

BMJ Open Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses

Brett D Thombs,^{1,2,3,4,5,6} Andrea Benedetti,^{3,4,7} Lorie A Kloda,⁸ Brooke Levis,^{1,3} Marleine Azar,^{1,3} Kira E Riehm,¹ Nazanin Saadat,¹ Pim Cuijpers,⁹ Simon Gilbody,¹⁰ John P A Ioannidis,^{11,12} Dean McMillan,¹⁰ Scott B Patten,^{13,14} Ian Shrier,^{1,3} Russell J Steele,^{1,15} Roy C Ziegelstein,¹⁶ Carmen G Loiselle,^{1,17} Melissa Henry,^{1,18} Zahinoor Ismail,^{13,17} Nicholas Mitchell,¹⁹ Marcello Tonelli²⁰

To cite: Thombs BD, Benedetti A, Kloda LA, *et al.* Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open* 2016;**6**:e011913. doi:10.1136/bmjopen-2016-011913

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-011913>).

Received 15 March 2016
Accepted 18 March 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Brett D Thombs;
brett.thombs@mcgill.ca

ABSTRACT

Introduction: The Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) has been recommended for depression screening in medically ill patients. Many existing HADS-D studies have used exploratory methods to select optimal cut-offs. Often, these studies report results from a small range of cut-off thresholds; cut-offs with more favourable accuracy results are more likely to be reported than others with worse accuracy estimates. When published data are combined in meta-analyses, selective reporting may generate biased summary estimates. Individual patient data (IPD) meta-analyses can address this problem by estimating accuracy with data from all studies for all relevant cut-off scores. In addition, a predictive algorithm can be generated to estimate the probability that a patient has depression based on a HADS-D score and clinical characteristics rather than dichotomous screening classification alone. The primary objectives of our IPD meta-analyses are to determine the diagnostic accuracy of the HADS-D to detect major depression among adults across all potentially relevant cut-off scores and to generate a predictive algorithm for individual patients. We are already aware of over 100 eligible studies, and more may be identified with our comprehensive search.

Methods and analysis: Data sources will include MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, PsycINFO and Web of Science. Eligible studies will have datasets where patients are assessed for major depression based on a validated structured or semistructured clinical interview and complete the HADS-D within 2 weeks (before or after). Risk of bias will be assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Bivariate random-effects meta-analysis will be

Strengths and limitations of this study

- The study will use individual patient data to estimate diagnostic accuracy for all relevant cut-off scores of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D). Using data from all patients at each cut-off score will overcome limitations related to selective cut-off reporting in primary study publications.
- The study will conduct analyses that exclude patients with current diagnoses of depression or who are undergoing mental health treatment at the time of study enrolment, as these patients would not be screened in clinical practice. This will overcome potential bias in primary diagnostic test accuracy studies where these patients are often included.
- The study will generate a predictive model to estimate the probability that a patient with a particular HADS-D score and relevant covariates typically available in clinical practice has depression. This will facilitate more informed clinical decision-making than can be done with standard diagnostic accuracy metrics.
- A potential limitation is that the success of the study depends on the ability to obtain the relevant individual patient data and to avoid selective availability of studies with better or worse accuracy results. We do not know the proportion of eligible datasets that will be possible to include in the study.

conducted for the full range of plausible cut-off values, and a predictive algorithm for individual patients will be generated.

Ethics and dissemination: The findings of this study will be of interest to stakeholders involved in research, clinical practice and policy.

BACKGROUND

Major depressive disorder (MDD) may be present in 10–20% of patients with acute and chronic medical conditions and is independently associated with poor prognosis.^{1–6} Healthcare teams in non-psychiatric settings, however, where the majority of depression care is provided, typically have little formal mental health training,⁷ and mental healthcare is often inconsistent. Many depressed patients are not diagnosed, and a high proportion of patients treated for depression do not meet diagnostic criteria.^{8–12} It has been suggested that routine screening for depression may be a solution,¹³ but this is controversial.^{14–18}

A necessary, but not sufficient, condition for depression screening to benefit patients is a screening tool with demonstrated high diagnostic accuracy. In order to correctly identify ideal cut-off screening thresholds and to obtain unbiased diagnostic accuracy estimates, several major problems common to studies of depression screening tools must be addressed. These include (1) the overreliance on primary studies with small sample sizes that selectively report results from only well-performing cut-off thresholds;^{19–20} (2) the inclusion in primary studies of patients already being treated for depression, even though these patients are not screened in actual practice, since screening is done to detect unrecognised cases;^{21–22} and (3) the lack of consideration of individual patient depression risk factors (eg, age, sex, inpatient vs outpatient care) in estimates of screening accuracy.²³

The Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) is the most commonly used screening tool in medically ill patients.^{24–26} Most existing studies of the accuracy of the HADS-D have (1) been conducted in samples too small to precisely estimate accuracy, (2) included already-diagnosed and treated patients and (3) selectively published accuracy results from cut-offs that perform well, but not other relevant cut-offs.^{25–29} For instance, in one meta-analysis of the diagnostic accuracy of the HADS-D,²⁷ the authors excluded 16 of 41 (39%) primary studies that examined the diagnostic accuracy of the HADS because those studies only reported results from study-specific optimal cut-offs, but not from standard cut-offs, which were pooled in the meta-analysis. The inability to include studies that did not publish more pessimistic results associated with standard HADS cut-offs likely led to inflated estimates of accuracy in the meta-analysis compared to real-world performance using standard cut-offs. In other meta-analyses,^{25–28–29} authors have reported results from a single bivariate accuracy model that included results from different cut-offs from different primary studies,

sometimes from the best-performing sample-specific cut-off in each study.

Individual patient data (IPD) meta-analysis can potentially address some of these problems. IPD meta-analysis involves using actual patient data obtained from researchers who conducted primary studies, rather than summary results from published or unpublished study reports.³⁰ The steps involved in conducting a systematic review with an IPD meta-analysis, in terms of defining a research question, establishing study inclusion and exclusion criteria, identifying and screening studies and analysing data, are similar to those in a traditional systematic review and meta-analysis and diverge only in analysing individual-level data rather than summary data.³¹ In the context of evaluating the diagnostic accuracy of depression screening tools, IPD meta-analysis has a number of potential advantages compared with a conventional meta-analysis. First, it can address bias from the selective reporting in publications of only well-performing cut-off thresholds since accuracy can be evaluated across all relevant cut-off scores. Second, it allows the systematic exclusion of already-treated patients, for whom the tool would not be used to screen for unidentified depression. Third, IPD meta-analysis with large numbers of patients and large numbers of depression cases allows the incorporation of study variables (eg, study setting, risk of bias factors) and individual factors that may influence screening accuracy (eg, age, sex, inpatient vs outpatient). In addition, a large IPD database allows the development of a predictive algorithm to generate estimates of the probability of having depression based on patient characteristics and actual HADS-D scores, rather than classifying patients as simply negative or positive based on screening results. This is an important consideration because a patient with a score of 0 on the HADS-D, for example, would almost certainly have a lower likelihood of having depression than a patient with a substantially higher, but subthreshold, score of 7, although typically both would be classified as negative screens.

A potential downside of IPD meta-analyses is that they are resource intensive. Furthermore, they can be biased if the primary datasets obtained are not representative.^{30–33} Currently, our team is conducting an IPD meta-analysis of the Patient Health Questionnaire depression screening tool, which is the first IPD meta-analysis of the diagnostic accuracy of a depression screening tool.³⁴ In that study, which is still in progress as of 14 March 2016, we had verified 76 eligible primary datasets and obtained usable data for 60 of them. This suggests that investigators are generally able and willing to provide primary data from studies of the diagnostic accuracy of depression screening tools for use in IPD meta-analyses. On the basis of preliminary searches, we are aware of at least 100 eligible studies on the HADS-D.

Thus, the objectives of this IPD meta-analysis are to determine the diagnostic accuracy of the HADS-D to detect major depression among patients in medical settings and to develop an algorithm to predict the

probability that individual patients have MDD based on HADS-D scores and patient characteristics.

METHODS AND ANALYSIS

This systematic review has been funded by the Canadian Institutes of Health Research (Funding Reference Number KRS-144045). The protocol has been registered in the PROSPERO prospective register of systematic reviews (CRD42015016761), and any changes to the study protocol will be registered as amendments with PROSPERO.

The IPD meta-analysis has been designed and will be conducted in accordance with best-practice standards as elaborated in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy³⁵ and other key sources.^{30 31 36} Results will be reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of individual patient data (PRISMA-IPD) statement.³⁷ To conduct the meta-analysis, we will seek primary datasets that allow us to compare HADS-D scores to MDD or major depressive episode (MDE) diagnostic status. Most primary studies use MDD as the reference standard, but some may use MDE, which is identical with respect to the symptoms of depression, but does not exclude patients with psychotic disorders or a history of manic episodes. If both are available, we will use MDD.

Sources of evidence

Our search strategy, which is based on strategies used in previous systematic reviews,^{34 38} was developed by a medical librarian and peer reviewed by another medical librarian. MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, PsycINFO (OvidSP platform) and Web of Science (Web of Knowledge platform) will be searched. The MEDLINE search strategy was validated by testing against already-identified publications from preliminary searches. The strategy was then adapted for PsycINFO and Web of Science. We limited our search strategy to these databases based on research showing that adding other databases (eg, EMBASE) when the MEDLINE search is highly sensitive does not identify additional eligible studies.³⁹ The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy³⁵ suggests combining concepts of the index test and the target conditions, but this was redundant for depression screening tools as these tests are limited to testing for depression. Thus, the search strategy for electronic databases was composed of two concepts: the index test of interest and studies of screening accuracy. There are no published search hedges designed specifically for mental health screening, but several key articles were consulted in developing search terms to ensure retrieval of relevant publications.^{40–42} Search strategies use a combination of subject headings, when available in the database, as well as keywords in the title, abstract or anywhere else in the record. See online supplementary

file 1 for detailed information on searches. To supplement electronic searches, reference lists of all included publications and relevant reviews will be scanned. In addition, a related articles search will be conducted for included papers indexed in MEDLINE using the PubMed 'related articles' search feature. We will also contact researchers who have published on the topic to obtain information about additional, unpublished studies. Search results will be initially uploaded into the citation management database RefWorks (RefWorks, RefWorks-COS, Bethesda, Maryland, USA), and the RefWorks duplicate check function will be used to identify citations retrieved from multiple sources. Unique citations will then be uploaded into the systematic review program DistillerSR (Evidence Partners, Ottawa, Canada), and DistillerSR will be used to store and track search results and to track results of the review process.

To identify relevant datasets, we will review articles in any language. Datasets will be sought for inclusion if patients completed the HADS-D and were assessed for MDD or MDE using Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM-III, DSM-IV or DSM-V) or International Classification of Diseases (ICD, V.10) criteria within 2 weeks (before or after) of completion of the HADS-D, since major depression criteria are for symptoms in the last 2 weeks. Diagnostic assessments must be based on a validated structured or semistructured interview (eg, Structured Clinical Interview for DSM,⁴³ Composite International Diagnostic Interview⁴⁴). Datasets where some patients were administered the screening tools within 2 weeks of the diagnostic interview and some patients were not will be included if the original data allow us to select patients administered the diagnostic interview and screening tools within the 2-week window. Data from studies where all patients are known to have psychiatric diagnoses, have been referred for mental health evaluation or are undergoing treatment for depression will be excluded, with the exception of patients treated for substance use disorders, for whom depression screening may be considered. The coding manual for inclusion and exclusion decisions is shown in online supplementary file 2.

Two investigators will independently review titles and abstracts for eligibility. If either reviewer determines that a study may be eligible based on title or abstract review, then a full-text article review will be completed. Disagreement between reviewers after full-text review will be resolved by consensus, including a third investigator as necessary. Translators will be consulted to evaluate titles/abstracts and articles for languages other than those for which team members are fluent. See online supplementary file 3 for a preliminary PRISMA flow of studies figure.

Transfer of data and dataset management

Authors of studies containing datasets that meet inclusion criteria will be contacted to invite them to contribute primary data for inclusion. Data will only be used

from studies that received ethics approval and all data that are transferred will be properly deidentified prior to transfer. All IPD that are obtained will be cleaned and coded to make patient data as uniform as possible across datasets, then entered into a single database. A preliminary codebook has been developed for coding data from original studies of the HADS-D. Actual data coding and transfer from original studies into the IPD database will be done by a supervised staff or trainee member of the team. Patient characteristics and screening accuracy results for each study using the cleaned datasets will be compared with those from the original datasets to identify any potential discrepancies.

In addition to obtaining original patient-level data, data will also be extracted from the published articles of included studies. We will crosscheck the published data with the original patient-level data obtained from each dataset, and any inconsistencies will be discussed with the original authors. Corrections will be made as necessary.

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool⁴⁵ will be used to assess risk of bias factors in primary studies. QUADAS-2 incorporates assessments of risk of bias across four core domains: patient selection, the index test, the reference standard, and the flow and timing of assessments. Two reviewers will independently assess risk of bias with any discrepancies resolved by consensus.

Data analysis

Analyses will estimate sensitivity and specificity, which will be used to generate estimates of positive and negative predictive value, which are more useful clinically. A bivariate random-effects meta-analysis will be fit, estimated via Gauss-Hermite adaptive quadrature, as described in Riley *et al.*⁴⁶ for the full range of plausible cut-off values. This approach models sensitivity and specificity simultaneously and accounts for variation in within-study precision.⁴⁶ Data from all primary studies will be analysed at the same time using a random-effects model so that sensitivity and specificity are assumed to vary across studies. For each cut-off, separately, this model will provide us with an overall summary sensitivity and specificity and an overall summary diagnostic OR. We will compare results that only include datasets that allow the exclusion of patients diagnosed with depression or receiving depression treatment (including antidepressants with reason unspecified) with results that also include studies where these data are not available. Additionally, a subgroup analysis will be conducted that includes only data from countries listed as 'very high development' on the United Nation's Human Development Index.⁴⁷

Heterogeneity will be quantified for each cut-off analysis by reporting the estimated variances of the random effects for sensitivity and specificity, as well as by

estimating R . R is the ratio of the estimated SD of the summary sensitivity (or specificity) from the random-effects model to the estimated SD of the summary sensitivity (or specificity) from the fixed-effects model.⁴⁸ We will explore underlying reasons for heterogeneity using patient-level and study-level factors. In diagnostic accuracy, this can easily be accomplished by including the factors or interaction terms in the random-effects model described above.⁴⁶ These analyses take advantage of the richness of IPD. When analysed at the patient level, accounting for correlation between patients from the same study, and for the correlation between sensitivity and specificity via the random-effects bivariate model, they are more powerful to detect interactions and not vulnerable to ecologic bias compared with traditional meta-analyses.^{49–53} At the patient level, covariates will include age (<60 years vs ≥ 60 years), sex and inpatient versus outpatient status. Study-level covariates, including risk of bias factors described in QUADAS-2,⁴⁵ will also be evaluated. QUADAS-2 factors include patient selection factors, blinding of reference standard to index test results, type of reference standard (eg, semistructured diagnostic interview, structured diagnostic interview) and timing of administration of index test and reference standard (eg, same day, delay of 1–7 days, delay of >7 days). Significance levels for the pre-specified interaction analyses will not be adjusted for multiple comparisons.

In addition to estimating sensitivity and specificity for each relevant cut-off, we will build a predictive model that uses the score on the screening questionnaire and any other key factors that account for substantial heterogeneity to estimate the probability that a patient has major depression. The model will be evaluated in terms of its calibration (eg, slope of linear predictor; are average, low and high predictions correct?) and discrimination (eg, c-statistic; how well are low-risk subjects distinguished from high-risk subjects?).⁵⁴ Validation with the same subjects used to develop a model results in overly optimistic performance. Internal validation will be assessed via the bootstrap method, which is preferable to split sample validation approaches (eg, developing the model in half the sample and evaluating it in the other half).⁵⁵ Although there are advantages to external validation, given the wide range of study samples that will be used, it would be unlikely that there would be another comparable dataset large enough for validation. Thus, assessment of internal validity via bootstrapping will help develop an understanding of how the model will likely perform in a clinical setting. Furthermore, by using the regression coefficients adjusted for optimism (ie, the shrinkage estimates), we will maximise actual accuracy. On the basis of our pilot work, it is anticipated that missing data will be minimal for the variables of primary interest. Regardless, multiple imputation will be done using chained equations^{54 56} to impute data for binary and continuous variables, considering study as a fixed effect in the imputation model.⁵⁷ This will allow

imputation for variables missing for entire studies as well those missing more sporadically.

Studies included in the IPD meta-analysis will be compared with eligible studies that do not provide data in terms of sensitivity and specificity, using published summary data from the studies that do not provide data. Depending on the number of missing studies, a sensitivity analysis may also be conducted that includes aggregate summary estimates of sensitivity and specificity from the studies that do not provide IPD in the main meta-analysis, along with data from studies that contribute to the IPD meta-analysis.⁴⁶ If there are a large number of studies that do not contribute primary data, this analysis may become the primary analysis.

ETHICS AND DISSEMINATION

This IPD meta-analysis does not require ethics approval, although only individual studies that obtained ethical clearance and informed consent will be included. The reasons that the IPD meta-analysis does not require ethics review are that the objectives of the IPD meta-analysis are consistent with the objectives of the primary studies, which already received ethics approval, and only anonymised data will be provided by the investigators of the original studies.

The main outcomes of the IPD meta-analysis reflect knowledge that will influence future research, clinical practice and policy. Strategies for effective dissemination and specific outputs will be based on research showing how to best tailor research outputs to different user groups,^{58–63} including research on improving the usefulness of reports of systematic review and meta-analyses for healthcare managers and policymakers.^{61 63} Dissemination will include publication of results in high-impact medical journals with open access, as well as presentations in seminars and symposia to policymakers, healthcare providers and researchers at national and international conferences.

If the predictive model performs well, we will create an online calculation tool that will be made freely available to estimate the probability that a given patient has major depression based on depression screening results and key patient characteristics. An example of a tool that is based on robust research evidence and effectively disseminated is the Fracture Risk Assessment Tool (FRAX) (<http://www.shef.ac.uk/FRAX/index.aspx>). The tool that will be made based on the results of our study will similarly be presented in an easy-to-use fashion with tablet and app versions. In addition, simpler nomogram-based presentations, which are user-friendly graphical depictions of positive and negative predictive value by prevalence, will be generated and made available.

Author affiliations

¹Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Québec, Canada

²Department of Psychiatry, McGill University, Montreal, Québec, Canada

³Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Québec, Canada

⁴Department of Medicine, McGill University, Montreal, Québec, Canada

⁵Department of Educational and Counselling Psychology, McGill University, Montreal, Québec, Canada

⁶Department of Psychology, McGill University, Montreal, Québec, Canada

⁷Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Québec, Canada

⁸Department of Libraries, Concordia University, Montreal, Québec, Canada

⁹Department of Clinical, Neuro and Developmental Psychology and EMGO Institute, VU University Amsterdam, Amsterdam, The Netherlands

¹⁰Department of Health Sciences, Hull York Medical School, University of York, York, UK

¹¹Department of Medicine, Health Research and Policy, Stanford Prevention Research Center, Stanford School of Medicine, Stanford, California, USA

¹²Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA

¹³Department of Community Health Sciences, University of Calgary, Calgary, Edmonton, Canada

¹⁴Department of Psychiatry, University of Calgary, Calgary, Edmonton, Canada

¹⁵Department of Mathematics and Statistics, McGill University, Montreal, Québec, Canada

¹⁶Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

¹⁷Department of Oncology, McGill University, Montreal, Québec, Canada

¹⁸Department of Clinical Neurosciences, University of Calgary, Calgary, Edmonton, Canada

¹⁹Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada

²⁰Department of Medicine, University of Calgary, Calgary, Edmonton, Canada

Twitter Follow Simon Gilbody at @SimonGilbody

Contributors BDT, AB, LAK, BL, MA, KER, NS, PC, SG, JPAI, DM, SBP, IS, RJS, RCZ, MH, ZI, CGL, NM and MT contributed to the conception and design of the systematic review and meta-analysis. BDT, AB, LAK, BL, MA, KER and NS will be involved in the acquisition of data. BDT, AB and BL will analyse the data. BDT, AB, LAK, BL, MA, KER, NS, PC, SG, JPAI, DM, SBP, IS, RJS, RCZ, MH, ZI, CGL, NM and MT will interpret the results. BDT and AB drafted this protocol. All authors provided critical revisions of the protocol and approved submission of the final manuscript. BDT is the guarantor.

Funding This research is supported by a grant from the Canadian Institutes of Health Research (CIHR; Funding Reference Number KRS-144045; PI Thombs). BDT receives support from an Investigator Award from the Arthritis Society. AB is supported by the Fonds de recherche du Québec—Santé (FRQS) researcher salary award. BL is supported by the Fonds de recherche du Québec—Santé (FRQS) doctoral award. MA is supported by a CIHR Frederick Banting and Charles Best Canadian Graduate Scholarships—Master's Award. SBP is a Senior Health Scholar with Alberta Innovates—Health Solutions. ZI receives funding from the Alzheimer Society Calgary via the Hotchkiss Brain Institute. CGL's work is supported by the Christine and Herschel Victor/Hope and Cope Research Chair in Psychosocial Oncology, McGill University. No funding body had any input into any aspect of this protocol.

Competing interests SBP received a research grant from a competition cosponsored by the Hotchkiss Brain Institute and Pfizer Canada. All other authors declare that they have no competing interests.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

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. Rudisch B, Nemeroff CB. Epidemiology of comorbid coronary artery disease and depression. *Biol Psychiatry* 2003;54:227–40.

2. Thombs BD, Bass EB, Ford DE, *et al.* Prevalence of depression in survivors of acute myocardial infarction. *J Gen Intern Med* 2006;21:30–8.
3. Evans DL, Charney DS, Lewis L, *et al.* Mood disorders in the medically ill: scientific review and recommendations. *Biol Psychiatry* 2005;58:175–89.
4. Dickens C, Creed F. The burden of depression in patients with rheumatoid arthritis. *Rheumatology (Oxford)* 2001;40:1327–30.
5. Ali S, Stone MA, Peters JL, *et al.* The prevalence of co-morbid depression in adults with type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23:1165–73.
6. Fann JR, Thomas-Rich AM, Katon WJ, *et al.* Major depression after breast cancer: a review of epidemiology and treatment. *Gen Hosp Psychiatry* 2008;30:112–26.
7. National Collaborating Centre for Mental Health. *Depression in adults with a chronic physical health problem: the NICE guideline on treatment and management*. Leicester, UK: The British Psychological Society & The Royal College of Psychiatrists, 2010.
8. Duhoux A, Fournier L, Gauvin L, *et al.* What is the association between quality of treatment for depression and patient outcomes? A cohort study of adults consulting in primary care. *J Affect Disord* 2013;151:265–74.
9. Duhoux A, Fournier L, Nguyen CT, *et al.* Guideline concordance of treatment for depressive disorders in Canada. *Soc Psychiatry Psychiatr Epidemiol* 2009;44:385–92.
10. Mojtabei R. Clinician-identified depression in community settings: concordance with structured-interview diagnoses. *Psychother Psychosom* 2013;82:161–9.
11. Mojtabei R, Olfson M. Proportion of antidepressants prescribed without a psychiatric diagnosis is growing. *Health Aff (Millwood)* 2011;30:1434–42.
12. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;374:609–19.
13. Siu AL, Bibbins-Domingo K, Grossman DC, *et al.*, U S Preventive Services Task Force. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA* 2016;315:380–7.
14. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014;348:g1253.
15. Thombs BD, Ziegelstein RC, Roseman M, *et al.* There are no randomized controlled trials that support the United States Preventive Services Task Force guideline on screening for depression in primary care: a systematic review. *BMC Med* 2014;12:13.
16. Thombs BD, Coyne JC, Cuijpers P, *et al.* Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012;184:413–18.
17. Gilbody S, Sheldon T, Wessely S. Should we screen for depression? *BMJ* 2006;332:1027–30.
18. Palmer SC, Coyne JC. Screening for depression in medical care: pitfalls, alternatives, and revised priorities. *J Psychosom Res* 2003;54:279–87.
19. Thombs BD, Rice DB. Sample sizes and precision of estimates of sensitivity and specificity from primary studies on the diagnostic accuracy of depression screening tools: a survey of recently published studies. *Int J Method Psych*. In press.
20. Levis B, Benedetti A, Levis AW, *et al.* Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual patient data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *Am J Epidemiol*. In press.
21. Rice DB, Thombs BD. Risk of bias from inclusion of currently diagnosed or treated patients in studies of depression screening tool accuracy: a cross-sectional analysis of recently published primary studies and meta-analyses. *PLoS ONE* 2016;11:e0150067.
22. Thombs BD, Arthurs E, El-Baalbaki G, *et al.* Risk of bias from inclusion of already diagnosed or treated patients in diagnostic accuracy studies of depression screening tools: a systematic review. *BMJ* 2011;343:d4825.
23. Joffres M, Jaramillo A, Dickinson J, *et al.* Recommendations on screening for depression in adults. *CMAJ* 2013;185:775–82.
24. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983;67:361–70.
25. 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.
26. 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* 2013;85:198–206.
27. 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.
28. Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. *J Affect Disord* 2010;126:335–48.
29. 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.
30. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
31. Stewart LA, Tierney JF, Clarke M. Chapter 18: reviews of individual patient data. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Chichester, England: Wiley-Blackwell; The Cochrane Collaboration, 2011:18.1–18.9.
32. Jaspers GJ, Degraeuwe PLJ. A failed attempt to conduct an individual patient data meta-analysis. *Syst Rev* 2014;3:97.
33. Parish E, Bloom T, Godlee F. Statins for people at low risk. *BMJ* 2015;351:h3908.
34. 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.
35. Deeks JJ, Bossuyt PM, Gatsonis C. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0*. The Cochrane Collaboration, 2009.
36. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Chichester, England: Wiley-Blackwell; The Cochrane Collaboration, 2011.
37. Stewart LA, Clarke M, Rovers M, *et al.* Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;313:1657–65.
38. Thombs BD, Benedetti A, Kloda LA, *et al.* Diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open* 2015;5:e009742.
39. Sampson M, Barrowman NJ, Moher D, *et al.* Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003;56:943–55.
40. Mann R, Gilbody SM. Should methodological filters for diagnostic test accuracy studies be used in systematic reviews of psychometric instruments? A case study involving screening for postnatal depression. *Syst Rev* 2012;1:9.
41. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391–7.
42. Wong SS, Wilczynski NL, Haynes RB, *et al.* Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003;2003:728–32.
43. First MB, Spitzer RL, Gibbon M, *et al.* *Structured clinical interview for DSM-IV axis I disorders—patient edition (SCID-IP, Version 2.0)*. New York: Biometrics Research Department, New York State Psychiatric Institute, 1996.
44. Wittchen HU. Reliability and validity studies of the WHO–Composite International Diagnostic Interview (CID): a critical review. *J Psychiatr Res* 1994;28:57–84.
45. Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
46. Riley RD, Dodd SR, Craig JV, *et al.* Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;27:6111–36.
47. United Nations Development Programme. Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World. New York, NY: United Nations Development Programme, 2013.
48. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
49. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007;26:2982–99.
50. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998;54:638–45.
51. Berlin JA, Santanna J, Schmid CH, *et al.* Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21:371–87.

52. Lambert PC, Sutton AJ, Abrams KR, *et al.* A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;55:86–94.
53. Schmid CH, Stark PC, Berlin JA, *et al.* Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004;57:683–97.
54. Steyerberg EW. *Clinical prediction models*. New York: Springer; 2009.
55. Steyerberg EW, Harrell FE Jr, Borsboom GJ, *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
56. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009;28:1982–98.
57. Riley RD, Takwoingi Y, Trikalinos T, *et al.* Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biomed Biostat* 2014;5:196.
58. Dobbins M, DeCorby K, Twiddy T. A knowledge transfer strategy for public health decision makers. *Worldviews Evid Based Nurs* 2004;1:120–8.
59. Graham ID, Logan J, Harrison MB, *et al.* Lost in knowledge translation: time for a map? *J Contin Educ Health Prof* 2006;26:13–24.
60. Haines A, Kuruvilla S, Borchert M. Bridging the implementation gap between knowledge and action for health. *Bull World Health Organ* 2004;82:724–31.
61. Dobbins M, Thomas H, O'Brien MA, *et al.* Use of systematic reviews in the development of new provincial public health policies in Ontario. *Int J Technol Assess Health Care* 2004;20:399–404.
62. Dobbins M, Jack S, Thomas H, *et al.* Public health decision-makers' informational needs and preferences for receiving research evidence. *Worldviews Evid Based Nurs* 2007;4:156–63.
63. Lavis J, Davies H, Oxman A, *et al.* Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy* 2005;10(Suppl 1):35–48.