Journal of Psychosomatic Research 84 (2016) 84–92

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

Methodological quality of meta-analyses of the diagnostic accuracy of depression screening tools

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

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

ARTICLE INFO

Article history:
Received 22 January 2016
Received in revised form 15 March 2016
Accepted 17 March 2016

Keywords:
AMSTAR
Depression
Diagnostic test accuracy
Meta-analyses
Screening
Quality

ABSTRACT

Objective: Concerns have been raised that primary studies of diagnostic accuracy of depression screening tools may exaggerate estimates of accuracy and that this could also influence the results of meta-analyses. No studies, however, have evaluated the quality of meta-analyses of depression screening tools. Our objective was to evaluate the quality of meta-analyses of the diagnostic accuracy of depression screening tools.
Methods: We searched MEDLINE and PsycINFO from January 1, 2005 through March 13, 2016 for recent meta-analyses in any language on the diagnostic accuracy of depression screening tools. Two reviewers independently assessed methodological quality using the AMSTAR tool with appropriate adaptations made for studies of diagnostic test accuracy.
Results: We identified 21 eligible meta-analyses. The majority provided a list of included studies (100%), included a comprehensive literature search (95%) and assessed risk of bias of included studies (71%). Meta-analyses less consistently included non-published evidence (38%), listed excluded studies (33%), incorporated risk of bias findings into conclusions (33%), and assessed selective cutoff reporting (29%). Meta-analyses rarely reported that duplicate study selection or data extraction occurred (14%), mentioned 'a priori' protocols (10%), or reported on conflicts of interest (0%) or funding sources (0%) of primary studies. Only 6 of 21 included meta-analyses complied with at least 7 of 14 adapted AMSTAR items.
Conclusions: The methodological quality of most meta-analyses of the diagnostic test accuracy of depression screening tools is suboptimal. Improving quality will reduce the risk of inaccurate estimates of accuracy and inappropriate inferences.

1. Introduction

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

The United States Preventative Services Task Force recently published an updated guideline which recommends universal depression screening in all adults, including all pregnant and postpartum women [1]. The UK National Screening Committee and the Canadian Task Force on Preventative Health Care (CTFPHC) recommend against depression screening due to the lack of evidence that depression screening would improve depressive symptoms or reduce the number of patients with depression [6,7]. In its 2013 guideline, the CTFPHC expressed specific concern that published studies of the diagnostic accuracy of depression screening tools may exaggerate accuracy estimates [7]. Numerous specialty medical societies recommend depression screening in inpatient and outpatient settings (e.g., cancer, diabetes, heart disease, stroke) [8–14], but these recommendations are not based on systematic evidence reviews.

Concerns have been raised about the quality of existing primary studies on depression screening tool accuracy. Many primary studies have

* Corresponding author at: Jewish General Hospital, 4333 Ste. Catherine Road, Montreal, Quebec H3T 1E4, Canada.
E-mail address: brett.thombs@mcgill.ca (B.D. Thombs).

http://dx.doi.org/10.1016/j.jpsychores.2016.03.013
0022-3999/

been conducted in samples too small to provide precise estimates [15]. As a result, cutoff scores identified as optimal vary dramatically across studies [16,17]. Many of these studies, however, selectively report accuracy results from a data-driven optimal cutoff and a small range of alternative cutoffs around it, and the cutoffs for which data are reported are not consistent across studies [18–21]. Another concern relates to the inclusion of patients already diagnosed or being treated for depression in these studies, even though these patients would not be screened in clinical practice. Approximately 95% of primary studies on the diagnostic accuracy of depression screening tools include already diagnosed or treated patients. Including these already diagnosed patients would overestimate the ability of a tool to identify previously unidentified patients who would be detected by a screening tool [22,23].

High-quality systematic reviews and meta-analyses can highlight shortcomings in primary studies. They can also provide guidance on how to improve research in order to address important health care questions. However, this can only occur to the degree that systematic reviews and meta-analyses are conducted rigorously, reflecting current standards for evidence synthesis [24,25]. No studies, however, have evaluated the quality of existing systematic reviews or meta-analyses of the diagnostic accuracy of depression screening tools.

The Assessing the Methodological Quality of Systematic Reviews (AMSTAR) tool is an 11-item checklist that was developed to assess the scientific quality and rigor of systematic reviews for treatment effects from randomized trials [26]. In the absence of an AMSTAR tool designed for studies of diagnostic test accuracy, we applied AMSTAR with some items adapted to reflect issues related to the assessment of the quality of diagnostic test accuracy systematic reviews and meta-analyses. The primary objective of our study was to evaluate the quality of meta-analyses of the diagnostic accuracy of depression screening tools published in journals indexed in the MEDLINE and PsycINFO databases.

2. Materials and methods

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

We searched MEDLINE and PsycINFO (both on the OvidSP platform) from January 1, 2005 through 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 [18], by adding search terms designed to restrict the results to meta-analyses. The strategy was then adapted for PsycINFO. A medical librarian adapted the meta-analysis search strategies and conducted the search. The complete search strategies used for MEDLINE and PsycINFO can be found in Appendix A.

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

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

2.2. Assessment of methodological quality

The methodological quality of the included papers was evaluated using an adapted version of the AMSTAR tool [26]. The AMSTAR checklist was developed to facilitate the conduct of high-quality reviews of treatment effects from randomized trials, and to provide a valid, reliable, and usable instrument to help differentiate between the methodological quality of systematic reviews using an 11-item checklist [26]. The response options for each item of the original AMSTAR checklist are: *yes*, *no*, *can't answer* and *not applicable*. Although developed for systematic reviews of randomized trials, many of the items are applicable to other designs, including systematic reviews and meta-analysis of diagnostic test accuracy. Previous studies have applied AMSTAR to systematic reviews of diagnostic test accuracy [27] and non-randomized observational studies [28], but in both cases, the authors pointed out the need for adaptations.

As there is currently no quality assessment tool for systematic reviews that include only primary studies of diagnostic test accuracy, we adapted the original AMSTAR tool so that items were applicable to diagnostic test accuracy studies (see Appendix B for details). The team that adapted the items for applicability in this study included members with expertise in evidence synthesis (IS, BDT, AB, LAK), information sciences for evidence synthesis (LAK), diagnostic test accuracy of depression screening tools (BDT, AB) and statistical analysis for diagnostic test accuracy meta-analyses (AB). We also consulted outside experts and referred to the Cochrane Handbook for Diagnostic Test Accuracy Meta-Analyses. Each original AMSTAR item was reviewed by team members, who considered ease of coding and applicability to studies of diagnostic test accuracy, then either accepted the item as appropriate or edited the item to better reflect practices in the conduct of systematic reviews of diagnostic test accuracy. In addition, a coding manual was developed with specific criteria for *yes* and *no* ratings, along with additional coding notes.

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