BMJ Open

Reporting Quality in Abstracts of Meta-Analyses of Depression Screening Tool Accuracy: A Review of Systematic Reviews and Meta-Analyses

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012867
Article Type:	Research
Date Submitted by the Author:	30-May-2016
Complete List of Authors:	Rice, Danielle; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Psychiatry Kloda, Lorie; Concordia University, Library Shrier, Ian; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Epidemiology, Biostatistics, and Occupational Health Thombs, Brett; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Psychiatry
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Mental health
Keywords:	Depression & mood disorders < PSYCHIATRY, PRISMA for Abstracts, diagnostic test accuracy, meta-analyses, screening

Assessing the quality of diagnostic test accuracy meta-analyses

Reporting Quality in Abstracts of Meta-Analyses of Depression Screening Tool Accuracy: A Review of Systematic Reviews and Meta-Analyses

Danielle B Rice[1,2], Lorie A. Kloda[3]; Ian Shrier[1,4] Brett D Thombs[1,2,4-8*]

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada

[2]Department of Psychiatry, McGill University, Montréal, Québec, Canada

[3]Library, McGill University, Montréal, Québec, Canada

[4]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Québec, Canada

[5]Department of Psychology, McGill University, Montréal, Québec, Canada

[6]Department of Medicine, McGill University, Montréal, Québec, Canada

[7]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada

[8]School of Nursing, McGill University, Montréal, Québec, Canada

Address for Correspondence: Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Saint Catherine Road; Montréal, Québec, Canada; H3T 1E4; Telephone: (514) 340-8222 ext. 5112; Fax: (514) 340-8124; Email: brett.thombs@mcgill.ca.

Keywords: PRISMA for Abstracts, depression, diagnostic test accuracy, meta-analyses, screening.

Word Count: 3,314

Assessing the quality of diagnostic test accuracy meta-analyses

ABSTRACT

Objective: Concerns have been raised regarding the quality and completeness of abstract reporting in evidence reviews, but this had not been evaluated in meta-analyses of the diagnostic accuracy of depression screening tools. Our objective was to evaluate reporting quality and completeness in abstracts of meta-analyses of depression screening tool accuracy, using the PRISMA for Abstracts tool.

Design: Cross-sectional study.

Inclusion Criteria: We searched MEDLINE and PsycINFO from January 1, 2005 through March 13, 2016 for recent meta-analyses in any language that compared a depression screening tool to a diagnosis based on clinical or validated diagnostic interview.

Data Extraction: Two reviewers independently assessed quality and completeness of abstract reporting using the PRISMA for Abstracts tool with appropriate adaptations made for studies of diagnostic test accuracy. Bivariate associations of number of PRISMA for Abstracts items complied with (1) journal abstract word limit and (2) AMSTAR scores of meta-analyses were also assessed.

Results: We identified 21 eligible meta-analyses. Only two of 21 included meta-analyses complied with at least half of adapted PRISMA for Abstracts items. The majority met criteria for reporting an appropriate title (95%), result interpretation (95%), and synthesis of results (76%). Meta-analyses less consistently reported databases searched (43%), associated search dates (33%) and strengths and limitations of evidence (19%). Most meta-analyses did not adequately report a clinically meaningful description of outcomes (14%), risk of bias (14%), included study characteristics (10%), study eligibility criteria (5%), registration information (5%), clear objectives (0%), report eligibility criteria (0%), or funding (0%). Overall meta-analyses quality scores were significantly associated with the number of PRISMA for Abstract scores items reported adequately (r = 0.45).

Assessing the quality of diagnostic test accuracy meta-analyses

Conclusions: Quality and completeness of reporting was found to be suboptimal. Journal editors should endorse PRISMA for Abstracts and allow for flexibility in abstract word counts to improve quality of abstracts.

STRENGTHS AND LIMITATIONS

- This is the first study to systematically evaluate the transparency and completeness of reporting in abstracts of meta-analyses of depression screening tools.
- Areas that require improvement were identified.
- Since there is not currently a PRISMA for Abstracts tool developed for reviews of diagnostic test accuracy, minor adaptations had to be made to the original tool.
- Our sample included a relatively small number of systematic reviews with meta-analyses.
- The lack of variability in the word limits of journal abstracts where included meta-analyses were published limited our ability to examine the association between PRISMA for Abstract ratings and abstract word limits.

INTRODUCTION

Researchers, clinicians and other consumers of research often rely primarily on information found in abstracts of systematic reviews.[1] Frequently, the abstract is the only part of an article that is read, making it the most frequently read part of biomedical articles after the title.[2] This may be due to time limitations, accessibility constraints, or language barriers.[2] For time-pressed readers or readers with limited access to a full-text article, the abstract must be able to stand alone in presenting a clear account of the methods, results, and conclusions that accurately reflect the core components of the full research report.[2] This goal, however, is infrequently achieved, as the quality and completeness of information provided in abstracts of systematic reviews are often suboptimal.[3-6]

The PRISMA for Abstracts tool was developed as an extension of the PRISMA statement,[2] with the goal of improving the quality and completeness of abstracts in systematic reviews, including meta-analyses.[2] The PRISMA for Abstracts checklist includes 12 items related to information that should be provided in systematic review abstracts, including title; objectives; eligibility criteria of included studies; information sources, including key databases and dates of searches; methods of assessing risk of bias; number and type of included studies; synthesis of results for main outcomes; description and direction of the effect; summary of strengths and limitations of evidence; general interpretation of results; source of funding; and registration number.

Only one previous study has used the PRISMA for Abstracts checklist to evaluate the quality and completeness of systematic review abstracts.[7] That study included 197 systematic review abstracts published in 2010 in the proceedings of nine leading international medical conferences that have conference abstracts that are searchable online. PubMed was then searched from 2010 to 2013 to identify subsequently published journal articles (N = 103).[7] In both published conference abstracts and published articles, 9 of the 12 PRISMA for Abstracts items were completed in less than

50% of abstracts reviewed. Poor reporting of abstracts has also been found in studies that have evaluated abstracts of meta-analyses and systematic reviews using other methods. We identified three studies, all from dentistry literature, that reviewed reporting in systematic reviews abstracts.[4-6] Two of the studies evaluated abstracts using a 16-item checklist derived from the full PRISMA statement, prior to the official PRISMA for Abstracts publication.[5, 6] The third study assessed abstract reporting based on the presence or absence of seven characteristics related to the meta-analyses results.[8] In all three studies, major deficiencies were identified.

Depression screening is an area where indirect evidence from diagnostic test accuracy studies has played an important role in policy and where the quality of reporting may be particularly important. Depression screening is controversial, and recommendations on screening are inconsistent.[9] Based on indirect evidence, including evidence on screening tool accuracy, the United States Preventative Services Task Force recently recommended universal depression screening in all adults.[10] Both the UK National Screening Committee and the Canadian Task Force on Preventative Health Care however, recommend against depression screening due to a lack of evidence from randomized controlled trials that depression screening would improve mental health outcomes.[11, 12]

No published studies have evaluated the completeness of reporting in abstracts of diagnostic test accuracy systematic reviews or meta-analyses of depression screening tool accuracy. The PRISMA for Abstracts guideline was developed for systematic reviews of interventions, and the authors suggested that modifications would be required to apply the checklist to DTA systematic reviews.[2] In the absence of a PRISMA for Abstracts tool designed for studies of DTA, we applied PRISMA for Abstracts with adaptations to some items in order to appropriately assess systematic reviews and meta-analyses of DTA studies. The primary objective of our study was to evaluate the

transparency and completeness of abstracts of meta-analyses of the diagnostic accuracy of depression screening tools that were published in journals indexed in the MEDLINE and PsycINFO databases, using PRISMA for Abstracts. Our secondary objective was to determine if the quality of the meta-analysis or the word count permitted by the journal of the meta-analyses were associated with PRISMA for Abstract scores.

METHODS

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 March 13, 2016 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,[13] 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 S1 Appendix.

We included publications of meta-analyses, but not systematic reviews without meta-analyses, in order to focus only 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). We excluded meta-analyses that did not use a clinical or diagnostic interview as the gold standard. 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 unique citation 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.

Assessment of reporting in abstracts

The reporting of abstracts was evaluated using a PRISMA for Abstracts tool, with some items adapted for applicability to studies of diagnostic test accuracy. The original PRISMA for Abstracts tool was developed to provide guidance on a minimum set of items necessary to provide a reasonably complete and transparent representation of a full article report.[2] The checklist was created to fit into headings mandated by journals and conference submissions, including title, background, methods, results, discussion and associated funding and registration information, but was designed with flexibility regarding the specific headings and where information should be listed. The PRISMA for Abstracts checklist was developed for systematic reviews of abstracts involving interventions, but

many of the items are applicable to other designs, including DTA systematic reviews and meta-analyses.

We adapted the original PRISMA for Abstracts tool to ensure that items were applicable to DTA studies. The team that adapted the PRISMA for Abstracts tool included members with expertise in evidence synthesis (IS, BT, LAK), information sciences for evidence synthesis (LAK) and DTA studies of depression screening tools (BDT). Each original PRISMA for Abstracts item was reviewed by team members, who considered ease of coding and applicability to DTA systematic reviews and meta-analyses, then either accepted the item as appropriate or edited the item to better reflect practices in the conduct of DTA systematic reviews. In addition, a coding manual was developed with specific criteria for *yes* and *no* ratings, along with additional coding notes (see S2 Appendix for details).

The adapted tool included 14 items because two of the original PRISMA for Abstracts items were divided into two parts. The two items that were divided did not undergo any additional changes. Item 3 was originally "Study and report characteristics used as criteria for inclusion" and was adapted to items 3a "Study characteristics used as inclusion criteria" and item 3b "Report characteristics used as inclusion criteria." Item 4, "Key databases searched and search dates", which involved reporting specific databases searched and the dates searched, was divided into 4a (key databases searched) and 4b (search dates). Of the original 12 items, seven were unaltered (1: title, 5: risk of bias, 6: included studies, 9: strengths and limitations of evidence, 10: interpretation, 11: funding, 12: registration). Three items (2: objectives, 7: synthesis of results, 8: description of effect) were slightly modified for applicability to DTA systematic review abstracts. The original item 2 refers to "the research question including components such as participants, interventions, comparators and outcomes". For increased relevance to DTA reviews, this item was revised to encompass the reference standard and index test

within the systematic review rather than the interventions and comparators found in intervention studies. Item 7 was adjusted to encompass results of the principle summary measures (e.g. sensitivity, specificity, positive predictive value, negative predictive value) that are reported in DTA studies. Lastly, the original item 8 refers to "the direction and size of the effect" and was adjusted to evaluate if the summary of accuracy estimates that are presented within DTA studies are presented in terms meaningful to clinicians.

Data extraction

For each meta-analysis publication, one investigator extracted author, year of publication, journal, journal impact factor for 2014, the abstract word limit of the journal where the meta-analysis was published, and previously published AMSTAR quality ratings.[14] Accuracy was verified by a second investigator. Two investigators independently rated each included meta-analysis using the adapted PRISMA for Abstracts checklist. 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. For publications that included meta-analyses of diagnostic accuracy and other measurement characteristics, only results relevant to diagnostic accuracy were extracted.

Statistical analyses

Bivariate associations between the (1) abstract word count permitted by the journal, and (2) AMSTAR scores of meta-analyses to the PRISMA for Abstracts scores were assessed with Pearson correlation coefficients. Analyses were conducted using SPSS version 22.0 (Chicago, IL), and statistical tests were two-sided with a $p < 0.05$ significance level. 95% confidence intervals (CI) were also calculated.

RESULTS

Article selection

The electronic database search yielded 1522 unique title and abstracts for review. Of these, 1492 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 30 articles that underwent full-text review, 9 were excluded because they were not meta-analyses of diagnostic accuracy of depression screening tools (see S3 Appendix), resulting in 21 eligible meta-analyses (see Figure 1).[15-35] Characteristics of included meta-analyses are shown in Table 1.

As shown in Table 2, of the 14 adapted PRISMA for Abstracts items, there were two items for which 20 of the 21 included meta-analyses received a *yes* rating: items 1 (title; 95%) and 10 (interpretation of results; 95%). One item received a *yes* rating in 16 of 21 meta-analyses (item 7, synthesis of results; 76%), and three items received a *yes* rating in 7 to 9 of 21 meta-analyses (33% to 43%): items 4a (databases searched), 4b (key search dates) and item 9 (strengths and limitations of evidence). Very few meta-analyses fulfilled criteria for a rating of *yes* for the remaining 8 items including item 8 (description of the outcomes; 14%), item 5 (risk of bias; 14%), item 6 (included studies; 10%), item 3a (eligibility criteria for study characteristics; 5%), item 12 (registration; 5%), item 2 (objectives; 0%), item 3b (eligibility criteria for report characteristics 0%), and item 11 (funding; 0%).

When considering item ratings for each meta-analysis, two of the 21 meta-analyses received a *yes* rating for 7 of the 14 adapted PRISMA for Abstracts.[15, 33] An additional seven meta-analyses received ratings of *yes* for 5[16, 17, 31, 34, 35] and 6 [18, 19] of the 14 PRISMA for Abstracts items.

The remaining 12 meta-analyses received *yes* ratings on between 2 and 4 of the 14 items (see Table 3).

Association of Journal Abstract Word Count and AMSTAR Scores with PRISMA For Abstract Scores

There was a significant positive association of AMSTAR scores with the number of *yes* ratings of PRISMA for Abstracts items (r =0.45, 95% CI = 0.02 to 0.74, p = 0.040). The abstract word count permitted by the journal was not significantly correlated to the PRISMA for Abstracts scores (r = -0.03, 95% CI = -0.45 to 0.41, p = 0.914). However, 20 out of 21 meta-analyses were published in journals that had word limits between 200 to 300 words.

DISCUSSION

The main findings of this study were that only 3 of 14 items from the adapted PRISMA for Abstracts tool received *yes* ratings in at least 50% of 21 meta-analyses of depression screening tools. The other 11 items were infrequently met. Furthermore, overall quality of reporting in the abstracts of the meta-analyses was poor, with only 2 of 21 meta-analyses rating *yes* for at least half of the PRISMA for Abstracts items. Overall quality ratings of the meta-analyses, based on AMSTAR, were associated with the number of PRISMA for Abstract items that were adequately reported.

Among meta-analyses evaluated in the present study, almost all met criteria for having a title that identified the report as systematic review or meta-analysis, for reporting the main results of the synthesis, and for providing a general interpretation of the results and important implications. In addition, 9 of 21 meta-analyses also provided a list of databases searched and 7 provided dates of coverage for the literature search and strengths and limitations of evidence. On the other hand, 3 or fewer meta-analyses received *yes* ratings for stating the methods used for assessing risk of bias, the number of included studies and participants, eligibility criteria for study characteristics, registration

information, and the description of summary estimates. No studies met criteria for the remaining 3 PRISMA for Abstracts items (complete study objectives, eligibility criteria for report characters, and funding information).

Beyond systematic reviews and meta-analyses, specific concerns have been raised about the quality of abstracts of primary studies of DTA. A 21-item tool was developed to assess whether abstracts of primary DTA studies are adequately informative, based on the reporting of essential methodological features and study results.[36] The tool was applied to a sample of 103 primary DTA studies published in 12 high-impact journals in 2012, and only 39 of the 103 primary studies that were evaluated received a rating of *adequate* for at least half of the items assessed. Specifically, the authors reported that fewer than 50% of included primary studies adequately reported the study population, setting, patient sampling, blinding, cutoffs used and confidence intervals around accuracy estimates.[36] The mean number of adequately reported items within abstracts was significantly lower for abstracts that had lower word counts.

Several authors have recommended that journal editors endorse abstract guidelines, such as the PRISMA for Abstracts tool, to help ensure that abstracts better address the needs of consumers of research,[2, 4, 7, 36] and, generally, journal endorsement of reporting guidelines improves the completeness of reporting.[37] The Consolidated Standards of Reporting Trials (CONSORT) reporting guidelines for abstracts of randomized controlled trials was published in 2009,[38] and a recent study found that journals that implement these guidelines have improved reporting in abstracts of randomized controlled trials.[39] As of April 6, 2016, only one of the journals where DTA meta-analyses included in the present study were published (Journal of General Internal Medicine) includes a statement specifically endorsing the PRISMA for Abstracts tool and a web link to the PRISMA for Abstracts tool in its author instructions. A second journal (Health Technology Assessments) required

authors to comply with general PRISMA guidelines in developing the abstract, but did not refer to the PRISMA for Abstracts statement or its items. No other journals mentioned PRISMA in relation to abstracts. All journals had word limits of between 200 and 300 words for abstracts with the exception of Health Technology Assessments, which allows 500 words. Health Technology Assessments is a UK National Institutes of Health Research journal that typically publishes extensive, multi-question systematic reviews. Currently, it is not likely to be feasible for authors to include all PRISMA for Abstracts recommended reporting items due to word count restraints typically imposed for biomedical journal abstracts. Thus, we recommend that journals endorse the use of the PRISMA for Abstracts checklist for formulating abstracts and that journals provide flexibility in word counts and the structure of abstract headings in order to comply with recommendations. This is already done in some journals (e.g., BMJ, PLOS Medicine).

As almost all of the meta-analyses that we evaluated were published prior to the development of the PRISMA for Abstracts tool, it could not have been expected that our sample of studies would have been able to follow the checklist when developing their abstracts. Our study provides baseline results representing DTA meta-analyses abstracts prior to the publication of PRISMA for Abstracts guidelines. This highlights areas where improvement is needed and will allow future studies to compare the reporting of abstracts after the PRISMA for Abstracts tool has been published and more widely endorsed.

Specific 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 for the adapted PRISMA for Abstracts items. 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, to meta-analyses on the diagnostic accuracy of depression screening tools that were not indexed in these two databases, or to meta-analyses of diagnostic accuracy in other conditions and other fields. Third, we reported results on an item-by-item basis for illustration purposes. Not all items, however, would be expected to influence the transparency and completeness of abstract reporting equally, and an evaluation of the quality of any given meta-analysis abstract would need to consider specific items individually. Finally, we adapted the PRISMA for Abstracts tool for this study, as it was developed for use in systematic reviews and meta-analyses of intervention studies. Ideally, however, a PRISMA for Abstracts tool would be developed specifically for reviews of diagnostic test accuracy. We also attempted to analyze the association between journal word limits and the PRISMA for abstract scores, however, 20 of 21 meta-analyses included in our study were published in journals with word limits of 200 to 300 words.

In conclusion, the present study found that only 2 of 21 existing meta-analyses of the diagnostic accuracy of depression screening tools met even half of the adapted PRISMA for Abstracts items related to quality and completeness of abstract reporting. Furthermore, the majority of the PRISMA for Abstracts items were rarely met in the meta-analyses we evaluated, including items related to study objectives, eligibility criteria for study characteristics, eligibility criteria for report characters, methods used for assessing risk of bias, the number of included studies and participants, the description of summary estimates, funding, and registration. Journal editors should endorse the PRISMA for Abstracts tool to improve upon the completeness of reporting in abstracts. They should also provide authors with flexibility in abstract headings and abstract word counts so that they can more realistically comply with PRISMA for Abstracts recommendations.

AUTHORS' CONTRIBUTIONS

DBR, LAK, IS and BDT were responsible for the study concept and design, drafted the study protocol, contributed to data extraction, contributed to drafting the manuscript, and approved the final manuscript. BDT is the guarantor.

FUNDING STATEMENT

Ms. Rice is supported by a Fonds de Recherché Santé Québec (FRSQ) Master's Award. Dr. Thombs receives support from an Investigator Award form the Arthritis Society. There was no specific funding for this study, and no funders had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Authors had full access to the data and can take responsibility for the integrity of the data and the accuracy of the data analysis.

COMPETING INTERESTS STATEMENT

The authors have read and understood the BMJ policy on declaration of interests and declare that they have no competing interests.

DATA SHARING STATEMENT

No additional data available. Full data extraction dataset is available in the Tables and Supplementary Data Files.

Assessing the quality of diagnostic test accuracy meta-analyses

REFERENCES

1. Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA* 1998;280:267-9.
2. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013;10:e1001419.
3. Beller EM, Glasziou PP, Hopewell S, et al. Reporting of effect direction and size in abstracts of systematic reviews. *JAMA* 2011;306:1981-2.
4. Faggion CM, Jr., Liu J, Huda F, et al. Assessment of the quality of reporting in abstracts of systematic reviews with meta-analyses in periodontology and implant dentistry. *J Periodontal Res* 2014;49:137-42.
5. Kiriakou J, Pandis N, Fleming PS, et al. Reporting quality of systematic review abstracts in leading oral implantology journals. *J Dent* 2013;41:1181-7.
6. Seehra J, Fleming PS, Polychronopoulou A, et al. Reporting completeness of abstracts of systematic reviews published in leading dental specialty journals. *Eur J Oral Sci* 2013;121:57-62.
7. Hopewell S, Boutron I, Altman DG, et al. Deficiencies in the publication and reporting of the results of systematic reviews presented at scientific medical conferences. *J Clin Epidemiol* 2015;68:1488-95.
8. Polychronopoulou A. The reporting quality of meta-analysis results of systematic review abstracts in periodontology and implant dentistry is suboptimal. *J Evid Based Dent Pract* 2014;14:209-10.
9. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014;348:g1253.

10. Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* 2016;315:380-7.
11. Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775-82.
12. Allaby M. Screening for depression: a report for the UK National Screening Committee (revised report). *UK National Screening Committee* 2010.
13. Thombs BD, Benedetti A, Kloda LA, 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.
14. Rice DB, Shrier I, Kloda LA, et al. Methodological quality of meta-analyses of the diagnostic accuracy of depression screening tools. *J Psychosom Res* 2016;84:84-92.
15. Gilbody S, Richards D, Brealey S, et al. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007;22:1596-602.
16. Brennan C, Worrall-Davies A, McMillan D, et al. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *J Psychosom Res* 2010;69:371-8.
17. Meader N, Mitchell AJ, Chew-Graham C, 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.
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.

Assessing the quality of diagnostic test accuracy meta-analyses

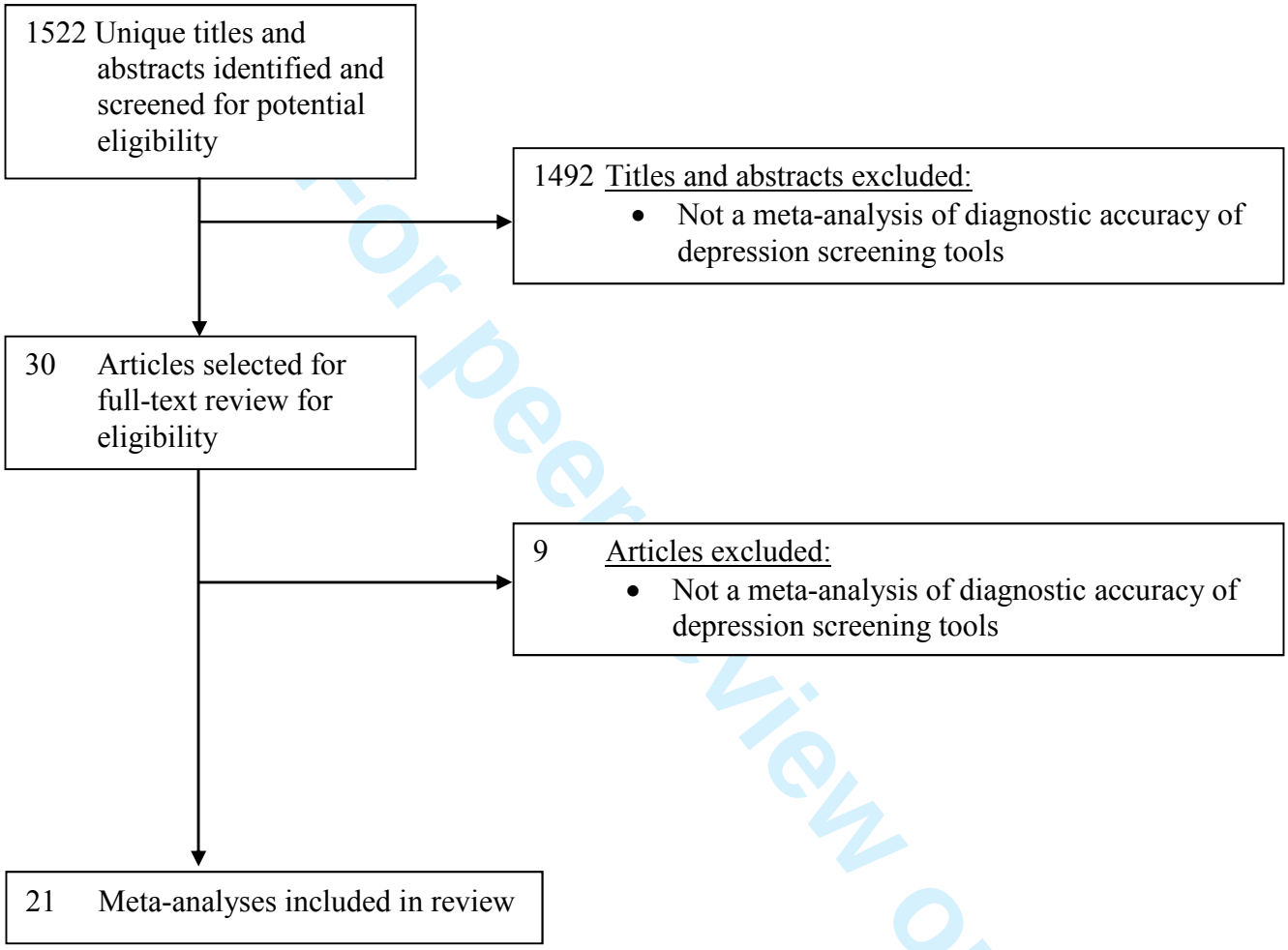
19. Wittkampf KA, Naeije L, Schene AH, et al. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007;29:388-95.
20. Mitchell AJ. Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. *Br J Cancer* 2008;98:1934-43.
21. Mitchell AJ, Bird V, Rizzo M, et al. 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;125:10-7.
22. Mitchell AJ, Bird V, Rizzo M, et al. 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.
23. 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.
24. 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.
25. Tsai AC. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr* 2014;66:503-11.
26. Tsai AC, Scott JA, Hung KJ, et al. Reliability and validity of instruments for assessing perinatal depression in African settings: systematic review and meta-analysis. *PLoS One* 2013;8:e82521.

27. 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. *Support Care Cancer* 2011;19:1899-908.
28. Hewitt C, Gilbody S, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009;13:1-145, 7-230.
29. Meader N, Moe-Byrne T, Llewellyn A, et al. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. *J Neurol Neurosurg Psychiatry* 2014;85:198-206.
30. 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. *J Affect Disord* 2012;140:149-60.
31. Bosanquet K, Bailey D, Gilbody S, et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open* 2015;5:e008913.
32. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *Gen Hosp Psychiatry* 2015;37:67-75.
33. Moriarty AS, Gilbody S, McMillan D, et al. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015;37:567-76.
34. Pocklington C, Gilbody S, Manea L, et al. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J of Geriatr Psychiatry* 2016;[Epub ahead of print].

35. Stockings E, Degenhardt L, Lee YY, et al. Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Affect Disord* 2015;174:447-63.
36. Korevaar DA, Cohen JF, Hooft L, et al. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. *J Clin Epidemiol* 2015;68:708-15.
37. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11:Mr000030.
38. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
39. Hopewell S, Ravaud P, Baron G, et al. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012;344:e4178.

Assessing the quality of diagnostic test accuracy meta-analyses

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



Assessing the quality of diagnostic test accuracy meta-analyses

Table 1. Characteristics of Included Meta-Analyses

First Author, Year of Publication	Journal (2014 Impact Factor)	Focus of Meta-Analysis	AMSTAR Scores	Journal Word Limit
Pocklington, 2016 [34]	Int J Geriatr Psychiatry (2.9)	Brief versions of the GDS in older patients	8 (57%)	250
Bosanquet, 2015 [31]	BMJ Open (2.3)	Whooley questions in any setting	9 (64%)	300
Moriarty, 2015 [33]	Gen Hosp Psychiatry (2.6)	PHQ-9 in any setting	9 (64%)	200
Stockings, 2015 [35]	J Affect Disord (3.4)	Screening tools in children and adolescents	4 (29%)	250
Manea, 2015 [32]	Gen Hosp Psychiatry (2.6)	PHQ-9 with algorithm scoring method in any setting	8 (57%)	200
Meader, 2014 [29]	J Neurol Neurosurg Psychiatry (6.8)	Screening tools in poststroke patients	6 (43%)	250
Tsai, 2014 [25]	JAIDS (4.6)	Screening tools in HIV-positive adults in Africa	5 (36%)	250
Tsai, 2013 [26]	PLoS One (3.2)	Screening tools in pregnancy or postpartum in Africa	6 (43%)	300
Mitchell, 2012 [30]	J Affect Disord (3.4)	Screening tools in cancer patients	4 (29%)	250
Manea, 2012 [18]	CMAJ (6.0)	PHQ-9 in any setting	10 (71%)	250
Meader, 2011 [17]	Br J Gen Pract (2.3)	Screening tools in patients with chronic health problems	5 (36%)	250
Vodermaier, 2011 [27]	Support Care Cancer (2.4)	HADS in cancer patients	6 (43%)	250
Brennan, 2010 [16]	J Psychosom Res (2.7)	HADS in any setting	5 (36%)	250
Mitchell, 2010a [22]	Am J Geriatr Psychiatry (4.2)	GDS in older patients	3 (21%)	250
Mitchell, 2010b [24]	J Affect Disord (3.4)	HADS in cancer and palliative settings	3 (21%)	250
Mitchell, 2010c [21]	J Affect Disord (3.4)	GDS in older primary care patients	3 (21%)	250

Assessing the quality of diagnostic test accuracy meta-analyses

Hewitt, 2009 [28]	Health Technol Assess (5.0)	Screening tools in women in pregnancy or postpartum	8 (57%)	500
Mitchell, 2008 [20]	Br J Cancer (4.8)	Short screening tools in cancer and palliative care	5 (36%)	200
Gilbody, 2007 [15]	J Gen Intern Med (3.4)	PHQ in medical settings	6 (43%)	300
Mitchell, 2007 [23]	Br J Gen Pract (2.3)	Ultra-short screening tools in primary care	4 (29%)	250
Wittkampf, 2007 [19]	Gen Hosp Psychiatry (2.6)	PHQ in any setting	6 (43%)	200

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

Assessing the quality of diagnostic test accuracy meta-analyses

Table 2. Adapted PRISMA for Abstracts Item Totals for the 16 Meta-Analyses Reviewed

Adapted PRISMA for Abstracts Item	Adapted Description	Proportion of Meta-Analyses with '*yes*' ratings (%)
Item 1	Title: Identify the report as a systematic review, meta-analyses or both.	20 (95%)
Item 2	Objectives: The research question including components such as participants, index test, reference standard and outcomes.	0 (0%)
Item 3a	Eligibility criteria: study characteristics used as criteria for inclusion.	1 (5%)
Item 3b	Eligibility criteria: report characteristics used as criteria for inclusion.	0 (0%)
Item 4a	Information sources: Key databases searched.	9 (43%)
Item 4b	Information sources: Key search dates.	7 (33%)
Item 5	Risk of bias: Methods of assessing risk of bias.	3 (14%)
Item 6	Included studies: Number and type of included studies and participants and relevant characteristics of studies.	2 (10%)
Item 7	Results of the principle summary measures (e.g., sensitivity and specificity, diagnostic odds ratio).	16 (76%)
Item 8	Description of outcomes: summary of accuracy outcomes in terms meaningful to clinicians and patients.	3 (14%)
Item 9	Strengths and Limitations of evidence: Brief summary of strengths and limitations of	7 (33%)

	evidence (e.g., inconsistency, imprecision, indirectness, or risk of bias, other supporting or conflicting evidence).	
Item 10	Interpretation: General interpretation of the results and important implications.	20 (95%)
Item 11	Funding: Primary source of funding for the review.	0 (0%)
Item 12	Registration: Registration number and registry name.	1 (5%)

Assessing the quality of diagnostic test accuracy meta-analyses

Table 3. PRISMA for Abstracts Item by Item Ratings

Reference	Item 1	Item 2	Item 3a	Item 3b	Item 4a	Item 4b	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Total 'Yes'
Pocklington, 2016 [34]	Yes	No	No	No	No	No	Yes	No	Yes	No	Yes	Yes	No	No	5 (36%)
Bosanquet, 2015 [31]	Yes	No	No	No	No	No	No	No	Yes	No	Yes	Yes	No	Yes	5 (36%)
Moriarty, 2015 [33]	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	No	7 (50%)
Stockings, 2015 [35]	Yes	No	No	No	Yes	No	No	No	Yes	No	Yes	Yes	No	No	5 (36%)
Manea, 2015 [32]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Meader, 2014 [29]	Yes	No	No	No	No	Yes	No	No	Yes	No	No	Yes	No	No	4 (29%)
Tsai, 2014 [25]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Tsai, 2013 [26]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)
Mitchell, 2012 [30]	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	4 (29%)
Manea, 2012 [18]	Yes	No	No	No	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	6 (43%)
Meader, 2011 [17]	Yes	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes	No	No	5 (36%)
Vodermaier, 2011 [27]	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	No	No	3 (21%)
Brennan, 2010 [16]	Yes	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes	No	No	5 (36%)
Mitchell, 2010a [22]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)
Mitchell, 2010b [24]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Mitchell, 2010c [21]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)

Assessing the quality of diagnostic test accuracy meta-analyses

Hewitt, 2009 [28]	No	No	No	No	Yes	No	Yes	No	Yes	No	No	Yes	No	No	4 (29%)
Mitchell, 2008 [20]	Yes	No	No	No	No	Yes	No	No	Yes	No	No	Yes	No	No	4 (29%)
Gilbody, 2007 [15]	Yes	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	7 (50%)
Mitchell, 2007 [23]	Yes	No	No	No	No	No	No	No	Yes	Yes	No	Yes	No	No	4 (29%)
Wittkampf, 2007 [19]	Yes	No	No	No	Yes	Yes	Yes	No	Yes	No	No	Yes	No	No	6 (43%)
Total 'Yes'	20 (95%)	0 (0%)	1 (5%)	0 (0%)	9 (43%)	7 (33%)	3 (14%)	2 (10%)	16 (76%)	3 (14%)	7 (33%)	20 (95%)	0 (0%)	1 (5%)	

Note: Item 1= title; Item 2= objectives; Item 3a= eligibility criteria study characteristics; Item 3b= eligibility criteria report characteristics; Item 4a= databases searched; Item 4b= search dates; Item 5= risk of bias; Item 6= included studies; Item 7= synthesis of results; Item 8= description of outcomes; Item 9= strengths and limitations of evidence; Item 10= interpretation; Item 11= funding; Item 12= registration.

S1 Appendix. Search Strategy

Ovid MEDLINE In-Process & Other Non-Indexed Citations and Ovid MEDLINE

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
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

S2 Appendix. Original and Adapted PRISMA for Abstract Tools

Item	Description	Adjusted for DTA	No	Yes	Items
Item 1	Title: Identify the report as a systematic review, meta-analyses or both.	----------	The title does not include words "systematic review" or "meta-analysis".	The title includes words "systematic review" or "meta-analysis".	
Item 2	Objectives: The research question including components such as participants, interventions, comparators and outcomes.	Objectives: The research question including components such as participants, index test, reference standard and outcomes.	There is no explicit statement of the questions being addressed or there is an objective statement, but it does not include a reference to each of the following (PIRO): participants, index text, reference standards and outcomes.	There is an explicit statement of the questions being addressed or an objectives statement with reference to each of the following (PIRO): participants, index test, reference standards, and outcomes.	A statement that refers to all PIRO components must be found in the objectives or research statement section of the abstract to be coded as "yes". Review questions may be narrowly focused or broad.
Item 3a	Eligibility criteria: study characteristics used as criteria for inclusion.	----------	One or more eligibility criteria are omitted and there is no statement that eligibility criteria had no restrictions.	Study eligibility criteria are stated for all PIRO study characteristics or there is a statement that there were no restrictions for eligibility criteria.	This item considers characteristics of the primary study itself. All PIRO components must be specified. If a study had no restrictions and includes any participants who completed a screening measure and assessment, this must be stated.
Item 3b	Eligibility criteria: report characteristics used as criteria for inclusion.	----------	Report characteristics considered for eligibility are not stated or there is a statement, but it does not address language of	Report characteristics considered for eligibility are clearly stated, including, at least, language of publication and publication status.	This item considers characteristics of the report of the primary study.

			publication and publication status.		
Item 4a	Information sources: Key databases searched.	----------	Does not list all databases searched.	Lists all databases searched (or at least 3 if more than 3 searched).	
Item 4b	Information sources: Key search dates.	----------	Does not state the dates of coverage of the search, including year and month of end date.	The dates of coverage of the search are provided, including year and month of end date.	
Item 5	Risk of bias: Methods of assessing risk of bias.	----------	There is no statement about how risk of bias was assessed, including the name of the tool for assessing risk of bias is listed. Alternatively, if risk of bias was not assessed, this is not stated.	There is a statement about how risk of bias was assessed, including the name of the tool used to assess bias is listed, or there is a statement that risk of bias was not assessed.	Assessments of study "quality" are coded as assessing risk of bias. All elements must be included to be coded "Yes".
Item 6	Included studies: Number and type of included studies and participants and relevant characteristics of studies.	----------	The number of primary studies, total number of participants and cutoffs included in the analyses are not provided.	The number of primary studies, total number of participants and cutoffs assessed are provided.	All elements must be included to be coded "Yes".
Item 7	Synthesis of results: Results for main outcomes (benefits and harms), preferably indicating the number of studies and participants for each. If meta-analysis was done, include summary measures and confidence	Results of the principle summary measures (e.g., sensitivity and specificity, diagnostic odds ratio).	Results of the principle summary measures are not provided (e.g., sensitivity and specificity, diagnostic odds ratio).	Results of the principle summary measures are provided (e.g., sensitivity and specificity, diagnostic odds ratio).	If sensitivity and specificity are the primary outcome measures, both must be reported to code "Yes".

	intervals.				
Item 8	Description of effect: direction of the effect and size of the effect in terms meaningful to clinicians and patients.	Description of outcomes: summary of accuracy outcomes in terms meaningful to clinicians and patients.	Results do not summarize the principle results in words and numbers, including the most accurate cutoff to use and how the screening tools would perform in practice, in terms meaningful to clinicians and patients (e.g. positive and negative predictive values or true and false positive rates).	Results summarize the principle results in words and numbers, including the most accurate cutoff to use and how the screening tools would perform in practice, in terms meaningful to clinicians and patients (e.g. positive and negative predictive values or true and false positive rates).	
Item 9	Strengths and Limitations of evidence: Brief summary of strengths and limitations of evidence (e.g., inconsistency, imprecision, indirectness, or risk of bias, other supporting or conflicting evidence).	----------	Relevant limitations of the diagnostic accuracy evidence are not noted (e.g., inconsistency, imprecision, indirectness, risk of bias, other supporting or conflicting evidence).	Relevant limitations of the diagnostic accuracy evidence are noted (e.g., inconsistency, imprecision, indirectness, risk of bias, other supporting or conflicting evidence).	
Item 10	Interpretation: General interpretation of the results and important implications.	----------	Authors do not provide a statement about clinical implications of results or need for more research if results suggest uncertainty.	Authors provide a statement about clinical implications of results or need for more research if results suggest uncertainty.	
Item 11	Funding: Primary source of funding for the review.	----------	Authors do not provide the sources of funding for the review or a statement that it was not funded.	Authors provide the source of funding for the review or a statement that it was not funded.	For a “yes”, funding information must be listed in the abstract that is available for viewing on an online database such as PubMed.

Item 12	Registration: Registration number and registry name.	----------	A registration number and/or registry name are not provided.	A registration number and registry name are provided.	Cochrane reviews are an exception to this requirement, as they are preceded by a peer reviewed protocol that is published in the Cochrane Library and can be downloaded from there. For a "yes", registration number and/or name must be listed in the abstract that is available for viewing on an online database such as PubMed.

S3 Appendix. 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, Gibbs FA, Williams JR, Tregear S. Peer Reviewed: Depression Screening and Treatment Among Nonpregnant Women of Reproductive Age in the United States, 1990-2010. Preventing chronic disease. 2011 Nov;8(6).
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(3):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(4):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;25(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 (Summ). 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(5):423-32.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	NA
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7-8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7-10
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-10
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	11, figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	14-15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	16

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

Reporting Quality in Abstracts of Meta-Analyses of Depression Screening Tool Accuracy: A Review of Systematic Reviews and Meta-Analyses

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012867.R1
Article Type:	Research
Date Submitted by the Author:	10-Oct-2016
Complete List of Authors:	Rice, Danielle; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Psychiatry Kloda, Lorie; Concordia University, Library Shrier, Ian; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Epidemiology, Biostatistics, and Occupational Health Thombs, Brett; Jewish General Hospital, Lady Davis Institute for Medical Research; McGill University, Psychiatry
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Mental health
Keywords:	Depression & mood disorders < PSYCHIATRY, PRISMA for Abstracts, diagnostic test accuracy, meta-analyses, screening

Assessing the quality of diagnostic test accuracy meta-analyses

Reporting Quality in Abstracts of Meta-Analyses of Depression Screening Tool Accuracy: A Review of Systematic Reviews and Meta-Analyses

Danielle B Rice[1,2], Lorie A. Kloda[3]; Ian Shrier[1,4] Brett D Thombs[1,2,4-8*]

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada

[2]Department of Psychiatry, McGill University, Montréal, Québec, Canada

[3]Library, McGill University, Montréal, Québec, Canada

[4]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Québec, Canada

[5]Department of Psychology, McGill University, Montréal, Québec, Canada

[6]Department of Medicine, McGill University, Montréal, Québec, Canada

[7]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada

[8]School of Nursing, McGill University, Montréal, Québec, Canada

Address for Correspondence: Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Saint Catherine Road; Montréal, Québec, Canada; H3T 1E4; Telephone: (514) 340-8222 ext. 5112; Fax: (514) 340-8124; Email: brett.thombs@mcgill.ca.

Keywords: PRISMA for Abstracts, depression, diagnostic test accuracy, meta-analyses, screening.

Word Count: 3,539

Assessing the quality of diagnostic test accuracy meta-analyses

ABSTRACT

Objective: Concerns have been raised regarding the quality and completeness of abstract reporting in evidence reviews, but this had not been evaluated in meta-analyses of diagnostic accuracy. Our objective was to evaluate reporting quality and completeness in abstracts of systematic reviews with meta-analyses of depression screening tool accuracy, using the PRISMA for Abstracts tool.

Design: Cross-sectional study.

Inclusion Criteria: We searched MEDLINE and PsycINFO from January 1, 2005 through March 13, 2016 for recent systematic reviews with meta-analyses in any language that compared a depression screening tool to a diagnosis based on clinical or validated diagnostic interview.

Data Extraction: Two reviewers independently assessed quality and completeness of abstract reporting using the PRISMA for Abstracts tool with appropriate adaptations made for studies of diagnostic test accuracy. Bivariate associations of number of PRISMA for Abstracts items complied with (1) journal abstract word limit and (2) AMSTAR scores of meta-analyses were also assessed.

Results: We identified 21 eligible meta-analyses. Only two of 21 included meta-analyses complied with at least half of adapted PRISMA for Abstracts items. The majority met criteria for reporting an appropriate title (95%), result interpretation (95%), and synthesis of results (76%). Meta-analyses less consistently reported databases searched (43%), associated search dates (33%) and strengths and limitations of evidence (19%). Most meta-analyses did not adequately report a clinically meaningful description of outcomes (14%), risk of bias (14%), included study characteristics (10%), study eligibility criteria (5%), registration information (5%), clear objectives (0%), report eligibility criteria (0%), or funding (0%). Overall meta-analyses quality scores were significantly associated with the number of PRISMA for Abstract scores items reported adequately (r = 0.45).

Assessing the quality of diagnostic test accuracy meta-analyses

Conclusions: Quality and completeness of reporting was found to be suboptimal. Journal editors should endorse PRISMA for Abstracts and allow for flexibility in abstract word counts to improve quality of abstracts.

STRENGTHS AND LIMITATIONS

- This is the first study to systematically evaluate the transparency and completeness of reporting in abstracts of systematic reviews with meta-analyses of depression screening tools.
- Areas that require improvement were identified.
- Since there is not currently a PRISMA for Abstracts tool developed for reviews of diagnostic test accuracy, minor adaptations had to be made to the original tool.
- Our sample included a relatively small number of systematic reviews with meta-analyses.
- The lack of variability in the word limits of journal abstracts where included systematic reviews with meta-analyses were published limited our ability to examine the association between PRISMA for Abstract ratings and abstract word limits.

INTRODUCTION

Researchers, clinicians and other consumers of research often rely primarily on information found in abstracts of systematic reviews.[1] Frequently, the abstract is the only part of an article that is read, making it the most frequently read part of biomedical articles after the title.[2] This may be due to time limitations, accessibility constraints, or language barriers.[2] For time-pressed readers or readers with limited access to a full-text article, the abstract must be able to stand alone in presenting a clear account of the methods, results, and conclusions that accurately reflect the core components of the full research report.[2] This goal, however, is infrequently achieved, as the quality and completeness of information provided in abstracts of systematic reviews are often suboptimal.[3-6]

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for Abstracts tool was developed as an extension of the PRISMA statement,[2] with the goal of improving the quality and completeness of abstracts in systematic reviews, including meta-analyses.[2] The PRISMA for Abstracts checklist includes 12 items related to information that should be provided in systematic review abstracts, including title; objectives; eligibility criteria of included studies; information sources, including key databases and dates of searches; methods of assessing risk of bias; number and type of included studies; synthesis of results for main outcomes; description and direction of the effect; summary of strengths and limitations of evidence; general interpretation of results; source of funding; and registration number.

Only one previous study has used the PRISMA for Abstracts checklist to evaluate the quality and completeness of abstracts for systematic reviews of trials.[7] That study included 197 systematic review abstracts published in 2010 in the proceedings of nine leading international medical conferences that have conference abstracts that are searchable online. PubMed was then searched from 2010 to 2013 to identify subsequently published journal articles (N = 103).[7] In both published

conference abstracts and published articles, 9 of the 12 PRISMA for Abstracts items were completed in less than 50% of abstracts reviewed. Poor reporting of abstracts has also been found in studies that have evaluated abstracts of meta-analyses and systematic reviews using other methods. We identified three studies, all from dentistry literature, that reviewed reporting of abstracts in systematic reviews of trials.[4-6] Two of the studies evaluated abstracts using a 16-item checklist derived from the full PRISMA statement, prior to the official PRISMA for Abstracts publication.[5, 6] The third study assessed abstract reporting based on the presence or absence of seven characteristics related to the meta-analyses results.[8] In all three studies, major deficiencies were identified.

Depression screening is an area where indirect evidence from diagnostic test accuracy (DTA) studies has played an important role in policy and where the quality of reporting may be particularly important. Depression screening is controversial, and recommendations on screening are inconsistent.[9] Based on indirect evidence, including evidence on screening tool accuracy, the United States Preventative Services Task Force recently recommended universal depression screening in all adults.[10] Both the UK National Screening Committee and the Canadian Task Force on Preventative Health Care however, recommend against depression screening due to a lack of evidence from randomized controlled trials that depression screening would improve mental health outcomes.[11, 12]

No published studies have evaluated the completeness of reporting in abstracts of diagnostic test accuracy systematic reviews or meta-analyses. The PRISMA for Abstracts guideline was developed for systematic reviews of interventions, and the authors suggested that modifications would be required to apply the checklist to DTA systematic reviews.[2] In the absence of a PRISMA for Abstracts tool designed for studies of DTA, we applied PRISMA for Abstracts with adaptations to some items in order to appropriately assess systematic reviews with meta-analyses of DTA studies of

depression screening tools. The primary objective of our study was to evaluate the transparency and completeness of abstracts of systematic reviews with meta-analyses of the diagnostic accuracy of depression screening tools that were published in journals indexed in the MEDLINE and PsycINFO databases, using PRISMA for Abstracts. Our secondary objective was to determine if the quality of the meta-analysis or the word count permitted by the journal of the systematic reviews with meta-analyses were associated with PRISMA for Abstract scores, as the feasibility of adhering to the PRISMA for Abstracts items may be compromised by abstract word count constraints set by journals.

METHODS

Identification of meta-analyses on the diagnostic accuracy of depression screening tools

The search strategy used for this study was originally conducted for a study assessing the quality of systematic reviews with meta-analyses of diagnostic test accuracy for depression screening tools.[13] We searched Medline and PsycINFO (both on the OvidSP platform) from January 1, 2005 through March 13, 2016 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,[14] 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 S1 Appendix.

We included publications of meta-analyses, but not systematic reviews without meta-analyses, in order to focus only 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). We excluded meta-analyses that did not use a clinical or diagnostic interview as the gold standard. 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 unique citation 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.

Assessment of reporting in abstracts

The reporting of abstracts was evaluated using a PRISMA for Abstracts tool, with some items adapted for applicability to studies of diagnostic test accuracy. The original PRISMA for Abstracts tool was developed to provide guidance on a minimum set of items necessary to provide a reasonably complete and transparent representation of a full article report.[2] The checklist was created to fit into headings mandated by journals and conference submissions, including title, background, methods,

results, discussion and associated funding and registration information, but was designed with flexibility regarding the specific headings and where information should be listed. The PRISMA for Abstracts checklist was developed for systematic reviews of abstracts involving interventions, but many of the items are applicable to other designs, including DTA systematic reviews and meta-analyses.

We adapted the original PRISMA for Abstracts tool to ensure that items were applicable to DTA studies. The team that adapted the PRISMA for Abstracts tool included members with expertise in evidence synthesis (IS, BT, LAK), information sciences for evidence synthesis (LAK) and DTA studies of depression screening tools (BDT). Each original PRISMA for Abstracts item was reviewed by team members, who considered ease of coding and applicability to DTA systematic reviews and meta-analyses, then either accepted the item as appropriate or edited the item to better reflect practices in the conduct of DTA systematic reviews. In addition, a coding manual was developed with specific criteria for *yes* and *no* ratings, along with additional coding notes (see S2 Appendix for details).

The adapted tool included 14 items because two of the original PRISMA for Abstracts items were divided into two parts. The two items that were divided did not undergo any additional changes. Item 3 was originally "Study and report characteristics used as criteria for inclusion" and was adapted to items 3a "Study characteristics used as inclusion criteria" and item 3b "Report characteristics used as inclusion criteria." Item 3 was divided into two parts in order to differentiate between characteristics for inclusion in primary studies (i.e., eligible participants. index tests, reference standards and outcomes), and characteristics for inclusion in the systematic review and meta-analyses (e.g., language and publication status of eligible reviews). Item 4, "Key databases searched and search dates", which involved reporting specific databases searched and the dates searched, was

divided into 4a (key databases searched) and 4b (search dates). Of the original 12 items, seven were unaltered (1: title, 5: risk of bias, 6: included studies, 9: strengths and limitations of evidence, 10: interpretation, 11: funding, 12: registration). Three items (2: objectives, 7: synthesis of results, 8: description of effect) were slightly modified for applicability to DTA systematic review abstracts. The original item 2 refers to "the research question including components such as participants, interventions, comparators and outcomes". For increased relevance to DTA reviews, this item was revised to encompass the reference standard and index test within the systematic review rather than the interventions and comparators found in intervention studies. Item 7 was adjusted to encompass results of the principle summary measures (e.g. sensitivity, specificity, positive predictive value, negative predictive value) that are reported in DTA studies. Lastly, the original item 8 refers to "the direction and size of the effect" and was adjusted to evaluate if the summary of accuracy estimates that are presented within DTA studies are presented in terms meaningful to clinicians.

Data extraction

For each meta-analysis publication, one investigator extracted author, year of publication, journal, journal impact factor for 2014, the abstract word limit of the journal where the meta-analysis was published (see S3 Appendix for details), and previously published A Measurement tool to Assess Systematic Reviews (AMSTAR) quality ratings.[13] Accuracy was verified by a second investigator. Two investigators independently rated each included systematic reviews with meta-analyses using the adapted PRISMA for Abstracts checklist. 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 systematic reviews with meta-analyses 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. For publications that included meta-analyses of diagnostic accuracy and other measurement characteristics, only results relevant to diagnostic accuracy were extracted.

Statistical analyses

Bivariate associations between the (1) abstract word count permitted by the journal, and (2) AMSTAR scores of meta-analyses to the PRISMA for Abstracts scores were assessed with Pearson correlation coefficients. Analyses were conducted using SPSS version 22.0 (Chicago, IL), and statistical tests were two-sided with a $p < 0.05$ significance level. 95% confidence intervals (CI) were also calculated.

RESULTS

Article selection

The electronic database search yielded 1522 unique title and abstracts for review. Of these, 1492 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 30 articles that underwent full-text review, 9 were excluded because they were not meta-analyses of diagnostic accuracy of depression screening tools (see S4 Appendix), resulting in 21 eligible systematic reviews with meta-analyses published between 2007 and 2016 (see Figure 1).[15-35] Characteristics of included systematic reviews with meta-analyses are shown in Table 1.

As shown in Table 2, of the 14 adapted PRISMA for Abstracts items, there were two items for which 20 of the 21 included meta-analyses received a *yes* rating: items 1 (title; 95%) and 10 (interpretation of results; 95%). One item received a *yes* rating in 16 of 21 meta-analyses (item 7, synthesis of results; 76%), and three items received a *yes* rating in 7 to 9 of 21 meta-analyses (33% to 43%): items 4a (databases searched), 4b (key search dates) and item 9 (strengths and limitations of evidence). Very few meta-analyses fulfilled criteria for a rating of *yes* for the remaining 8 items

including item 8 (description of the outcomes; 14%), item 5 (risk of bias; 14%), item 6 (included studies; 10%), item 3a (eligibility criteria for study characteristics; 5%), item 12 (registration; 5%), item 2 (objectives; 0%), item 3b (eligibility criteria for report characteristics 0%), and item 11 (funding; 0%).

When considering item ratings for each meta-analysis, two of the 21 meta-analyses received a *yes* rating for 7 of the 14 adapted PRISMA for Abstracts.[15, 33] An additional seven meta-analyses received ratings of *yes* for 5[16, 17, 31, 34, 35] and 6 [18, 19] of the 14 PRISMA for Abstracts items. The remaining 12 meta-analyses received *yes* ratings on between 2 and 4 of the 14 items (see Table 3).

Association of Journal Abstract Word Count and AMSTAR Scores with PRISMA For Abstract Scores

There was a significant positive association of AMSTAR scores with the number of *yes* ratings of PRISMA for Abstracts items (r =0.45, 95% CI = 0.02 to 0.74, p = 0.040). The abstract word count permitted by the journal was not significantly correlated to the PRISMA for Abstracts scores (r = -0.03, 95% CI = -0.45 to 0.41, p = 0.914). However, 20 out of 21 meta-analyses were published in journals that had word limits between 200 to 300 words.

DISCUSSION

The main findings of this study were that only 3 of 14 items from the adapted PRISMA for Abstracts tool received *yes* ratings in at least 50% of 21 systematic reviews with meta-analyses of depression screening tools. The other 11 items were infrequently met. Furthermore, overall quality of reporting in the abstracts of the systematic reviews with meta-analyses was poor, with only 2 of 21 meta-analyses rating *yes* for at least half of the PRISMA for Abstracts items. Overall quality ratings

of the systematic reviews with meta-analyses, based on AMSTAR, were associated with the number of PRISMA for Abstract items that were adequately reported.

Among meta-analyses evaluated in the present study, almost all met criteria for having a title that identified the report as systematic review or meta-analysis, for reporting the main results of the synthesis, and for providing a general interpretation of the results and important implications. In addition, 9 of 21 systematic reviews with meta-analyses also provided a list of databases searched and 7 provided dates of coverage for the literature search and strengths and limitations of evidence. On the other hand, 3 or fewer systematic reviews with meta-analyses received *yes* ratings for stating the methods used for assessing risk of bias, the number of included studies and participants, eligibility criteria for study characteristics, registration information, and the description of summary estimates. No studies met criteria for the remaining 3 PRISMA for Abstracts items (complete study objectives, eligibility criteria for report characters, and funding information).

Beyond systematic reviews and meta-analyses, specific concerns have been raised about the quality of abstracts of primary studies of DTA. A 21-item tool was developed to assess whether abstracts of primary DTA studies are adequately informative, based on the reporting of essential methodological features and study results.[36] The tool was applied to a sample of 103 primary DTA studies published in 12 high-impact journals in 2012, and only 39 of the 103 primary studies that were evaluated received a rating of *adequate* for at least half of the items assessed. Specifically, the authors reported that fewer than 50% of included primary studies adequately reported the study population, setting, patient sampling, blinding, cutoffs used and confidence intervals around accuracy estimates.[36] The mean number of adequately reported items within abstracts was significantly lower for abstracts that had lower word counts.

Assessing the quality of diagnostic test accuracy meta-analyses

Several authors have recommended that journal editors endorse abstract guidelines, such as the PRISMA for Abstracts tool, to help ensure that abstracts better address the needs of consumers of research,[2, 4, 7, 36] and, generally, journal endorsement of reporting guidelines improves the completeness of reporting.[37] The Consolidated Standards of Reporting Trials (CONSORT) reporting guidelines for abstracts of randomized controlled trials was published in 2009,[38] and a recent study found that journals that implement these guidelines have improved reporting in abstracts of randomized controlled trials.[39] As of April 6, 2016, only one of the journals where DTA meta-analyses included in the present study were published (Journal of General Internal Medicine) includes a statement specifically endorsing the PRISMA for Abstracts tool and a web link to the PRISMA for Abstracts tool in its author instructions. A second journal (Health Technology Assessments) required authors to comply with general PRISMA guidelines in developing the abstract, but did not refer to the PRISMA for Abstracts statement or its items. No other journals mentioned PRISMA in relation to abstracts. All journals had word limits of between 200 and 300 words for abstracts with the exception of Health Technology Assessments, which allows 500 words. Health Technology Assessments is a UK National Institutes of Health Research journal that typically publishes extensive, multi-question systematic reviews. Currently, it is not likely to be feasible for authors to include all PRISMA for Abstracts recommended reporting items due to word count restraints typically imposed for biomedical journal abstracts. Thus, we recommend that journals endorse the use of the PRISMA for Abstracts checklist for formulating abstracts and that journals provide flexibility in word counts and the structure of abstract headings in order to comply with recommendations. This is already done in some journals (e.g., BMJ, PLOS Medicine).

As almost all of the systematic reviews with meta-analyses that we evaluated were published prior to the development of the PRISMA for Abstracts tool, it could not have been expected that our

sample of studies would have been able to follow the checklist when developing their abstracts. Our study provides direction for evaluating PRISMA for Abstract adherence in reviews and meta-analyses in the field of DTA. Further, our study highlights areas where improvement is needed, specifically in systematic reviews with meta-analyses of DTA of depression screening, and will allow future DTA reviews to apply our coding manual, and compare the reporting of abstracts after the PRISMA for Abstracts tool has been more widely endorsed.

Specific limitations should be considered when interpreting the results of our study. First, we did not perform a pilot test of out tool. Adjustments were made to our coding manual during the initial part of our meta-analysis scoring and, as such, we were unable to calculate an interrater agreement statistic for the adapted PRISMA for Abstracts items. 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, to meta-analyses on the diagnostic accuracy of depression screening tools that were not indexed in these two databases, or to meta-analyses of diagnostic accuracy in other conditions and other fields. Third, we reported results on an item-by-item basis for illustration purposes. Not all items, however, would be expected to influence the transparency and completeness of abstract reporting equally, and an evaluation of the quality of any given meta-analysis abstract would need to consider specific items individually. Finally, we adapted the PRISMA for Abstracts tool for this study, as it was developed for use in systematic reviews and meta-analyses of intervention studies. Ideally, however, a PRISMA for Abstracts tool would be developed specifically for reviews of diagnostic test accuracy. We also attempted to analyze the association between journal word limits and the PRISMA for abstract scores, however, 20 of 21 meta-analyses included in our study were published in journals with word limits of 200 to 300 words.

In conclusion, the present study found that only 2 of 21 existing meta-analyses of the diagnostic accuracy of depression screening tools met even half of the adapted PRISMA for Abstracts items related to quality and completeness of abstract reporting. Furthermore, the majority of the PRISMA for Abstracts items were rarely met in the meta-analyses we evaluated, including items related to study objectives, eligibility criteria for study characteristics, eligibility criteria for report characters, methods used for assessing risk of bias, the number of included studies and participants, the description of summary estimates, funding, and registration. Journal editors should endorse the PRISMA for Abstracts tool to improve upon the completeness of reporting in abstracts. When PRISMA for Abstracts is updated, it should consider the number of words that may be necessary to comply with recommendations. Journal editors should either provide authors with flexibility in abstract headings and abstract word counts, or match their abstract word limit with that recommendation so that authors can more realistically comply with PRISMA for Abstracts recommendations.

AUTHORS' CONTRIBUTIONS

DBR, LAK, IS and BDT were responsible for the study concept and design, drafted the study protocol, contributed to data extraction, contributed to drafting the manuscript, and approved the final manuscript. BDT is the guarantor.

FUNDING STATEMENT

Ms. Rice is supported by a Fonds de Recherché Santé Québec (FRSQ) Master's Award. Dr. Thombs receives support from an Investigator Award from the Arthritis Society. There was no specific funding for this study, and no funders had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Authors had full access to the data and can take responsibility for the integrity of the data and the accuracy of the data analysis.

COMPETING INTERESTS STATEMENT

The authors have read and understood the BMJ policy on declaration of interests and declare that they have no competing interests. The authors of this study did not contribute to any of the included studies that were evaluated.

DATA SHARING STATEMENT

No additional data available. Full data extraction dataset is available in the Tables and Supplementary Data Files.

REFERENCES

1. Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA* 1998;280:267-9.
2. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013;10:e1001419.
3. Beller EM, Glasziou PP, Hopewell S, et al. Reporting of effect direction and size in abstracts of systematic reviews. *JAMA* 2011;306:1981-2.
4. Faggion CM, Jr., Liu J, Huda F, et al. Assessment of the quality of reporting in abstracts of systematic reviews with meta-analyses in periodontology and implant dentistry. *J Periodontal Res* 2014;49:137-42.
5. Kiriakou J, Pandis N, Fleming PS, et al. Reporting quality of systematic review abstracts in leading oral implantology journals. *J Dent* 2013;41:1181-7.
6. Seehra J, Fleming PS, Polychronopoulou A, et al. Reporting completeness of abstracts of systematic reviews published in leading dental specialty journals. *Eur J Oral Sci* 2013;121:57-62.
7. Hopewell S, Boutron I, Altman DG, et al. Deficiencies in the publication and reporting of the results of systematic reviews presented at scientific medical conferences. *J Clin Epidemiol* 2015;68:1488-95.
8. Polychronopoulou A. The reporting quality of meta-analysis results of systematic review abstracts in periodontology and implant dentistry is suboptimal. *J Evid Based Dent Pract* 2014;14:209-10.
9. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014;348:g1253.

10. Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* 2016;315:380-7.
11. Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775-82.
12. Allaby M. Screening for depression: a report for the UK National Screening Committee (revised report). *UK National Screening Committee* 2010.
13. Rice DB, Shrier I, Kloda LA, et al. Methodological quality of meta-analyses of the diagnostic accuracy of depression screening tools. *J Psychosom Res* 2016;84:84-92.
14. Thombs BD, Benedetti A, Kloda LA, 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.
15. Gilbody S, Richards D, Brealey S, et al. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007;22:1596-602.
16. Brennan C, Worrall-Davies A, McMillan D, et al. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *J Psychosom Res* 2010;69:371-8.
17. Meader N, Mitchell AJ, Chew-Graham C, 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.
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. Wittkampf KA, Naeije L, Schene AH, et al. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007;29:388-95.
20. Mitchell AJ. Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. *Br J Cancer* 2008;98:1934-43.
21. Mitchell AJ, Bird V, Rizzo M, et al. 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;125:10-7.
22. Mitchell AJ, Bird V, Rizzo M, et al. 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.
23. 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.
24. 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.
25. Tsai AC. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr* 2014;66:503-11.
26. Tsai AC, Scott JA, Hung KJ, et al. Reliability and validity of instruments for assessing perinatal depression in African settings: systematic review and meta-analysis. *PLoS One* 2013;8:e82521.

27. 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. *Support Care Cancer* 2011;19:1899-908.
28. Hewitt C, Gilbody S, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009;13:1-145, 7-230.
29. Meader N, Moe-Byrne T, Llewellyn A, et al. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. *J Neurol Neurosurg Psychiatry* 2014;85:198-206.
30. 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. *J Affect Disord* 2012;140:149-60.
31. Bosanquet K, Bailey D, Gilbody S, et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open* 2015;5:e008913.
32. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *Gen Hosp Psychiatry* 2015;37:67-75.
33. Moriarty AS, Gilbody S, McMillan D, et al. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015;37:567-76.
34. Pocklington C, Gilbody S, Manea L, et al. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J of Geriatr Psychiatry* 2016;[Epub ahead of print].

35. Stockings E, Degenhardt L, Lee YY, et al. Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Affect Disord* 2015;174:447-63.
36. Korevaar DA, Cohen JF, Hooft L, et al. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. *J Clin Epidemiol* 2015;68:708-15.
37. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11:Mr000030.
38. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
39. Hopewell S, Ravaud P, Baron G, et al. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012;344:e4178.

Assessing the quality of diagnostic test accuracy meta-analyses

Table 1. Characteristics of Included Meta-Analyses

First Author, Year of Publication	Journal (2014 Impact Factor)	Focus of Meta-Analysis	AMSTAR Scores	Journal Word Limit
Pocklington, 2016 [34]	Int J Geriatr Psychiatry (2.9)	Brief versions of the GDS in older patients	8 (57%)	250
Bosanquet, 2015 [31]	BMJ Open (2.3)	Whooley questions in any setting	9 (64%)	300
Moriarty, 2015 [33]	Gen Hosp Psychiatry (2.6)	PHQ-9 in any setting	9 (64%)	200
Stockings, 2015 [35]	J Affect Disord (3.4)	Screening tools in children and adolescents	4 (29%)	250
Manea, 2015 [32]	Gen Hosp Psychiatry (2.6)	PHQ-9 with algorithm scoring method in any setting	8 (57%)	200
Meader, 2014 [29]	J Neurol Neurosurg Psychiatry (6.8)	Screening tools in poststroke patients	6 (43%)	250
Tsai, 2014 [25]	JAIDS (4.6)	Screening tools in HIV-positive adults in Africa	5 (36%)	250
Tsai, 2013 [26]	PLoS One (3.2)	Screening tools in pregnancy or postpartum in Africa	6 (43%)	300
Mitchell, 2012 [30]	J Affect Disord (3.4)	Screening tools in cancer patients	4 (29%)	250
Manea, 2012 [18]	CMAJ (6.0)	PHQ-9 in any setting	10 (71%)	250
Meader, 2011 [17]	Br J Gen Pract (2.3)	Screening tools in patients with chronic health problems	5 (36%)	250
Vodermaier, 2011 [27]	Support Care Cancer (2.4)	HADS in cancer patients	6 (43%)	250
Brennan, 2010 [16]	J Psychosom Res (2.7)	HADS in any setting	5 (36%)	250
Mitchell, 2010a [22]	Am J Geriatr Psychiatry (4.2)	GDS in older patients	3 (21%)	250
Mitchell, 2010b [24]	J Affect Disord (3.4)	HADS in cancer and palliative settings	3 (21%)	250
Mitchell, 2010c [21]	J Affect Disord (3.4)	GDS in older primary care patients	3 (21%)	250

Assessing the quality of diagnostic test accuracy meta-analyses

Hewitt, 2009 [28]	Health Technol Assess (5.0)	Screening tools in women in pregnancy or postpartum	8 (57%)	500
Mitchell, 2008 [20]	Br J Cancer (4.8)	Short screening tools in cancer and palliative care	5 (36%)	200
Gilbody, 2007 [15]	J Gen Intern Med (3.4)	PHQ in medical settings	6 (43%)	300
Mitchell, 2007 [23]	Br J Gen Pract (2.3)	Ultra-short screening tools in primary care	4 (29%)	250
Wittkampf, 2007 [19]	Gen Hosp Psychiatry (2.6)	PHQ in any setting	6 (43%)	200

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

Assessing the quality of diagnostic test accuracy meta-analyses

Table 2. Adapted PRISMA for Abstracts Item Totals for the 21 Meta-Analyses Reviewed

Adapted PRISMA for Abstracts Item	Adapted Description	Proportion of Meta-Analyses with '*yes*' ratings (%)
Item 1	Title: Identify the report as a systematic review, meta-analyses or both.	20 (95%)
Item 2	Objectives: The research question including components such as participants, index test, reference standard and outcomes.	0 (0%)
Item 3a	Eligibility criteria: study characteristics used as criteria for inclusion.	1 (5%)
Item 3b	Eligibility criteria: report characteristics used as criteria for inclusion.	0 (0%)
Item 4a	Information sources: Key databases searched.	9 (43%)
Item 4b	Information sources: Key search dates.	7 (33%)
Item 5	Risk of bias: Methods of assessing risk of bias.	3 (14%)
Item 6	Included studies: Number and type of included studies and participants and relevant characteristics of studies.	2 (10%)
Item 7	Results of the principle summary measures (e.g., sensitivity and specificity, diagnostic odds ratio).	16 (76%)
Item 8	Description of outcomes: summary of accuracy outcomes in terms meaningful to clinicians and patients.	3 (14%)
Item 9	Strengths and Limitations of evidence: Brief summary of strengths and limitations of	7 (33%)

Assessing the quality of diagnostic test accuracy meta-analyses

	evidence (e.g., inconsistency, imprecision, indirectness, or risk of bias, other supporting or conflicting evidence).	
Item 10	Interpretation: General interpretation of the results and important implications.	20 (95%)
Item 11	Funding: Primary source of funding for the review.	0 (0%)
Item 12	Registration: Registration number and registry name.	1 (5%)

Assessing the quality of diagnostic test accuracy meta-analyses

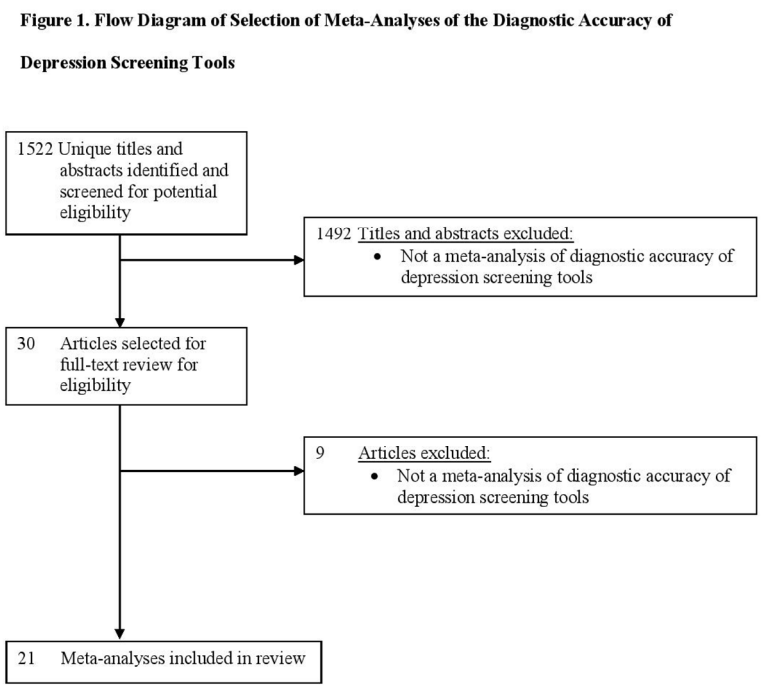
Table 3. PRISMA for Abstracts Item by Item Ratings

Reference	Item 1	Item 2	Item 3a	Item 3b	Item 4a	Item 4b	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Total 'Yes'
Pocklington, 2016 [34]	Yes	No	No	No	No	No	Yes	No	Yes	No	Yes	Yes	No	No	5 (36%)
Bosanquet, 2015 [31]	Yes	No	No	No	No	No	No	No	Yes	No	Yes	Yes	No	Yes	5 (36%)
Moriarty, 2015 [33]	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	No	7 (50%)
Stockings, 2015 [35]	Yes	No	No	No	Yes	No	No	No	Yes	No	Yes	Yes	No	No	5 (36%)
Manea, 2015 [32]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Meader, 2014 [29]	Yes	No	No	No	No	Yes	No	No	Yes	No	No	Yes	No	No	4 (29%)
Tsai, 2014 [25]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Tsai, 2013 [26]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)
Mitchell, 2012 [30]	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	4 (29%)
Manea, 2012 [18]	Yes	No	No	No	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	6 (43%)
Meader, 2011 [17]	Yes	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes	No	No	5 (36%)
Vodermaier, 2011 [27]	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	No	No	3 (21%)
Brennan, 2010 [16]	Yes	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes	No	No	5 (36%)
Mitchell, 2010a [22]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)
Mitchell, 2010b [24]	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	2 (14%)
Mitchell, 2010c [21]	Yes	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No	3 (21%)

Assessing the quality of diagnostic test accuracy meta-analyses

Hewitt, 2009 [28]	No	No	No	No	Yes	No	Yes	No	Yes	No	No	Yes	No	No	4 (29%)
Mitchell, 2008 [20]	Yes	No	No	No	No	Yes	No	No	Yes	No	No	Yes	No	No	4 (29%)
Gilbody, 2007 [15]	Yes	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	7 (50%)
Mitchell, 2007 [23]	Yes	No	No	No	No	No	No	No	Yes	Yes	No	Yes	No	No	4 (29%)
Wittkampf, 2007 [19]	Yes	No	No	No	Yes	Yes	Yes	No	Yes	No	No	Yes	No	No	6 (43%)
Total 'Yes'	20 (95%)	0 (0%)	1 (5%)	0 (0%)	9 (43%)	7 (33%)	3 (14%)	2 (10%)	16 (76%)	3 (14%)	7 (33%)	20 (95%)	0 (0%)	1 (5%)	

Note: Item 1= title; Item 2= objectives; Item 3a= eligibility criteria study characteristics; Item 3b= eligibility criteria report characteristics; Item 4a= databases searched; Item 4b= search dates; Item 5= risk of bias; Item 6= included studies; Item 7= synthesis of results; Item 8= description of outcomes; Item 9= strengths and limitations of evidence; Item 10= interpretation; Item 11= funding; Item 12= registration.



Flow Diagram of Selection of Meta-Analyses of the Diagnostic Accuracy of Depression Screening Tools
Figure 1
107x139mm (300 x 300 DPI)

S1 Appendix. Search Strategy

Ovid MEDLINE In-Process & Other Non-Indexed Citations and Ovid MEDLINE

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
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

S2 Appendix. Original and Adapted PRISMA for Abstract Tools

Item	Description	Adjusted for DTA	No	Yes	Items
Item 1	Title: Identify the report as a systematic review, meta-analyses or both.	----------	The title does not include words "systematic review" or "meta-analysis".	The title includes words "systematic review" or "meta-analysis".	
Item 2	Objectives: The research question including components such as participants, interventions, comparators and outcomes.	Objectives: The research question including components such as participants, index test, reference standard and outcomes.	There is no explicit statement of the questions being addressed or there is an objective statement, but it does not include a reference to each of the following (PIRO): participants, index text, reference standards and outcomes.	There is an explicit statement of the questions being addressed or an objectives statement with reference to each of the following (PIRO): participants, index test, reference standards, and outcomes.	A statement that refers to all PIRO components must be found in the objectives or research statement section of the abstract to be coded as "yes". Review questions may be narrowly focused or broad.
Item 3a*	Eligibility criteria: study characteristics used as criteria for inclusion.	----------	One or more eligibility criteria are omitted and there is no statement that eligibility criteria had no restrictions.	Study eligibility criteria are stated for all PIRO study characteristics or there is a statement that there were no restrictions for eligibility criteria.	This item considers characteristics of the primary study itself. All PIRO components must be specified. If a study had no restrictions and includes any participants who completed a screening measure and assessment, this must be stated.
Item 3b*	Eligibility criteria: report characteristics used as criteria for inclusion.	----------	Report characteristics considered for eligibility are not stated or there is a statement, but it does not address language of	Report characteristics considered for eligibility are clearly stated, including, at least, language of publication and publication status.	This item considers characteristics of the report of the primary study.

			publication and publication status.		
Item 4a†	Information sources: Key databases searched.	----------	Does not list all databases searched.	Lists all databases searched (or at least 3 if more than 3 searched).	
Item 4b†	Information sources: Key search dates.	----------	Does not state the dates of coverage of the search, including year and month of end date.	The dates of coverage of the search are provided, including year and month of end date.	
Item 5	Risk of bias: Methods of assessing risk of bias.	----------	There is no statement about how risk of bias was assessed, including the name of the tool for assessing risk of bias is listed. Alternatively, if risk of bias was not assessed, this is not stated.	There is a statement about how risk of bias was assessed, including the name of the tool used to assess bias is listed, or there is a statement that risk of bias was not assessed.	Assessments of study "quality" are coded as assessing risk of bias. All elements must be included to be coded “Yes”.
Item 6	Included studies: Number and type of included studies and participants and relevant characteristics of studies.	----------	The number of primary studies, total number of participants and cutoffs included in the analyses are not provided.	The number of primary studies, total number of participants and cutoffs assessed are provided.	All elements must be included to be coded “Yes”.
Item 7	Synthesis of results: Results for main outcomes (benefits and harms), preferably indicating the number of studies and participants for each. If meta-analysis was done, include summary measures and confidence	Results of the principle summary measures (e.g., sensitivity and specificity, diagnostic odds ratio).	Results of the principle summary measures are not provided (e.g., sensitivity and specificity, diagnostic odds ratio).	Results of the principle summary measures are provided (e.g., sensitivity and specificity, diagnostic odds ratio).	If sensitivity and specificity are the primary outcome measures, both must be reported to code "Yes".

	intervals.				
Item 8	Description of effect: direction of the effect and size of the effect in terms meaningful to clinicians and patients.	Description of outcomes: summary of accuracy outcomes in terms meaningful to clinicians and patients.	Results do not summarize the principle results in words and numbers, including the most accurate cutoff to use and how the screening tools would perform in practice, in terms meaningful to clinicians and patients (e.g. positive and negative predictive values or true and false positive rates).	Results summarize the principle results in words and numbers, including the most accurate cutoff to use and how the screening tools would perform in practice, in terms meaningful to clinicians and patients (e.g. positive and negative predictive values or true and false positive rates).	
Item 9	Strengths and Limitations of evidence: Brief summary of strengths and limitations of evidence (e.g., inconsistency, imprecision, indirectness, or risk of bias, other supporting or conflicting evidence).	----------	Relevant limitations of the diagnostic accuracy evidence are not noted (e.g., inconsistency, imprecision, indirectness, risk of bias, other supporting or conflicting evidence).	Relevant limitations of the diagnostic accuracy evidence are noted (e.g., inconsistency, imprecision, indirectness, risk of bias, other supporting or conflicting evidence).	
Item 10	Interpretation: General interpretation of the results and important implications.	----------	Authors do not provide a statement about clinical implications of results or need for more research if results suggest uncertainty.	Authors provide a statement about clinical implications of results or need for more research if results suggest uncertainty.	
Item 11	Funding: Primary source of funding for the review.	----------	Authors do not provide the sources of funding for the review or a statement that it was not funded.	Authors provide the source of funding for the review or a statement that it was not funded.	For a "yes", funding information must be listed in the abstract that is available for viewing on an online database such as PubMed.

Item 12	Registration: Registration number and registry name.	----------	A registration number and/or registry name are not provided.	A registration number and registry name are provided.	Cochrane reviews are an exception to this requirement, as they are preceded by a peer reviewed protocol that is published in the Cochrane Library and can be downloaded from there. For a "yes", registration number and/or name must be listed in the abstract that is available for viewing on an online database such as PubMed.

* = if either item 3a or 3b is coded as "no", the original item 3 would also be coded as "no".
† = if either item 4a or 4b is coded as "no", the original item 4 would also be coded as "no".

S4 Appendix. Sources of Journal Word Limits

Author, Year	Name of Journal	URL Website	Abstract Word Count Limit
Pocklington, 2016	Int J Geriatr Psychiatry	http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1166/homepage/ForAuthors.html	250
Bosanquet, 2015	BMJ Open	http://bmjopen.bmj.com/site/about/guidelines.xhtml	300
Moriarty, 2015	Gen Hosp Psychiatry	https://www.elsevier.com/journals/general-hospital-psychiatry/0163-8343/guide-for-authors	200
Stockings, 2015	J Affect Disord	https://www.elsevier.com/journals/journal-of-affective-disorders/0165-0327/guide-for-authors#39000	250
Manea, 2015	Gen Hosp Psychiatry	https://www.elsevier.com/journals/general-hospital-psychiatry/0163-8343/guide-for-authors	200
Meader, 2014	J Neurol Neurosurg Psychiatry	http://jnnp.bmj.com/site/about/guidelines.xhtml	250
Tsai, 2014	JAIDS	http://edmgr.ovid.com/jaids/accounts/ifauth.htm	250
Tsai, 2013	PLoS One	http://journals.plos.org/plosone/s/submission-guidelines#loc-abstract	300
Mitchell, 2012	J Affect Disord	http://www.elsevier.com/journals/journal-of-affective-disorders/0165-0327?generatepdf=true	250

Manea, 2012	CMAJ	http://www.cmaj.ca/site/authors/preparing.xhtml#_Preparing_text_for	250
Meader, 2011	Br J Gen Pract	http://bjgp.org/authors/writing-for-bjgp-research	250
Vodermaiier 2011	Support Care Cancer	http://www.springer.com/medicine/oncology/journal/520	250
Brennan, 2010	J Psychosom Res	http://www.elsevier.com/journals/journal-of-psychosomatic-research/0022-3999/guide-for-authors	250
Mitchell, 2010a	Am J Geriatr Psychiatry	http://www.editorialmanager.com/jgp/account/InstructionsForAuthors.pdf	250
Mitchell, 2010b	J Affect Disord	http://www.elsevier.com/journals/journal-of-affective-disorders/0165-0327?generatepdf=true	250
Mitchell, 2010c	J Affect Disord	http://www.elsevier.com/journals/journal-of-affective-disorders/0165-0327?generatepdf=true	250
Hewitt, 2009	Health Technol Assess	http://www.journalslibrary.nihr.ac.uk/information-for-authors/abstract	500
Mitchell, 2008	Br J Cancer	http://www.nature.com/bjc/authors/submit.html#manuscript-format	200
Mitchell, 2007	J Gen Intern Med	http://bjgp.org/authors/writing-for-bjgp-research	250
Wittkampf, 2007	Br J Gen Pract	http://www.elsevier.com/journals/general-hospital-psychiatry/0163-8343/guide-for-authors#39001	200
Gilbody, 2007	Gen Hosp Psychiatry	http://www.jgimed.org/authors/jgim%20instructions%20for%20authors.pdf	300

S3 Appendix. 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, Gibbs FA, Williams JR, Tregear S. Peer Reviewed: Depression Screening and Treatment Among Nonpregnant Women of Reproductive Age in the United States, 1990-2010. Preventing chronic disease. 2011 Nov;8(6).
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(3):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(4):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;25(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 (Summ). 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(5):423-32.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	NA
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7-8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7-10
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-10
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	11, figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	14-15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	16

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097