BMJ Open

Diagnostic Accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5, and Geriatric Depression Scale-4 for Detecting Major Depression: Protocol for a Systematic Review and Individual Participant Data Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026598
Article Type:	Protocol
Date Submitted by the Author:	10-Sep-2018
Complete List of Authors:	Benedetti, Andrea; McGill University, Departments of Medicine and of Epidemiology, Biostatistics & Occupational Health; McGill University Health Centre, Respiratory Epidemiology and Clinical Research Unit Wu, Yin; McGill University, Departments of Psychiatry and of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Levis, Brooke; McGill University, Department of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Wilchesky, Machelle ; McGill University, Departments of Medicine and of Family Medicine; Jewish General Hospital, Lady Davis Institute for Medical Research Boruff, Jill; McGill University, Schulich Library of Science and Engineering Ioannidis, John; Stanford University, USA, Stanford Prevention Research Center, Departments of Medicine and of Health Research & Policy Patten, Scott ; University of Calgary, Departments of Community Health Sciences and of Psychiatry Cuijpers, Pim; VU University Amsterdam, Department of Clinical, Neuro and Developmental Psychology and Amsterdam Public Health research institute Shrier, Ian; McGill University, Department of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Gilbody, Simon; The University of York, Department of Health Sciences Ismail, Zahinoor; University of Calgary, Departments of Community Health Sciences, of Psychiatry and of Clinical Neurosciences McMillan, Dean; University of York, Department of Health Sciences Mitchell, Nicholas; University of Alberta, Department of Psychiatry Ziegelstein, Roy; Johns Hopkins University School of Medicine, Department of Medicine Thombs, Brett; McGill University, Psychiatry
Keywords:	depression, Geriatric Depression Scale, GDS, diagnostic test accuracy, individual participant data meta-analysis

Diagnostic Accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5, and Geriatric Depression Scale-4 for Detecting Major Depression: Protocol for a Systematic Review and Individual Participant Data Meta-Analysis

Andrea Benedetti, PhD[1-3]; Yin Wu, PhD[2,4,5]; Brooke Levis, MSc[2,5]; Machelle Wilchesky, PhD[3,5,6]; Jill Boruff, MLIS[7]; John P. A. Ioannidis, MD, DSc[8,9]; Scott B. Patten, MD, PhD[10,11]; Pim Cuijpers, PhD[12]; Ian Shrier, MD, PhD[2,5]; Simon Gilbody, DPhil, FRCPsych[13]; Zahinoor Ismail, MD FRCPC[10,11,14]; Dean McMillan, PhD[13]; Nicholas Mitchell, MD, MSc[15]; Roy C. Ziegelstein, MD[16]; Brett D. Thombs, PhD[2-5,17,18]

[1]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Quebec, Canada; Departments of [2]Epidemiology, Biostatistics and Occupational Health; [3]Medicine; and [4]Psychiatry; McGill University, Montreal, Quebec, Canada; [5]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; [6]Department of Family Medicine, McGill University, Montreal, Quebec, Canada; [7]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Quebec, Canada; [8]Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford School of Medicine; Stanford, California, USA; [9]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA; Departments of [10]Community Health Sciences and [11]Psychiatry, University of Calgary, Calgary, Canada; [12]Department of Clinical, Neuro and Developmental Psychology and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, The

Netherlands; [13]Hull York Medical School and Department of Health Sciences, University of York, York, UK; [14]Department of Clinical Neurosciences, University of Calgary, Calgary, Edmonton, Canada; [15]Department of Psychiatry, University of Alberta, Edmonton, Canada; [16]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Departments of [17]Psychology and [18]Educational and Counselling Psychology; McGill University, Montreal, Quebec, Canada.

Addresses for Correspondence:

Andrea Benedetti, PhD; 5252 de Maisonneuve Blvd. West; Montreal, Quebec, Canada H4A 3S5; Tel (514) 934-1934 ext. 32161; E-mail: andrea.benedetti@mcgill.ca

Brett D. Thombs, PhD; 4333 Cote Ste Catherine Road; Montreal, Quebec, Canada H3T 1E4; Tel (514) 340-8222 ext. 25112; E-mail: brett.thombs@mcgill.ca.

Key Words: depression; Geriatric Depression Scale; GDS; diagnostic test accuracy; individual participant data meta-analysis.

Word Count: 3,002

ABSTRACT

Introduction: The thirty-item Geriatric Depression Scale (GDS-30) and the shorter GDS-15, GDS-5 and GDS-4 are recommended as depression screening tools for elderly individuals. Existing meta-analyses on the diagnostic accuracy of the GDS have not been able to conduct subgroup analyses, have included patients already identified as depressed who would not be screened in practice, and have not accounted for possible bias due to selective reporting of results from only better-performing cutoffs in primary studies. Individual participant data meta-analysis (IPDMA), which involves a standard systematic review, then a synthesis of individual participant data, rather than summary results, could address these limitations. The objective of our IPDMA is to generate accuracy estimates to detect major depression for all possible cutoffs of each version of the GDS among studies using different reference standards, separately, and among participant subgroups based on age, sex, dementia diagnosis, and care settings. In addition, we will use a modelling approach to generate individual participant probabilities for major depression based on GDS scores (rather than a dichotomous cutoff) and participant characteristics (e.g., sex, age, dementia status, care setting).

Methods and Analysis: Individual participant data comparing GDS scores to a major depression diagnosis based on a validated structured or semi-structured diagnostic interview will be sought via a systematic review. Data sources will include Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and Web of Science. Bivariate random-effects models will be used to estimate diagnostic accuracy parameters for each cutoff of the different versions of the GDS. Pre-specified subgroup analyses will be conducted. Risk of bias will be assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

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

Systematic Review Registration: PROSPERO (CRD42018104329)

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This study will use individual participant data to estimate diagnostic accuracy for all relevant cutoff scores of the different versions of the Geriatric Depression Scale (GDS). Using data from all participants at each cutoff score will overcome limitations related to selective cutoff reporting in primary study publications.
- This study will conduct analyses that exclude patients with current diagnoses of depression or who are undergoing mental health treatment, including antidepressants, 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.
- This study will include subgroup analyses of diagnostic accuracy across different reference standards and by participant characteristics (e.g., sex, age, dementia status, care setting).
- A potential limitation is that the success of the study depends on the ability to obtain the relevant individual participant 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.

BACKGROUND

Major depression is present in 5-10% of the geriatric population internationally.[1,2] Effective treatments for depression are available, but identification is often haphazard. Physicians may fail to recognize up to half of all patients with depression, and most patients with depression do not receive minimally adequate care.[3-4] At the same time, there is a high rate of overdiagnosis and overtreatment, and the majority of patients who are treated do not meet diagnostic criteria.[5-7] Diagnosis of elderly individuals can be particularly difficult for clinicians due to factors such as cognitive impairments, social stigma, medical comorbidity, and atypical or vague clinical presentation.[1,8,9]

Some Canadian and international geriatric care organizations recommend screening elderly adults for depression,[10-13] but the Canadian Task Force for Preventive Health Care (CTFPHC), for instance, does not recommend depression screening, including for geriatric individuals.[14] The CTFPHC has expressed concern that published studies may overstate the accuracy of depression screening tools and that screening could lead to high rates of false positive tests, and still not improve depression outcomes.[14]

The thirty-item Geriatric Depression Scale (GDS-30), and the GDS-15, GDS-5 and GDS-4, which are fifteen-item, five-item and four-item subsets of the GDS-30, are commonly recommended as depression screening tools for elderly individuals.[15-17] As with other depression screening tools, primary studies on the diagnostic accuracy of the different versions of the GDS have been limited by (1) small samples; (2) the selective reporting of results for cutoffs when they perform well in a given sample, but not when they perform poorly; (3) the inclusion of patients already known by clinicians to have depression; and (4) the inability to conduct subgroup analyses (e.g., different age groups, dementia diagnosis, care settings) due to

small sample sizes. Conventional meta-analyses of the GDS or short versions of the GDS that have synthesized published summary data have not been able to conduct subgroup analyses or exclude already-diagnosed patients,[15,16,18] and concerns have been raised about bias in these meta-analyses due to selective cutoff reporting in primary studies that could not be addressed.[18]

Individual participant data meta-analysis (IPDMA), which involves a standard systematic review, followed by synthesis of actual participant data from primary studies, rather than aggregating summary data, can address these problems by including actual participant data from all studies.[19,20] In the context of evaluating the diagnostic accuracy of depression screening tools, IPDMA has three major advantages compared to conventional meta-analyses. First, for the conventional binary screening approach, IPDMA can address bias from the selective publication of diagnostic accuracy results for well-performing cutoffs from small studies since accuracy can be evaluated across all relevant cutoffs for all participants. Second, IPDMA allows the appropriate exclusion of already-diagnosed or already-treated patients when primary studies have data on existing diagnoses and treatment. Third, an IPDMA with large numbers of participants and major depression cases would allow subgroup analyses by study-level factors (e.g., study setting, risk of bias factors) and individual factors that may influence screening accuracy (e.g., age, sex, dementia diagnosis). Finally, a large IPD database would allow the development of a predictive algorithm to generate estimates of the probability of having major depression based on participant characteristics and actual GDS scores, rather than binary classifications of individuals as simply negative or positive based on screening results. This is important because, for instance, an individual with a score of 0 on the GDS-30, may have a lower likelihood of having depression than an individual with a substantially higher, but sub-cutoff, score of 10. Using a dichotomous

cutoff method, however, both would be classified as negative screens and assigned the same probability of having depression.

One of the downsides of IPDMAs is that they are resource intensive. In addition, if the primary datasets obtained are not representative of all primary studies, the IPDMA could be biased.[19-22] In a previous IPDMA of the Patient Health Questionnaire-9 (PHQ-9) screening tool, which was the first IPDMA of the diagnostic accuracy of a depression screening tool,[23] we were able to synthesize 58 of 72 eligible primary datasets (17,357 participants, 2,312 major depression cases). 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 IPDMA. A preliminary PubMed search for the GDS verified the existence of enough primary studies (more than 100 potentially eligible datasets that appear to have at least 30,000 participants, 4,000 cases) to make IPDMA feasible for the GDS.

Thus, the objectives of this IPDMA are to evaluate the diagnostic accuracy of the GDS-30, GDS-15, GDS-5 and GDS-4 among studies using different reference standards, separately; among participant subgroups based on age, sex, dementia diagnosis, and care settings; and excluding participants identified as already-diagnosed or treated for depression. Furthermore, a prediction model will be generated.

METHODS AND ANALYSIS

This systematic review has been funded by the Canadian Institutes of Health Research (Funding Reference Number PJT-156365). The protocol has been registered in the PROSPERO prospective register of systematic reviews (CRD42018104329), and any changes to the study protocol will be registered as amendments with PROSPERO.

The IPDMA 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 Accuracy[24] and other key sources.[19,20,25] Results will be reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement and the PRISMA-IPD statement.[26,27] The IPDMA protocol does not deviate substantively from previous IPDMA protocols that we have developed and published for other depression screening tools.[23,28,29]

Sources of Evidence

The search strategy was developed by a medical librarian and was adapted from a search strategy developed for a similar systematic review to obtain datasets for IPDMA of the PHQ-9 depression screening tool,[23] which was peer-reviewed using the Peer Review of the Electronic Search Strategy (PRESS) standard.[30] The search strategy is also similar to strategies that we have used for systematic reviews and IPDMA of the Hospital Anxiety and Depression Scale and Edinburgh Postnatal Depression Scale.[28,29]

We will search Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO (OvidSP platform), and Web of Science (Web of Knowledge platform). The Medline search strategy for the GDS was validated by testing against already-identified publications from our preliminary search. 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 (e.g., EMBASE) when the Medline search is highly sensitive does not identify additional eligible studies.[31] The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[24] 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 comprised 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 key articles were consulted in developing search terms.[32-34] See Supplementary File 1 for detailed information on searches. To supplement electronic searches, we will search reference lists of included publications and relevant reviews, conduct a related articles search using the PubMed "related articles" feature, and query authors of included studies for unpublished studies. Search results will be uploaded into the citation management database RefWorks (RefWorks-COS, Bethesda, MD, 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.

Selection of Eligible Studies

To conduct the meta-analysis, we will seek primary datasets that allow us to compare GDS scores to major depression diagnostic status. Datasets from articles in any language will be sought for inclusion if they compare results from any version of the GDS to diagnoses of major depressive disorder (MDD) or major depressive episode (MDE) made with a validated diagnostic interview, administered within 2 weeks of the GDS and based on Diagnostic and Statistical Manual (DSM) or International Classification of Diseases criteria (ICD), which are similar to DSM criteria and generally used outside of North America.

The two-week criterion was set because that is the duration of symptoms required for a diagnosis of major depression. Datasets where some participants were administered the screening tool within 2 weeks of the diagnostic interview and some participants were not will be included

if the original data allows us to identify and select eligible participants. 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 participants with psychotic disorders or a history of manic episodes. If both are available, we will record both and prioritize DSM over ICD and MDE over MDD in analyses. Data from studies where all participants 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 participants treated for substance use disorders, for whom depression screening may be considered. The coding manual for inclusion and exclusion decisions is shown in Supplementary File 2.

Two investigators will review articles independently for eligibility. If either reviewer determines that a study may be eligible based on title or abstract review, 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 used to evaluate titles/abstracts and articles for languages other than those for which team members are fluent. See Supplementary File 3 for a preliminary PRISMA flow of studies figure.

Transfer of Data and Data 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 de-identified prior to transfer. Participant data will be cleaned and coded for uniformity across datasets using an already developed codebook, similar to codebooks used in our previous IPDMAs.[23,28,29] 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. Participant characteristics and screening

accuracy results for each study using the cleaned datasets will be compared to those from the original datasets to identify any potential discrepancies.

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

Quality Assessment

Two reviewers will independently use the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool[35] to assess risk of bias in primary studies. QUADAS-2 incorporates assessments of risk of bias across four core domains: participant selection, the index test, the reference standard, and the flow and timing of assessments. Two reviewers will independently assess risk of bias, and any discrepancies will be resolved by consensus.

Data Analysis

Analyses will estimate sensitivity and specificity separately at each cutoff by bivariate random-effects meta-analysis models as described in Riley et al.[36] For each GDS version, we will fit these models, estimated via Gaussian Hermite adaptive quadrature, for the full range of plausible GDS cutoff values.[36] This approach models sensitivity and specificity simultaneously and accounts for the precision of estimates within studies.[36] Data from all included primary studies will be analyzed simultaneously with a random-effects model as sensitivity and specificity are assumed to vary across primary studies. We will also construct a pooled ROC curve and identify the optimal cutoff.[36] 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. For assessment of each version, we will include studies that report total scores for the specific GDS version or individual GDS item scores from longer versions of the GDS which could be used to calculate total scores for the shorter version. We will consider imputation if a large part of data is missing.

In a previous IPDMA with the PHQ-9,[37] we found that reference standards appeared to perform differently. The Mini International Neuropsychiatric Interview (MINI) is fully structured, but was designed for very rapid administration and described as its authors as being over-inclusive as a result. We found that, controlling for depressive symptom scores, the MINI classified approximately twice as many participants with major depression as other fully structured interviews.[38,39] Compared to semi-structured interviews, which are intended to be done by experienced diagnosticians and involve some clinical judgment (e.g., Structured Clinical Interview for DSM Disorders) fully structured interviews (MINI excluded), diagnosed more participants with low symptom levels as depressed and fewer participants with higher symptom levels. Fully structured interviews can be delivered by lay interviewers and are intended to achieve a high level of standardization, but may sacrifice accuracy.[40-43] Thus, we will assess possible differences and evaluate sensitivity and specificity separately by reference standard.

In secondary analyses, to the extent that there are sufficient data, we will investigate subgroups according to age, sex, dementia status and severity, dementia subtype, number of medical comorbidities (with specific comorbidities integrated to the extent possible), care setting and risk of bias. QUADAS-2 factors that will be considered include patient selection factors, blinding of reference standard to index test results, and timing between administration of index test and reference standard (e.g., 0 to 7 days, 7 to 14 days). 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.[44]

If there is a sufficient number of studies with published diagnostic accuracy data for major depression that are eligible but do not provide data, studies included in the IPDMA will be compared to 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 individual participant data in the main meta-analysis, along with data from studies that contribute to the IPDMA.[36] If there are a large number of studies that do not contribute primary data, this analysis may become the primary analysis.

Clinical predictive models have not been used previously to generate individualized probabilities that an individual has major depression based on screening tool scores and participant characteristics. There is a rich tradition of using predictive models for risk scores or classifying patients based on diagnostic tests, and our approach will build upon those traditions.[45-50] To do this, we will develop binary predictive models that use GDS scores as well as key participant characteristics (e.g., sex, age, dementia status, care setting) to estimate the probability and associated 95% CI that an individual has major depression. We will estimate logistic mixed models and then integrate over the distribution of the random effects as described in Pavlou et al. and Skrondal et al.[54,55] Continuous variables (GDS score and age) will be modeled using flexible semi-parametric methods (e.g., regression splines). We will consider the inclusion of interaction terms. The models will be evaluated in terms of their overall performance (Nagelkerke's R^2, Brier score), calibration (e.g., slope of linear predictor; are

average, low and high predictions correct) and discrimination (e.g., c-statistic; discrimination slope: can we separate subjects with and without major depression).[45,46] Validation with the same subjects used to develop a model results in overly optimistic performance. We will assess internal validation via the bootstrap method, which has been shown to be preferable to split sample validation approaches.[47] Although there are advantages to external validation, given the wide range of study populations that we will be using, it would be unlikely that there would be another comparable data set large enough for validation. Thus, assessment of internal validity via bootstrapping will allow us to understand how our model may perform in a clinical setting, and by adjusting our regression coefficients for optimism, the performance of our model will be as accurate as possible. In sensitivity analyses, we will explore including each item from the GDS questionnaires as a separate predictor variable, rather than only the total score.

ETHICS AND DISSEMINATION

This IPDMA does not require ethics approval, although only individual studies that obtained ethical clearance and informed consent will be included. The reason that the IPDMA does not require ethics review is that the objectives of the IPDMA are consistent with the objectives of the primary studies, which already received ethics approval, and only anonymized data will be provided by the investigators of the original studies.

The main outcomes of the IPDMA 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,[18,48-52] including research on improving the usefulness of reports of systematic review and meta-analyses for health care managers and policy makers.[50,52] Dissemination will include

publication of results in high-impact medical journals with open access, as well as presentations in seminars and symposia to policy-makers, health care providers, and researchers at national and international conferences.

If the predictive model performs well, a free and easy-to use online calculation tool will be created to incorporate individual characteristics into accuracy estimates and provide users of our research with probabilities that individual patients have depression based on their GDS score and key characteristics. The calculator will be similar to other successful tools, such as the FRAX® Fracture Risk Assessment Tool (http://www.shef.ac.uk/FRAX/index.aspx). The tool that will be made from the results of this study will be modeled on this tool and presented with tablet and app versions.

AUTHORS' CONTRIBUTIONS

AB, YW, BL, MW, JB, JPAI, SBP, PC, IS, SG, ZI, DM, NM, RCZ, and BDT contributed to the conception and design of the systematic review and meta-analysis. JB developed the database search strategy. AB, YW, BL, MW, JB, and BDT will be involved in acquisition of data. AB, YW, BL, and BDT will analyze the data. All authors will contribute to the interpretation of results. AB, YW, and BDT drafted this protocol. All authors provided critical revisions of the protocol and approved submission of the final manuscript. AB is the guarantor.

FUNDING STATEMENT

This research is supported by a grant from the Canadian Institutes of Health Research (CIHR; Funding Reference Number PJT-156365; PIs = Benedetti, Thombs, Wilchesky). Drs. Benedetti and Thombs are supported by the Fonds de recherche du Québec - Santé (FRQS)

researcher salary award. Dr. Wu is supported by an Utting Postdoctoral Fellowship from the Jewish General Hospital, Montreal, Quebec. Ms. Levis is supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award. Dr. Patten is a Senior Health Scholar with Alberta Innovates, Health Solutions. Dr. Ismail receives funding from the Alzheimer Society Calgary via the Hotchkiss Brain Institute. Dr. Wilchesky is supported by the Donald Berman Maimonides Medical Research Foundation. No funding body had any input into any aspect of this protocol.

COMPETING INTERESTS STATEMENT

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

REFERENCES

1. Fiske A, Wetherell JL, Gatz M. Depression in older adults. *Annu Rev Clin Psychol* 2009;5:363-89.
2. Luppa M, Sikorski C, Luck T, et al. Age- and gender-specific prevalence of depression in latest-life – Systematic review and meta-analysis. *J Affect Disord* 2012;136:212-221.
3. Duhoux A, Fournier L, Nguyen CT, Roberge P, Beveridge R. Guideline concordance of treatment for depressive disorders in Canada. *Soc Psychiatry Psychiatr Epidemiol* 2009;44:385-392.
4. Duhoux A, Fournier L, Gauvin L, Roberge P. What is the association between quality of treatment for depression and patient outcomes? A cohort study of adults consulting in primary care. *J Affect Disord* 2013;151:265-274.
5. Mojtabai R, Olfson M. Proportion of antidepressants prescribed without a psychiatric diagnosis is growing. *Health Aff* 2011;30:1434-1442.
6. Mojtabai R. Clinician-identified depression in community settings: Concordance with structured interview diagnoses. *Psychother Psychosom* 2013;82:161-169.
7. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. *JAMA Intern Med* [Epub ahead of print].
8. Jeste DV, Alexopoulos GS, Bartels SJ, et al. Consensus statement on the upcoming crisis in geriatric mental health: research agenda for the next 2 decades. *Arch Gen Psychiatry* 1999;56:848-853.
9. Weeks R, Kalucy R, Hill J. Chapter 3: Depression in Elderly Patients. In: Jackson S, Jansen P, Mangoni A, eds. *Prescribing for Elderly Patients.* John Wiley & Sons, Ltd; 2009.

10. Registered Nurses Association of Ontario. *Screening for Delirium, Dementia, and Depression in Older Adults.* Toronto, Ontario: Registered Nurses Association of Ontario; 2003.

11. American Medical Directors Association (AMDA). *Depression in the long-term care setting.* Columbia, Maryland: American Medical Directors Association (AMDA); 2011.

12. American Geriatrics Society, American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral symptoms associated with dementia. *J Am Geriatr Soc*. 2003;51:1287-1298.

13. Buchanan D, Tourigny-Rivard M, Cappeliez P, et al. National guidelines for seniors' mental health: the assessment and treatment of depression. *Can J Geriatr* 2006;9:52-58.

14. Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775-782.

15. 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;18:1066-1077.

16. 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.

17. Montorio I, Izal M. The Geriatric Depression Scale: a review of its development and utility. *Int Psychogeriatr* 1996;8:103-112.

18. Pocklington C, Gilbody S, Manea L, McMillan D. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J Geriatr Psychiatry* 2016;31:837-857.

19. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.

20. 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.

21. Jaspers GJ, Degraeuwe PLJ. A failed attempt to conduct an individual patient data meta-analysis. *Systematic Reviews* 2014;3:97.

22. Parish E, Bloom T, Godlee F. Statins for people at low risk. *BMJ* 2015;351:h3908.

23. 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. *Systematic Reviews* 2014;3:124.

24. Deeks JJ, Bossuyt PM, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration; 2009.

25. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Chichester, England: Wiley-Blackwell; The Cochrane Collaboration 2011.

26. McInnes MD, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018; 319:388-96.

27. 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.

28. Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the Hospital Anxiety and Depression Scale (HADS) for detecting major depression: Protocol for a systematic review and individual patient data meta-analyses. *BMJ Open* 2016; 6:e011913.

29. Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: Protocol for a systematicrReview and individual patient data meta-analyses. *BMJ Open* 2015;5:e009742.

30. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiology* 2016;75:40-6.

31. Sampson M, Barrowman NJ, Moher D, et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003;56:943-55.

32. 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;1:9,4053-1-9.

33. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391-7.

34. Wong SS, Wilczynski NL, Haynes RB, et al. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003:728-32.

35. 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.

36. 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.

37. Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *Br J Psychiatry* 2018;212:377-85.

38. Lecrubier Y, Sheehan DV, Weiller E et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;12:224–231.

39. Sheehan DV, Lecrubier Y, Sheehan KH et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry* 1997;12:232–241.

40. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol Medicine* 2001;31:1001–13.

41. Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychol Med* 1999;29:1013-20.

42. Nosen E, Woody SR. Chapter 8: Diagnostic assessment in research. In, McKay D. *Handbook of research methods in abnormal and clinical psychology.* Sage; 2008.

43. Kurdyak PA, Gnam WH. Small signal, big noise: performance of the CIDI depression module. *Can J Psychiatry* 2005;50:851-56.

44. 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.

45. Steyerberg EW. *Clinical Prediction Models*. New York: Springer; 2009.

46. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* 2010;21:128-38.

47. 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.

48. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-76.

49. Meijer A, Roseman M, Milette K, et al. Depression screening and patient outcomes in cancer: A systematic review. *PLOS ONE* 2011;6:e27181.

50. 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.

51. 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 [updated March 2011]. Cochrane Collaboration, 2011.

52. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: A database survey. *BMJ* 2012;344:d7762.

SUPPLEMENTARY FILE 1: Search Strategies

GDS screening accuracy

Researcher: Brett Thombs

Search strategy peer reviewed:

Database searches conducted:

Reference/related list searches conducted

Search Terms	
GDS* **Geriatric Depression Scale***	

Search Terms: SCREENING ACCURACY Medline "filter"	PsycINFO "filter"	Web of Science "filter"
Mass Screening/ Psychiatric Status Rating Scales/ Predictive Value of Tests/ Reproducibility of Results/ Exp "Sensitivity and Specificity"/ Psychometrics/ Prevalence/ Reference Values/ Reference Standards/ Exp Diagnostic Errors/ Mental Disorders/diagnosis Mood Disorders/diagnosis Depressive Disorder/diagnosis Depressive Disorder, Major/diagnosis Depression, Postpartum/diagnosis Depression/diagnosis validation studies.pt. comparative study.pt screen*.af. prevalence.af. predictive value*.af.	Diagnosis/ Medical Diagnosis/ Psychodiagnosis/ Misdiagnosis/ Screening/ Health Screening/ Screening Tests/ Prediction/ Cutting Scores/ Psychometrics/ Test Validity/ screen*.af. predictive value*.af. detect*.ti. sensitiv*.ti. valid*.ti. revalid*.ti. accura*.ti. psychometric*.ti. specificit*.ab. cut?off*.ab. cut* score*.ab. cut?point*.ab.	TS= screen* prevalence "predictive value*" detect* sensitiv* valid* revalid* predict* accura* psychometric* identif* specificit* cutoff* "cut off*" "cut* score*" cutpoint* "cut point*" "threshold score*" "reference standard*" "reference test*" "index test*" "gold standard"

detect*.ti.	threshold score*.ab.
sensitiv*.ti.	reference standard*.ab.
valid*.ti.	reference test*.ab.
revalid*.ti.	index test*.ab
predict*.ti.	gold standard.ab.
accura*.ti.	
psychometric*.ti.	~~Psychological Assessment/~~
identif*.ti	~~Psychiatric Evaluation/~~
specificit*.ab.	~~Testing/~~
cut?off*.ab.	~~Test Interpretation/~~
cut* score*.ab.	~~Rating Scales/~~
cut?point*.ab.	~~prevalence.af.~~
threshold score*.ab.	~~predict*.ti.~~
reference standard*.ab.	~~identif*.ti~~
reference test*.ab.	
index test*.ab.	
gold standard.ab.	

Diagnostic accuracy (prediction) filters:

Some diagnostic accuracy terms were adapted from the following 2 hedges, demonstrated to be the two best strategies in [Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS One; 2012;7(2):e32844]:

Ingui BJ, Rogers MA (2001) Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 8: 391–397.

Wong SS, Wilczynski NL, Haynes RB, Ramkissoonsingh R (2003) Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. AMIA Annu Symp Proc 728–732.

Searches conducted: <July 16, 2018>

Platform	Databases(s)	Results	Saved (account)	Remarks
OvidSP	Ovid MEDLINE(R), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid OLDMEDLINE(R) 1946 to Present (MEDLINE search)	2960	DepressionSR/ depress	Screening Accuracy—Rebuild AND GDS
	PsycInfo	5686		
Clarivate	Web of Science	4166		
	Total database search results in EndNote	12812		
	Total deduplicated in EndNote			

De-Duplication using RefWorks: July 25th, 2018

Platform	**Database(s)**	**Results (No. Of Citations)**	**After de-Duplication (No. of citations)**	**Saved (account)**	**Remarks**
OvidSP	Ovid MEDLINE(R), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid OLDMEDLINE(R) 1946 to Present (MEDLINE search)	2960	2940 (Refworks shows 2972, which is not possible. Must be some error in RefWorks. 2940 was obtained by subtracting the no. of citation in psycInfo and WoS from the Total no. of citations)	DepressionSR/ depress	Screening Accuracy—Rebuild AND GDS
	PsycInfo	5686	4648		
Clarivate	Web of Science	4166	1876		

Total database search results in EndNote	12812	9464 (9496, if you calculate using 2972)		

NOTE: When the citations were uploaded onto the DEPRESSD – GDS account on Distiller SR, Distiller detected the overlap in citations between the different databases (i.e. MEDLINE, Psycinfo and WoS), which RefWorks was not able to detect (which had lead to the 2972 no. of citations in the MEDLINE folder in RefWorks and total number of citations to 9496) and skipped them. So the Total number of citation uploaded on DEPRESSD-GDS on Distiller is 9464.

NOTE: 30/07/2018 - De-Duplication - Before De-duplication in DistillerSR - 9464 citations.114 additional duplicates were removed by DistillerSR, leaving 9350 unique citations.

Web of Science

TS=(screen* OR prevalence OR "predictive value*" OR detect* OR sensitiv* OR valid* OR revalid* OR predict* OR accura* OR psychometric* OR identif* OR specificit* OR cutoff* OR "cut off*" OR "cut* score*" OR cutpoint* OR "cut point*" OR "threshold score*" OR "reference standard*" OR "reference test*" OR "index test*" OR "gold standard" OR "reliab*")
AND

TS=(GDS* OR “Geriatric Depression Scale*”)

PsycInfo

1	geriatric depression scale*.mp.	8643
2	GDS*.mp.	1741
3	1 or 2	9118
4	Diagnosis/	38494
5	Medical Diagnosis/	6672
6	Psychodiagnosis/	11207
7	Misdiagnosis/	515
8	Screening/	8719
9	Health Screening/	2931
10	Screening Tests/	4974

11 Prediction/	14366
12 Cutting Scores/	303
13 Psychometrics/	51843
14 Test Validity/	62121
15 screen*.af.	246719
16 predictive value*.af.	30082
17 detect*.ti.	14839
18 sensitiv*.ti.	15767
19 valid*.ti.	41010
20 revalid*.ti.	47
21 accura*.ti.	7580
22 psychometric*.ti.	11717
23 specificit*.ab.	30160
24 cut?off*.ab.	5626
25 cut* score*.ab.	2594
26 cut?point*.ab.	265
27 threshold score*.ab.	174
28 reference standard*.ab.	502
29 reference test*.ab.	90
30 index test*.ab.	73
31 gold standard.ab.	4174
32 or/4-31	453786
33 3 and 32	5686

MEDLINE

1 Geriatric depression scale*.af.
2 GDS*.af.
3 1 or 2
4 Mass Screening/
5 Psychiatric Status Rating Scales/
6 "Predictive Value of Tests"/
7 "Reproducibility of Results"/
8 exp "Sensitivity and Specificity"/
9 Psychometrics/
10 Prevalence/

11 Reference Values/
12 Reference Standards/
13 exp Diagnostic Errors/
14 validation studies.pt.
15 comparative study.pt.
16 screen*.af.
17 prevalence.af.
18 predictive value*.af.
19 detect*.ti.
20 sensitiv*.ti.
21 valid*.ti.
22 revalid*.ti.
23 predict*.ti.
24 accura*.ti.
25 psychometric*.ti.
26 identif*.ti.
27 specificit*.ab.
28 cut?off*.ab.
29 cut* score*.ab.
30 cut?point*.ab.
31 threshold score*.ab.
32 reference standard*.ab.
33 reference test*.ab.
34 index test*.ab.
35 gold standard.ab.
36 Mental disorders/di, pc
37 Mood disorders/di, pc
38 depressive disorder/di, pc
39 depressive disorder, major/di, pc
40 depression, postpartum/di, pc
41 depression/di, pc
42 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41
43 3 and 42

SUPPLEMENTARY FILE 2: Coding Manual

Title/Abstract Screening

1. **Exclude if no original human data or it is a case study.**

   Exclude if it is clear from the title and abstract that the article is not an original report of primary data, but for example a letter, editorial, systematic review or meta-analysis, or if it is a case series or single case study. Studies reporting only on animal, cellular, or genetic data are also excluded. Studies that report results in conference abstracts are eligible for inclusion.

2. **Exclude if study did not involve administration of the GDS-30, GDS-15, GDS-5 or GDS-4.**

   Exclude if there is no mention in the title or abstract of any of these versions of the Geriatric Depression Scale (GDS).

3. **Exclude if there is no assessment of major depression.**

   Exclude studies if it is clear from the title and abstract that a clinical interview for depression was not conducted. Only studies that assess adults for a DSM diagnosis of current (30-day or actual presence) MDD/MDE or ICD diagnosis of a current major depressive episode will be included. Studies that include broader diagnostic categories, such as other depressive (e.g., minor depression, dysthymia) or anxiety disorders, are eligible for inclusion only if they may have separate classifications of adults with MDD or major depressive episode in the primary data. It is unlikely that studies can be excluded at the title/abstract level based on differential diagnosis (e.g., major versus major + minor depression).

4. **Exclude if studies do not use a validated diagnostic interview to assess major depression.**

   Only studies that assess adults for a DSM diagnosis of current (30-day or actual presence) MDD/MDE or ICD diagnosis of a current major depressive episode using a validated structured or semi-structured diagnostic interview will be included. Examples of validated diagnostic interviews and other assessment tools that are not validated diagnostic interviews are listed below. Studies that clearly only used a self-report questionnaire to classify patients as depressed are excluded. If studies appear to have conducted a clinical interview to diagnose depression based on the title/abstract review, but it is not clear if a validated diagnostic interview was used, they should be included for full-text review.

   Examples of validated diagnostic interviews:
   Composite International Diagnostic Interview (CIDI)
   Diagnostic Interview Schedule (DIS)

Diagnostic Interview Schedule for Children (DISC)
Diagnostisches Interview bei psychischen Störungen im Kindes (Kinder-DIPS)
Mini-International Neuropsychiatric Interview (MINI)
Schedule for Affective Disorders and Schizophrenia (SADS)
Schedules for Clinical Assessment in Neuropsychiatry (SCAN)
Structured Clinical Interview for DSM (SCID)

Examples of assessment tools that are not validated diagnostic interviews:
Any self-report measure completed by patients
Hamilton Depression Rating Scale (HAM-D, HDRS)
Montgomery Asberg Depression Rating Scale (MADRS)
Primary Care Evaluation of Mental Disorders (PRIME-MD)
WHO Major Depression Inventory
International Diagnostic Checklist for ICD-10

5. Exclude if GDS and diagnostic interview not administered within 2 weeks of each other.

Studies are excluded if it is clear based on the title and abstract that the GDS and diagnostic interview were not administered within two weeks of one another, such as in a longitudinal study that administered one at one time point and the other at a different time point.

6. Exclude if sample selection is based on the presence of distress or depression.

Studies of patients who are pre-selected as possibly distressed or depressed (e.g., based on clinician's judgment or screening instrument cut-off) prior to administration of the study screening tool and diagnostic interview are excluded. Studies of patients receiving psychiatric treatment or with psychiatric diagnoses are excluded with the exception of studies of substance or alcohol abuse patients. Studies in which only part of the sample is selected based on distress or depression may be eligible if data for patients not selected due to distress levels can be obtained. If only patients above a cutoff score on the GDS are administered the diagnostic interview, the study is excluded. If, however, a proportion of patients both above and below the GDS cutoff are administered the interview, the study would be included.

7. Exclude if not adults.

Studies are excluded if it is clear from the title/abstract that the study sample does not include adults aged 18 and over. Studies with mixed population samples are eligible for inclusion if data for adults can be obtained. However, studies that assess only pediatric, adolescent, school or undergraduate samples will not be included, even if some participants are at least 18 years old.

Full Text Review

1. **Exclude if no original human data or it is a case study.**

   Exclude if the article is not an original report of primary data, but for example a letter, editorial, systematic review or meta-analysis, or it is a case series or single case study. Studies reporting only on animal, cellular, or genetic data are also excluded. Studies that report results in conference abstracts are eligible for inclusion.

2. **Exclude if study did not involve administration of the GDS-30, GDS-15, GDS-5 or GDS-4.**

   Exclude if patients were not administered the GDS-30, GDS-15, GDS-5 or GDS-4.

3. **Exclude if there is no assessment of major depression.**

   Exclude studies if there is not a clinical interview to diagnose current (30-day or actual) MDD based on DSM or a current major depressive episode based on ICD. Studies that include broader diagnostic categories, such as other depressive (e.g., minor depression, dysthymia) or anxiety disorders, are eligible for inclusion only if they have classified adults with MDD or major depressive episode in the primary data.

   Examples of inclusion / exclusion of different depression diagnoses:
   DSM-IV-TR:
   Include: Major Depression.
   Exclude: Dysthymic Disorder, Minor Depression (at least two depressive symptoms are present for two weeks).

   ICD-10:
   Include: mild, moderate, severe, recurrent depressive episodes.
   Exclude: recurrent brief depressive disorder (requires a depressive episode with symptomatic criteria, but lasting less than 2 weeks and requires that the episodes occur at least once per month for 12 consecutive months).

   RESEARCH DIAGNOSTIC CRITERIA (RDC):
   Include: Major Depressive Disorder.

   DSM-III:
   Include: Major depression.
   Exclude: Dysthymic Disorder, atypical affective disorders.

4. **Exclude if studies do not use a validated diagnostic interview to assess major depression.**

   Only studies that assess adults for a DSM diagnosis of current (30-day or actual

presence) MDD or ICD diagnosis of a current major depressive episode using a validated structured or semi-structured diagnostic interview will be included. Examples of validated diagnostic interviews and other assessment tools that are not validated diagnostic interviews are listed below. Studies that clearly only used a self-report questionnaire to classify patients as depressed are excluded.

Examples of validated diagnostic interviews:
Composite International Diagnostic Interview (CIDI)
Diagnostic Interview Schedule (DIS)
Diagnostic Interview Schedule for Children (DISC)
Diagnostisches Interview bei psychischen Störungen im Kindes (Kinder-DIPS)
Mini-International Neuropsychiatric Interview (MINI)
Schedule for Affective Disorders and Schizophrenia (SADS)
Schedules for Clinical Assessment in Neuropsychiatry (SCAN)
Structured Clinical Interview for DSM (SCID)

Examples of assessment tools that are not validated diagnostic interviews:
Any self-report measure completed by patients
Hamilton Depression Rating Scale (HAM-D, HDRS)
Montgomery Asberg Depression Rating Scale (MADRS)
Primary Care Evaluation of Mental Disorders (PRIME-MD)
International Diagnostic Checklist for ICD-10

5. Exclude if GDS and diagnostic interview not administered within 2 weeks of each other.

Studies are excluded if the GDS and diagnostic interview were not administered within two weeks of one another. 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 allows us to select patients administered the diagnostic interview and screening tools within the two-week window.

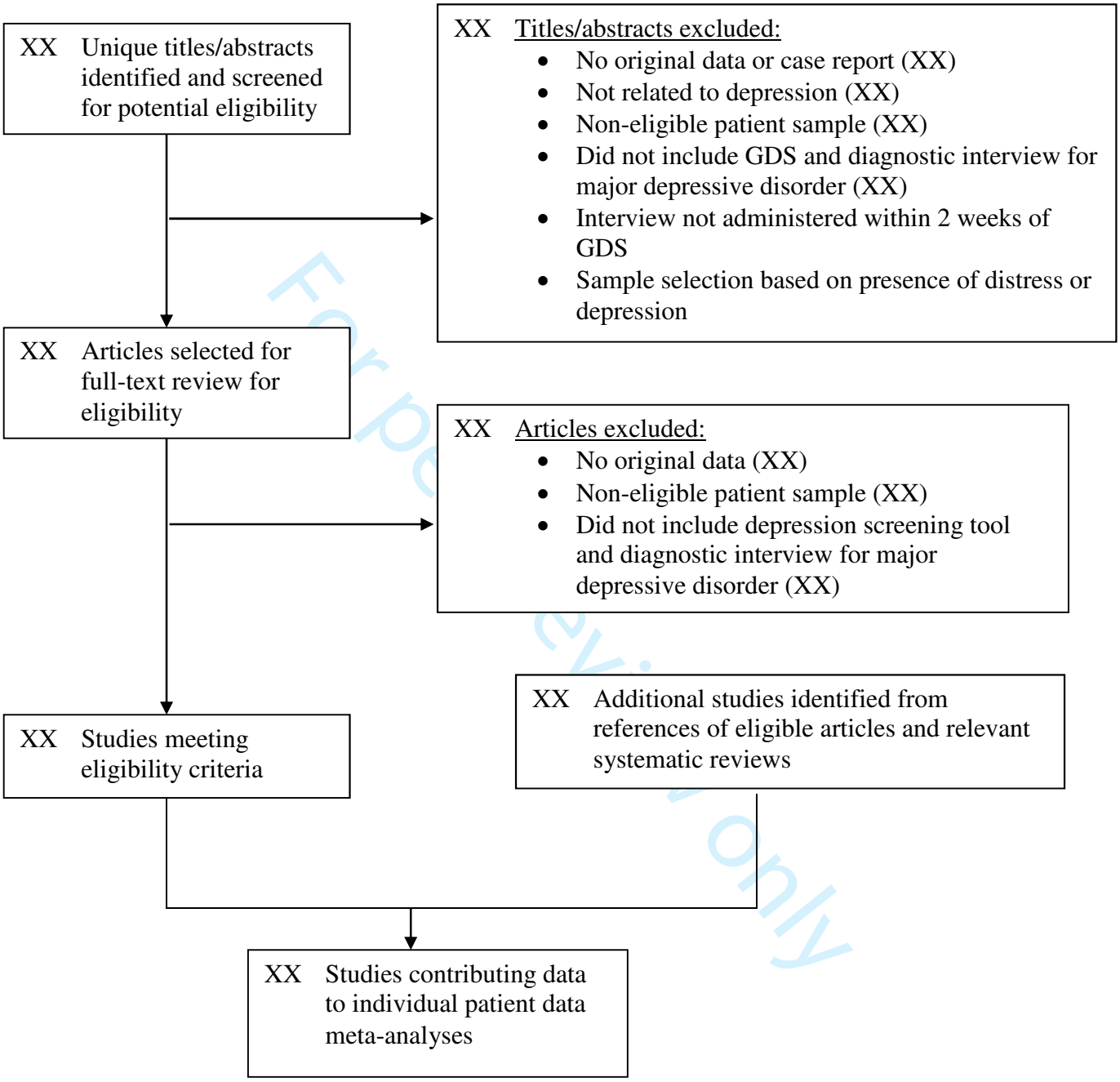
6. Exclude if sample selection is based on the presence of distress or depression.

Studies of patients who are pre-selected as possibly distressed or depressed (e.g., based on clinician's judgment or screening instrument cut-off) prior to administration of the study screening tool and diagnostic interview are excluded. Studies of patients receiving psychiatric treatment or with psychiatric diagnoses are excluded with the exception of studies of substance or alcohol abuse patients. Studies in which only part of the sample is selected based on distress or depression may be eligible if data for patients not selected due to distress levels can be obtained. If only patients above a cutoff score on the GDS are administered the diagnostic interview, the study is excluded. If, however, a proportion of patients both above and below the GDS cutoff are administered the interview, the study would be included.

7. Exclude if not adults.

Studies are excluded if the study sample does not include adults aged 18 and over. Studies with mixed population samples are eligible for inclusion if data for adults can be obtained. However, studies that assess only pediatric, adolescent, school or undergraduate samples will not be included, even if some participants are at least 18 years old.

SUPPLEMENTARY FILE 3: Draft Flow Diagram of Study Selection Process



PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Page No	Checklist item
			ADMINISTRATIVE INFORMATION
Title:			
Identification	1a	1	Identify the report as a protocol of a systematic review
Update	1b	N/A	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	4	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:			
Contact	3a	1-2	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	16	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	9	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:			
Sources	5a	16	Indicate sources of financial or other support for the review
Sponsor	5b	16	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	16-17	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
			INTRODUCTION
Rationale	6	6-8	Describe the rationale for the review in the context of what is already known
Objectives	7	8	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
			METHODS
Eligibility criteria	8	10-11	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	9-10	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	24-29	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:			
Data	11a	11-12	Describe the mechanism(s) that will be used to manage records and data throughout the review

management			
Selection process	11b	10-11	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	11-12	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	13-14	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	12-15	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	12	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	11-12	Describe criteria under which study data will be quantitatively synthesised
	15b	12-15	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)
	15c	12-14	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	N/A	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	N/A	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	N/A	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

Diagnostic Accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5, and Geriatric Depression Scale-4 for Detecting Major Depression: Protocol for a Systematic Review and Individual Participant Data Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026598.R1
Article Type:	Protocol
Date Submitted by the Author:	08-Oct-2018
Complete List of Authors:	Benedetti, Andrea; McGill University, Departments of Medicine and of Epidemiology, Biostatistics & Occupational Health; McGill University Health Centre, Respiratory Epidemiology and Clinical Research Unit Wu, Yin; McGill University, Departments of Psychiatry and of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Levis, Brooke; McGill University, Department of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Wilchesky, Machelle ; McGill University, Departments of Medicine and of Family Medicine; Jewish General Hospital, Lady Davis Institute for Medical Research Boruff, Jill; McGill University, Schulich Library of Science and Engineering Ioannidis, John; Stanford University, USA, Stanford Prevention Research Center, Departments of Medicine and of Health Research & Policy Patten, Scott ; University of Calgary, Departments of Community Health Sciences and of Psychiatry Cuijpers, Pim; VU University Amsterdam, Department of Clinical, Neuro and Developmental Psychology and Amsterdam Public Health research institute Shrier, Ian; McGill University, Department of Epidemiology, Biostatistics & Occupational Health; Jewish General Hospital, Lady Davis Institute for Medical Research Gilbody, Simon; The University of York, Department of Health Sciences Ismail, Zahinoor; University of Calgary, Departments of Community Health Sciences, of Psychiatry and of Clinical Neurosciences McMillan, Dean; University of York, Department of Health Sciences Mitchell, Nicholas; University of Alberta, Department of Psychiatry Ziegelstein, Roy; Johns Hopkins University School of Medicine, Department of Medicine Thombs, Brett; McGill University, Psychiatry
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Diagnostics, Mental health
Keywords:	depression, Geriatric Depression Scale, GDS, diagnostic test accuracy,

	individual participant data meta-analysis

Diagnostic Accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5, and Geriatric Depression Scale-4 for Detecting Major Depression: Protocol for a Systematic Review and Individual Participant Data Meta-Analysis

Andrea Benedetti, PhD[1-3]; Yin Wu, PhD[2,4,5]; Brooke Levis, MSc[2,5]; Machelle Wilchesky, PhD[3,5,6]; Jill Boruff, MLIS[7]; John P. A. Ioannidis, MD, DSc[8,9]; Scott B. Patten, MD, PhD[10,11]; Pim Cuijpers, PhD[12]; Ian Shrier, MD, PhD[2,5]; Simon Gilbody, DPhil, FRCPsych[13]; Zahinoor Ismail, MD FRCPC[10,11,14]; Dean McMillan, PhD[13]; Nicholas Mitchell, MD, MSc[15]; Roy C. Ziegelstein, MD[16]; Brett D. Thombs, PhD[2-5,17,18]

[1]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Quebec, Canada; Departments of [2]Epidemiology, Biostatistics and Occupational Health; [3]Medicine; and [4]Psychiatry; McGill University, Montreal, Quebec, Canada; [5]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; [6]Department of Family Medicine, McGill University, Montreal, Quebec, Canada; [7]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Quebec, Canada; [8]Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford School of Medicine; Stanford, California, USA; [9]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA; Departments of [10]Community Health Sciences and [11]Psychiatry, University of Calgary, Calgary, Canada; [12]Department of Clinical, Neuro and Developmental Psychology and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, The

Netherlands; [13]Hull York Medical School and Department of Health Sciences, University of York, York, UK; [14]Department of Clinical Neurosciences, University of Calgary, Calgary, Edmonton, Canada; [15]Department of Psychiatry, University of Alberta, Edmonton, Canada; [16]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Departments of [17]Psychology and [18]Educational and Counselling Psychology; McGill University, Montreal, Quebec, Canada.

Addresses for Correspondence:

Andrea Benedetti, PhD; 5252 de Maisonneuve Blvd. West; Montreal, Quebec, Canada H4A 3S5; Tel (514) 934-1934 ext. 32161; E-mail: andrea.benedetti@mcgill.ca

Brett D. Thombs, PhD; 4333 Cote Ste Catherine Road; Montreal, Quebec, Canada H3T 1E4; Tel (514) 340-8222 ext. 25112; E-mail: brett.thombs@mcgill.ca.

Key Words: depression; Geriatric Depression Scale; GDS; diagnostic test accuracy; individual participant data meta-analysis.

Word Count: 3,018

ABSTRACT

Introduction: The thirty-item Geriatric Depression Scale (GDS-30) and the shorter GDS-15, GDS-5 and GDS-4 are recommended as depression screening tools for elderly individuals. Existing meta-analyses on the diagnostic accuracy of the GDS have not been able to conduct subgroup analyses, have included patients already identified as depressed who would not be screened in practice, and have not accounted for possible bias due to selective reporting of results from only better-performing cutoffs in primary studies. Individual participant data meta-analysis (IPDMA), which involves a standard systematic review, then a synthesis of individual participant data, rather than summary results, could address these limitations. The objective of our IPDMA is to generate accuracy estimates to detect major depression for all possible cutoffs of each version of the GDS among studies using different reference standards, separately, and among participant subgroups based on age, sex, dementia diagnosis, and care settings. In addition, we will use a modelling approach to generate individual participant probabilities for major depression based on GDS scores (rather than a dichotomous cutoff) and participant characteristics (e.g., sex, age, dementia status, care setting).

Methods and Analysis: Individual participant data comparing GDS scores to a major depression diagnosis based on a validated structured or semi-structured diagnostic interview will be sought via a systematic review. Data sources will include Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and Web of Science. Bivariate random-effects models will be used to estimate diagnostic accuracy parameters for each cutoff of the different versions of the GDS. Pre-specified subgroup analyses will be conducted. Risk of bias will be assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

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

Systematic Review Registration: PROSPERO (CRD42018104329)

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This study will use individual participant data to estimate diagnostic accuracy for all relevant cutoff scores of the different versions of the Geriatric Depression Scale (GDS). Using data from all participants at each cutoff score will overcome limitations related to selective cutoff reporting in primary study publications.
- This study will conduct analyses that exclude patients with current diagnoses of depression or who are undergoing mental health treatment, including antidepressants, 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.
- This study will include subgroup analyses of diagnostic accuracy across different reference standards and by participant characteristics (e.g., sex, age, dementia status, care setting).
- A potential limitation is that the success of the study depends on the ability to obtain the relevant individual participant 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.

BACKGROUND

Major depression is present in 5-10% of the geriatric population internationally.[1,2] Effective treatments for depression are available, but identification is often haphazard. Physicians may fail to recognize up to half of all patients with depression, and most patients with depression do not receive minimally adequate care.[3-4] At the same time, there is a high rate of overdiagnosis and overtreatment, and the majority of patients who are treated do not meet diagnostic criteria.[5-7] Diagnosis of elderly individuals can be particularly difficult for clinicians due to factors such as cognitive impairments, social stigma, medical comorbidity, and atypical or vague clinical presentation.[1,8,9]

Some Canadian and international geriatric care organizations recommend screening elderly adults for depression,[10-13] but the Canadian Task Force for Preventive Health Care (CTFPHC), for instance, does not recommend depression screening, including for geriatric individuals.[14] The CTFPHC has expressed concern that published studies may overstate the accuracy of depression screening tools and that screening could lead to high rates of false positive tests, and still not improve depression outcomes.[14]

The thirty-item Geriatric Depression Scale (GDS-30), and the GDS-15, GDS-5 and GDS-4, which are fifteen-item, five-item and four-item subsets of the GDS-30, are commonly recommended as depression screening tools for elderly individuals.[15-17] As with other depression screening tools, primary studies on the diagnostic accuracy of the different versions of the GDS have been limited by (1) small samples; (2) the selective reporting of results for cutoffs when they perform well in a given sample, but not when they perform poorly; (3) the inclusion of patients already known by clinicians to have depression; and (4) the inability to conduct subgroup analyses (e.g., different age groups, dementia diagnosis, care settings) due to

small sample sizes. Conventional meta-analyses of the GDS or short versions of the GDS that have synthesized published summary data have not been able to conduct subgroup analyses or exclude already-diagnosed patients,[15,16,18] and concerns have been raised about bias in these meta-analyses due to selective cutoff reporting in primary studies that could not be addressed.[18]

Individual participant data meta-analysis (IPDMA), which involves a standard systematic review, followed by synthesis of actual participant data from primary studies, rather than aggregating summary data, can address these problems by including actual participant data from all studies.[19,20] In the context of evaluating the diagnostic accuracy of depression screening tools, IPDMA has three major advantages compared to conventional meta-analyses. First, for the conventional binary screening approach, IPDMA can address bias from the selective publication of diagnostic accuracy results for well-performing cutoffs from small studies since accuracy can be evaluated across all relevant cutoffs for all participants. Second, IPDMA allows the appropriate exclusion of already-diagnosed or already-treated patients when primary studies have data on existing diagnoses and treatment. Third, an IPDMA with large numbers of participants and major depression cases would allow subgroup analyses by study-level factors (e.g., study setting, risk of bias factors) and individual factors that may influence screening accuracy (e.g., age, sex, dementia diagnosis). Finally, a large IPD database would allow the development of a predictive algorithm to generate estimates of the probability of having major depression based on participant characteristics and actual GDS scores, rather than binary classifications of individuals as simply negative or positive based on screening results. This is important because, for instance, an individual with a score of 0 on the GDS-30, may have a lower likelihood of having depression than an individual with a substantially higher, but sub-cutoff, score of 10. Using a dichotomous

cutoff method, however, both would be classified as negative screens and assigned the same probability of having depression.

One of the downsides of IPDMAs is that they are resource intensive. In addition, if the primary datasets obtained are not representative of all primary studies, the IPDMA could be biased.[19-22] In a previous IPDMA of the Patient Health Questionnaire-9 (PHQ-9) screening tool, which was the first IPDMA of the diagnostic accuracy of a depression screening tool,[23] we were able to synthesize 58 of 72 eligible primary datasets (17,357 participants, 2,312 major depression cases). 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 IPDMA. A preliminary PubMed search for the GDS verified the existence of enough primary studies (more than 100 potentially eligible datasets that appear to have at least 30,000 participants, 4,000 cases) to make IPDMA feasible for the GDS.

Thus, the objectives of this IPDMA are to evaluate the diagnostic accuracy of the GDS-30, GDS-15, GDS-5 and GDS-4 among studies using different reference standards, separately; among participant subgroups based on age, sex, dementia diagnosis, and care settings; and excluding participants identified as already-diagnosed or treated for depression. Furthermore, a prediction model will be generated.

METHODS AND ANALYSIS

This systematic review has been funded by the Canadian Institutes of Health Research (Funding Reference Number PJT-156365). The protocol has been registered in the PROSPERO prospective register of systematic reviews (CRD42018104329), and any changes to the study protocol will be registered as amendments with PROSPERO.

The IPDMA 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 Accuracy[24] and other key sources.[19,20,25] Results will be reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement and the PRISMA-IPD statement.[26,27] The IPDMA protocol does not deviate substantively from previous IPDMA protocols that we have developed and published for other depression screening tools.[23,28,29]

Sources of Evidence

The search strategy was developed by a medical librarian and was adapted from a search strategy developed for a similar systematic review to obtain datasets for IPDMA of the PHQ-9 depression screening tool,[23] which was peer-reviewed using the Peer Review of the Electronic Search Strategy (PRESS) standard.[30] The search strategy is also similar to strategies that we have used for systematic reviews and IPDMA of the Hospital Anxiety and Depression Scale and Edinburgh Postnatal Depression Scale.[28,29]

We will search Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO (OvidSP platform), and Web of Science (Web of Knowledge platform). The Medline search strategy for the GDS was validated by testing against already-identified publications from our preliminary search. 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 (e.g., EMBASE) when the Medline search is highly sensitive does not identify additional eligible studies.[31] The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[24] 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 comprised 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 key articles were consulted in developing search terms.[32-34] See Supplementary File 1 for detailed information on searches. To supplement electronic searches, we will search reference lists of included publications and relevant reviews, conduct a related articles search using the PubMed "related articles" feature, and query authors of included studies for unpublished studies. Search results will be uploaded into the citation management database RefWorks (RefWorks-COS, Bethesda, MD, 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.

Selection of Eligible Studies

To conduct the meta-analysis, we will seek primary datasets that allow us to compare GDS scores to major depression diagnostic status. Datasets from articles in any language will be sought for inclusion if they compare results from any version of the GDS to diagnoses of major depressive disorder (MDD) or major depressive episode (MDE) made with a validated diagnostic interview, administered within 2 weeks of the GDS and based on Diagnostic and Statistical Manual (DSM) or International Classification of Diseases criteria (ICD), which are similar to DSM criteria and generally used outside of North America.

The two-week criterion was set because that is the duration of symptoms required for a diagnosis of major depression. Datasets where some participants were administered the screening tool within 2 weeks of the diagnostic interview and some participants were not will be included

if the original data allows us to identify and select eligible participants. 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 participants with psychotic disorders or a history of manic episodes. If both are available, we will record both and prioritize DSM over ICD and MDE over MDD in analyses. Data from studies where all participants 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 participants treated for substance use disorders, for whom depression screening may be considered. The coding manual for inclusion and exclusion decisions is shown in Supplementary File 2.

Two investigators will review articles independently for eligibility. If either reviewer determines that a study may be eligible based on title or abstract review, 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 used to evaluate titles/abstracts and articles for languages other than those for which team members are fluent. See Supplementary File 3 for a preliminary PRISMA flow of studies figure.

Transfer of Data and Data 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 de-identified prior to transfer. Participant data will be cleaned and coded for uniformity across datasets using an already developed codebook, similar to codebooks used in our previous IPDMAs.[23,28,29] 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. Participant characteristics and screening

accuracy results for each study using the cleaned datasets will be compared to those from the original datasets to identify any potential discrepancies.

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

Quality Assessment

Two reviewers will independently use the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool[35] to assess risk of bias in primary studies. QUADAS-2 incorporates assessments of risk of bias across four core domains: participant selection, the index test, the reference standard, and the flow and timing of assessments. Two reviewers will independently assess risk of bias, and any discrepancies will be resolved by consensus.

Data Analysis

Analyses will estimate sensitivity and specificity separately at each cutoff by bivariate random-effects meta-analysis models as described in Riley et al.[36] For each GDS version, we will fit these models, estimated via Gaussian Hermite adaptive quadrature, for the full range of plausible GDS cutoff values.[36] This approach models sensitivity and specificity simultaneously and accounts for the precision of estimates within studies.[36] Data from all included primary studies will be analyzed simultaneously with a random-effects model as sensitivity and specificity are assumed to vary across primary studies. We will also construct a pooled ROC curve and identify the optimal cutoff.[36] 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. For assessment of each version, we will include studies that report total scores for the specific GDS version or individual GDS item scores from longer versions of the GDS which could be used to calculate total scores for the shorter version. We will consider imputation if a large part of data is missing.

In a previous IPDMA with the PHQ-9,[37] we found that reference standards appeared to perform differently. The Mini International Neuropsychiatric Interview (MINI) is fully structured, but was designed for very rapid administration and described as its authors as being over-inclusive as a result. We found that, controlling for depressive symptom scores, the MINI classified approximately twice as many participants with major depression as other fully structured interviews.[38,39] Compared to semi-structured interviews, which are intended to be done by experienced diagnosticians and involve some clinical judgment (e.g., Structured Clinical Interview for DSM Disorders) fully structured interviews (MINI excluded), diagnosed more participants with low symptom levels as depressed and fewer participants with higher symptom levels. Fully structured interviews can be delivered by lay interviewers and are intended to achieve a high level of standardization, but may sacrifice accuracy.[40-43] Thus, we will assess possible differences and evaluate sensitivity and specificity separately by reference standard.

In secondary analyses, to the extent that there are sufficient data, we will investigate subgroups according to age, sex, dementia status and severity, dementia subtype, number of medical comorbidities (with specific comorbidities integrated to the extent possible), care setting and risk of bias. QUADAS-2 factors that will be considered include patient selection factors, blinding of reference standard to index test results, and timing between administration of index test and reference standard (e.g., 0 to 7 days, 7 to 14 days). 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.[44]

If there is a sufficient number of studies with published diagnostic accuracy data for major depression that are eligible but do not provide data, studies included in the IPDMA will be compared to 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 individual participant data in the main meta-analysis, along with data from studies that contribute to the IPDMA.[36] If there are a large number of studies that do not contribute primary data, this analysis may become the primary analysis.

Clinical predictive models have not been used previously to generate individualized probabilities that an individual has major depression based on screening tool scores and participant characteristics. There is a rich tradition of using predictive models for risk scores or classifying patients based on diagnostic tests, and our approach will build upon those traditions.[45-50] To do this, we will develop binary predictive models that use GDS scores as well as key participant characteristics (e.g., sex, age, dementia status, care setting) to estimate the probability and associated 95% CI that an individual has major depression. We will estimate logistic mixed models and then integrate over the distribution of the random effects as described in Pavlou et al. and Skrondal et al.[51,52] Continuous variables (GDS score and age) will be modeled using flexible semi-parametric methods (e.g., regression splines). We will consider the inclusion of interaction terms. The models will be evaluated in terms of their overall performance (Nagelkerke's R^2, Brier score), calibration (e.g., slope of linear predictor; are

average, low and high predictions correct) and discrimination (e.g., c-statistic; discrimination slope: can we separate subjects with and without major depression).[45,46] Validation with the same subjects used to develop a model results in overly optimistic performance. We will assess internal validation via the bootstrap method, which has been shown to be preferable to split sample validation approaches.[47] Although there are advantages to external validation, given the wide range of study populations that we will be using, it would be unlikely that there would be another comparable data set large enough for validation. Thus, assessment of internal validity via bootstrapping will allow us to understand how our model may perform in a clinical setting, and by adjusting our regression coefficients for optimism, the performance of our model will be as accurate as possible. In sensitivity analyses, we will explore including each item from the GDS questionnaires as a separate predictor variable, rather than only the total score.

Patient and Public Involvement

Patients and members of the public were not involved in the study.

ETHICS AND DISSEMINATION

This IPDMA does not require ethics approval, although only individual studies that obtained ethical clearance and informed consent will be included. The reason that the IPDMA does not require ethics review is that the objectives of the IPDMA are consistent with the objectives of the primary studies, which already received ethics approval, and only anonymized data will be provided by the investigators of the original studies.

The main outcomes of the IPDMA 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,[53-58] including research on improving the usefulness of reports of systematic review and meta-

analyses for health care managers and policy makers.[56,58] Dissemination will include publication of results in high-impact medical journals with open access, as well as presentations in seminars and symposia to policy-makers, health care providers, and researchers at national and international conferences.

If the predictive model performs well, a free and easy-to use online calculation tool will be created to incorporate individual characteristics into accuracy estimates and provide users of our research with probabilities that individual patients have depression based on their GDS score and key characteristics. The calculator will be similar to other successful tools, such as the FRAX® Fracture Risk Assessment Tool (http://www.shef.ac.uk/FRAX/index.aspx). The tool that will be made from the results of this study will be modeled on this tool and presented with tablet and app versions.

AUTHORS' CONTRIBUTIONS

AB, YW, BL, MW, JB, JPAI, SBP, PC, IS, SG, ZI, DM, NM, RCZ, and BDT contributed to the conception and design of the systematic review and meta-analysis. JB developed the database search strategy. AB, YW, BL, MW, JB, and BDT will be involved in acquisition of data. AB, YW, BL, and BDT will analyze the data. All authors will contribute to the interpretation of results. AB, YW, and BDT drafted this protocol. All authors provided critical revisions of the protocol and approved submission of the final manuscript. AB is the guarantor.

FUNDING STATEMENT

This research is supported by a grant from the Canadian Institutes of Health Research (CIHR; Funding Reference Number PJT-156365; PIs = Benedetti, Thombs, Wilchesky). Drs.

Benedetti and Thombs are supported by the Fonds de recherche du Québec - Santé (FRQS) researcher salary award. Dr. Wu is supported by an Utting Postdoctoral Fellowship from the Jewish General Hospital, Montreal, Quebec. Ms. Levis is supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award. Dr. Patten is a Senior Health Scholar with Alberta Innovates, Health Solutions. Dr. Ismail receives funding from the Alzheimer Society Calgary via the Hotchkiss Brain Institute. Dr. Wilchesky is supported by the Donald Berman Maimonides Medical Research Foundation. No funding body had any input into any aspect of this protocol.

COMPETING INTERESTS STATEMENT

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

REFERENCES

1. Fiske A, Wetherell JL, Gatz M. Depression in older adults. *Annu Rev Clin Psychol* 2009;5:363-89.
2. Luppa M, Sikorski C, Luck T, et al. Age- and gender-specific prevalence of depression in latest-life – Systematic review and meta-analysis. *J Affect Disord* 2012;136:212-221.
3. Duhoux A, Fournier L, Nguyen CT, Roberge P, Beveridge R. Guideline concordance of treatment for depressive disorders in Canada. *Soc Psychiatry Psychiatr Epidemiol* 2009;44:385-392.
4. Duhoux A, Fournier L, Gauvin L, Roberge P. What is the association between quality of treatment for depression and patient outcomes? A cohort study of adults consulting in primary care. *J Affect Disord* 2013;151:265-274.
5. Mojtabai R, Olfson M. Proportion of antidepressants prescribed without a psychiatric diagnosis is growing. *Health Aff* 2011;30:1434-1442.
6. Mojtabai R. Clinician-identified depression in community settings: Concordance with structured interview diagnoses. *Psychother Psychosom* 2013;82:161-169.
7. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. *JAMA Intern Med* [Epub ahead of print].
8. Jeste DV, Alexopoulos GS, Bartels SJ, et al. Consensus statement on the upcoming crisis in geriatric mental health: research agenda for the next 2 decades. *Arch Gen Psychiatry* 1999;56:848-853.
9. Weeks R, Kalucy R, Hill J. Chapter 3: Depression in Elderly Patients. In: Jackson S, Jansen P, Mangoni A, eds. *Prescribing for Elderly Patients.* John Wiley & Sons, Ltd; 2009.

10. Registered Nurses Association of Ontario. *Screening for Delirium, Dementia, and Depression in Older Adults.* Toronto, Ontario: Registered Nurses Association of Ontario; 2003.

11. American Medical Directors Association (AMDA). *Depression in the long-term care setting.* Columbia, Maryland: American Medical Directors Association (AMDA); 2011.

12. American Geriatrics Society, American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral symptoms associated with dementia. *J Am Geriatr Soc*. 2003;51:1287-1298.

13. Buchanan D, Tourigny-Rivard M, Cappeliez P, et al. National guidelines for seniors' mental health: the assessment and treatment of depression. *Can J Geriatr* 2006;9:52-58.

14. Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775-782.

15. 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;18:1066-1077.

16. 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.

17. Montorio I, Izal M. The Geriatric Depression Scale: a review of its development and utility. *Int Psychogeriatr* 1996;8:103-112.

18. Pocklington C, Gilbody S, Manea L, McMillan D. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J Geriatr Psychiatry* 2016;31:837-857.

19. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.

20. 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.

21. Jaspers GJ, Degraeuwe PLJ. A failed attempt to conduct an individual patient data meta-analysis. *Systematic Reviews* 2014;3:97.

22. Parish E, Bloom T, Godlee F. Statins for people at low risk. *BMJ* 2015;351:h3908.

23. 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. *Systematic Reviews* 2014;3:124.

24. Deeks JJ, Bossuyt PM, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration; 2009.

25. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Chichester, England: Wiley-Blackwell; The Cochrane Collaboration 2011.

26. McInnes MD, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018; 319:388-96.

27. 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.

28. Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the Hospital Anxiety and Depression Scale (HADS) for detecting major depression: Protocol for a systematic review and individual patient data meta-analyses. *BMJ Open* 2016; 6:e011913.

29. Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: Protocol for a systematicrReview and individual patient data meta-analyses. *BMJ Open* 2015;5:e009742.

30. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiology* 2016;75:40-6.

31. Sampson M, Barrowman NJ, Moher D, et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003;56:943-55.

32. 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;1:9,4053-1-9.

33. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391-7.

34. Wong SS, Wilczynski NL, Haynes RB, et al. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003:728-32.

35. 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.

36. 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.

37. Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *Br J Psychiatry* 2018;212:377-85.

38. Lecrubier Y, Sheehan DV, Weiller E et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;12:224–231.

39. Sheehan DV, Lecrubier Y, Sheehan KH et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry* 1997;12:232–241.

40. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol Medicine* 2001;31:1001–13.

41. Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychol Med* 1999;29:1013-20.

42. Nosen E, Woody SR. Chapter 8: Diagnostic assessment in research. In, McKay D. *Handbook of research methods in abnormal and clinical psychology.* Sage; 2008.

43. Kurdyak PA, Gnam WH. Small signal, big noise: performance of the CIDI depression module. *Can J Psychiatry* 2005;50:851-56.

44. 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.

45. Steyerberg EW. *Clinical Prediction Models*. New York: Springer; 2009.

46. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* 2010;21:128-38.

47. 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.

48. Dalton JE, Kattan MW. Recent advances in evaluating the prognostic value of a marker. *Scand J Clin Lab Invest* 2010; 242 (suppl):59–62.

49. Zhou Xh, McClish DK, Obuchowski NA. Statistical Methods in Diagnostic Medicine, 2nd ed. Wiley: Hoboken, NJ, 2011.

50. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer: New York, 2001

51. Pavlou M, Ambler G, Seaman S, Omar RZ. A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes. *BMC Medical Research Methodology* 2015;15:59.

52. Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models. *J R Statist Soc A* 2009;172:659-687.

53. Dobbins M, DeCorby K, Twiddy T. A knowledge transfer strategy for public health decision makers. *Worldviews Evid Based Nurs* 2004;1:120-8.

54. 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.

55. Haines A, Kuruvilla S, Borchert M. Bridging the implementation gap between knowledge and action for health. *Bull World Health Organ* 2004;82:724-31.

56. 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.

57. 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.

58. 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.

SUPPLEMENTARY FILE 1: Search Strategies

GDS screening accuracy

Researcher: Brett Thombs

Search strategy peer reviewed:

Database searches conducted:

Reference/related list searches conducted

Search Terms	
GDS* **Geriatric Depression Scale***	

Search Terms:

SCREENING ACCURACY

Medline "filter"	PsycINFO "filter"	Web of Science "filter"
Mass Screening/	Diagnosis/	TS=
Psychiatric Status Rating Scales/	Medical Diagnosis/	screen*
Predictive Value of Tests/	Psychodiagnosis/	prevalence
Reproducibility of Results/	Misdiagnosis/	"predictive value*"
Exp "Sensitivity and Specificity"/	Screening/	detect*
Psychometrics/	Health Screening/	sensitiv*
Prevalence/	Screening Tests/	valid*
Reference Values/	Prediction/	revalid*
Reference Standards/	Cutting Scores/	predict*
Exp Diagnostic Errors/	Psychometrics/	accura*
Mental Disorders/diagnosis	Test Validity/	psychometric*
Mood Disorders/diagnosis	screen*.af.	identif* specificit*
Depressive Disorder/diagnosis	predictive value*.af.	cutoff*
Depressive Disorder, Major/diagnosis	detect*.ti.	"cut off*"
Depression, Postpartum/diagnosis	sensitiv*.ti.	"cut* score*"
Depression/diagnosis	valid*.ti.	cutpoint*
	revalid*.ti.	"cut point*"
validation studies.pt.	accura*.ti.	"threshold score*"
comparative study.pt	psychometric*.ti.	"reference standard*"
	specificit*.ab.	"reference test*"
screen*.af.	cut?off*.ab.	"index test*"
prevalence.af.	cut* score*.ab.	"gold standard"
predictive value*.af.	cut?point*.ab.	

detect*.ti.	threshold score*.ab.
sensitiv*.ti.	reference standard*.ab.
valid*.ti.	reference test*.ab.
revalid*.ti.	index test*.ab
predict*.ti.	gold standard.ab.
accura*.ti.	
psychometric*.ti.	~~Psychological Assessment/~~
identif*.ti	~~Psychiatric Evaluation/~~
specificit*.ab.	~~Testing/~~
cut?off*.ab.	~~Test Interpretation/~~
cut* score*.ab.	~~Rating Scales/~~
cut?point*.ab.	~~prevalence.af.~~
threshold score*.ab.	~~predict*.ti.~~
reference standard*.ab.	~~identif*.ti~~
reference test*.ab.	
index test*.ab.	
gold standard.ab.	

Diagnostic accuracy (prediction) filters:

Some diagnostic accuracy terms were adapted from the following 2 hedges, demonstrated to be the two best strategies in [Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS One; 2012;7(2):e32844]:

Ingui BJ, Rogers MA (2001) Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 8: 391–397.

Wong SS, Wilczynski NL, Haynes RB, Ramkissoonsingh R (2003) Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. AMIA Annu Symp Proc 728–732.

Searches conducted: <July 16, 2018>

Platform	Databases(s)	Results	Saved (account)	Remarks
OvidSP	Ovid MEDLINE(R), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid OLDMEDLINE(R) 1946 to Present (MEDLINE search)	2960	DepressionSR/ depress	Screening Accuracy—Rebuild AND GDS
	PsycInfo	5686		
Clarivate	Web of Science	4166		
	Total database search results in EndNote	12812		
	Total deduplicated in EndNote			

De-Duplication using RefWorks: July 25th, 2018

Platform	**Database(s)**	**Results (No. Of Citations)**	**After de-Duplication (No. of citations)**	**Saved (account)**	**Remarks**
OvidSP	Ovid MEDLINE(R), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid OLDMEDLINE(R) 1946 to Present (MEDLINE search)	2960	2940 (Refworks shows 2972, which is not possible. Must be some error in RefWorks. 2940 was obtained by subtracting the no. of citation in psycInfo and WoS from the Total no. of citations)	DepressionSR/ depress	Screening Accuracy—Rebuild AND GDS
	PsycInfo	5686	4648		
Clarivate	Web of Science	4166	1876		

Total database search results in EndNote	12812	9464 (9496, if you calculate using 2972)		

NOTE: When the citations were uploaded onto the DEPRESSD – GDS account on Distiller SR, Distiller detected the overlap in citations between the different databases (i.e. MEDLINE, Psycinfo and WoS), which RefWorks was not able to detect (which had lead to the 2972 no. of citations in the MEDLINE folder in RefWorks and total number of citations to 9496) and skipped them. So the Total number of citation uploaded on DEPRESSD-GDS on Distiller is 9464.

NOTE: 30/07/2018 - De-Duplication - Before De-duplication in DistillerSR - 9464 citations.114 additional duplicates were removed by DistillerSR, leaving 9350 unique citations.

Web of Science

TS=(screen* OR prevalence OR "predictive value*" OR detect* OR sensitiv* OR valid* OR revalid* OR predict* OR accura* OR psychometric* OR identif* OR specificit* OR cutoff* OR "cut off*" OR "cut* score*" OR cutpoint* OR "cut point*" OR "threshold score*" OR "reference standard*" OR "reference test*" OR "index test*" OR "gold standard" OR "reliab*")
AND

TS=(GDS* OR “Geriatric Depression Scale*”)

PsycInfo

1	geriatric depression scale*.mp.	8643
2	GDS*.mp.	1741
3	1 or 2	9118
4	Diagnosis/	38494
5	Medical Diagnosis/	6672
6	Psychodiagnosis/	11207
7	Misdiagnosis/	515
8	Screening/	8719
9	Health Screening/	2931
10	Screening Tests/	4974
11	Prediction/	14366

12 Cutting Scores/	303
13 Psychometrics/	51843
14 Test Validity/	62121
15 screen*.af.	246719
16 predictive value*.af.	30082
17 detect*.ti.	14839
18 sensitiv*.ti.	15767
19 valid*.ti.	41010
20 revalid*.ti.	47
21 accura*.ti.	7580
22 psychometric*.ti.	11717
23 specificit*.ab.	30160
24 cut?off*.ab.	5626
25 cut* score*.ab.	2594
26 cut?point*.ab.	265
27 threshold score*.ab.	174
28 reference standard*.ab.	502
29 reference test*.ab.	90
30 index test*.ab.	73
31 gold standard.ab.	4174
32 or/4-31	453786
33 3 and 32	5686

MEDLINE

1 Geriatric depression scale*.af.
2 GDS*.af.
3 1 or 2
4 Mass Screening/
5 Psychiatric Status Rating Scales/
6 "Predictive Value of Tests"/
7 "Reproducibility of Results"/
8 exp "Sensitivity and Specificity"/
9 Psychometrics/
10 Prevalence/
11 Reference Values/

12 Reference Standards/
13 exp Diagnostic Errors/
14 validation studies.pt.
15 comparative study.pt.
16 screen*.af.
17 prevalence.af.
18 predictive value*.af.
19 detect*.ti.
20 sensitiv*.ti.
21 valid*.ti.
22 revalid*.ti.
23 predict*.ti.
24 accura*.ti.
25 psychometric*.ti.
26 identif*.ti.
27 specificit*.ab.
28 cut?off*.ab.
29 cut* score*.ab.
30 cut?point*.ab.
31 threshold score*.ab.
32 reference standard*.ab.
33 reference test*.ab.
34 index test*.ab.
35 gold standard.ab.
36 Mental disorders/di, pc
37 Mood disorders/di, pc
38 depressive disorder/di, pc
39 depressive disorder, major/di, pc
40 depression, postpartum/di, pc
41 depression/di, pc
42 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41
43 3 and 42

SUPPLEMENTARY FILE 2: Coding Manual

Title/Abstract Screening

1. **Exclude if no original human data or it is a case study.**

   Exclude if it is clear from the title and abstract that the article is not an original report of primary data, but for example a letter, editorial, systematic review or meta-analysis, or if it is a case series or single case study. Studies reporting only on animal, cellular, or genetic data are also excluded. Studies that report results in conference abstracts are eligible for inclusion.

2. **Exclude if study did not involve administration of the GDS-30, GDS-15, GDS-5 or GDS-4.**

   Exclude if there is no mention in the title or abstract of any of these versions of the Geriatric Depression Scale (GDS).

3. **Exclude if there is no assessment of major depression.**

   Exclude studies if it is clear from the title and abstract that a clinical interview for depression was not conducted. Only studies that assess adults for a DSM diagnosis of current (30-day or actual presence) MDD/MDE or ICD diagnosis of a current major depressive episode will be included. Studies that include broader diagnostic categories, such as other depressive (e.g., minor depression, dysthymia) or anxiety disorders, are eligible for inclusion only if they may have separate classifications of adults with MDD or major depressive episode in the primary data. It is unlikely that studies can be excluded at the title/abstract level based on differential diagnosis (e.g., major versus major + minor depression).

4. **Exclude if studies do not use a validated diagnostic interview to assess major depression.**

   Only studies that assess adults for a DSM diagnosis of current (30-day or actual presence) MDD/MDE or ICD diagnosis of a current major depressive episode using a validated structured or semi-structured diagnostic interview will be included. Examples of validated diagnostic interviews and other assessment tools that are not validated diagnostic interviews are listed below. Studies that clearly only used a self-report questionnaire to classify patients as depressed are excluded. If studies appear to have conducted a clinical interview to diagnose depression based on the title/abstract review, but it is not clear if a validated diagnostic interview was used, they should be included for full-text review.

   Examples of validated diagnostic interviews:
   Composite International Diagnostic Interview (CIDI)
   Diagnostic Interview Schedule (DIS)

Diagnostic Interview Schedule for Children (DISC)
Diagnostisches Interview bei psychischen Störungen im Kindes (Kinder-DIPS)
Mini-International Neuropsychiatric Interview (MINI)
Schedule for Affective Disorders and Schizophrenia (SADS)
Schedules for Clinical Assessment in Neuropsychiatry (SCAN)
Structured Clinical Interview for DSM (SCID)

Examples of assessment tools that are not validated diagnostic interviews:
Any self-report measure completed by patients
Hamilton Depression Rating Scale (HAM-D, HDRS)
Montgomery Asberg Depression Rating Scale (MADRS)
Primary Care Evaluation of Mental Disorders (PRIME-MD)
WHO Major Depression Inventory
International Diagnostic Checklist for ICD-10

5. Exclude if GDS and diagnostic interview not administered within 2 weeks of each other.

Studies are excluded if it is clear based on the title and abstract that the GDS and diagnostic interview were not administered within two weeks of one another, such as in a longitudinal study that administered one at one time point and the other at a different time point.

6. Exclude if sample selection is based on the presence of distress or depression.

Studies of patients who are pre-selected as possibly distressed or depressed (e.g., based on clinician's judgment or screening instrument cut-off) prior to administration of the study screening tool and diagnostic interview are excluded. Studies of patients receiving psychiatric treatment or with psychiatric diagnoses are excluded with the exception of studies of substance or alcohol abuse patients. Studies in which only part of the sample is selected based on distress or depression may be eligible if data for patients not selected due to distress levels can be obtained. If only patients above a cutoff score on the GDS are administered the diagnostic interview, the study is excluded. If, however, a proportion of patients both above and below the GDS cutoff are administered the interview, the study would be included.

7. Exclude if not adults.

Studies are excluded if it is clear from the title/abstract that the study sample does not include adults aged 18 and over. Studies with mixed population samples are eligible for inclusion if data for adults can be obtained. However, studies that assess only pediatric, adolescent, school or undergraduate samples will not be included, even if some participants are at least 18 years old.

Full Text Review

1. **Exclude if no original human data or it is a case study.**

   Exclude if the article is not an original report of primary data, but for example a letter, editorial, systematic review or meta-analysis, or it is a case series or single case study. Studies reporting only on animal, cellular, or genetic data are also excluded. Studies that report results in conference abstracts are eligible for inclusion.

2. **Exclude if study did not involve administration of the GDS-30, GDS-15, GDS-5 or GDS-4.**

   Exclude if patients were not administered the GDS-30, GDS-15, GDS-5 or GDS-4.

3. **Exclude if there is no assessment of major depression.**

   Exclude studies if there is not a clinical interview to diagnose current (30-day or actual) MDD based on DSM or a current major depressive episode based on ICD. Studies that include broader diagnostic categories, such as other depressive (e.g., minor depression, dysthymia) or anxiety disorders, are eligible for inclusion only if they have classified adults with MDD or major depressive episode in the primary data.

   Examples of inclusion / exclusion of different depression diagnoses:
   DSM-IV-TR:
   Include: Major Depression.
   Exclude: Dysthymic Disorder, Minor Depression (at least two depressive symptoms are present for two weeks).

   ICD-10:
   Include: mild, moderate, severe, recurrent depressive episodes.
   Exclude: recurrent brief depressive disorder (requires a depressive episode with symptomatic criteria, but lasting less than 2 weeks and requires that the episodes occur at least once per month for 12 consecutive months).

   RESEARCH DIAGNOSTIC CRITERIA (RDC):
   Include: Major Depressive Disorder.

   DSM-III:
   Include: Major depression.
   Exclude: Dysthymic Disorder, atypical affective disorders.

4. **Exclude if studies do not use a validated diagnostic interview to assess major depression.**

   Only studies that assess adults for a DSM diagnosis of current (30-day or actual

presence) MDD or ICD diagnosis of a current major depressive episode using a validated structured or semi-structured diagnostic interview will be included. Examples of validated diagnostic interviews and other assessment tools that are not validated diagnostic interviews are listed below. Studies that clearly only used a self-report questionnaire to classify patients as depressed are excluded.

Examples of validated diagnostic interviews:
Composite International Diagnostic Interview (CIDI)
Diagnostic Interview Schedule (DIS)
Diagnostic Interview Schedule for Children (DISC)
Diagnostisches Interview bei psychischen Störungen im Kindes (Kinder-DIPS)
Mini-International Neuropsychiatric Interview (MINI)
Schedule for Affective Disorders and Schizophrenia (SADS)
Schedules for Clinical Assessment in Neuropsychiatry (SCAN)
Structured Clinical Interview for DSM (SCID)

Examples of assessment tools that are not validated diagnostic interviews:
Any self-report measure completed by patients
Hamilton Depression Rating Scale (HAM-D, HDRS)
Montgomery Asberg Depression Rating Scale (MADRS)
Primary Care Evaluation of Mental Disorders (PRIME-MD)
International Diagnostic Checklist for ICD-10

5. Exclude if GDS and diagnostic interview not administered within 2 weeks of each other.

Studies are excluded if the GDS and diagnostic interview were not administered within two weeks of one another. 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 allows us to select patients administered the diagnostic interview and screening tools within the two-week window.

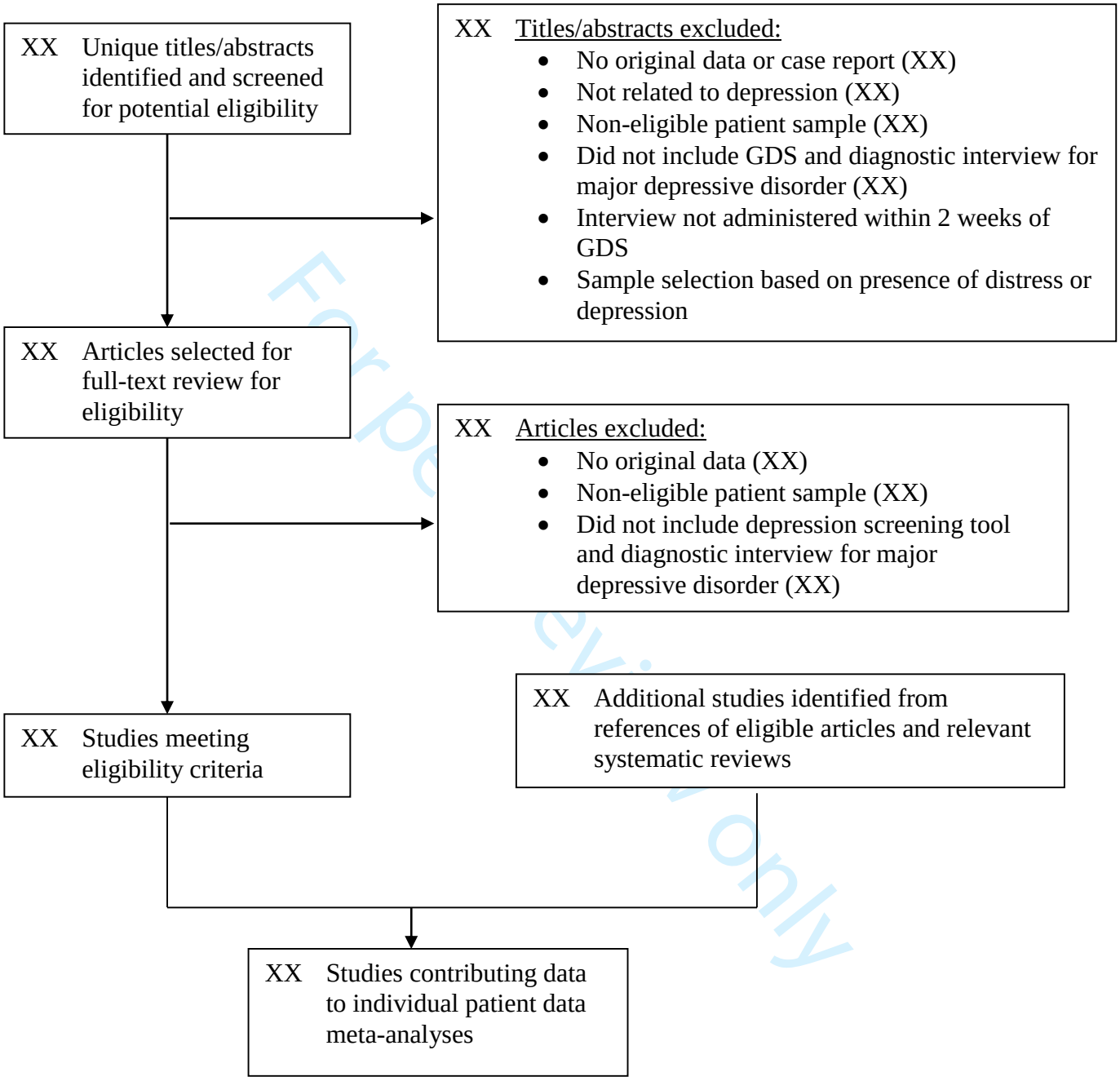
6. Exclude if sample selection is based on the presence of distress or depression.

Studies of patients who are pre-selected as possibly distressed or depressed (e.g., based on clinician's judgment or screening instrument cut-off) prior to administration of the study screening tool and diagnostic interview are excluded. Studies of patients receiving psychiatric treatment or with psychiatric diagnoses are excluded with the exception of studies of substance or alcohol abuse patients. Studies in which only part of the sample is selected based on distress or depression may be eligible if data for patients not selected due to distress levels can be obtained. If only patients above a cutoff score on the GDS are administered the diagnostic interview, the study is excluded. If, however, a proportion of patients both above and below the GDS cutoff are administered the interview, the study would be included.

7. Exclude if not adults.

Studies are excluded if the study sample does not include adults aged 18 and over. Studies with mixed population samples are eligible for inclusion if data for adults can be obtained. However, studies that assess only pediatric, adolescent, school or undergraduate samples will not be included, even if some participants are at least 18 years old.

SUPPLEMENTARY FILE 3: Draft Flow Diagram of Study Selection Process



PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Page No	Checklist item
			ADMINISTRATIVE INFORMATION
Title:			
Identification	1a	1	Identify the report as a protocol of a systematic review
Update	1b	N/A	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	4	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:			
Contact	3a	1-2	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	16	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	9	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:			
Sources	5a	16	Indicate sources of financial or other support for the review
Sponsor	5b	16	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	16-17	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
			INTRODUCTION
Rationale	6	6-8	Describe the rationale for the review in the context of what is already known
Objectives	7	8	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
			METHODS
Eligibility criteria	8	10-11	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	9-10	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	24-29	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:			
Data	11a	11-12	Describe the mechanism(s) that will be used to manage records and data throughout the review

management			
Selection process	11b	10-11	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	11-12	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	13-14	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	12-15	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	12	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	11-12	Describe criteria under which study data will be quantitatively synthesised
	15b	12-15	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)
	15c	12-14	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	N/A	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	N/A	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	N/A	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.