening tool to depression case classifications based on a cutoff from another self-report screening tool or based on chart records of depression status, but not a clinical or diagnostic interview, were excluded. We also excluded meta-analyses of only measurement properties of depression screening tools other than diagnostic accuracy (e.g., general validity, reliability) if they did not also include a meta-analysis of diagnostic accuracy. Publications that included meta-

analyses of the diagnostic accuracy of screening tools for depression and for other disorders, such as anxiety disorders, separately, were eligible for inclusion, but only results for screening for depression were considered.

Search results were initially downloaded into the citation management database RefWorks (RefWorks, RefWorks-COS, Bethesda, MD, USA), duplicates were removed and the unique records were transferred into the systematic review program DistillerSR (Evidence Partners, Ottawa, Canada). DistillerSR was used to identify duplicate citations and to track results of the review process. Two investigators independently reviewed citations for eligibility. If either reviewer deemed a citation potentially eligible based on a review of the title and abstract, we carried out a full-text review of the article. Any disagreement between reviewers after full-text evaluation was resolved by consensus, including consultation with an independent third reviewer if necessary.

1.2.2 Assessment of Methodological Quality

The methodological quality of the included papers was evaluated using an adapted version of the AMSTAR tool [23]. The original AMSTAR checklist was developed to facilitate the conduct of high-quality reviews of treatment effects from randomized trials, and to provide a valid, reliable, and usable instrument to help differentiate between the methodological quality of systematic reviews using an 11-item checklist [23]. The response options for each item of the original AMSTAR checklist are: *yes, no, can't answer* and *not applicable*. Although developed for systematic reviews of randomized trials, many of the items are applicable to other designs, including systematic reviews and meta-analysis of diagnostic test accuracy.

We adapted the original AMSTAR tool to ensure that items were applicable to diagnostic test accuracy studies (see Appendix B for details). The team that adapted the tool included

members with expertise in evidence synthesis (IS, BDT, AB, LAK), information sciences for evidence synthesis (LAK), diagnostic test accuracy of depression screening tools (BDT, AB) and statistical analysis for diagnostic test accuracy meta-analyses (AB). We also consulted outside experts and referred to the Cochrane Handbook for Diagnostic Test Accuracy Meta-Analyses. Each original AMSTAR item was reviewed by team members, who considered ease of coding and applicability to studies of diagnostic test accuracy, then either accepted the item as appropriate or edited the item to better reflect practices in the conduct of systematic reviews of diagnostic test accuracy. In addition, a coding manual was developed with specific criteria for *yes* and *no* ratings, along with additional coding notes.

The adapted tool included 14-items because three of the 11 items in the original AMSTAR tool were divided into two parts. The three items that were divided did not undergo any additional changes. Item 5 was originally, “Was a list of studies (included and excluded) provided?” and was adapted to items 5a “Was a list of included studies provided?” and 5b “Was a list of excluded studies provided?” Item 9, “Were the methods used to combine the findings of studies appropriate?”, which incorporated both the meta-analysis model and heterogeneity assessment was divided into 9a (appropriate methods to combine studies) and 9b (heterogeneity appropriately assessed). Item 11 on conflicts of interest was revised to reflect funding of the review and primary studies (11a) and other potential conflicts of interest (11b). There were an additional five items that were unaltered (1:‘a priori’ design, 2: duplicate study selection, 3: comprehensive literature search, 4: publication status, 6: characteristics of included studies). Two items were only slightly modified in wording to incorporate the concept of risk of bias assessment, in addition to quality (7: scientific quality or risk of bias assessed, 8: scientific quality or risk of bias used in conclusions). One item, item 10 was altered more substantially.

The original AMSTAR item referred to publication bias. For increased relevancy to diagnostic test accuracy, this item was revised to consider selective cutoff reporting in the primary studies included in each meta-analyses, which is an important concern in studies of diagnostic test accuracy [17-19]. The adapted AMSTAR for systematic reviews of diagnostic test accuracy items included scoring response options of *yes* and *no*, which were coded based on information reported in the published article and associated supplementary material, where applicable. A coding manual was developed to help clarify scoring and eliminate the use of the *can't answer* response.

1.2.3 Data Extraction

One investigator independently extracted data from each included meta-analysis publication into a standardized database. For each meta-analysis publication, we extracted author, year of publication, journal, and journal impact factor for 2014.

For publications that included meta-analyses of diagnostic accuracy and other measurement characteristics, only results relevant to diagnostic accuracy were extracted.

Two coders independently rated each included meta-analysis using the adapted AMSTAR tool. Disagreements between reviewers were discussed and resolved by consensus after consultation with an independent third reviewer, as necessary. When there was difficulty determining whether a meta-analysis met criteria for a *yes* coding on any item, the adapted item was discussed by three team members and revised for better clarity, as necessary.

1.3 RESULTS

The electronic database search yielded 1296 unique title and abstracts for review. Of these, 1273 were excluded after title and abstract review because they did not report results from a meta-analysis or because the study was not related to the diagnostic accuracy of a depression

screening tool. Of the 23 articles that underwent full-text review, 7 were excluded (see Appendix C), resulting in 16 eligible meta-analyses [18, 19, 24-37] (see Figure 1). Characteristics of included meta-analyses are shown in Table 1.

As shown in Table 2, of the 14 adapted AMSTAR items, there were five for which at least half of the 16 meta-analyses received a *yes* rating, including items 5a (list of included studies; 100%), 3 (comprehensive literature search; 94%), 7 (scientific quality or risk of bias assessed; 69%), 9a (appropriate methods to combine studies; 56%) and 9b (heterogeneity appropriately assessed; 56%). Four items, received a *yes* rating for between 25% and 45% of meta-analyses, including items 6 (characteristics of included studies; 44%), 5b (list of excluded studies; 31%), 4 (publication status; 25%) and 10 (selective cutoff reporting bias assessed; 25%). Very few meta-analyses fulfilled criteria for a rating of *yes* for items 2 (duplicate study selection; 13%), 8 (scientific quality or risk of bias used in conclusions; 13%), 1 (‘a priori’ design; 6%), 11a (conflict of interest (funding); 0%) and 11b (other conflict of interest; 0%).

When considering item ratings for each meta-analysis, two of the 16 meta-analyses received a *yes* rating for 8 [25] and 10 [18] of the 14 adapted AMSTAR items. The other meta-analyses received *yes* ratings on between 3 and 6 of the 14 items (see Table 2).

1.4 DISCUSSION

The main findings of this study were that (1) few items on an adapted AMSTAR tool were met consistently by existing meta-analyses of the diagnostic test accuracy of depression screening tools published in journals indexed in MEDLINE and PsycINFO and (2) the overall quality of these meta-analyses was suboptimal, with only 2 of 16 meta-analyses rating *yes* for at least half of the adapted AMSTAR items.

Among meta-analyses reviewed in the present study, almost all met criteria for performing a comprehensive literature search (94%) and providing a list of included studies (100%). A majority of meta-analyses also used appropriate methods to combine results (56%), assessed heterogeneity (56%), and assessed the risk of bias within included primary studies (69%). On the other hand, only 4 of 16 meta-analyses received *yes* ratings for searching for unpublished studies and assessing the presence of possible selective cutoff reporting bias. Criteria for other adapted AMSTAR items were rarely met, including items related to the existence of an 'a priori' design for the meta-analysis (6%), duplicate study selection (13%), incorporating study quality or risk of bias ratings into conclusions (13%); none of the meta-analyses reported conflict of interest related to the funding of primary studies or other conflicts of interest.

Of the 16 meta-analyses reviewed, only 2 had ratings of *yes* for more than half of all quality items, suggesting that there is room for substantial improvement in the methods and reporting of meta-analyses that evaluate the diagnostic accuracy of depression screening tools. Of particular concern, very few meta-analyses incorporated ratings of quality or risk of bias into conclusions for users of meta-analysis results. Thus, even when important deficiencies were identified in primary studies, this was rarely considered in meta-analysis conclusions.

Similarly, possible bias from selective cutoff reporting was infrequently addressed. Selective cutoff reporting refers to the practice of determining which accuracy results to report based on the relative accuracy of the cutoffs, instead of reporting results for all potentially relevant cutoffs. Since the cutoffs that are most accurate for a screening tool can vary dramatically across studies [16, 38], meta-analyses are sometimes left with incomplete datasets that tend to exclude low estimates of accuracy [16, 18, 19]. In meta-analyses that estimate accuracy for a range of relevant cutoffs, this can lead to paradoxical findings that would be mathematically impossible if

all data were available, such as estimates of sensitivity that increase as cutoff severity increases [17, 18]. In other meta-analyses, which have evaluated accuracy at only a single standard cutoff, this has resulted in the elimination of some primary studies when the the standard cutoff is not reported due to poor performance [24] or in the substitution of data from different, better-performing cutoffs when results from the standard cutoff were not available [26, 27]. Only 4 of the 16 meta-analyses we reviewed mentioned the possibility of selective cutoff reporting bias, and none described the problem in detail or adjusted for it in any way.

Although many studies have used the AMSTAR tool to review the quality of meta-analyses of interventions, we identified only one other study that has applied AMSTAR to systematic reviews of diagnostic test accuracy. That study evaluated 24 systematic reviews of diagnostic tools used for Alzheimer's disease dementia and other dementias [39], and, in that study, only 3 of 24 included systematic reviews received a rating of *yes* for at least half of the AMSTAR items, whereas the other 21 received *yes* ratings for between 2 and 5 of the 11 items from the original AMSTAR tool. The authors reported that few systematic reviews reported 'a priori' protocols, incorporated risk of bias into the conclusions, accurately pooled primary data, or assessed publication bias appropriately.

Compliance with several of the AMSTAR items was low in both the Alzheimer's and dementia review [39] and the present study, including items on 'a priori' protocols and the incorporation of risk of bias into the study conclusion. For the item that reflected having an 'a priori' protocol with pre-specified methodological plans, 1 of 24 systematic reviews in the Alzheimer's and dementia review and 1 of 16 meta-analyses in the present study were rated *yes*. For a rating of *yes* in the present study, our adapted AMSTAR tool required only a statement that a review and meta-analysis protocol had been established prior to initiation of the meta-analysis.

Ideally, all systematic reviews and meta-analyses would be registered in a publically accessible registry. Registration of systematic reviews in such a registry was not possible until 2011, however, and only 4 of the 16 meta-analyses reviewed in the present study could have registered 'a priori' [26, 29, 34, 35] based on the dates of their searches. Currently, systematic reviews and meta-analyses are easily registered in the PROSPERO international prospective register of systematic reviews [40, 41]. It is hoped that the availability of publically accessible protocols will help to reduce the likelihood of biased post hoc decisions and avoid unnecessary duplication of research efforts [41, 42].

For the AMSTAR item reflecting quality and risk of bias of included primary studies, 10 of 24 Alzheimer's and dementia systematic reviews and 11 of 16 depression screening tool meta-analyses assessed the quality and/or risk of bias in included studies in some manner. However, findings from these assessments were seldom incorporated into the conclusions of meta-analyses with only 8 of the 24 Alzheimer's and dementia systematic reviews and 2 of the 16 depression screening tool meta-analyses incorporating findings from quality or bias results into study conclusions. The failure to incorporate risk of bias considerations into conclusions in systematic reviews and meta-analyses of diagnostic test accuracy has been identified as a concern previously. A 2014 study [43] evaluated 63 diagnostic test accuracy systematic reviews found in MEDLINE or EMBASE that were published in 2012 in any journal. That study found that 93% used a quality or risk of bias tool to evaluate included primary studies, but only 9% linked quality or risk of bias assessments to conclusions in the abstract or full article [43].

The findings of the present study have important implications for consumers of research and for future research in this area. Researchers and clinicians interpreting findings should consider that accuracy estimates presented in many existing meta-analyses may be overly positive due to

missing data from primary studies where selective cutoff reporting occurs and the failure to incorporate quality and risk of bias findings into review conclusions, in particular. This suggests that when interpreting these findings, users should be more circumspect than authors of meta-analyses have in some cases suggested. Future research should aim to address adapted AMSTAR items when developing and reporting reviews. Specifically, authors of meta-analyses should attend to missing cutoff data from primary studies by considering methods such as contacting authors for additional data, or undertaking individual patient data meta-analyses, which involves gathering all data from original primary studies as opposed to using summary results from published reports. Individual patient data meta-analysis can address bias from the selective publication of well-performing cutoff thresholds from small studies since accuracy can be evaluated across all relevant cutoff scores [17]. Additionally, authors should carefully evaluate the quality and risk of bias of primary studies being included in evidence syntheses and should address limitations of this evidence in review conclusions.

Several limitations should be considered when interpreting the results of our study. First, since adjustments were made to our coding manual during the initial part of our meta-analysis scoring, we were unable to calculate an interrater agreement statistic. Thus, although the original AMSTAR tool has demonstrated good interrater agreement, this has not been established for the adapted AMSTAR tool that was applied in the present study. However, changes from the original AMSTAR tool were minor. Thus, it would seem unlikely that the adapted tool would perform substantively differently, although this should be tested. Second, our sample included a relatively small number of systematic reviews with meta-analyses that were indexed in MEDLINE and PsycINFO. It is not clear to what degree our findings would be applicable to systematic reviews without meta-analyses or to meta-analyses that were not

indexed in these two databases, if there are any on this topic. Third, the lack of clearly established best practice analysis methods for meta-analyses of diagnostic test accuracy made implementation of the AMSTAR tool less straightforward. For example heterogeneity is a common issue in diagnostic test accuracy reviews, but at present there is not general agreement on how heterogeneity should be evaluated in meta-analyses of diagnostic test accuracy [44]. In the case of heterogeneity, our adapted item was designed to generate a code of *yes* for any reasonable method of assessment that has not been identified as inappropriate based on expert sources, particularly the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [45].

In sum, the present study found that only 2 of 16 existing meta-analyses of the diagnostic accuracy of depression screening tools met even half of the adapted AMSTAR items related to methodological quality. Furthermore, there were a number of adapted AMSTAR items that were rarely met by the included meta-analyses, including items related to the presence of 'a priori' research protocols, duplicate study selection and data extraction, selective cutoff reporting, the incorporation of risk of bias assessments in study conclusions, and the reporting of conflicts of interest. Consumers of research should be aware that some of these limitations in current meta-analyses might result in overly enthusiastic accuracy estimates. Future systematic reviews with meta-analyses should improve upon methodological quality to more accurately and transparently synthesize evidence on depression screening tool accuracy.

REFERENCES

1. U.S. Preventive Services Task Force. Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151:784-92.
2. National Collaborating Center for Mental Health. The NICE guideline on the management and treatment of depression in adults (Updated edition). United Kingdom: National Institute for Health and Clinical Excellence; 2010.
3. Duhoux A, Fournier L, Nguyen CT, Roberge P, Beveridge R. Guideline concordance of treatment for depressive disorders in Canada. Soc Psychiatry Psychiatr Epidemiol 2009;44:385-92.
4. Duhoux A, Fournier L, Gauvin L, Roberge P. What is the association between quality of treatment for depression and patient outcomes? A cohort study of adults consulting in primary care. J Affect Disord 2013;151:265-74.
5. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? BMJ 2014;348:g1253.
6. Allaby M. Screening for Depression: A Report for the UK National Screening Committee (Revised report). United Kingdom: UK National Screening Committee; 2010.
7. Joffres M, Jaramillo A, Dickinson J, Lewin G, Pottie K, Shaw E, Connor Gorber S, Tonelli M. Canadian Task Force on Preventive Health Care: Recommendations on screening for depression in adults. CMAJ 2013;185:775–82.
8. Colquhoun DM, Bunker SJ, Clarke DM, Glozier N, Hare DL, Hickie IB, Tatoulis J, Thompson DR, Tofler GH, Wilson A, Branagan MG. Screening, referral and treatment for depression in patients with coronary heart disease. Med J Aust 2013;198:483–4.

9. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee of the Canadian Diabetes Advisory Board. Canadian Diabetes Association 2013 clinical practice guidelines for the prevention and management of diabetes in Canada. Can J Diabetes 2013;37:S1–S212
10. National Institute for Clinical Excellence. Guideline on cancer services: Improving supportive and palliative care for adults with cancer. 2004. Available at http://www.nice.org.uk/guidance/csgsp/resources/supportive-and-palliative-care-the-manual-2. Accessed on July 18, 2015.
11. National Comprehensive Cancer Network. Distress Management. NCCN Clinical Practice Guidelines in Oncology. 2008. Available at http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed on July 18, 2015.
12. Holland JC, Andersen B, Breitbart WS, Buchmann LO, Compas B, Deschields TL, et al. Distress management. J Natl Compr Canc Netw 2013;11:190-09.
13. Lichtman JH, Bigger JT Jr, Blumenthal JA, Frasure-Smith K, Kaufmann PG, Lésperance F, et al. Depression and coronary heart disease: recommendations for screening, referral, and treatment: a science advisory from the American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, Council on Clinical Cardiology, Council on Epidemiology and Prevention, and Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Psychiatric Association. Circulation 2008;118:1768-75.
14. Eskes GA, Lanctot KL, Herrmann N, Lindsay P, Bayley M, Bouvier L, et al. Canadian Stroke Best Practice Recommendations: Mood, cognition and fatigue following stroke practice guidelines, updates 2015. Int J Stroke 2015;[E pub ahead of print].

15. Thombs BD, de Jonge P, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ, et al. Depression screening and patient outcomes in cardiovascular care: a systematic review. JAMA 2008;300:2161-71.
16. Meijer A, Roseman M, Milette K, Coyne JC, Stefanek ME, Ziegelstein RC, et al. Depression screening and patient outcomes in cancer: a systematic review. PLoS One 2011;6:e27181.
17. Thombs BD, Benedetti A, Kloda LA, Levis B, Nicolau I, Cuijpers P, et al. The diagnostic accuracy of the Patient Health Questionnaire-2 (PHQ-2), Patient Health Questionnaire-8 (PHQ-8), and Patient Health Questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. Syst Rev 2014;3:124.
18. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012;184:E191-6
19. Brennan C, Worrall-Davies A, McMillan D, Gilbody S, House A. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. J Psychosom Res 2010;69:371-8.
20. Thombs BD, Arthurs E, El-Baalbaki G, Meijer A, Ziegelstein RC, Steele RJ. Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. BMJ 2011;343:d4825.
21. Higgins J, Altman D. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration; 2008.
22. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane

Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration; 2009.

23. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007;2:e1350.
24. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Int Med. 2007;22:1596-1602.
25. Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess 2009;13:7-230.
26. Meader N, Moe-Byrne T, Llewellyn A, Mitchell AJ. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. J Neurol Neurosurg Psychiatry 2014;85:198-206.
27. Meader N, Mitchell AJ, Chew-Graham C, Goldberg D, Rizzo M, Bird V, et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. Br J Gen Pract 2011;61:e808-20.
28. 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;126:335-48.
29. Mitchell AJ, Meader N, Davies E, Clover K, Carter GL, Loscalzo MJ, 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. J Affect Disord 2012;140:149-60.

30. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007;57:144-51.
31. Mitchell AJ, Bird V, Rizzo M, Meader N. Which version of the Geriatric Depression Scale is most useful in medical settings and nursing homes? Diagnostic validity meta-analysis. Am J Geriatr Psychiatry 2010;18:1066-77.
32. Mitchell AJ, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: a meta-analysis of GDS30 and GDS15. J Affect Disord 2010;125:10-17.
33. Mitchell AJ. Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. Br J Gen Pract 2008;98:1934-43.
34. Tsai AC, Scott JA, Hung KJ, Zhu JQ, Matthews LT, Psaros C, et al. Reliability and validity of instruments for assessing perinatal depression in African settings: systematic review and meta-analysis. PLoS One 2013;8:e82521.
35. Tsai AC. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. JAIDS 2014;66:503-11.
36. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29:388-95.
37. Vodermaier A, Millman RD. Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. Supp Care Cancer 2011;19:1899-1908.

38. Thombs BD, de Jonge P, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ et al. Depression screening and patient outcomes in cardiovascular care: a systematic review. JAMA 2008;300:2161-71

39. Avero-Rodriguez I, Segura O, Sola I, Bonfill X, Sanchez E, Alonso-Coello P. Diagnostic tools for Alzheimer's disease dementia and other dementias: an overview of diagnostic test accuracy (DTA) systematic reviews. BMC Neurol 2014;14:183.

40. PROSPERO - international prospective register of systematic reviews. http://www.crd.york.ac.uk/prospero/. Accessed on July 18, 2015.

41. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. Syst Rev 2012;1:4053-1-2.

42. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. PROSPERO at one year: An evaluation of its utility. Syst Rev 2013;2:4053-2-4.

43. Ochodo EA, van Enst WA, Naaktgeboren CA, de Groot JA, Hooft L, Moons KG, et al. Incorporating quality assessments of primary studies in the conclusions of diagnostic accuracy reviews: a cross-sectional study. BMC Med Res Methodol 2014;13:33

44. Deeks JJ, Macaskill P, Irwig L: The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.

45. Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11:Interpreting results and drawing conclusions. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 0.9. The Cochrane Collaboration; 2013. Available from: http://srdta.cochrane.org/. Accessed on July 18, 2015.

Figure 1. Flow Diagram of Selection of Meta-Analyses of the Diagnostic Accuracy of Depression Screening Tools

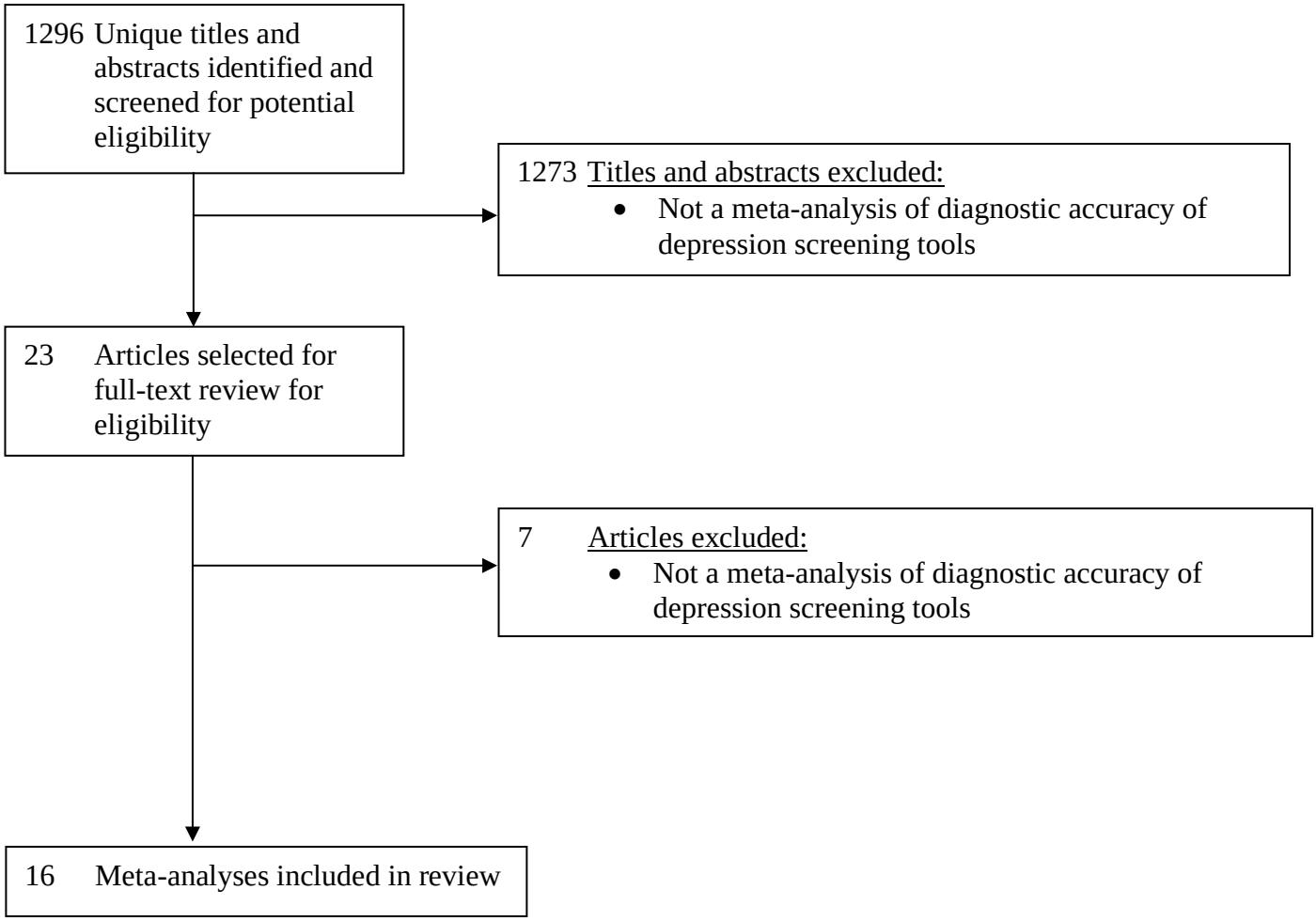

Table 1. Characteristics of Included Meta-Analyses

First Author, Year of Publication	Journal (2014 Impact Factor)	Focus of Meta-Analysis
Meader, 2014[26]	J Neurol Neurosurg Psychiatry (6.8)	Screening tools in poststroke patients
Tsai, 2014[35]	JAIDS (4.6)	Screening tools in HIV-positive adults in Africa
Tsai, 2013[34]	PLoS One (3.2)	Screening tools in pregnancy or postpartum in Africa
Mitchell, 2012[29]	J Affect Disord (3.4)	Screening tools in cancer patients
Manea, 2012[18]	CMAJ (6.0)	PHQ-9 in any setting
Meader, 2011[27]	Br J Gen Pract (2.3)	Screening tools in patients with chronic physical health problems
Vodermaier, 2011[37]	Support Care Cancer (2.4)	HADS in cancer patients
Brennan, 2010[19]	J Psychosom Res (2.7)	HADS in any setting
Mitchell, 2010a[31]	Am J Geriatr Psychiatry (4.2)	GDS in older patients
Mitchell, 2010b[28]	J Affect Disord (3.4)	HADS in cancer and palliative settings
Mitchell, 2010c[32]	J Affect Disord (3.4)	GDS in older primary care patients
Hewitt, 2009[25]	Health Technol Assess (5.0)	Screening tools in women in pregnancy or postpartum
Mitchell, 2008[33]	Br J Cancer (4.8)	One and two-question screening tools in cancer and palliative care
Gilbody, 2007[24]	J Gen Intern Med (3.4)	PHQ in medical settings
Mitchell, 2007[30]	Br J Gen Pract (2.3)	Ultra-short screening tools in primary care
Wittkampf, 2007[36]	Gen Hosp Psychiatry (2.6)	PHQ in any setting

GDS= Geriatric Depression Scale; HADS= Hospital Anxiety and Depression Scale; PHQ= Patient Health Questionnaire.

Table 2. Adapted AMSTAR Ratings

First Author, Year	Item 1	Item 2	Item 3	Item 4	Item 5A	Item 5B	Item 6	Item 7	Item 8	Item 9A	Item 9B	Item 10	Item 11A	Item 11B	Total 'Yes' (%)
Meader, 2014[26]	No	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	6 (43%)
Tsai, 2014[35]	No	Yes	Yes	No	Yes	No	No	Yes	No	Yes	Yes	No	No	No	6 (43%)
Tsai, 2013[34]	No	No	Yes	No	Yes	No	No	Yes	No	Yes	Yes	No	No	No	5 (36%)
Mitchell, 2012[18]	No	No	Yes	No	Yes	Yes	No	Yes	No	No	No	No	No	No	4 (29%)
Manea, 2012[18]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	10 (71%)
Meader, 2011[27]	Yes	No	Yes	No	Yes	No	No	Yes	No	Yes	No	No	No	No	5 (36%)
Vodermair, 2011[37]	No	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No	6 (43%)
Brennan, 2010[19]	No	No	Yes	No	Yes	No	No	No	No	Yes	Yes	Yes	No	No	5 (36%)
Mitchell, 2010a[31]	No	No	Yes	No	Yes	No	No	Yes	No	No	No	No	No	No	3 (21%)
Mitchell, 2010b[28]	No	No	No	No	Yes	Yes	No	Yes	No	No	No	No	No	No	3 (21%)
Mitchell, 2010c[32]	No	No	Yes	No	Yes	No	Yes	No	No	No	No	No	No	No	3 (21%)
Hewitt, 2009[25]	No	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No	8 (57%)
Mitchell, 2008[33]	No	No	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No	No	No	5 (36%)
Gilbody, 2007[24]	No	No	Yes	No	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No	6 (43%)
Mitchell, 2007[30]	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	4 (29%)
Witkampf, 2007[36]	No	No	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	6 (43%)
Total 'Yes' (%)	1 (6%)	2 (13%)	15 (94%)	4 (25%)	16 (100%)	5 (31%)	7 (44%)	11 (69%)	2 (13%)	9 (56%)	9 (56%)	4 (25%)	0 (0%)	0 (0%)	

Note: Item 1= 'a priori' design provided; Item 2= duplicate study selection; Item 3= comprehensive literature search; Item 4= publication status; Item 5A= list of included studies; Item 5B= list of excluded studies; Item 6= characteristics of included studies; Item 7= scientific quality or risk of bias assessed; Item 8= scientific quality or risk of bias used to in conclusions; Item 9A= appropriate methods to combine studies; Item 9B = heterogeneity appropriately assessed described; Item 10 = selective cutoff reporting bias assessed; Item 11A = conflict of interest (funding); Item 11B = other conflicts of interest.

Appendix A. Search Strategy

Ovid MEDLINE In-Process & Other Non-Indexed Citations and Ovid MEDLINE 1996 to October 31, 2014

1. Mass Screening/
2. Psychiatric Status Rating Scales/
3. "Predictive Value of Tests"/
4. "Reproducibility of Results"/
5. exp "Sensitivity and Specificity"/
6. Psychometrics/
7. Prevalence/
8. Reference Values/
9. Reference Standards/
10. exp Diagnostic Errors/
11. validation studies.pt.
12. comparative study.pt.
13. screen*.af.
14. prevalence.af.
15. predictive value*.af.
16. detect*.ti.
17. sensitiv*.ti.
18. valid*.ti.
19. revalid*.ti.
20. predict*.ti.
21. accura*.ti.
22. psychometric*.ti.
23. identif*.ti.
24. specificit*.ab.
25. cut?off*.ab.
26. cut* score*.ab.
27. cut?point*.ab.
28. threshold score*.ab.
29. reference standard*.ab.
30. reference test*.ab.
31. index test*.ab.
32. gold standard.ab.
33. or/1-32
34. Depression/
35. Depressive Disorder/
36. Depressive Disorder, Major/
37. Depressive Disorder, Postpartum/
38. depress*.tw.
39. or/34-38
40. Meta-Analysis/
41. meta-analysis as topic/
42. meta analysis.pt.

43. meta analy*.tw.
44. or/40-43
45. 33 and 39 and 44
46. limit 45 to yr="2005 -Current"

PsycINFO 2002 to October Week 5 2014

1. Diagnosis/
2. Medical Diagnosis/
3. Psychodiagnosis/
4. Misdiagnosis/
5. Screening/
6. Health Screening/
7. Screening Tests/
8. Prediction/
9. Cutting Scores/
10. Psychometrics/
11. Test Validity/
12. screen*.af.
13. predictive value*.af.
14. detect*.ti.
15. sensitiv*.ti.
16. valid*.ti.
17. revalid*.ti.
18. accura*.ti.
19. psychometric*.ti.
20. specificit*.ab.
21. cut?off*.ab.
22. cut* score*.ab.
23. cut?point*.ab.
24. threshold score*.ab.
25. reference standard*.ab.
26. reference test*.ab.
27. index test*.ab.
28. gold standard.ab.
29. or/1-28
30. major depression/
31. exp "Depression (Emotion)"/
32. postpartum depression/
33. depress*.tw.
34. or/30-33
35. meta analysis/
36. "1200".md.
37. meta analy*.tw.
38. or/35-37
39. 29 and 34 and 38
40. limit 39 to yr="2005 -Current"

Appendix B. Original and Adapted AMSTAR Tools

Original AMSTAR Item	Original Description	Adapted AMSTAR Item	Adapted Description	Yes	No	Notes
Item 1	Was an 'a priori' design provided?	**Item 1**	Was an 'a priori' design provided?	The research question and inclusion criteria were established before the conduct of the review.	It is not documented that the research question and inclusion criteria were established before the conduct of the review.	Need to refer to a protocol, ethics approval, or pre-determined/'a priori' published research objectives to score a "yes".
Item 2	Was there duplicate study selection and data extraction?	**Item 2**	Was there duplicate study selection and data extraction?	There were at least two individuals who did study selection and data extraction *independently*. A consensus procedure for disagreements was also in place.	There were fewer than two individuals who selected the studies or extracted data *independently* or there was no consensus procedure for disagreements.	Code "yes" if two people did study selection, 2 people did data extraction and there was a consensus process or one person checked the other's work. Articles must include all elements to be coded "Yes".
Item 3	Was a comprehensive literature search performed?	**Item 3**	Was a comprehensive literature search performed?	At least two electronic sources were searched with years and databases used (e.g., Central, EMBASE, and MEDLINE) specified. Searches should be supplemented by at least one	Fewer than two electronic source were searched, search strategies were not provided, or a supplementary search strategy was not used.	If at least 2 sources + one supplementary strategy used, code "yes". Cochrane CENTRAL and specialized registries count as two sources; a grey literature search counts as supplementary.

				supplementary method (e.g., consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, or by reviewing the references in the studies found).		
Item 4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	**Item 4**	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	The authors state that they searched for reports regardless of their publication type.	The authors do not state that they searched for reports regardless of their publication type or they indicate that reported were excluded based on publication status.	If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SINGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, authors must specify that they were searching for grey/unpublished literature.

Item 5	Was a list of studies (included and excluded) provided?	**Item 5a**	Was a list of included studies provided?	A list of all included studies was provided.	A list of all included studies was not provided.	Acceptable if the included studies are referenced with sufficient detail to locate the primary studies. If there is an electronic link to the list but the link is dead, select "no."
----------		**Item 5b**	Was a list of excluded studies provided?	A list of excluded studies was provided.	A list of excluded studies was not provided.	This does not refer to a list of all studies excluded at title/abstract or full-text review stages. Rather, this refers to a list of potentially relevant studies that is provided to clarify for readers why they were excluded. Code "yes" if there is any such list of excluded studies.
Item 6	Were the characteristics of the included studies provided?	**Item 6**	Were the characteristics of the included studies provided?	Data from the original studies is provided on characteristics of participants (e.g., age, sex, medical diagnoses), index test, reference standard, number of total participants and number of cases.	Data from the original studies on participant characteristics, index test, reference standard, number of total participants, or number of cases is not provided.	Acceptable if not in table format as long as all the details are provided. If any item is missed, the answer is "no". For key participant characteristics to be rated "yes", must include information on at least age, sex,

						and medical diagnosis, if appropriate.
Item 7	Was the scientific quality of the included studies assessed and documents?	**Item 7**	Was the scientific quality or risk of bias of the included studies assessed and documented?	'A priori' methods of assessment of risk of bias or quality (e.g., QUADAS-2) are provided and results are reported for each included study.	'A priori' methods of assessment of risk of bias or quality (e.g., QUADAS-2) are not provided or only a summary score for all studies combined is provided.	Can include use of a quality scoring tool or checklist (e.g., QUADAS-2) or a description of a set of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).
Item 8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	**Item 8**	Was the scientific quality or risk of bias of the included studies used appropriately in formulating conclusions?	The results of the methodological rigor and scientific quality are considered in the discussion or conclusions of the review or formulating recommendations.	The results of the methodological rigor and scientific quality are not considered in the discussion or conclusion of the review or in formulating recommendations.	For a code of "yes", must directly link quality to interpretation of accuracy results or implications for using test. Only mentioning study quality as a limitation with no other integration is coded "no." Cannot score "yes" for this question if scored

						"no" for question 7.
Item 9	Were the methods used to combine the findings of studies appropriate?	**Item 9a**	Were the methods used to combine the findings of studies appropriate?	Results were pooled using an appropriate model for meta-analyses of diagnostic test accuracy (e.g., hierarchical models, including HSROC and bivariate models).	Results for sensitivity and specificity were pooled separately on a univariate basis or were pooled using a model that does not provide estimates of heterogeneity between studies (e.g, Moses-Littenberg).	
----------		**Item 9b**	Was heterogeneity appropriately assessed and described?	Heterogeneity is addressed using appropriate graphical or statistical methods and described.	Heterogeneity is not addressed, not described, or is addressed using inappropriate methods.	Refer to the Cochrane Diagnostic Test Accuracy Handbook, Chapter 10. Univariate tests for heterogeneity in sensitivity and specificity and estimates of the I^2 statistic are not considered appropriate methods.
Item 10	Was the likelihood of publication bias assessed?	**Item 10**	Was the likelihood of selective cutoff reporting bias assessed?	An assessment is provided of whether the unavailability of diagnostic accuracy outcomes for some cutoffs in some studies may have influenced pooled results.	There is no assessment of whether diagnostic accuracy outcomes were available for all cutoffs considered in all studies, or whether unavailability of	To code "yes", for each cutoff score for which pooled accuracy results are analyzed, must indicate if all eligible studies reported results for that cutoff. If results were not

					outcomes for some cutoffs in some studies may have influenced pooled results.	available for all studies for cutoffs analyzed, must describe what was done (e.g. excluded studies, substituted results from another cutoff), and must discuss potential bias.
Item 11	Was the conflict of interest included?	**Item 11a**	Was the conflict of interest included (funding)?	Sources of support are clearly acknowledged in both the systematic review and each of the included studies.	Sources of support are either not clearly acknowledged for the systematic review or each of the included studies.	To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.
----------		**Item 11b**	Was a statement of other potential conflicts of interest included?	A statement of the present or absence of other conflicts of interest, including author-industry financial ties or employment, was provided for the systematic review and each of the included studies.	There was not a statement of the presence or absence of other conflicts of interest, including author-industry financial ties or employment, for either the systematic review or the included studies.	To get a "yes," must report presence or absence of other conflicts of interest for the systematic review and present, absence or non-reporting for the included studies.

Appendix C. List of Excluded Studies

1. Akena D, Joska J, Obuku EA, Amos T, Musisi S, Stein DJ. Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries: a systematic review. BMC Psychiatry 2012;12:187.
2. Farr SL, Dietz PM, Williams JR, Gibbs FA, Tregear S. Depression screening and treatment among non pregnant women of reproductive age in the United States, 1990-2010. Prev Chronic Dis 2011;8:A122.
3. Ziegler L, Hill K, Neilly L, Bennett MI, Higginson IJ, Murray SA, 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;41:619-36.
4. Mitchell AJ. Short screening tools for cancer-related distress: a review and diagnostic validity meta-analysis. J Natl Compr Canc Netw 2010;8:487-94.
5. Mirkhil S, Kent PM. The diagnostic accuracy of brief screening questions for psychosocial risk factors of poor outcome from an episode of pain: a systematic review. Clin J Pain 2009;4:340-8.
6. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess 2005;119:1-8.
7. Yirmiya R, Bab I. Major depression is a risk factor for low bone mineral density: a meta analysis. Biol Psychiatry 2009;66:423-32.