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 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for Abstracts tool.

Design: Cross-sectional study.

Inclusion Criteria: We searched MEDLINE and PsycINFO from 1 January 2005 through 13 March 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 the 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) A Measurement Tool to Assess Systematic Reviews (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 Abstracts scores items reported (r=0.45).

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

INTRODUCTION

Researchers, clinicians and other consumers of research often rely primarily on information found in abstracts of systematic reviews. 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. This may be due to time limitations, accessibility constraints or language barriers. 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. This goal, however, is infrequently achieved, as the quality and completeness of information provided in abstracts of systematic reviews are often suboptimal.
the PRISMA statement, with the goal of improving the quality and completeness of abstracts in systematic reviews, including meta-analyses. 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. 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). In published conference abstracts and published articles, nine of the 12 PRISMA for Abstracts items were completed in <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. 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. The third study assessed abstract reporting based on the presence or absence of seven characteristics related to the meta-analyses results. 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. Based on indirect evidence, including evidence on screening tool accuracy, the US Preventative Services Task Force recently recommended universal depression screening in all adults. The UK National Screening Committee and the Canadian Task Force on Preventative Healthcare, however, recommend against depression screening due to a lack of evidence from randomised controlled trials that depression screening would improve mental health outcomes.

No published studies have evaluated the completeness of reporting in abstracts of DTA 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. 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 Abstracts 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 DTA for depression screening tools. We searched MEDLINE and PsycINFO (both on the OvidSP platform) from 1 January 2005 through 13 March 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, 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 online supplementary 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 (eg, sensitivity, specificity, diagnostic odds ratio) of one or more depression screening tools compared with a reference standard diagnosis of depression based on a clinical interview or validated diagnostic interview (eg, 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, Maryland, 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 DTA. 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. 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, BDT, 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 online supplementary 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 5 was originally ‘Study and report characteristics used as criteria for inclusion’ and was adapted to item 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 (ie, eligible participants, index tests, reference standards and outcomes), and characteristics for inclusion in the systematic review and meta-analyses (eg, 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 text within the systematic review rather than the interventions and comparators found in intervention studies. Item 7 was adjusted to encompass results of the principal summary measures (eg, sensitivity, specificity, positive predictive value, negative predictive value) that are reported in DTA studies. Finally, 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 online supplementary S3 appendix for details) and previously published A Measurement Tool to Assess Systematic Reviews (AMSTAR) quality ratings. Accuracy was verified by a second investigator. Two investigators independently rated each included systematic review 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 meta-analysis publication 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 V22.0 (Chicago, Illinois, USA); statistical tests were two-sided with a p<0.05 significance level and 95% CIs 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 online supplementary S4 appendix), resulting in 21 eligible systematic reviews with meta-analyses published between 2007 and 2016 (see figure 1). 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 seven to nine of 21 meta-analyses (33–43%): 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 eight 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 seven of the 14 adapted PRISMA for Abstracts items. An additional seven meta-analyses received ratings of yes for 516 17 31 34 35 and 618 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 and 300 words.

DISCUSSION

The main findings of this study were that only three 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 two 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 Abstracts 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, three 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 three 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. 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 <50% of included primary studies adequately reported the study population, setting, patient sampling, blinding, cut-offs used and CIs around accuracy estimates. 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 and, generally, journal endorsement of reporting guidelines improves the completeness of reporting. The Consolidated Standards of Reporting Trials (CONSORT) reporting guidelines for abstracts of randomised controlled trials was published in 2009, and a recent study found that journals that implement these guidelines have improved reporting in abstracts of randomised controlled trials. As of 6 April 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 weblink 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

| First Author, year of publication | Journal (2014 impact factor) | Focus of meta-analysis | AMSTAR scores | Journal word limit |
|-----------------------------------|-----------------------------|------------------------|---------------|-------------------|
| Pocklington, 2016               | Int J Geriatr Psychiatry (2.9) | Brief versions of the GDS in older patients | 8 (57%) | 250 |
| Bosanquet, 2015                 | BMJ Open (2.3)              | Whooley questions in any setting | 9 (64%) | 300 |
| Moriarty, 2015                  | Gen Hosp Psychiatry (2.6)   | PHQ-9 in any setting    | 9 (64%) | 200 |
| Stockings, 2015                 | J Affect Disorder (3.4)     | Screening tools in children and adolescents | 4 (29%) | 250 |
| Manea, 2015                     | Gen Hosp Psychiatry (2.6)   | PHQ-9 with algorithm scoring method in any setting | 8 (57%) | 200 |
| Meader, 2014                    | J Neurol Neurosurg Psychiatry (6.8) | Screening tools in poststroke patients | 6 (43%) | 250 |
| Tsai, 2014                      | JAIDS (4.6)                 | Screening tools in HIV-positive adults in Africa | 5 (36%) | 250 |
| Tsai, 2013                      | PLOS One (3.2)              | Screening tools in pregnancy or postpartum in Africa | 6 (43%) | 300 |
| Mitchell, 2012                  | J Affect Disorder (3.4)     | Screening tools in patients with cancer | 4 (29%) | 250 |
| Manea, 2012                     | CMAJ (6.0)                  | PHQ-9 in any setting | 10 (71%) | 250 |
| Meader, 2011                    | Br J Gen Pract (2.3)        | Screening tools in patients with chronic health problems | 5 (36%) | 250 |
| Vodermaier, 2011                | Support Care Cancer (2.4)   | HADS in cancer patients | 6 (43%) | 250 |
| Brennan, 2010                   | J Psychosom Res (2.7)       | HADS in any setting | 5 (36%) | 250 |
| Mitchell, 2010a                 | Am J Geriatr Psychiatry (4.2) | GDS in older patients | 3 (21%) | 250 |
| Mitchell, 2010b                 | J Affect Disorder (3.4)     | HADS in cancer and palliative settings | 3 (21%) | 250 |
| Mitchell, 2010c                 | J Affect Disorder (3.4)     | GDS in older primary care patients | 3 (21%) | 250 |
| Hewitt, 2009                    | Health Technol Assess (5.0) | Screening tools in women in pregnancy or postpartum | 8 (57%) | 500 |
| Mitchell, 2006                  | Br J Cancer (4.8)           | Short screening tools in cancer and palliative care | 5 (36%) | 200 |
| Gilbody, 2007                   | J Gen Intern Med (3.4)      | PHQ in medical settings | 6 (43%) | 300 |
| Mitchell, 2007                  | Br J Gen Pract (2.3)        | Ultra-short screening tools in primary care | 4 (29%) | 250 |
| Witkampf, 2007                  | Gen Hosp Psychiatry (2.6)   | PHQ in any setting | 6 (43%) | 200 |

AMSTAR, A Measurement Tool to Assess Systematic Reviews; GDS, Geriatric Depression Scale; HADS, Hospital Anxiety and Depression Scale; PHQ, Patient Health Questionnaire.
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, multiquestion 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 (eg, *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 Abstracts 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 our 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 inter-rater 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

### 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 principal summary measures (eg, sensitivity and specificity, diagnostic OR). | 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 evidence (eg, inconsistency, imprecision, indirectness or risk of bias, other supporting or conflicting evidence). | 7 (33%) |
| 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%) |

PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
| 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  | Yes    | No     | No      | No      | No      | No      | Yes    | No     | Yes    | No     | Yes    | No      | No      | No       | 5 (36%)   |
| Bosanquet, 2015    | Yes    | No     | No      | No      | No      | No      | Yes    | No     | Yes    | No     | Yes    | No      | Yes     | No        | 5 (36%)   |
| Moriarty, 2015     | Yes    | No     | Yes     | No      | No      | No      | Yes    | No     | Yes    | No     | Yes    | No      | Yes     | No        | 7 (50%)   |
| Stockings, 2015    | Yes    | No     | No      | No      | No      | No      | Yes    | No     | Yes    | No     | Yes    | No      | No      | No        | 5 (36%)   |
| Manea, 2015        | Yes    | No     | No      | No      | No      | No      | No     | No     | No     | No     | Yes    | No      | No      | No        | 2 (14%)   |
| Meader, 2014       | Yes    | No     | No      | No      | No      | Yes     | No     | No     | Yes    | No     | No     | No      | No      | No        | 4 (29%)   |
| Tsai, 2014         | Yes    | No     | No      | No      | No      | No      | No     | No     | No     | No     | Yes    | No      | No      | No        | 2 (14%)   |
| Tsai, 2013         | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | No     | No     | Yes     | No      | No        | 3 (21%)   |
| Mitchell, 2012     | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | Yes    | Yes    | Yes     | Yes     | No        | 4 (29%)   |
| Manea, 2012        | Yes    | No     | No      | Yes     | Yes     | No      | Yes    | Yes    | No     | Yes    | No     | No      | No      | No        | 6 (43%)   |
| Meader, 2011       | Yes    | No     | No      | Yes     | Yes     | No      | No     | No     | Yes    | No     | Yes    | Yes     | No      | No        | 5 (36%)   |
| Vodermaier, 2011   | Yes    | No     | No      | Yes     | No      | No      | No     | Yes    | No     | No     | No     | No      | No      | No        | 3 (21%)   |
| Brennan, 2010      | Yes    | No     | No      | Yes     | Yes     | No      | No     | No     | No     | Yes    | No     | No      | No      | No        | 5 (36%)   |
| Mitchell, 2010a    | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | No     | No     | Yes     | No      | No        | 3 (21%)   |
| Mitchell, 2010b    | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | Yes    | Yes    | Yes     | Yes     | No        | 2 (14%)   |
| Mitchell, 2010c    | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | No     | No     | Yes     | Yes     | No        | 3 (21%)   |
| Hewitt, 2009       | No     | No     | No      | Yes     | No      | Yes     | No     | No     | Yes    | No     | Yes    | No      | No      | No        | 4 (29%)   |
| Mitchell, 2008     | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | No     | No     | Yes     | Yes     | No        | 4 (29%)   |
| Gilbody, 2007      | Yes    | No     | No      | Yes     | Yes     | No      | No     | No     | Yes    | Yes    | Yes    | Yes     | Yes     | No        | 7 (50%)   |
| Mitchell, 2007     | Yes    | No     | No      | No      | No      | No      | No     | No     | Yes    | Yes    | Yes    | Yes     | Yes     | No        | 4 (29%)   |
| Wittkampf, 2007    | Yes    | No     | No      | No      | No      | No      | No     | Yes    | Yes    | No     | Yes    | No      | 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%)    |

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.

PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
specifically for reviews of DTA. We also attempted to analyse the association between journal word limits and the PRISMA for Abstracts scores; however, 20 of 21 meta-analyses included in our study were published in journals with word limits of 200–300 words.

In conclusion, the present study found that only two of 21 existing meta-analyses of the diagnostic accuracy of depression screening tools met at least 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 on 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.

Author affiliations
1Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada
2Department of Psychiatry, McGill University, Montréal, Québec, Canada
3Library, Concordia University, Montréal, Québec, Canada
4Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
5Department of Psychology, McGill University, Montréal, Québec, Canada
6Department of Medicine, McGill University, Montréal, Québec, Canada
7Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada
8School of Nursing, McGill University, Montréal, Québec, Canada

Contributors 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 DBR is supported by a Fonds de Recherche Santé Québec (FRSQ) Master’s Award. BDT receives support from an Investigator Award from the Arthritis Society. There was no specific funding for this study.

Disclaimer 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 take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available. Full data extraction data set is available in the tables and supplementary data files.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

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, Glassziou PP, Altman DG, et al. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. PLoS Med 2013;10:e1001419.
3. Beller EM, Glassziou 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 Periodontol Res 2014;49:137–42.
5. Kriakou 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, Bouton 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. Thoms BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care?. BMJ 2014;348:g1223.
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, Shnier I, Kloda LA, et al. Methodological quality of meta-analyses of the diagnostic accuracy of depression screening tests. J Psychosom Res 2016;84:84–92.
14. Thoms 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;29:388–94.
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;125:10–17.
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 Psychiatr* 2014;85:196–206.

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;31:837–57.

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.