Selective cutoff reporting in studies of the accuracy of depression screening tools: comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis

Dipika Neupane

Department of Epidemiology, Biostatistics and Occupational Health

McGill University, Montreal

July 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Epidemiology (Thesis)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CI	Confidence Interval
C-DIS	Computerized Diagnostic Interview Schedule
CIDI	Composite International Diagnostic Interview
CIHR	Canadian Institutes of Health Research
CIS-R	Clinical Interview Schedule – Revised
DEPRESSD	DEPRESsion Screening Data
DIGS	Diagnostic Interview for Genetic Studies
DIS	Diagnostic Interview Schedule
DISH	Depression Interview and Structured Hamilton
DSM	Diagnostic and Statistical Manual of Mental Disorders
EPDS	Edinburgh Postnatal Depression Scale
FRQ-S	Fonds de recherche du Québec - Santé
ICD	International Classification of Diseases
IPD	Individual Participant Data
IPDMA	Individual Participant Data Meta-Analysis
MDD	Major Depressive Disorder
MDE	Major Depressive Episode
MINI	Mini International Neuropsychiatric Interview
PHQ-9	Patient Health Questionnnaire-9
ROC	Receiver Operating Characteristic
SADS	Schedule for Affective Disorders and Schizophrenia
SCAN	Schedule for Clinical Assessment in Neuropsychiatry

SCID	Structured Clinical Interview for DSM Disorders
STARD	Standards for Reporting of Diagnostic Accuracy Studies

ABSTRACT

Background: Selectively reporting accuracy results from only well-performing cutoffs could result in biased accuracy estimates in meta-analyses of diagnostic test accuracy studies. A previous individual participant data meta-analysis (IPDMA) of 13 Patient Health Questionnaire-9 (PHQ-9) diagnostic accuracy studies observed bias due to selective reporting of cutoffs. But the extent of bias may differ depending on the availability of a well-defined standard cutoff.

Objectives: Bias in accuracy estimates and cutoff reporting patterns was compared for the PHQ-9 (well-defined standard cutoff 10, i.e. score ≥ 10) and Edinburgh Postnatal Depression Scale (EPDS; no standard cutoff, common cutoffs 10 to 13).

Methods: Medline, Medline In-Process & Other Non-Indexed Citations and PsycINFO via OvidSP, and Web of Science via ISI Web of Knowledge were searched from January 2000 to February 2015 (PHQ-9) and inception to June 2016 (EPDS) for studies that published at least one cutoff with the PHQ-9 or EPDS. Separately, for the PHQ-9 and EPDS, bivariate random effects meta-analysis was used to compare accuracy estimates based on published cutoffs only versus all cutoffs from all studies. The number of published cutoffs below and above the standard or common cutoffs was compared in relation to study-specific "optimal" cutoffs.

Results: In the IPDMA, 30 unique PHQ-9 diagnostic accuracy studies (11,773 participants and 1,587 major depression cases) and 19 unique EPDS diagnostic accuracy studies (3,637 participants and 531 major depression cases) were included. Compared to IPDMA, PHQ-9 sensitivity estimates based on published cutoffs were underestimated for cutoffs below 10

(median difference: -0.06) and overestimated for cutoffs above 10 (median difference: 0.07). EPDS sensitivity estimates were similar for cutoffs below 10 (median difference: 0.01) but higher for published cutoffs above 13 (median difference: 0.14). Mean cutoff of all cutoffs reported among PHQ-9 studies with optimal cutoffs below 10 was 8.8 compared to 11.8 for studies with optimal cutoffs above 10. 18 of 19 EPDS studies had optimal cutoffs below 13; those with below 10 did not report more cutoffs below 10 (mean cutoff: 9.9), but those with above 10 reported more cutoffs above 10 (mean cutoff: 11.8).

Conclusion: Selective cutoff reporting and resulting bias in accuracy estimates were more pronounced for the PHQ-9 than the EPDS. Researchers evaluating diagnostic accuracy of screening tools should report accuracy results for all relevant cutoffs.

RÉSUMÉ

Contexte: La communication sélective de résultats de précision provenant uniquement de seuils bien performants pourrait entraîner des estimations de précision biaisées dans les méta-analyses des études d'exactitude des tests de diagnostic. Une précédente méta-analyse des données individuelles des participants (IPDMA) de 13 études de précision diagnostique du questionnaire de santé des patients-9 (PHQ-9) a observé un biais dû à la déclaration sélective des seuils. Mais la mesure du biais peut différer en fonction de la disponibilité d'un seuil standard bien défini.

Objectifs: Les biais dans les estimations de précision et les profils de rapport de seuil ont été comparés pour le PHQ-9 (seuil standard bien défini ≥ 10) et pour l'Édimbourg Postnatal Depression Scale (EPDS; pas de seuil standard, seuils communs 10 à 13).

Méthodes: Medline, Medline In-Process & Other Non-Indexed Citations et PsycINFO via OvidSP, et Web of Science par ISI Web of Knowledge ont été recherchés de janvier 2000 à février 2015 (PHQ-9) et du début à juin 2016 (EPDS) pour des études qui publié au moins un seuil avec le PHQ-9 ou EPDS. Séparément, pour le PHQ-9 et l'EPDS, une méta-analyse bivariée à effets aléatoires a été utilisée pour comparer les estimations de précision basées uniquement sur les seuils publiés et tous les seuils de toutes les études. Le nombre de seuils publiés en dessous et au-dessus des seuils standard ou communs a été comparé par rapport aux seuils «optimaux» spécifiques à l'étude.

Résultats: Dans l'IPDMA, 30 études de précision diagnostique PHQ-9 uniques (11 773 participants et 1 587 cas de dépression majeure) et 19 études de précision diagnostique EPDS uniques (3 637 participants et 531 cas de dépression majeure) ont été incluses. Comparativement à l'IPDMA, les estimations de sensibilité au PHQ-9 basées sur les seuils publiés ont été sous-estimées pour les seuils inférieurs à 10 (différence médiane: -0,06) et surestimées pour les seuils supérieurs à 10 (différence médiane: 0,07). Les estimations de sensibilité à l'EPDS étaient similaires pour les seuils inférieurs à 10 (différence médiane: 0,01) mais plus élevées pour les seuils publiés supérieurs à 13 (différence médiane: 0,14). Le seuil moyen de tous les seuils signalés dans les études PHQ-9 avec des seuils optimaux inférieurs à 10 était de 8,8 par rapport à 11,8 pour les études avec des seuils optimaux supérieurs à 10. 18 des 19 études EPDS avaient des seuils optimaux inférieurs à 13; ceux avec moins de 10 n'ont pas rapporté plus de seuils en dessous de 10 (seuil moyen: 9,9), mais ceux avec plus de 10 ont rapporté plus de seuils au-dessus de 10 (seuil moyen: 11,8).

Conclusion: Les rapports de valeurs seuil sélectives et le biais résultant dans les estimations de précision étaient plus prononcés pour le PHQ-9 que pour l'EPDS. Les chercheurs qui évaluent la précision diagnostique des outils de dépistage devraient rapporter des résultats de précision pour tous les seuils pertinents.

ACKNOWLEDGEMENTS

I would like to thank several people for their support and contribution without whom the thesis could not be completed. A massive thanks to my supervisors, Dr. Brett Thombs and Dr. Andrea Benedetti for creating an excellent learning environment and training me for my career as a researcher. Thanks to Dr. Brooke Levis for her original ideas, mentorship and answers to all my questions which made tough coding much easier.

I would like to thank DEPRESSD project team members, for the quality of work they do for the project and happy working environment they create everyday. Thanks to the DEPRESSD project steering committee for providing valuable feedback on my protocol. I am extremely thankful to all the participants of the primary datasets and the data-contributors who provided data to the project. I would like to thank Julia Nordlund for translating my abstract.

I would like to express my gratitude to the Department of Epidemiology, Biostatistics and Occupational Health for all the knowledge I received, without which I would not be able to grasp the concepts required for completion of this project. Thanks to the McGill Faculty of Medicine for providing me the studentship awards which helped me financially in completion of the thesis.

I would like to thank my parents, Deepkala Neupane and Dipendra Keshari Neupane, and family for their support, love and encouragement.

Finally, a heartfelt thanks to Parash Mani Bhandari for helping me in each step of my thesis and for inspiring me, loving me and caring for me.

PREFACE AND CONTRIBUTIONS OF AUTHORS

The thesis is presented in manuscript-based format. It compares the selective cutoff reporting in studies of the accuracy of the depression screening tools with different characteristics, Patient Health Questionnaire-9 (PHQ-9) and Edinburgh Postnatal Depression Scale (EPDS).

Chapter 1 and 2 include introduction and literature review. I drafted these two chapters. Dr. Brett D. Thombs and Dr. Andrea Benedetti critically reviewed it.

Chapter 3 presents the manuscript prepared for submission to the *International Journal of Epidemiology*. This manuscript utilized data from DEPRESSD project, led by my supervisors Dr. Brett D. Thombs and Dr. Andrea Benedetti, which consists of trainees, staff, steering committee members, knowledge users and data contributors. This manuscript uses the methods and design originally used in a manuscript by Dr. Brooke Levis. I am the first author of the manuscript included in this thesis. I drafted the manuscript with contributions from Dr. Brooke Levis, Parash Mani Bhandari, Dr. Brett D. Thombs and Dr. Andrea Benedetti. All co-authors agreed to include the manuscript in this thesis. Contributions of all authors are provided in the manuscript in Chapter 3.

Chapter 4 provides a brief discussion of the thesis, which includes key findings, clinical and research implications, potential limitations and conclusion from the study. I drafted this chapter. Dr. Brett D. Thombs and Dr. Andrea Benedetti critically reviewed it.

At the end, I present references and an appendix that includes methodology and results of the manuscript.

CHAPTER 1. INTRODUCTION

Depression screening refers to using validated questionnaires to identify patients who may have depression, among those not previously diagnosed, to further assess them and, if required, treat them for depression.[1,2] Depression screening is controversial. In 2013, the Canadian Task Force for Preventive Health Care recommended against depression screening in primary care, raising a concern that the diagnostic accuracy results reported in the publications of depression screening studies may be over-estimated compared to real practice.[3] In primary diagnostic accuracy studies, results are often published for cutoffs that have high accuracy estimates in that particular study but not from other cutoffs that have less optimistic accuracy estimates.[4] Because of this tendency to report only the cutoffs around the best performing cutoff, a meta-analysis including these primary studies would likely produce biased accuracy estimates.

Selective reporting is a potential source of bias which arises due to reporting of only the most favorable outcomes.[5,6] In diagnostic test accuracy studies, selective cutoff reporting occurs when accuracies are calculated for multiple cutoffs, but the decision on which cutoffs to report is made depending upon the results.[4] When only the best-performing cutoffs are reported, the resulting accuracy estimates will overestimate the true accuracy of the screening tool. Primary depression screening studies also tend to report "standard" cutoff or cutoffs around the "standard" cutoff.[4,7,8] The "standard" cutoffs are usually obtained from early studies that included small number of participants and major depression cases; hence it cannot be confidently used as a best cutoff.[9-11] The limitation due to small sample size in obtaining the best cutoff can be addressed using a meta-analysis. But, results from a meta-analysis including primary studies that selectively reported cutoffs will also be biased.

Only one previous study,[4] which was based on 13 studies with 4,589 participants and 1,037 major depression cases, has investigated pattern of selective cutoff reporting. That study obtained accuracy estimates from an individual participant data meta-analysis (IPDMA) and compared it to accuracy estimates obtained from meta-analysis of published cutoffs. It found that the sensitivity estimates were under-estimated for cutoffs below 10 (i.e. cutoff of ≥ 10), over-estimated for cutoffs above 10 and similar for standard cutoff of 10. This pattern was observed because primary studies tended to report cutoffs lower or higher than 10 depending upon the sensitivity of the PHQ-9 at cutoff of 10. It was not known whether the findings would be similar with a larger number of studies and participants. Moreover, in the previous study the pattern of selective cutoff reporting and comparison between IPDMA and aggregate data meta-analysis of published results was assessed only for the PHQ-9. The pattern of selective reporting may be different for other screening tools and may depend on how the standard cutoff is defined for the screening tool. For the PHQ-9, a cutoff of 10 is a well-defined standard cutoff that is used consistently.[9,10,12-14] The EPDS, on the other hand, which is the most commonly used screening tool among women in pregnancy and postpartum period,[15,16] does not have a well-defined standard cutoff. Different studies commonly report accuracies using cutoffs between 10 and 13 to identify major depression.[16,17]

In the present study, the objective was to examine how the presence or absence of well-defined standard cutoff may affect selective cutoff reporting. IPDMA on larger set of PHQ-9 studies was used and compared to the findings with EPDS, which does not have a well-defined standard cutoff. IPDMA was performed to synthesize results from all cutoffs for each included primary study and, separately, results from only cutoffs with published accuracy estimates in the original primary studies. Specific objectives were to (1) compare sensitivity and specificity based

on all cutoffs from all primary studies versus data from only cutoffs for which estimates were published in the primary studies; (2) explore cutoff reporting patterns with reference to the identified optimal cutoff in each primary study.

CHAPTER 2. LITERATURE REVIEW

2.1 Depression screening

Depression is the leading cause of disability among adults and is common among pregnant and postpartum mothers.[18] Over 300 million people are now living with depression; the percentage of people living with depression increased by more than 18% between 2005 and 2015.[19] Diagnosing depression in primary care settings is difficult because in addition to classic symptoms of depressed mood, patients often present with multiple comorbidities and somatic symptoms such as changes in appetite, changes in sleep, digestive problems and sexual dysfunction.[20] Because of the difficulty in diagnosing depression, primary care providers use self reported questionnaires for depression screening to identify patients who may have depression. Depression screening involves using a screening instrument, usually self-reported questionnaire, to identify patients who may have depression but who are not already diagnosed as having depression, so they can be further assessed by health care providers and treated if necessary.[1,2]

2.2 Depression screening tools

The PHQ-9,[12] a nine-item questionnaire, is the most commonly used depression screening tool in primary care and other medical settings.[13,14] The total score is 27, with higher scores representing more severe symptoms of depression. The standard cutoff of PHQ-9 is 10,[9,10,12-14] which was identified by the first validation study (N participants = 580, N major depression = 41).[9,10]

The EPDS is the most commonly used screening tool in pregnancy and postpartum.[15,16] EPDS has 10 items and the maximum score is 30. The first EPDS study, from 1987, which was based on a sample that included only 24 definite or probable major depression cases, suggested

cutoffs of 10 or 13 could be used.[11] But because the standard cutoff is not clear, studies commonly use cutoffs 10 to 13 as the standard, with 13 being the most commonly used.[16,17]

2.3 Accuracy of depression screening tools

In depression screening, diagnostic test accuracy is the ability of a test to distinguish between patients who have major depression (Major Depressive Disorder (MDD) or Major Depressive Episode (MDE)) from those who do not. Results from the depression screening tool, whose diagnostic test accuracy is to be assessed, are compared with the results from a validated diagnostic interview that is designed to reflect Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Disease (ICD) criteria for major depression. The test accuracy is often expressed as test's sensitivity (the probability that the patients with major depression will be correctly identified as depressed by the screening tool) and specificity (the probability that the patients without major depression will be correctly identified as not depressed by the screening tool).[21,22]

2.4 Selective cutoff reporting and aggregate data meta-analysis

Depression screening tools measure symptoms of participants in continuous or ordinal scale and provide a cumulative total score. To make a decision whether a participant may be depressed or not using a depression screening tool, a cutoff needs to be defined; the participants above the cutoff will be considered as positive screens.[23] Although selecting a screening cutoff in practice should evaluate the consequences of true and false positive screens, many studies identify an "optimal" cutoff by selecting a cutoff that maximizes both sensitivity and specificity. When the true "optimal" cutoff for the screening tool is not known, accuracy estimates for all relevant cutoffs should be reported. But, if authors make the decision whether to report particular cutoffs

after performing the test and report only those that maximize the sensitivity and specificity, the accuracy of the test will likely be over-estimated.[23]

In the context of selective cutoff reporting, results from aggregate data meta-analyses including primary studies that report cutoffs selectively will also be biased. A 2012 meta-analysis of the PHQ-9[7] accuracy studies discussed that the accuracy of cutoffs could not be compared properly in their study because different studies reported different cutoffs. Hence, all studies could not be meta-analyzed for all cutoffs. Due to selective cutoff reporting the sensitivity increased with the increase in cutoff from 9 to 11, which is mathematically impossible. Another study[4] obtained individual participant data (IPD) from 13 of 16 studies included in the 2012 meta-analysis[7] and compared results for all cutoffs based on IPDMA to results from meta-analysis of published cutoffs only. The study found that (a) estimates of sensitivity differed between the published and the IPD datasets with cutoffs lower than the standard cutoff of 10 underestimating and cutoffs higher than 10 overestimating, but the standard cutoff about the same; (b) that this could be explained by the reporting pattern; in the studies in which the PHQ-9 was poorly sensitive at the cutoff of 10, cutoff less than 10 was identified as optimal and the studies tended to publish accuracy estimates for cutoff 10 and below whereas, in the studies in which the PHQ-9 was highly sensitive at the cutoff of 10, cutoff greater than 10 was identified as optimal and the studies tended to publish accuracy estimates for cutoff of 10 and above. Thus, compared to the IPD dataset, for published dataset, sensitivity was underestimated for cutoffs below 10 and overestimated for cutoffs above 10.

2.5 Individual participant data meta-analysis in diagnostic test accuracy studies

In aggregate data meta-analysis of diagnostic accuracy studies, aggregate study level accuracy estimates (i.e., sensitivity and specificity) are synthesized; analyzing only results from

cutoffs reported by the authors. The limitation of not being able to include data from all studies for all cutoffs can be overcome using IPDMA, in which individual level patient data for each study are obtained and used for analysis.[24] Thus, accuracy from all cutoffs from all studies can be compared to identify the best cutoff with maximum sensitivity and specificity. Because of the benefit to include all the data from all primary studies it is considered as the gold standard in evidence synthesis by the Cochrane Collaboration.[25]

CHAPTER 3. MANUSCRIPT

The following manuscript presents the research that was done to achieve the objectives mentioned in Chapter 1:

Neupane D, Levis B, Bhandari PM, Thombs BD, Benedetti A, and the DEPRESsion Screening Data (DEPRESSD) Collaboration. *Selective cutoff reporting in studies of the accuracy of the PHQ-9 and EPDS depression screening tools: comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis***.**

This manuscript has been prepared for submission to the *International Journal of Epidemiology*.

Title

Selective cutoff reporting in studies of the accuracy of the PHQ-9 and EPDS depression screening tools: comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis

Authors

Dipika Neupane[1,2], Brooke Levis[1-3], Parash Mani Bhandari[1,2], Brett D Thombs[1,2,4-8,*], Andrea Benedetti[2,5,9,*], and the DEPRESsion Screening Data (DEPRESSD) Collaboration

Authors' affiliations

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

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

[3]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK

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

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

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

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

[8]Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada

[9]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada

* Co-senior authors.

DEPRESSD collaboration members: Ying Sun, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Chen He, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Yin Wu, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Ankur Krishnan, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Zelalem Negeri, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Mahrukh Imran, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Danielle B. Rice, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Kira E. Riehm, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Nazanin Saadat, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Marleine Azar, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Tatiana A. Sanchez, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Matthew J. Chiovitti, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Alexander W. Levis, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Jill T. Boruff, Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Québec, Canada; Pim Cuijpers, Department of Clinical, Neuro and Developmental Psychology, EMGO Institute, Vrije Universiteit Amsterdam, the Netherlands;

Simon Gilbody, Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; John P. A. Ioannidis, Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA; Lorie A. Kloda, Library, Concordia University, Montréal, Québec, Canada; Dean McMillan, Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; Scott B. Patten, Departments of Community Health Sciences and Psychiatry, University of Calgary, Calgary, Canada; Ian Shrier, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Roy C. Ziegelstein, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Liane Comeau, International Union for Health Promotion and Health Education, École de santé publique de l'Université de Montréal, Montréal, Québec, Canada; Nicholas D. Mitchell, Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada; Marcello Tonelli, Department of Medicine, University of Calgary, Calgary, Alberta, Canada; Simone N. Vigod, Women's College Hospital and Research Institute, University of Toronto, Toronto, Ontario, Canada; Dickens H. Akena, Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda; Rubén Alvarado, School of Public Health, Faculty of Medicine, Universidad de Chile, Santiago, Chile; Bruce Arroll, Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand; Muideen O. Bakare, Child and Adolescent Unit, Federal Neuropsychiatric Hospital, Enugu, Nigeria; Hamid R. Baradaran, Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran; Cheryl Tatano Beck, University of Connecticut School of Nursing, Mansfield, Connecticut, USA; Charles H. Bombardier, Department of Rehabilitation Medicine, University of Washington,

Seattle, Washington, USA; Adomas Bunevicius, Neuroscience Institute, Lithuanian University of Health Sciences, Kaunas, Lithuania; Gregory Carter, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; Marcos H. Chagas, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Linda H. Chaudron, Departments of Psychiatry, Pediatrics, Obstetrics and Gynecology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA; Rushina Cholera, Department of Pediatrics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA; Kerrie Clover, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; Yeates Conwell, Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA; Tiago Castro e Couto, Federal University of Uberlândia, Brazil; Janneke M. de Man-van Ginkel, Julius Center for Health Sciences and Primary Care, Department of Nursing Science, University Medical Center Utrecht – University Utrecht, Utrecht, the Netherlands; Jaime Delgadillo, Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK; Jesse R. Fann, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; Nicolas Favez, Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland; Daniel Fung, Department of Child & Adolescent Psychiatry, Institute of Mental Health, Singapore; Lluïsa Garcia-Esteve, Perinatal Mental Health Unit CLINIC-BCN. Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Bizu Gelaye, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; Felicity Goodyear-Smith, Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand; Thomas Hyphantis, Department of Psychiatry, Faculty of Medicine, School of Health Sciences, University of Ioannina, Ioannina,

Greece; Masatoshi Inagaki, Department of Psychiatry, Faculty of Medicine, Shimane University, Shimane, Japan; Khalida Ismail, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neurosciences, King's College London Weston Education Centre, London, UK; Nathalie Jetté, Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada; Dina Sami Khalifa, Department of Community Medicine, Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway; Mohammad E. Khamseh, Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran; Jane Kohlhoff, School of Psychiatry, University of New South Wales, Kensington, Australia; Zoltán Kozinszky, Department of Obstetrics and Gynecology, Danderyd Hospital, Stockholm, Sweden; Laima Kusminskas, Private Practice, Hamburg, Germany; Shen-Ing Liu, Programme in Health Services & Systems Research, Duke-NUS Medical School, Singapore; Manote Lotrakul, Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; Sonia R. Loureiro, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Bernd Löwe, Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Sherina Mohd Sidik, Cancer Resource & Education Centre, and Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia; Sandra Nakić Radoš, Department of Psychology, Catholic University of Croatia, Zagreb, Croatia; Flávia L. Osório, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Susan J. Pawlby, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; Brian W. Pence, Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at

Chapel Hill, Chapel Hill, North Carolina, USA; Tamsen J. Rochat, MRC/Developmental Pathways to Health Research Unit, School of Clinical Medicine, University of Witwatersrand, South Africa; Alasdair G. Rooney, Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburg, Edinburgh, Scotland, UK; Deborah J. Sharp, Centre for Academic Primary Care, Bristol Medical School, University of Bristol, UK; Lesley Stafford, Centre for Women's Mental Health, Royal Women's Hospital, Parkville, Australia; Kuan-Pin Su, Mind-Body Interface Laboratory and Department of Psychiatry, China Medical University Hospital, Taiwan; Sharon C. Sung, Department of Child & Adolescent Psychiatry, Institute of Mental Health, Singapore; Meri Tadinac, Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia; S. Darius Tandon, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; Pavaani Thiagayson, Institute of Mental Health, Singapore; Annamária Töreki, Department of Emergency, University of Szeged, Hungary; Anna Torres-Giménez, Perinatal Mental Health Unit CLINIC-BCN. Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Alyna Turner, School of Medicine and Public Health, University of Newcastle, New South Wales, Newcastle, Australia; Christina M. van der Feltz-Cornelis, Department of Health Sciences, HYMS, University of York, York, UK; Johann M. Vega-Dienstmaier, Facultad de Medicina Alberto Hurtado, Universidad Peruana Cayetano Heredia, Lima, Perú; Paul A. Vöhringer, Department of Psychiatry and Mental Health, Clinical Hospital, Universidad de Chile, Santiago, Chile; Jennifer White, Department of Physiotherapy, School of Primary and Allied Health Care, Monash University, Melbourne, Australia; Mary A. Whooley, Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA; Kirsty Winkley, Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, UK; Mitsuhiko Yamada,

Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawa-Higashi, Kodaira, Tokyo, Japan

Corresponding authors:

Andrea Benedetti, PhD; Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve, Montréal, Quebec, H4A 3S5, Canada; Tel (514) 934-1934 ext. 32161; E-mail: andrea.benedetti@mcgill.ca

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

ABSTRACT

Background: Selectively reporting accuracy results from only well-performing cutoffs in studies of diagnostic or screening tests may result in biased estimates when synthesized. We compared selective cutoff reporting in studies on Patient Health Questionnaire-9 (PHQ-9; well-defined standard cutoff 10, i.e. cutoff ≥10) and Edinburgh Postnatal Depression Scale (EPDS; no standard cutoff, common cutoffs 10 to 13) accuracy.

Methods: We analyzed individual participant data from primary studies. Separately, for the PHQ-9 and EPDS, we used bivariate random effects meta-analysis to compare accuracy estimates from published versus all cutoffs. We also compared the number of published cutoffs below and above the standard or common cutoffs in relation to study-specific "optimal" cutoffs.

Results: For the PHQ-9 (30 studies, N = 11,773), published results underestimated sensitivity compared to results for all cutoffs for cutoffs below 10 (median difference: -0.06) and overestimated for cutoffs above 10 (median difference: 0.07). EPDS (19 studies, N = 3,637) sensitivity estimates were similar for cutoffs below 10 (median difference: 0.01) but higher for published cutoffs above 13 (median difference: 0.14). Mean cutoff of all cutoffs reported among PHQ-9 studies with optimal cutoffs below 10 was 8.8 compared to 11.8 for studies with optimal cutoffs above 10. 18 of 19 EPDS studies had optimal cutoffs below 13; those below 10 did not report more cutoffs below 10 (mean: 9.9), but those with above 10 reported more above 10 (mean: 11.8).

Conclusion: Selective cutoff reporting was more pronounced for the PHQ-9 than EPDS. Researchers evaluating diagnostic accuracy should report results for all relevant cutoffs.

Key words: diagnostic test accuracy, individual participant data meta-analysis, meta-analysis, selective cutoff reporting, publication bias

INTRODUCTION

Selective reporting occurs when authors make decisions regarding publication of study results based on whether or not outcomes are favorable.[1] In studies of the accuracy of ordinal or continuous tests, selective cutoff reporting occurs when accuracy results are published for one or more cutoffs that maximize sensitivity and specificity in a particular study but not for other relevant cutoffs.[2,3] Selective cutoff reporting can lead to overestimation of diagnostic accuracy in primary studies and in meta-analyses that synthesize results from primary studies with selectively reported results.[4]

Only one previous study has investigated patterns of selective cutoff reporting in diagnostic test accuracy studies.[2] That study obtained individual participant data (IPD) from 13 primary studies included in a published meta-analysis of the accuracy of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool and compared results for all cutoffs from all included studies to results from published cutoffs only. Estimates of sensitivity differed substantially between published and IPD datasets for cutoffs lower and higher than the standard cutoff of 10 (i.e. ≥ 10) but were similar at the standard cutoff. This was because most studies published results for the standard cutoff, but authors tended to publish results from cutoffs lower or higher than 10 depending on whether the PHQ-9 was relatively poorly sensitive but specific (lower cutoffs published) or highly sensitive but poorly specific (higher cutoffs published) in their dataset.

A cutoff of 10 is used as the standard cutoff for screening for major depression with the PHQ-9[5-9] and maximizes combined sensitivity and specificity,[10] but standard cutoffs are less well-defined for other commonly used depression screening tools. The Edinburgh Postnatal Depression Scale (EPDS), is the most commonly used screening tool among women in

pregnancy and postpartum.[11,12] Different studies describe cutoffs between 10 and 13 as standard, with 13 being most commonly used.[12,13] A recent individual participant data meta-analysis (IPDMA) of the diagnostic accuracy of the EPDS[14] found that a cutoff of 11 maximized combined sensitivity and specificity.

The degree to which there is an agreed upon standard cutoff for a screening tool may influence selective cutoff reporting. Thus, the aim of the present study was to evaluate selective cutoff reporting with a substantially larger set of PHQ-9 studies than was used in the previous study[2] and to compare results to the EPDS, which does not have a well-defined standard cutoff. Specific objectives were to use IPDMA with the PHQ-9 and EPDS, separately, to (1) compare sensitivity and specificity based on all cutoffs from all primary studies versus data from only cutoffs for which accuracy estimates were published in the primary studies; and (2) explore cutoff reporting patterns with reference to the identified optimal cutoff in each primary study.

METHODS

We analyzed data accrued for IPDMAs on PHQ-9 and EPDS diagnostic accuracy (PROSPERO CRD42014010673, CRD42015024785), and protocols were published for each IPDMA.[15,16] The protocol for the present study, which was not part of the main IPDMA protocols, was published separately (https://osf.io/vw3bz/). The protocol described only the EPDS analysis, and we subsequently added the PHQ-9 to be able to compare screening tools with and without well-defined standard cutoffs. As this study involved only analysis of previously collected de-identified data and because all included studies were required to have obtained ethics approval and informed consent, the Research Ethics Committee of the Jewish General Hospital determined that ethics approval was not required.

Study eligibility

Datasets from articles in any language were eligible for the main IPDMAs if (1) they used the PHQ-9 or EPDS; (2) they included diagnostic classification for current Major Depressive Disorder (MDD) or Major Depressive Episode (MDE) using Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Diseases (ICD) criteria based on a validated diagnostic interview; (3) the interview and PHQ-9 or EPDS were administered within two weeks of each other; (4) participants were ≥ 18 years and not recruited from school-based settings (PHQ-9) or ≥ 18 years and pregnant or within 12 months postpartum (EPDS); and (5) participants were not recruited from psychiatric settings or because they had symptoms of depression, since screening is done to identify previously unrecognized cases. Datasets where not all participants were eligible were included if primary data allowed selection of eligible participants.

Many primary studies in the main IPDMA databases that contributed eligible datasets never published estimates of screening accuracy. Thus, for the present study, we restricted analyses to primary studies with publications that included sensitivity and specificity estimates for at least one PHQ-9 or EPDS cutoff for identifying major depression. We excluded studies if the sample size from the published primary study differed by > 10% from the sample included in our IPDMA datasets. Sample sizes from original primary studies and the IPDMA databases differed in some cases because, for instance, we excluded participants who were included in the original studies if there were > 2 weeks between their index test and reference standard administrations or if they were < 18 years old. We also excluded primary studies with publications that reported accuracy results only for diagnostic classifications broader than major depression (e.g., “any depressive disorder”) if the number of cases in the published article and IPDMA datasets differed by > 10%.

Search strategy and study selection

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations and PsycINFO via OvidSP, and Web of Science via ISI Web of Knowledge from January 1, 2000 to February 7, 2015 (Supplementary Methods 1) for the PHQ-9 and from inception to June 10, 2016 (Supplementary Methods 2) for the EPDS, using peer-reviewed search strategies.[17] We also reviewed reference lists of relevant reviews and queried contributing authors about non-published studies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA) for de-duplication and then into DistillerSR (Evidence Partners, Ottawa, Canada).

Two investigators independently reviewed titles and abstracts. If either deemed a study potentially eligible, full-text review was done by two investigators, independently, with disagreements resolved by consensus, consulting a third investigator when necessary. Translators were consulted for languages other than those for which team members were fluent.

Data contribution, extraction, and synthesis

Authors of eligible datasets were emailed invitations to contribute de-identified primary data at least three times, as necessary, then we emailed co-authors and attempted phone contact. For each study, we compared published results with results from raw datasets and resolved any discrepancies in consultation with primary study investigators. For defining major depression, we considered MDD or MDE based on DSM or ICD. If more than one was reported, we prioritized MDE over MDD and DSM over ICD. For studies with multiple time points, we included data from only the time point with the most participants. To facilitate comparison between published results and IPDMA results, we applied sampling weights in the IPDMA only when accuracy results reported in the original published study were calculated using weights.

Statistical analyses

We replicated the statistical analyses used in the previous study of selective cutoff reporting with the PHQ-9.[2] For both the PHQ-9 and EPDS, we estimated sensitivity and specificity from cutoffs up to 5 points below and above cutoffs used as standard (PHQ-9 cutoff 10, range 5 to 15; EPDS cutoffs 10 to 13, range 5 to 18). We compared meta-analyses results from data using only cutoffs for which accuracy estimates were published in the primary studies (the *published dataset*) and using data from all cutoffs from all studies (the *full dataset*).

For both sets of meta-analyses, for each cutoff, bivariate random-effects models were estimated via Gauss-Hermite quadrature.[18] This approach models sensitivity and specificity simultaneously, accounting for the inherent correlation between them and the precision of estimates within studies.

Differences in sensitivity and specificity estimates using *published* versus *full datasets*

In order to examine differences in results produced by meta-analyses based on *published* and *full datasets*, we constructed separate pooled receiver operator characteristic (ROC) curves. In addition, 95% confidence intervals (CI) for the differences between methods in sensitivity and specificity at each cutoff were constructed via bootstrap,[19,20] resampling at the study and subject level with 1000 iterations for each cutoff. We calculated the median absolute difference in estimated sensitivity and specificity across evaluated cutoffs.

Reporting patterns

We assessed whether primary studies tended to preferentially report low or high cutoffs depending on the study's sample-specific optimal cutoff. For each primary study, we identified the optimal cutoff that the authors explicitly described as optimal or using a similar term. If the authors did not identify an optimal cutoff, we used the cutoff that maximized Youden's J (sensitivity + specificity – 1).[21] For each study, we plotted the optimal cutoff, along with all other

cutoffs for which results were published. We noted whether the reported cutoffs tended to be low or high compared to the standard cutoff (PHQ-9: 10) or set of commonly used cutoffs (EPDS: 10 to 13). For studies with optimal cutoffs below and above the standard or commonly used cutoffs, separately, we calculated the mean of the cutoffs reported.

RESULTS

Identification of eligible studies

PHQ-9

For the main PHQ-9 IPDMA, 58 studies were included.[10] Of these, 28 studies were excluded from the present study because they did not publish diagnostic accuracy results for any PHQ-9 cutoffs or because the number of participants or major depression cases in the IPD dataset differed by more than 10% from the published studies or could not be determined (see Supplementary Figure 1a for primary study numbers included and excluded at each review stage and Supplementary Tables (1a-4a) for information on excluded studies). Thus, 30 unique studies (total N = 11,773, major depression N = 1587 (13%)) were included (see Supplementary Table 5a for study characteristics). Of the 30 studies, 7 reported only a single cutoff and 23 reported more than one cutoff. Of the 23 with multiple cutoffs reported, 18 identified an optimal cutoff in the published study; of those, 16 (89%) were described as based on Youden's J (N = 8) or equivalent to Youden's calculated from published cutoffs but did not have an explanation (N = 8).

EPDS

The original IPDMA dataset included 49 studies. Of these, 30 studies were excluded because they did not publish accuracy results or because published and IPDMA datasets differed by more than 10% for total sample or number of cases (see Supplementary Figure 1b and

Supplementary Tables (1b-4b)). Thus, 19 unique studies (total N = 3,637, major depression N = 531(15%)) were included (see Supplementary Table 5b). Of the 13 studies that reported more than one cutoff, 12 identified an optimal cutoff; of those 9 (75%) were based on Youden's J (N = 2) or did not have an explanation but matched what would have been obtained using Youden's J calculated from published cutoffs (N =7).

Differences in sensitivity and specificity estimates based on *published* versus *full datasets*

Table 1 shows sensitivity and specificity for the PHQ-9 and EPDS at each cutoff for the *published* and *full datasets* with the ROC plots in Figures 1 and 2.

PHQ-9

For the PHQ-9 (see Table 2), the difference between estimated sensitivity (*published – full dataset*) ranged from -0.09 to 0.10 (median 0.06). For cutoffs below 10, estimated sensitivity was lower for the *published dataset* (-0.02 to -0.09; median -0.06) with 95% CIs including zero but inclining more towards negative, whereas estimated specificity was higher (0.01 to 0.14; median 0.03) with 95% CIs including zero. For the standard cutoff 10, the differences in sensitivity and specificity were -0.01 (95% CI: -0.05, 0.01), and 0.01 (95% CI: 0.00, 0.04), respectively. For cutoffs above 10, estimated sensitivity was higher for the *published dataset* (0.00 to 0.10; median 0.07) with CIs including zero but inclining more towards positive, and estimated specificity was similar (0.00 to 0.02; median 0.01) with CIs including zero.

EPDS

For the EPDS (see Table 2), the difference between estimated sensitivity ranged from -0.04 to 0.20 (median 0.03) with all 95% CIs including zero. For cutoffs below 10, estimated sensitivity (-0.04 to 0.01; median 0.01), and estimated specificity (0.01 to 0.03; median 0.01) were similar for the *published and full datasets*. For cutoffs of 10 to 13, estimated sensitivity

differed by -0.02 to 0.03 (median 0.02), and estimated specificity differed by 0.00 to 0.02 (median 0.01). For cutoffs above 13, estimated sensitivity was higher for the *published dataset* (0.05 to 0.20; median 0.14), and estimated specificity was similar or lower (0.00 to -0.08; median 0.00).

Reporting patterns

PHQ-9

Figure 3 shows the pattern of reporting with respect to optimal cutoffs for included PHQ-9 studies; 9 studies had optimal cutoffs below 10, 14 equal to 10, 6 greater than 10 and 1 study had optimal cutoffs of both 10 and 12. Studies for which the PHQ-9 was poorly sensitive at the cutoff 10 (sensitivity 0.27 – 0.74),[22-30] had optimal cutoffs that were below 10. These studies tended to report more cutoffs below 10 than above 10 (mean of reported cutoffs 8.8). Studies for which the PHQ-9 was highly sensitive at cutoff 10 (sensitivity 0.85 – 1.00),[31-36] had optimal cutoffs that were greater than 10. These studies tended to report more cutoffs above 10 than below 10 (mean of reported cutoffs 11.8).

EPDS

Figure 4 shows the pattern of reporting cutoffs for the EPDS; 5 studies had optimal cutoffs below 10, 13 between 10 and 13, and 1 greater than 13. Studies for which the EPDS was poorly sensitive at cutoff 10 (sensitivity: 0.43 – 0.73),[37-41] had optimal cutoffs that were less than 10 (mean of reported cutoffs 9.9). Studies for which EPDS was highly sensitive at cutoff 10 (sensitivity: 0.82 – 1.00),[42-53] had optimal cutoffs greater than 10. These studies tended to report more cutoffs above 10 than below 10 (mean of reported cutoffs 11.8). All of these studies had optimal cutoffs between 10 and 13 with one exception, a study reported accuracy only for cutoff 13 even though sensitivity was low at this cutoff (sensitivity: 0.35).[54]

DISCUSSION

We compared cutoff reporting patterns and bias due to selective cutoff reporting between screening instruments with and without a clearly defined standard cutoff. We performed meta-analyses of published cutoffs and compared results to meta-analysis of all cutoffs using individual participant data from studies on the screening accuracy of the PHQ-9 and EPDS. Patterns suggesting selective cutoff reporting were identified for both the PHQ-9 and EPDS, but selective cutoff reporting and bias were more pronounced for the PHQ-9, which has a clearly defined standard cutoff, than for the EPDS, which does not have a clearly defined standard cutoff.

For the PHQ-9, compared to meta-analysis of the *full dataset,* which included results for all relevant cutoffs for all included studies, estimates of specificity using the *published dataset*, which included results from published cutoffs only, were similar; however, sensitivity was underestimated in the *published dataset* for PHQ-9 cutoffs below 10, similar for the standard PHQ-9 cutoff 10, and overestimated for cutoffs above 10. The underestimation of sensitivity for cutoffs below 10 and the overestimation of sensitivity for cutoffs above 10 can be explained by the cutoff reporting patterns in primary studies. Studies in which the PHQ-9 was poorly sensitive but more specific identified optimal cutoffs below 10 as optimal and tended to publish accuracy estimates for cutoffs below 10, whereas studies in which the PHQ-9 was highly sensitive but less specific identified optimal cutoffs above 10 and tended to publish accuracy estimates for cutoffs above 10.

For the EPDS, compared to the *full dataset,* estimated specificity using the *published dataset* was similar across all cutoffs; however estimated sensitivity was similar for cutoffs less than 10 and for the most commonly reported cutoffs 10 to 13, but overestimated for cutoffs

above 13. This may also be explained by the reporting pattern. For the EPDS, unlike the PHQ-9, only primary studies in which EPDS was highly sensitive at cutoff 10 tended to report more cutoffs above 10. Studies with poor sensitivity that reported optimal cutoffs below 10 reported results from cutoffs above 10 more often than comparable studies with the PHQ-9. This may be because the PHQ-9 has a single standard cutoff of 10, whereas for the EPDS it may be an expectation that results for commonly used cutoffs of 10 to 13 are reported.

Findings in context

The first validation study of the PHQ-9, which was done in 2001, included a sample with only 41 major depression cases and identified 10 as the standard cutoff.[5,7] Meta-analyses have subsequently verified that PHQ-9 cutoff 10 maximizes combined sensitivity and specificity.[10] Consequently, most PHQ-9 diagnostic accuracy studies have reported accuracy estimates for cutoff 10,[2,3] but studies have selectively reported accuracy estimates for cutoffs other than 10 depending upon the sensitivity of PHQ-9 at the cutoff 10. Only one previous study, an IPDMA which included 13 studies, 4589 participants and 1037 major depression cases, has examined selective cutoff reporting in screening instruments (for the PHQ-9).[2] The previous study found that when only published cutoffs were considered, the estimates of sensitivity were underestimated for cutoffs lower than 10, overestimated for cutoffs greater than 10, but similar at the standard cutoff of 10, which was explained by the cutoff reporting pattern. We replicated the analysis in the present study with much larger sample of 30 studies, 11,773 participants and 1587 cases and found that though the reporting patterns were similar, the magnitude of bias was lower in the present study. In the previous study, when the cutoff increased markedly from 9 to 10 and 10 to 11, the sensitivity also increased, which is not possible if all data are analyzed. In the

present study, the sensitivity increased when cutoff increased from 10 to 11, but the increment was minimal.

This was the first study to examine selective cutoff reporting with the EPDS. The first EPDS study, from 1987, which was based on a sample that included only 24 definite or probable major depression cases, suggested cutoffs of 10 or 13 could be used.[55] In the absence of a clearly defined standard cutoff, studies conducted since the original study have often reported a range of cutoffs from 10 to 13.[12,13] A recent IPDMA meta-analysis found that 11 maximized combined sensitivity and specificity.[14] Consequently, given the range of cutoffs that are often reported, selective cutoff reporting appears to be less pronounced as compared to the PHQ-9.

Clinical and research implications

To avoid the bias in reporting cutoffs, authors of primary studies should report accuracy estimates for all possible cutoffs. The Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist requires that estimates of diagnostic accuracy and precision, as well as the cross tabulation of the index test and the reference standard should be reported.[56] The checklist should also recommend reporting accuracy estimates for all possible cutoffs within the range of relevancy for ordinal index tests.

In the presence of selective cutoff reporting, meta-analyses based on accuracy estimates from published cutoffs only may result in biased estimates of diagnostic accuracy. In a 2012 aggregate-data meta-analysis, which meta-analyzed published cutoffs only, the sensitivity increased with the increase in cutoff from 9 to 11, a mathematical impossiblity.[57] When there are missing data from some cutoffs in primary studies, accuracy estimates in meta-analyses can be corrected by using modelling techniques[58] or by doing IPDMA, which has some advantages, but is highly resource intensive.[59-62]

Strengths and limitations

One major strength of this study is that we compared two depression screening instruments with different characteristics using IPDMA. We explored how the presence of a clearly defined standard cutoff versus the absence of such a standard may be associated with cutoff reporting patterns and bias due to selective cutoff reporting. A potential limitation is that we were not able to include data from all eligible studies; 14 of 69 (20%) eligible PHQ-9 studies and 24 of 72 (33%) eligible EPDS studies did not provide the data in the main IPDMA. However, 4 of the 14 (29%) PHQ-9 studies and 10 of the 24 (42%) EPDS studies that did not provide data did not publish diagnostic accuracy results for their databases, so they would not have been eligible for the present study. Another limitation could be that for the studies that did not specify any optimal cutoff, we calculated the optimal cutoff based on Youden's J. Those studies may not have considered the cutoff that maximized Youden's J as optimal. However, Youden's J appears to be the most typical method of identifying optimal cutoff thresholds for depression screening measures. A previous study reported that 11 of 13 publications on EPDS accuracy used Youden's J to define the optimal cutoff.[63] In the present study, 16 of 18 (89%) PHQ-9 studies and 9 of 12 (75%) EPDS studies with multiple reported cutoffs that identified an optimal cutoff used Youden's J or identified an optimal cutoff that was equivalent to the Youden's J optimal cutoff.

Conclusion

Selective cutoff reporting and resulting bias in accuracy estimates were more pronounced with the PHQ-9, which has a clearly defined standard cutoff, than with the EPDS, for which a range of cutoff thresholds are commonly reported, but there is not a clear single standard cutoff. For the PHQ-9, when studies appeared to choose cutoffs for reporting selectively depending

upon the sensitivity at the standard cutoff, synthesis of accuracy results from published cutoffs led to underestimation of sensitivity below the standard cutoff and overestimation of sensitivity above the standard cutoff. This phenomenon appears to be diluted for EPDS when the standard cutoff is not clearly defined and there is a range of commonly used and reported cutoffs, because the primary studies tend to report range of cutoffs around the true optimal cutoff. To reduce bias in estimates of diagnostic test accuracy of screening instruments in evidence syntheses, researchers conducting primary studies should report accuracy estimates or a contingency table of results for all relevant cutoffs. Alternatively, researchers should make primary data available so that others can estimate the sensitivity and specificity for all relevant cutoffs. Researchers who conduct meta-analyses should use modelling approaches to overcome possible biases from selective cutoff reporting or should use an IPDMA approach.

FUNDING

This work was supported by the Canadian Institutes of Health Research [CIHR; KRS-134297, KRS 140994]. Ms. Neupane was supported by G.R. Caverhill Fellowship from the Faculty of Medicine, McGill University. Dr. Levis was supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award and a Fonds de recherche du Québec - Santé (FRQS) Postdoctoral Award. Mr. Bhandari was supported by a studentship from the Research Institute of the McGill University Health Centre. Drs. Thombs and Benedetti were supported by FRQS researcher salary awards. Dr. Wu was supported by a FRQS Postdoctoral Training Fellowship. Ms. Rice was supported by a Vanier Canada Graduate Scholarship. Ms. Riehm and Ms. Saadat were supported by CIHR Frederick Banting and Charles Best Canada Graduate Scholarship master's awards. Ms. Azar and Mr. Levis were supported by FRQS Masters Training Awards. The primary study by Alvarado et al. was supported by the Ministry of Health of Chile. Collection of data for the study by Arroll et al. was supported by a project grant from the Health Research Council of New Zealand. The primary study by Khamseh et al. was supported by a grant (M-288) from Tehran University of Medical Sciences. The primary study by Beck et al. was supported by the Patrick and Catherine Weldon Donaghue Medical Research Foundation and the University of Connecticut Research Foundation. The primary study by Bombardier et al. was supported by the Department of Education, National Institute on Disability and Rehabilitation Research, Spinal Cord Injury Model Systems: University of Washington [grant no. H133N060033], Baylor College of Medicine [grant no. H133N060003], and University of Michigan [grant no. H133N060032]. Prof. Robertas Bunevicius, MD, PhD (1958-2016) was Principal Investigator of the primary study by Bunevicius et al, but passed away and was unable to participate in this project. The primary study by Chaudron et al. was

supported by a grant from the National Institute of Mental Health [grant K23 MH64476]. Dr. Cholera was supported by a United States National Institute of Mental Health (NIMH) grant [5F30MH096664], and the United States National Institutes of Health (NIH) Office of the Director, Fogarty International Center, Office of AIDS Research, National Cancer Center, National Heart, Blood, and Lung Institute, and the NIH Office of Research for Women's Health through the Fogarty Global Health Fellows Program Consortium [1R25TW00934001] and the American Recovery and Reinvestment Act. Dr. Conwell received support from NIMH [R24MH071604] and the Centers for Disease Control and Prevention [R49 CE002093]. The primary study by Couto et al. was supported by the National Counsel of Technological and Scientific Development (CNPq) [Grant no. 444254/2014-5] and the Minas Gerais State Research Foundation (FAPEMIG) [Grant no. APQ-01954-14]. Collection of data for the primary study by Delgadillo et al. was supported by grant from St. Anne's Community Services, Leeds, United Kingdom. Collection of data for the primary study by Fann et al. was supported by grant RO1 HD39415 from the US National Center for Medical Rehabilitation Research. The primary study by Tissot et al. was supported by the Swiss National Science Foundation [grant 32003B 125493]. The primary study by Garcia-Esteve et al. was supported by grant 7/98 from the Ministerio de Trabajo y Asuntos Sociales, Women's Institute, Spain. Data for the primary study by Gelaye et al. was supported by grant from the NIH [T37 MD001449]. The primary study by Inagaki et al. was supported by the Ministry of Health, Labour and Welfare, Japan. The primary study by Twist et al. was funded by the UK National Institute for Health Research under its Programme Grants for Applied Research Programme [grant reference number RP-PG-0606-1142]. The primary study by Phillips et al. was supported by a scholarship from the National Health and Medical and Research Council (NHMRC). The primary study by Liu et al. (2011) was funded by

a grant from the National Health Research Institute, Republic of China [NHRI-EX97-9706PI]. The primary study by Lotrakul et al. was supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand (grant number 49086). Dr. Bernd Löwe received research grants from Pfizer, Germany, and from the medical faculty of the University of Heidelberg, Germany [project 121/2000] for the study by Gräfe et al. The primary study by Mohd Sidik et al. was funded under the Research University Grant Scheme from Universiti Putra Malaysia, Malaysia and the Postgraduate Research Student Support Accounts of the University of Auckland, New Zealand. The primary study by Nakić Radoš et al. was supported by the Croatian Ministry of Science, Education, and Sports [134-0000000-2421]. The primary study by Pawlby et al. was supported by a Medical Research Council UK Project Grant [number G89292999N]. Collection of primary data for the study by Pence et al. was provided by NIMH [R34MH084673]. The primary study by Rochat et al. was supported by grants from the University of Oxford [HQ5035], the Tuixen Foundation (9940), the Wellcome Trust [082384/Z/07/Z and 071571], and the American Psychological Association. Dr. Rochat receives salary support from a Wellcome Trust Intermediate Fellowship [211374/Z/18/Z]. The primary study by Rooney et al. was funded by the United Kingdom National Health Service Lothian Neuro-Oncology Endowment Fund. Dr. Stafford received PhD scholarship funding from the University of Melbourne. The primary study by Su et al. was supported by grants from the Department of Health [DOH94F044 and DOH95F022] and the China Medical University and Hospital [CMU94-105, DMR-92-92 and DMR94-46]. The primary study by Tandon et al. was funded by the Thomas Wilson Sanitarium. Collection of data for the studies by Turner et al. (2012) were funded by a bequest from Jennie Thomas through the Hunter Medical Research Institute. The study by van Steenbergen-Weijenburg et al. was funded by Innovatiefonds

Zorgverzekeraars. The primary study by Vega-Dienstmaier et al. was supported by Tejada Family Foundation, Inc, and Peruvian-American Endowment, Inc. Dr Vöhringer was supported by the Fund for Innovation and Competitiveness of the Chilean Ministry of Economy, Development and Tourism, through the Millennium Scientific Initiative [grant number IS130005]. The primary study by Thombs et al. was done with data from the Heart and Soul Study. The Heart and Soul Study was funded by the Department of Veterans Epidemiology Merit Review Program, the Department of Veterans Affairs Health Services Research and Development service, the National Heart Lung and Blood Institute [R01 HL079235], the American Federation for Aging Research, the Robert Wood Johnson Foundation, and the Ischemia Research and Education Foundation. Collection of data for the primary study by Gjerdingen et al. was supported by grants from the NIMH [R34 MH072925, K02 MH65919, P30 DK50456]. No other authors reported funding for primary studies or for their work on the present study.

COMPETING INTERESTS

All authors have completed the ICJME uniform disclosure form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years with the following exceptions: Dr. Tonelli declares that he has received a grant from Merck Canada, outside the submitted work. Dr. Vigod declares that she receives royalties from UpToDate, outside the submitted work. Dr. Beck declares that she receives royalties for her Postpartum Depression Screening Scale published by Western Psychological Services. Dr. Inagaki declares that he has received a grant from Novartis Pharma, and personal fees from Meiji, Mochida, Takeda, Novartis, Yoshitomi, Pfizer, Eisai, Otsuka, MSD, Technomics, and Sumitomo Dainippon, all

outside of the submitted work. Dr. Ismail declares that she has received honorarium for speaker fees for educational lectures for Sanofi, Sunovion, Janssen and Novo Nordisk. All authors declare no other relationships or activities that could appear to have influenced the submitted work. No funder had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

(1) Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010; **340:** c365.

(2) Levis B, Benedetti A, Levis AW, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *Am J Epidemiol* 2017; **185:** 954-64.

(3) Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015; **37:** 567-76.

(4) Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem* 2008; **54:** 729-37.

(5) Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16:** 606-13.

(6) Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002; **32:** 509-15.

(7) Spitzer RL, Kroenke K, Williams JB, Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999; **282:** 1737-44.

(8) Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007; **29:** 388-95.

(9) Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007; **22:** 1596-602.

(10) Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 2019; **365:** l1476.

(11) Howard LM, Molyneaux E, Dennis CL, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. *Lancet* 2014; **384:** 1775-88.

(12) Hewitt C, Gilbody S, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009; **13:** 1-230.

(13) O'Connor E, Rossom RC, Henninger M, Groom HC, Burda BU. Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2016; **315:** 388-406.

(14) Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, DEPRESSD EPDS Collaboration. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for Screening to Detect Major Depression Among Pregnant and Postpartum Women: an Individual Participant Data Meta-Analysis. *Under Review*.

(15) Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the Patient Health Questionnaire-2 (PHQ-2), Patient Health Questionnaire-8 (PHQ-8), and Patient Health Questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *Syst Rev* 2014; **3:** 124.

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

(17) McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016; **75:** 40-6.

(18) Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008; **27:** 6111-36.

(19) Van der Leeden R, Meijer E, Busing F. Resampling Multilevel Models. In: de Leeuw J, Meijer E, editors. Handbook of Multilevel Analysis: New York: Springer New York; 2008.

(20) Van der Leeden R, Busing F, Meijer E. Bootstrap methods for two-level models. Technical Report PRM 97-04: Leiden University, Department of Psychology, Leiden; 1997.

(21) Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3:** 32-5.

(22) Inagaki M, Ohtsuki T, Yonemoto N, et al. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: a cross-sectional study. *Gen Hosp Psychiatry* 2013; **35:** 592-7.

(23) Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry* 2007; **29:** 417-24.

(24) Sung SC, Low CC, Fung DS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: a comparison of the nine-item Patient Health

Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). *Asia Pac Psychiatry* 2013; **5:** 249-58.

(25) Razykov I, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research Group. Utility of the Patient Health Questionnaire-9 to assess suicide risk in patients with systemic sclerosis. *Arthritis Care Res (Hoboken)* 2013; **65:** 753-8.

(26) Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 2010; **8:** 348-53.

(27) Pence BW, Gaynes BN, Atashili J, et al. Validity of an interviewer-administered patient health questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *J Affect Disord* 2012; **143:** 208-13.

(28) Lambert SD, Clover K, Pallant JF, et al. Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis. *J Natl Compr Canc Netw* 2015; **13:** 1203-11.

(29) Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry* 2008; **8:** 46-244X.

(30) Turner A, Hambridge J, White J, et al. Depression screening in stroke: a comparison of alternative measures with the structured diagnostic interview for the diagnostic and statistical manual of mental disorders, fourth edition (major depressive episode) as criterion standard. *Stroke* 2012; **43:** 1000-5.

(31) Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR. Validity of the Patient Health Questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. *Arch Phys Med Rehabil* 2012; **93:** 1838-45.

(32) Delgadillo J, Payne S, Gilbody S, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *J Affect Disord* 2011; **134:** 266-71.

(33) Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78:** 131-40.

(34) Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil* 2005; **20:** 501-11.

(35) Twist K, Stahl D, Amiel SA, Thomas S, Winkley K, Ismail K. Comparison of depressive symptoms in type 2 diabetes using a two-stage survey design. *Psychosom Med* 2013; **75:** 791-7.

(36) Khamseh ME, Baradaran HR, Javanbakht A, Mirghorbani M, Yadollahi Z, Malek M. Comparison of the CES-D and PHQ-9 depression scales in people with type 2 diabetes in Tehran, Iran. *BMC Psychiatry* 2011; **11:** 61-244X.

(37) Bakare MO, Okoye JO, Obindo JT. Introducing depression and developmental screenings into the national programme on immunization (NPI) in southeast Nigeria: an experimental cross-sectional assessment. *Gen Hosp Psychiatry* 2014; **36:** 105-12.

(38) Chaudron LH, Szilagyi PG, Tang W, et al. Accuracy of depression screening tools for identifying postpartum depression among urban mothers. *Pediatrics* 2010; **125:** e609-17.

(39) Radoš SN, Tadinac M, Herman R. Validation study of the Croatian version of the Edinburgh Postnatal Depression Scale (EPDS). *Suvrem Psihol* 2013; **16:** 203-8.

(40) Thiagayson P, Krishnaswamy G, Lim ML, et al. Depression and anxiety in Singaporean high-risk pregnancies - prevalence and screening. *Gen Hosp Psychiatry* 2013; **35:** 112-6.

(41) Toreki A, Ando B, Kereszturi A, et al. The Edinburgh Postnatal Depression Scale: translation and antepartum validation for a Hungarian sample. *Midwifery* 2013; **29:** 308-15.

(42) Castro E Couto T, Martins Brancaglion MY, Nogueira Cardoso M, et al. What is the best tool for screening antenatal depression? *J Affect Disord* 2015; **178:** 12-7.

(43) Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *J Affect Disord* 2003; **75:** 71-6.

(44) Tandon SD, Cluxton-Keller F, Leis J, Le HN, Perry DF. A comparison of three screening tools to identify perinatal depression among low-income African American women. *J Affect Disord* 2012; **136:** 155-62.

(45) Bunevicius A, Kusminskas L, Pop VJ, Pedersen CA, Bunevicius R. Screening for antenatal depression with the Edinburgh Depression Scale. *J Psychosom Obstet Gynaecol* 2009; **30:** 238-43.

(46) Khalifa DS, Glavin K, Bjertness E, Lien L. Postnatal depression among Sudanese women: prevalence and validation of the Edinburgh Postnatal Depression Scale at 3 months postpartum. *Int J Womens Health* 2015; **7:** 677-84.

(47) Phillips J, Charles M, Sharpe L, Matthey S. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. *J Affect Disord* 2009; **118:** 101-12.

(48) Alvarado R, Jadresic E, Guajardo V, Rojas G. First validation of a Spanish-translated version of the Edinburgh postnatal depression scale (EPDS) for use in pregnant women. A Chilean study. *Arch Womens Ment Health* 2015; **18:** 607-12.

(49) Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nurs Res* 2001; **50:** 242-50.

(50) Rochat TJ, Tomlinson M, Newell ML, Stein A. Detection of antenatal depression in rural HIV-affected populations with short and ultrashort versions of the Edinburgh Postnatal Depression Scale (EPDS). *Arch Womens Ment Health* 2013; **16:** 401-10.

(51) Su KP, Chiu TH, Huang CL, et al. Different cutoff points for different trimesters? The use of Edinburgh Postnatal Depression Scale and Beck Depression Inventory to screen for depression in pregnant Taiwanese women. *Gen Hosp Psychiatry* 2007; **29:** 436-41.

(52) Toreki A, Ando B, Dudas RB, et al. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. *Midwifery* 2014; **30:** 911-8.

(53) Vega-Dienstmaier JM, Mazzotti Suarez G, Campos Sanchez M. Validation of a Spanish version of the Edinburgh Postnatal Depression Scale. *Actas Esp Psiquiatr* 2002; **30:** 106-11.

(54) Pawlby S, Sharp D, Hay D, O'Keane V. Postnatal depression and child outcome at 11 years: the importance of accurate diagnosis. *J Affect Disord* 2008; **107:** 241-5.

(55) Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987; **150:** 782-6.

(56) Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351:** h5527.

(57) Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012; **184:** E191-6.

(58) Benedetti A, Levis B, Rücker G, et al. An empirical comparison of three methods for multiple cut-off diagnostic test meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool using published data versus individual level data. *Under Review*.

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

(60) Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR, International Meta-analysis of HIV Host Genetics. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002; **156:** 204-10.

(61) Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; **25:** 76-97.

(62) Cochrane Methods: IPD Meta-analysis. Frequently Asked Questions. https://methods.cochrane.org/ipdma/frequently-asked-questions#9. Accessed January 14, 2020.

(63) Bhandari PM, Levis B, Neupane D, et al. Bias in diagnostic accuracy estimates due to data-driven cutoff selection: simulation study using individual participant data from 49 studies on the diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS). *Under Review*.

(64) Thombs BD, Ziegelstein RC, Whooley MA. Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *J Gen Intern Med* 2008; **23:** 2014-7.

(65) Gelaye B, Tadesse MG, Williams MA, Fann JR, Vander Stoep A, Andrew Zhou XH. Assessing validity of a depression screening instrument in the absence of a gold standard. *Ann Epidemiol* 2014; **24:** 527-31.

(66) Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med* 2009; **7:** 63-70.

(67) Sherina MS, Arroll B, Goodyear-Smith F. Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. *Med J Malaysia* 2012; **67:** 309-15.

(68) Rooney AG, McNamara S, Mackinnon M, et al. Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. *Neuro Oncol* 2013; **15:** 122-9.

(69) de Man-van Ginkel, J M, Hafsteinsdottir T, Lindeman E, Burger H, Grobbee D, Schuurmans M. An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke* 2012; **43:** 854-6.

(70) Cholera R, Gaynes BN, Pence BW, et al. Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *J Affect Disord* 2014; **167:** 160-6.

(71) Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA. Diagnostic accuracy, internal consistency, and convergent validity of the Greek version of the patient health questionnaire 9 in diagnosing depression in rheumatologic disorders. *Arthritis Care Res (Hoboken)* 2011; **63:** 1313-21.

(72) Amoozegar F, Patten SB, Becker WJ, et al. The prevalence of depression and the accuracy of depression screening tools in migraine patients. *Gen Hosp Psychiatry* 2017; **48:** 25-31.

(73) Richardson TM, He H, Podgorski C, Tu X, Conwell Y. Screening depression aging services clients. *Am J Geriatr Psychiatry* 2010; **18:** 1116-23.

(74) Liu SI, Yeh ZT, Huang HC, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry* 2011; **52:** 96-101.

(75) Akena D, Joska J, Obuku EA, Stein DJ. Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care* 2013; **25:** 1245-52.

(76) Vohringer PA, Jimenez MI, Igor MA, et al. Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. *J Med Screen* 2013; **20:** 118-24.

(77) Chagas MH, Tumas V, Rodrigues GR, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age Ageing* 2013; **42:** 645-9.

(78) de Lima Osorio F, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care* 2009; **45:** 216-27.

(79) van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Serv Res* 2010; **10:** 235-6963.

(80) Tissot H, Favez N, Frascarolo-Moutinot F, Despland J. Assessing postpartum depression: Evidences for the need of multiple methods. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 2015; **65:** 61-6.

Table 1. Comparison of accuracy results from IPDMA of PHQ-9 and EPDS with the *published dataset* only versus the *full dataset*

PHQ-9

	Published dataset							*Full dataset* *30 studies; N = 11,773; MD cases = 1,587*			
Cutoff	**No. of studies**	**No. of patients**	**No of MD cases**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	5	1,663	367	0.91	0.86, 0.94	0.68	0.55, 0.79	0.97	0.94, 0.98	0.54	0.48, 0.60
6	6	2,193	377	0.87	0.77, 0.93	0.72	0.61, 0.82	0.96	0.92, 0.97	0.62	0.56, 0.68
7	6	2,050	438	0.87	0.75, 0.93	0.72	0.60, 0.81	0.94	0.90, 0.97	0.69	0.63, 0.74
8	12	5,798	720	0.87	0.78, 0.92	0.77	0.70, 0.82	0.92	0.87, 0.95	0.75	0.70, 0.79
9	14	5,283	766	0.85	0.76, 0.91	0.81	0.75, 0.85	0.87	0.81, 0.91	0.80	0.76, 0.84
10	26	10,593	1,378	0.82	0.74, 0.88	0.86	0.83, 0.89	0.83	0.76, 0.88	0.85	0.81, 0.88
11	15	5,292	767	0.83	0.72, 0.91	0.88	0.83, 0.92	0.76	0.69, 0.82	0.88	0.85, 0.91
12	16	6,188	832	0.73	0.63, 0.81	0.91	0.87, 0.94	0.69	0.62, 0.75	0.91	0.88, 0.93
13	9	2,104	455	0.70	0.59, 0.79	0.95	0.87, 0.98	0.60	0.54, 0.67	0.93	0.91, 0.95
14	5	1,231	277	0.63	0.47, 0.76	0.96	0.89, 0.99	0.54	0.47, 0.61	0.95	0.93, 0.96
15	6	3,546	374	0.47	0.37, 0.59	0.97	0.97, 0.98	0.47	0.40, 0.54	0.96	0.95, 0.97

EPDS

	Published dataset							*Full dataset* *19 studies; N = 3,637; MD cases = 531*			
Cutoff	**No. of studies**	**No. of patients**	**No. of MD cases**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	4	830	52	0.98	0.84, 1.00	0.38	0.18, 0.62	0.98	0.95, 0.99	0.36	0.29, 0.43
6	4	830	52	0.98	0.86, 1.00	0.46	0.23, 0.70	0.97	0.93, 0.98	0.45	0.37, 0.53
7	7	1,413	122	0.93	0.84, 0.97	0.56	0.41, 0.70	0.94	0.89, 0.97	0.55	0.47, 0.62
8	8	1,586	158	0.89	0.78, 0.95	0.64	0.50, 0.77	0.91	0.85, 0.94	0.63	0.55, 0.71
9	12	2,473	306	0.83	0.76, 0.88	0.74	0.65, 0.82	0.87	0.81, 0.91	0.71	0.63, 0.78
10	10	1,881	174	0.80	0.72, 0.86	0.79	0.70, 0.86	0.82	0.76, 0.87	0.79	0.72, 0.84
11	13	2,462	277	0.83	0.72, 0.90	0.83	0.76, 0.89	0.80	0.72, 0.86	0.85	0.79, 0.90
12	11	2,039	216	0.73	0.57, 0.85	0.87	0.80, 0.92	0.72	0.63, 0.80	0.89	0.84, 0.92

13	16	2,698	411	0.67	0.57, 0.75	0.93	0.89, 0.96	0.65	0.56, 0.74	0.93	0.89, 0.95
14	8	1616	148	0.63	0.52, 0.73	0.95	0.89, 0.98	0.58	0.49, 0.67	0.95	0.92, 0.97
15[#]	5	952	95	0.64	0.53, 0.73	0.96	0.90, 0.99	0.50	0.43, 0.58	0.96	0.94, 0.98
16	3	682	65	0.61	0.47, 0.73	0.98	0.78, 1.00	0.41	0.35, 0.49	0.98	0.96, 0.99
17[##]	1	306	19	0.47	0.25, 0.71	0.91	0.87, 0.94	0.33	0.27, 0.41	0.99	0.97, 0.99
18[##]	1	306	19	0.37	0.17, 0.61	0.95	0.92, 0.97	0.26	0.21, 0.33	0.99	0.98, 1.00

Abbreviations: CI: Confidence Interval; EPDS: Edinburgh Postnatal Depression Scale; IPDMA: Individual Participant Data Meta-analysis; MD: Major Depression.
#For this cutoff, the default optimizer in glmer failed, thus bobyqa was used instead.
[##]For these cutoffs, one sample proportion test with continuity correction was used to estimate sensitivity and specificity and confidence intervals.

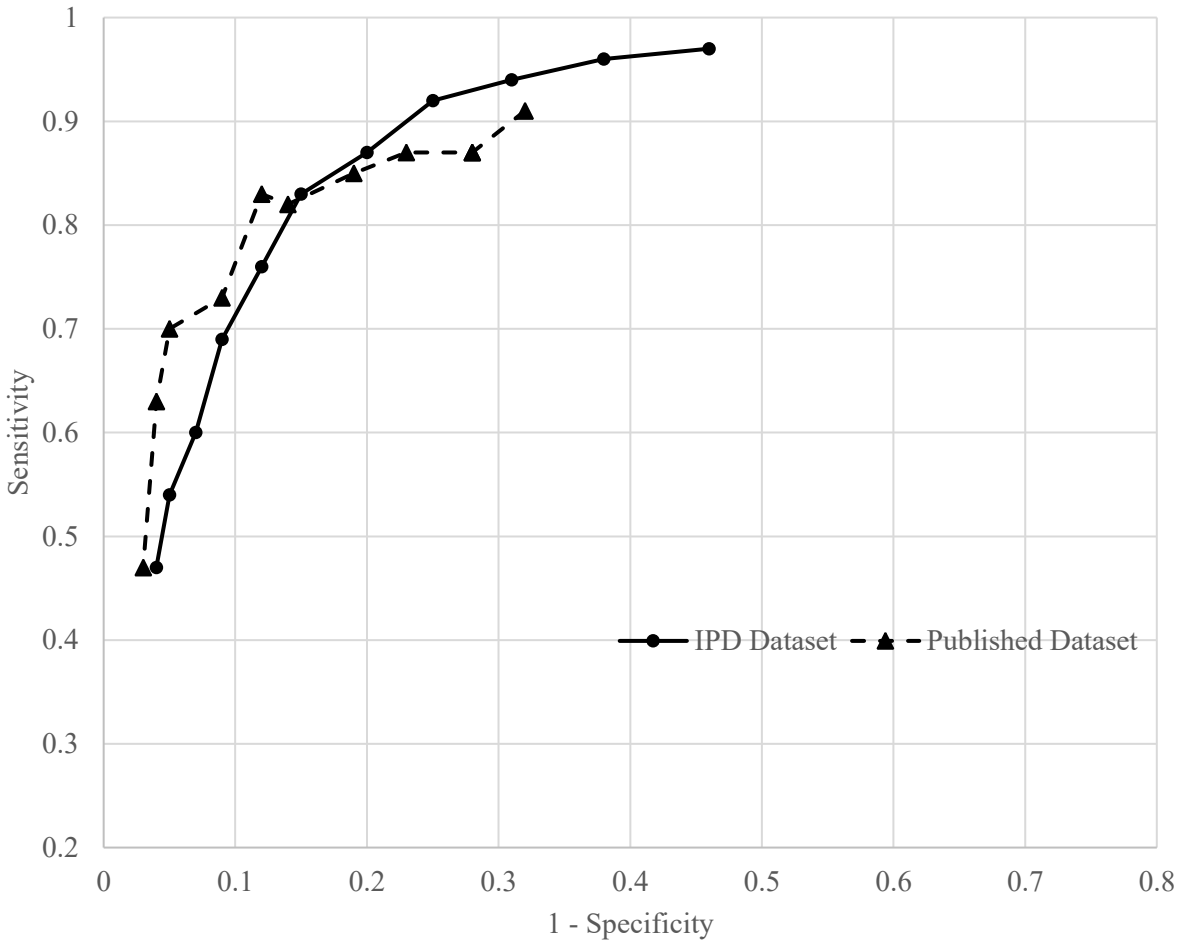


Figure 1. Receiver operating characteristic (ROC) curves for the diagnostic accuracy of Patient Health Questionnaire-9 (PHQ-9).

The points in the ROC curves indicate each of the PHQ-9 cutoffs between 5 (right) and 15 (left).

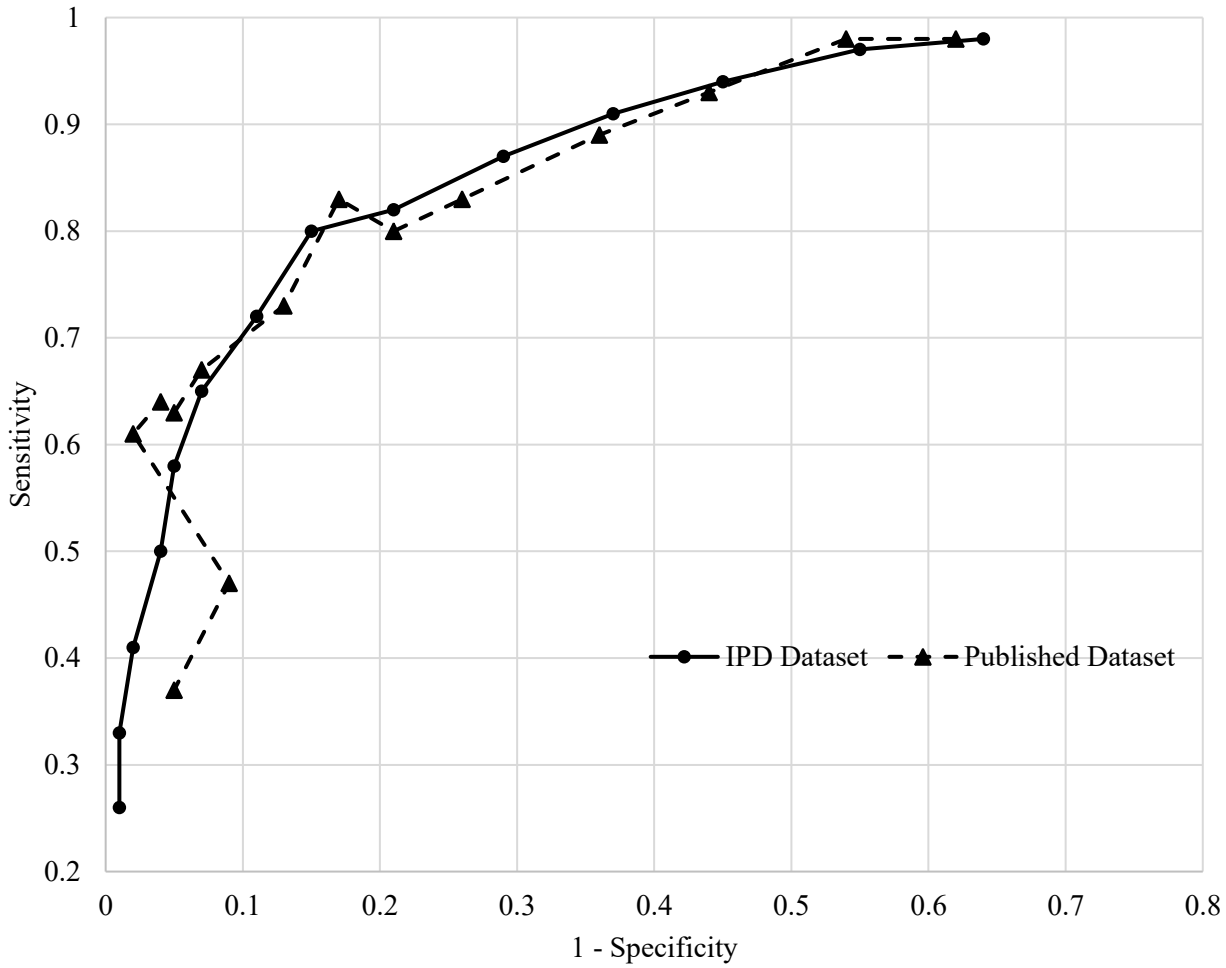


Figure 2. Receiver operating characteristic (ROC) curves for the diagnostic accuracy of Edinburgh Postnatal Depression Scale (EPDS).

The points in the ROC curves indicate each of the EPDS cutoffs between 5 (right) and 18 (left).

Table 2. Differences in estimated sensitivity and specificity using the *published dataset* versus the *full dataset* for PHQ-9 and EPDS

PHQ-9

Cutoff	% of participants included in published results for each cutoff		Differences in estimates using *published dataset* versus *full dataset* (*published - full*)			
			Sensitivity		Specificity	
	% patients	% MD cases	Estimated difference	Bootstrap 95% CI	Estimated difference	Bootstrap 95% CI
5	14	23	-0.06	-0.13, 0.00	0.14	0.02, 0.26
6	19	24	-0.09	-0.18, -0.01	0.10	0.00, 0.20
7	17	28	-0.07	-0.20, 0.00	0.03	-0.09, 0.15
8	49	45	-0.05	-0.14, 0.02	0.02	-0.03, 0.08
9	45	48	-0.02	-0.11, 0.05	0.01	-0.04, 0.05
10	90	87	-0.01	-0.05, 0.01	0.01	0.00, 0.04
11	45	48	0.07	0.00, 0.13	0.00	-0.03, 0.03
12	53	52	0.04	-0.03, 0.09	0.00	-0.02, 0.03
13	18	29	0.10	-0.02, 0.20	0.02	-0.04, 0.05
14	10	17	0.09	-0.07, 0.23	0.01	-0.04, 0.04
15	30	24	0.00	-0.12, 0.13	0.01	0.00, 0.03

EPDS

Cutoff	% of participants included in published results for each cutoff		Differences in estimates using *published dataset* versus *full dataset (published - full*)			
			Sensitivity		Specificity	
	% patients	% MD cases	Estimate difference	Bootstrap 95% CI	Estimate difference	Bootstrap 95% CI
5	23	10	0.00	-0.06, 0.04	0.02	-0.16, 0.21
6	23	10	0.01	-0.04, 0.05	0.01	-0.19, 0.21
7	39	23	-0.01	-0.10, 0.07	0.01	-0.12, 0.15
8	44	30	-0.02	-0.13, 0.07	0.01	-0.12, 0.13
9	68	58	-0.04	-0.12, 0.04	0.03	-0.05, 0.12
10	52	33	-0.02	-0.17, 0.09	0.00	-0.08, 0.08

11	68	52	0.03	-0.06, 0.11	-0.02	-0.08, 0.03
12	56	41	0.01	-0.19, 0.16	-0.02	-0.09, 0.03
13	74	77	0.02	-0.07, 0.09	0.00	-0.02, 0.02
14	44	28	0.05	-0.15, 0.20	0.00	-0.07, 0.04
15	26	18	0.14	-0.03, 0.32	0.00	-0.09, 0.03
16	19	12	0.20	-0.03, 0.39	0.00	-0.08, 0.03
17	8	4	0.14	-	-0.08	-
18	8	4	0.11	-	-0.04	-

Abbreviations: CI: Confidence Interval, EPDS: Edinburg Postnatal Depression Scale, PHQ-9: Patient Health Questionnaire-9
For PHQ-9, 15 iterations (1.5%) that did not produce difference estimates were removed prior to determining the bootstrap CI.
For EPDS, 284 iterations (28.4%) for cutoffs 5-6, 60 iterations (6%) for cutoffs 7-15 and 275 iterations (27.5%) for cutoff 16 that did not produce difference estimates were removed prior to determining bootstrap CIs.
Bootstrap CIs were not constructed for EPDS cutoffs 17 and 18 because only one study published accuracy results for these cutoffs.

Author	Published cutoffs for PHQ-9											No. of published cutoff for each study	Mean of reported cutoffs	Sensitivity at cutoff 10	Specificity at cutoff 10
	5	6	7	8	9	10	11	12	13	14	15				
Inagaki, 2013[22]	O											10	8.50	0.55	0.98
Stafford, 2007[23]		O										3	7.00	0.54	0.91
Sung, 2013[24]		O										1	6.00	0.67	0.91
Thombs, 2008[64]		O										6	5.50	0.54	0.90
Pence, 2012[27]				O								3	10.00	0.27	0.94
Arrol, 2010[26]				O								4	11.25	0.74	0.91
Turner, 2012[30]					O							3	8.67	0.69	0.78
Lambert, 2015[28]					O							4	11.80	0.71	0.82
Lotrakul, 2008[29]					O							10	10.50	0.74	0.85
Gelaye, 2014[65]						O						3	10.00	0.53	0.78
Gjerdingen, 2009[66]						O						1	10.00	0.74	0.91
Mohd Sidik, 2012[67]						O						1	10.00	0.77	0.87
Rooney, 2013[68]						O						4	9.50	0.79	0.86
de Man-van Ginkel, 2012[69]						O						1	10.00	0.80	0.78
Cholera, 2014[70]						O						3	10.0	0.81	0.83
Hyphantis, 2011[71]						O						11	9.50	0.81	0.87
Amoozegar, 2017[72]						O						6	12.50	0.82	0.79
Richardson, 2010[73]						O						6	9.50	0.82	0.86
Liu, 2011[74]						O						3	10.00	0.86	0.94
Akena, 2013[75]						O						6	10.50	0.91	0.89
Vöhringer, 2013[76]						O						1	10.00	0.93	0.77
Chagas, 2013[77]						O						4	9.50	1.00	0.83
Osório, 2009[78]						O						6	15.50	1.00	0.98
van Steenbergen-Weijenburg, 2010[79]						O		O				5	10.00	0.92	0.65
Bombardier, 2012[31]							O					4	10.50	1.00	0.80
Fann, 2005[34]								O				2	11.00	0.88	0.90
Delgadillo, 2011[32]								O				1	12.00	0.94	0.42
Löwe, 2004[33]								O				3	12.00	0.97	0.76
Twist, 2013[35]								O				5	12.00	0.98	0.64
Khamseh, 2011[36]									O			1	13.00	0.85	0.66
No. of studies that published each cutoff	5	6	6	12	14	26	15	16	9	5	6				

Figure 3. **Pattern of cutoff reporting for PHQ-9 studies.**

O represents the optimal cutoff for PHQ-9 explicitly stated in the studies except for Inagaki 2013, Pence 2012, Arroll 909, Cholera 2014, Amoozegar 2017, which did not identify an optimal cutoff. For those, Youden's J optimal was calculated from published accuracies. For Gjerdingen 2009 and Vöhringer 2013, only one cutoff was reported without stating whether it was optimal or not.
van Steenbergen- Weijenburg reported 10 and 12 as optimal cutoffs
Studies that reported accuracies for cutoffs beyond presented in the table: Inagaki 2013 reported the accuracy for cutoffs 4-13, Thombs 2008 reported the accuracy for cutoffs 1-10, Lambert 2015 reported the accuracy for cutoffs 5,9,10,15,20, Hyphantis 2011 reported the accuracy for cutoffs 4-16, Osorio 2009 reported the accuracy for cutoffs 10-21.
All the reported cutoffs were included while calculating the mean of reported cutoffs though they are not shown in the figure.

Author	Published cutoffs for EPDS														No. of published cutoff for each study	Mean of reported cutoffs	Sensitivity at cutoff 10	Specificity at cutoff 10
	5	6	7	8	9	10	11	12	13	14	15	16	17	18				
Töreki, 2013[41]					O										10	9.50	0.43	0.93
Nakić Radoš, 2013[39]					O										8	10.50	0.60	0.82
Bakare, 2014[37]					O										1	9.00	0.66	0.89
Chaudron, 2010[38]					O										2	11.00	0.73	0.84
Thiagayson, 2013[40]					O										6	9.50	0.73	0.74
Tissot, 2015[80]						O									5	11.00	0.50	0.75
Couto, 2015[42]							O								7	11.00	0.86	0.68
Tandon, 2012[44]							O								2	12.00	0.92	0.81
Garcia-Esteve, 2003[43]							O								1	11.00	1.00	0.59
Philips, 2009[47]								O							3	12.00	0.88	0.66
Khalifa, 2015[46]								O							11	8.00	0.89	0.68
Bunevicius, 2009[45]								O							7	12.00	0.92	0.87
Töreki, 2014[52]									O						12	10.50	1.00	0.91
Pawlby, 2008[54]									O						1	13.00	0.61	0.94
Alvarado, 2015[48]									O						10	11.50	0.82	0.82
Beck, 2001[49]									O						1	13.00	0.83	0.86
Su, 2007[51]									O						1	13.00	0.91	0.70
Rochat, 2013[50]									O						1	13.00	0.94	0.50
Vega-Dienstmaier, 2002[53]										O					14	13.50	0.89	0.45
No. of studies that published each cutoff	4	4	7	8	12	10	13	11	16	8	5	3	1	1				

Figure 4. Pattern of cutoff reporting for EPDS studies.

O represents the optimal cutoff for EPDS explicitly stated in the studies except for Philips 2009, which did not identify an optimal cutoff. For those, Youden's J optimal was calculated from published accuracies. Youden's J optimal calculated from published accuracies. For Bakare 2014, Pawlby 2007, Beck 2001 only one cutoff was reported without stating whether it was optimal or not.
Studies that reported accuracies for cutoffs beyond presented in the table: Khalifa 2015 reported accuracy for cutoffs 1-15, Vega-Dienstmaier 2002 reported the accuracy for cutoffs 1-26.
All the reported cutoffs were included while calculating the mean of reported cutoffs though they are not shown in the figure.

CHAPTER 4. DISCUSSION

4.1 Key findings

Bias in accuracy estimates and cutoff reporting patterns were compared between two depression screening tools; the PHQ-9, which has a clearly defined standard cutoff, and the EPDS, which does not have a clearly defined standard cutoff. For both the PHQ-9 and the EPDS, the results from meta-analysis of published cutoffs only was compared to meta-analysis of all cutoffs from all studies. Sensitivity for the PHQ-9 was underestimated for cutoffs below the standard cutoff of 10, similar for standard cutoff of 10 and overestimated for cutoffs above the standard cutoff of 10. Sensitivity for the EPDS was similar for cutoffs below 10, similar for commonly reported cutoffs 10 to 13 but overestimated for cutoffs above 13. This may be explained by the reporting pattern that the PHQ-9 studies that had optimal cutoff below 10 reported more cutoffs below 10, and studies that had optimal cutoff above 10 reported more cutoffs above 10. But, in the case of the EPDS only the studies that had optimal cutoff above 10 reported more cutoffs above 10, and the opposite was not observed. This may be because the PHQ-9 has a single standard cutoff of 10, whereas for the EPDS it may be an expectation that results for commonly used cutoff 10 to 13 are reported. Overall, the bias in accuracy estimate was observed in both the screening tools, but it was more pronounced for the PHQ-9 than for the EPDS.

This thesis includes the first study to provide evidence of a differential bias in accuracy estimates for different screening tools, depending on whether a clearly defined standard cutoff is available or not.

4.2 Clinical and research implications

The results from screening test accuracy studies may not represent the true accuracy in clinical practice when cutoffs are selectively reported. Users of results from these primary studies should interpret results from these studies with caution. However, to avoid bias at the first step,

primary studies should report accuracy estimates for all possible cutoffs, regardless of the presence or absence of well-defined standard cutoff for the index test. Standard for Reporting of Diagnostic Accuracy Studies (STARD) checklist requires reporting accuracy and precision estimates and crosstabulation of index tests and reference standards.[23] The STARD checklist should also make it mandatory to report accuracy estimates for all relevant cutoffs.

Results from aggregate data meta-analysis including primary studies that have selectively reported the cutoffs may also be biased. In this scenario, IPDMA approach can be used to estimate accuracy across all cutoffs for all studies using participant level data. But, IPDMA approach is highly labor intensive.[24,26-28] Alternatively bias due to selective cutoff reporting can be corrected statistically using modelling techniques that fill in missing cutoffs using published cutoffs. In a study by Benedetti et al, the accuracy estimates generated from published cutoffs after applying a modelling technique were similar to accuracy estimates generated from IPDMA.[29]

4.3 Limitations

One limitation of the study reported in this thesis is that 14 of 69 (20%) PHQ-9 studies and 24 of 72 (33%) EPDS eligible studies did not provide data for the IPDMA dataset. Of these, 4 of the 14 PHQ-9 (29%) and 10 of the 24 (42%) EPDS studies did not publish accuracy estimates for any cutoff so they would not have been eligible for the present study.

Another limitation is that, for the studies that did not report any cutoff as "optimal", an optimal cutoff using Youden's J method was calculated. Those studies may not have considered Youden's J optimal as the "optimal" cutoff. However, a previous study found that Youden's J method is the most common method of identifying optimal cutoff in EPDS accuracy studies.[30] In the present study, majority of studies (16 of 18 (89%) PHQ-9 and 9 of 12 (75%) EPDS studies) that reported multiple cutoffs and identified an optimal cutoff used Youden's J or identified an

optimal cutoff that was equivalent to the Youden's J optimal cutoff obtained from published cutoffs.

4.4 Conclusion

Selective cutoff reporting was more pronounced for the PHQ-9, which has a well-defined standard cutoff, than for the EPDS, for which there is not a well-defined standard cutoff. Sensitivity estimates were found to be under-estimated below the standard cutoff and over-estimated above the standard cutoff due to selective cutoff reporting. To reduce this bias, primary studies should report accuracy estimates for all relevant cutoffs or contingency table of results of index test and reference standard for all relevant cutoffs. Researchers performing meta-analysis of published cutoffs should use modelling techniques to correct for the bias due to selective cutoff reporting or should use an IPDMA approach.

REFERENCES

(1) Thombs BD, Coyne JC, Cuijpers P, et al. Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012; **184:** 413-8.

(2) Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014; **348:** g1253.

(3) Thombs BD, Arthurs E, El-Baalbaki G, Meijer A, Ziegelstein RC, Steele RJ. Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *BMJ* 2011; **343:** d4825.

(4) Levis B, Benedetti A, Levis AW, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *Am J Epidemiol* 2017; **185:** 954-64.

(5) Higgins J, Green S. Handbook for systematic reviews of interventions version 5.1. 0 [updated March 2011]. *The Cochrane Collaboration* 2011.

(6) Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010; **340:** c365.

(7) Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012; **184:** E191-6.

(8) Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015; **37:** 567-76.

(9) Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16:** 606-13.

(10) Spitzer RL, Kroenke K, Williams JB, Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999; **282:** 1737-44.

(11) Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987; **150:** 782-6.

(12) Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002; **32:** 509-15.

(13) Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007; **29:** 388-95.

(14) Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007; **22:** 1596-602.

(15) Howard LM, Molyneaux E, Dennis CL, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. *Lancet* 2014; **384:** 1775-88.

(16) Hewitt C, Gilbody S, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009; **13:** 1-230.

(17) O'Connor E, Rossom RC, Henninger M, Groom HC, Burda BU. Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2016; **315:** 388-406.

(18) Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA* 2016; **315:** 380-7.

(19) World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: World Health Organization; 2017.

(20) Ferenchick EK, Ramanuj P, Pincus HA. Depression in primary care: part 1-screening and diagnosis. *BMJ* 2019; **365:** l794.

(21) Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; **149:** 889-97.

(22) Bossuyt PM. Interpreting diagnostic test accuracy studies. *Semin Hematol* 2008; **45:** 189-95.

(23) Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351:** h5527.

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

(25) Stewart LA, Tierney JF, Clarke M. Reviews of individual patient data. In: Stewart LA, Tierney JF, Clarke M, editors. Cochrane Handbook for Systematic Reviews of Interventions. England, UK: Wiley Online Library; 2008. p. 547-558.

(26) Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR, International Meta-analysis of HIV Host Genetics. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002; **156:** 204-10.

(27) Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; **25:** 76-97.

(28) Cochrane Methods: IPD Meta-analysis. Frequently Asked Questions. https://methods.cochrane.org/ipdma/frequently-asked-questions#9. Accessed January 14, 2020.

(29) Benedetti A, Levis B, Rücker G, et al. An empirical comparison of three methods for multiple cut-off diagnostic test meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool using published data versus individual level data. *Under Review*.

(30) Bhandari PM, Levis B, Neupane D, et al. Bias in diagnostic accuracy estimates due to data-driven cutoff selection: simulation study using individual participant data from 49 studies on the diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS). *Under Review*.

APPENDIX

Supplementary file for the manuscript in chapter 3

Supplementary Table 4b. Characteristics of primary studies that were excluded in the present study because the difference in sample size or MD cases between the IPDMA dataset and published dataset was >10% for EPDS (N=9)

Supplementary Table 5a. Characteristics of primary studies for PHQ-9 included in the present study (N=30)

Supplementary Table 5b. Characteristics of primary studies for EPDS included in the present study (N=19)

Supplementary Methods 1a. Search strategies for PHQ-9

MEDLINE (OvidSP)
1. PHQ*.af.
2. patient health questionnaire*.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. Mental Disorders/di, pc [Diagnosis, Prevention & Control]
15. Mood Disorders/di, pc [Diagnosis, Prevention & Control]
16. Depressive Disorder/di, pc [Diagnosis, Prevention & Control]
17. Depressive Disorder, Major/di, pc [Diagnosis, Prevention & Control]
18. Depression, Postpartum/di, pc [Diagnosis, Prevention & Control]
19. Depression/di, pc [Diagnosis, Prevention & Control]
20. validation studies.pt.
21. comparative study.pt.
22. screen*.af.
23. prevalence.af.
24. predictive value*.af.
25. detect*.ti.
26. sensitiv*.ti.
27. valid*.ti.
28. revalid*.ti.
29. predict*.ti.
30. accura*.ti.
31. psychometric*.ti.
32. identif*.ti.
33. specificit*.ab.
34. cut?off*.ab.
35. cut* score*.ab.
36. cut?point*.ab.
37. threshold score*.ab.
38. reference standard*.ab.
39. reference test*.ab.
40. index test*.ab.
41. gold standard.ab.
42. or/4-41
43. 3 and 42
44. limit 43 to yr=”2000-Current”

PsycINFO (OvidSP)
1. PHQ*.af.
2. patient health questionnaire*.af.
3. 1 or 2
4. Diagnosis/
5. Medical Diagnosis/
6. Psychodiagnosis/
7. Misdiagnosis/
8. Screening/
9. Health Screening/
10. Screening Tests/
11. Prediction/
12. Cutting Scores/
13. Psychometrics/
14. Test Validity/
15. screen*.af.
16. predictive value*.af.
17. detect*.ti.
18. sensitiv*.ti.
19. valid*.ti.
20. revalid*.ti.
21. accura*.ti.
22. psychometric*.ti.
23. specificit*.ab.
24. cut?off*.ab.
25. cut* score*.ab.
26. cut?point*.ab.
27. threshold score*.ab.
28. reference standard*.ab.
29. reference test*.ab.
30. index test*.ab.
31. gold standard.ab.
32. or/4-31
33. 3 and 32
38. Limit 33 to "2000 to current"

Web of Science (Web of Knowledge)
#1: TS=(PHQ* OR "Patient Health Questionnaire*")
#2: 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")
#1 AND #2

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2014

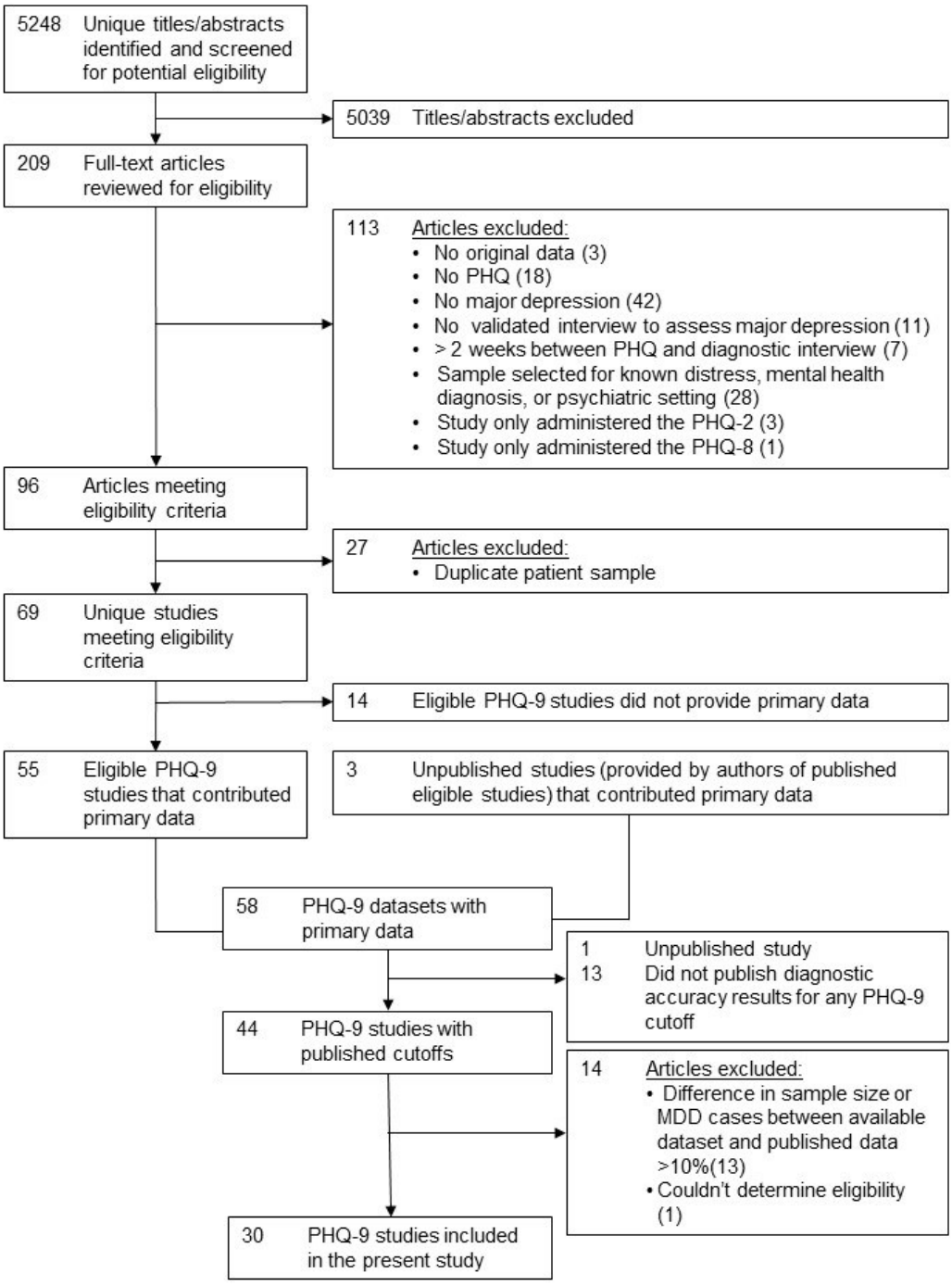
Supplementary Methods 1b. Search strategies for EPDS

MEDLINE (OvidSP)
1. EPDS.af.
2. Edinburgh Postnatal Depression.af.
3. Edinburgh Depression Scale.af.
4. or/1-3
5. Mass Screening/
6. Psychiatric Status Rating Scales/
7. "Predictive Value of Tests"/
8. "Reproducibility of Results"/
9. exp "Sensitivity and Specificity"/
10. Psychometrics/
11. Prevalence/
12. Reference Values/
13. Reference Standards/
14. exp Diagnostic Errors/
15. Mental Disorders/di, pc [Diagnosis, Prevention & Control]
16. Mood Disorders/di, pc [Diagnosis, Prevention & Control]
17. Depressive Disorder/di, pc [Diagnosis, Prevention & Control]
18. Depressive Disorder, Major/di, pc [Diagnosis, Prevention & Control]
19. Depression, Postpartum/di, pc [Diagnosis, Prevention & Control]
20. Depression/di, pc [Diagnosis, Prevention & Control]
21. validation studies.pt.
22. comparative study.pt.
23. screen*.af.
24. prevalence.af.
25. predictive value*.af.
26. detect*.ti.
27. sensitiv*.ti.
28. valid*.ti.
29. revalid*.ti.
30. predict*.ti.
31. accura*.ti.
32. psychometric*.ti.
33. identif*.ti.
34. specificit*.ab.
35. cut?off*.ab.
36. cut* score*.ab.
37. cut?point*.ab.
38. threshold score*.ab.
39. reference standard*.ab.
40. reference test*.ab.
41. index test*.ab.
42. gold standard.ab.
43. or/5-42
44. 4 and 43

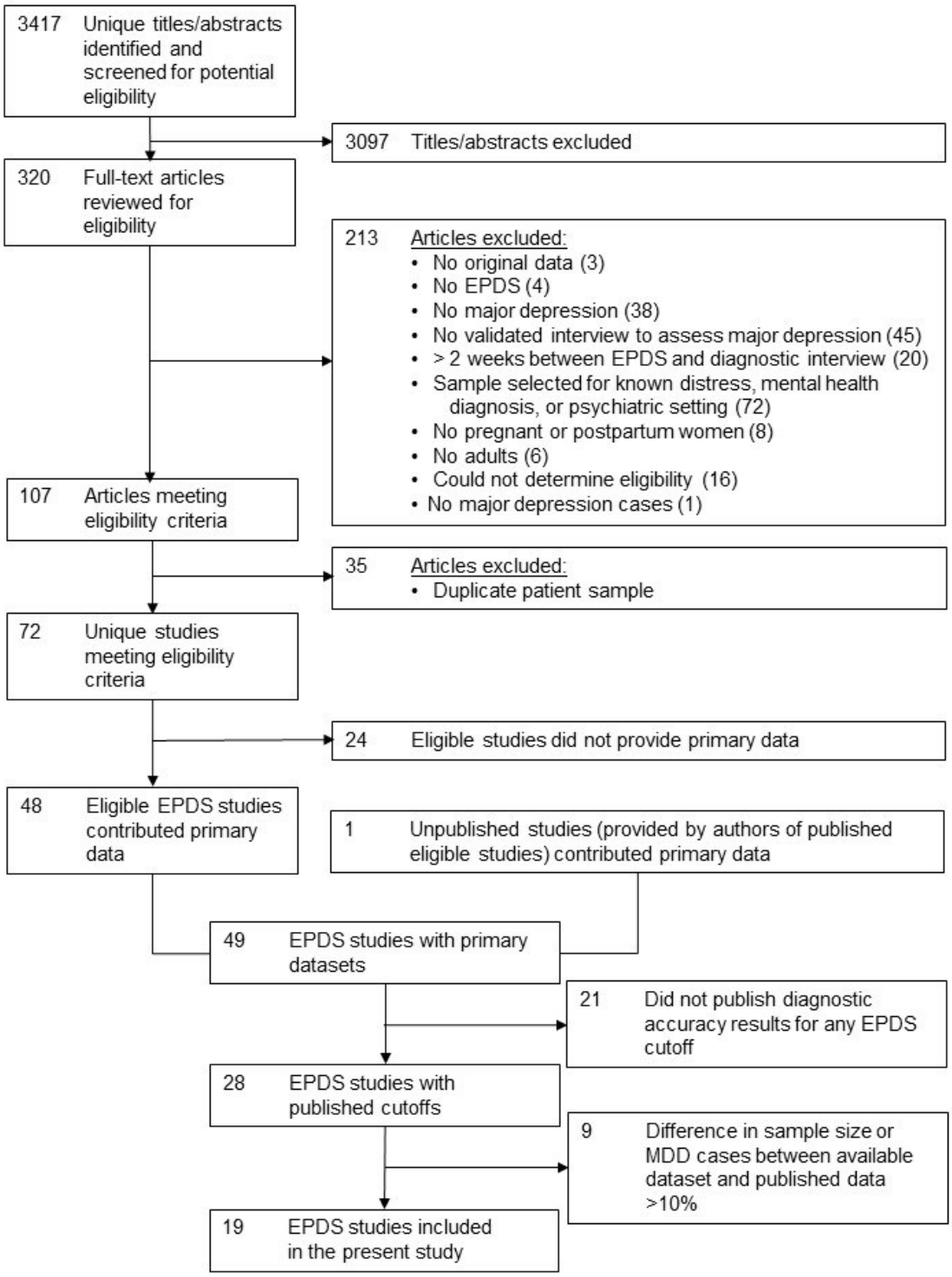
PsycINFO (OvidSP)
1. EPDS.af.
2. Edinburgh Postnatal Depression.af.
3. Edinburgh Depression Scale.af.
4. or/1-3
5. Diagnosis/
6. Medical Diagnosis/
7. Psychodiagnosis/
8. Misdiagnosis/
9. Screening/
10. Health Screening/
11. Screening Tests/
12. Prediction/
13. Cutting Scores/
14. Psychometrics/
15. Test Validity/
16. screen*.af.
17. predictive value*.af.
18. detect*.ti.
19. sensitiv*.ti.
20. valid*.ti.
21. revalid*.ti.
22. accura*.ti.
23. psychometric*.ti.
24. specificit*.ab.
25. cut?off*.ab.
26. cut* score*.ab.
27. cut?point*.ab.
28. threshold score*.ab.
29. reference standard*.ab.
30. reference test*.ab.
31. index test*.ab.
32. gold standard.ab.
33. or/5-32
34. 4 and 33

Web of Science (Web of Knowledge)
#1. TS=(EPDS OR "Edinburgh Postnatal Depression" OR "Edinburgh Depression Scale")
#2. 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*")
#2 AND #1
Databases=SCI-EXPANDED, SSCI, A&HCI

Supplementary Figure 1a. Flow diagram of study selection process for PHQ-9



Supplementary Figure 1b. Flow diagram of study selection process for EPDS



Supplementary Table 1a. Reasons for exclusion of all articles excluded at the full-text level for the main IPDMA of the PHQ-9 (N=113)

Reference	Reason for exclusion
Albert NM, Moser DK, Nutter B, Pozuelo L. Are PHQ-9 and PHQ-2 Depression score cutoffs the best cutoffs for determining significant depression in Pts with HF and Mild-Moderate Symptoms? *Journal of Cardiac Failure*. 2009;**15**:S114-S114.	Major depression not assessed
Allgaier AK, Pietsch K, Fruhe B, et al. Depression in pediatric care: Is the WHO-Five Well-Being Index a valid screening instrument for children and adolescents? *General Hospital Psychiatry.* 2012;**34**:234-241.	PHQ not administered
Armstrong G, Nuken A, Samson L, et al. Quality of life, depression, anxiety and suicidal ideation among men who inject drugs in Delhi, India. *BMC Psychiatry*. 2013;**13**:151-151.	Major depression not assessed
Arroll B, Goodyear-Smith F, Kerse N, et al. The prevalence of depression among Maori patients in Auckland general practice. *Journal of Primary Health Care*. 2009;**1**:26-29.	Major depression not assessed
Berghofer A, Hartwich A, Bauer M, et al. Efficacy of a systematic depression management program in high utilizers of primary care: a randomized trial. *BMC Health Services Research*. 2012;**12**:298.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Buehler B, Kocalevent R, Berger R, et al. Treatment situation of long-term unemployed with psychological disorders. *Nervenarzt*. 2013;**84**:603-607.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Cannon DS, Tiffany ST, Coon H, et al. The PHQ-9 as a brief assessment of lifetime major depression. *Psychological Assessment.* 2007;**19**:247-251.	Major depression not assessed
Carballeira Y, Dumont P, Borgacci S, et al. Criterion validity of the French version of Patient Health Questionnaire (PHQ) in a hospital department of internal medicine. *Psychology & Psychotherapy: Theory, Research & Practice*. 2007;**80**:69-77.	No validated interview to assess major depression
Cassin S, Sockalingam S, Hawa R, et al. Psychometric properties of the Patient Health Questionnaire (PHQ-9) as a depression screening tool for bariatric surgery candidates. *Psychosomatics*. 2013;**54**:352-358.	> 2 weeks between PHQ and diagnostic interview
Chen S, Chiu H, Xu B, et al. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. I*nternational Journal of Geriatric Psychiatry.* 2010;**25**:1127-1133.	> 2 weeks between PHQ and diagnostic interview
Choi Y, Mayer TG, Williams MJ, Gatchel RJ. What is the best screening test for depression in chronic spinal pain patients? *Spine Journal: Official Journal of the North American Spine Society.* 2014;**14**:1175-1182.	> 2 weeks between PHQ and diagnostic interview
Corapcioglu A, Ozer GU. Adaptation of revised Brief PHQ (Brief-PHQ-r) for diagnosis of depression, panic disorder and somatoform disorder in primary healthcare settings. *International Journal of Psychiatry in Clinical Practice*. 2004;**8**:11-18.	No validated interview to assess major depression
Creed F. The relationship between somatic symptoms, health anxiety, and outcome in medical out-patients. *Psychiatric Clinics of North America*. 2011;**34**:545-564.	PHQ not administered
Davis K, Pearlstein T, Stuart S, O'Hara M, Zlotnick C. Analysis of brief screening tools for the detection of postpartum depression: comparisons of the PRAMS 6-item instrument, PHQ-9, and structured interviews. *Archives of Women's Mental Health*. 2013;**16**:271-277.	Sample selected for known distress, mental health diagnosis, or psychiatric setting

de Man-van Ginkel J, Floor G, Marieke S, Eline L, Thora H. Early detection of post stroke depression: a clinimetric evaluation of the PHQ-9. *Journal of Clinical Nursing*. 2010;**19**:88-88.	Major depression not assessed
Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the Patient Health Questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosomatic Medicine.* 2001;**63**:679-686.	No validated interview to assess major depression
Esler D, Johnston F, Thomas D, Davis B. The validity of a depression screening tool modified for use with Aboriginal and Torres Strait Islander people. *Australian & New Zealand Journal of Public Health.* 2008;**32**:317-321.	No validated interview to assess major depression
Fine TH, Contractor AA, Tamburrino M, et al. Validation of the telephone-administered PHQ-9 against the in-person administered SCID-I major depression module. *Journal of Affective Disorders*. 2013;**150**:1001-1007.	PHQ not administered
Galek A, Erbsloeh-Moeller B, Koellner V, et al. Mental disorders in patients with fibromyalgia syndrome. Screening in centres of different medical specialties. *Schmerz*. 2013;**27**:296-304.	Major depression not assessed
Gawlik S, Waldeier L, Mueller M, et al. Subclinical depressive symptoms during pregnancy and birth outcome-a pilot study in a healthy German sample. *Archives of Womens Mental Health*. 2013;**16**:93-100.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gellis ZD. Depression screening in medically ill homecare elderly. *Best Practices in Mental Health: An International Journal*. 2010;**6**:1-16.	PHQ not administered
Gibbons RD, Hooker G, Finkelman MD, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *Journal of Clinical Psychiatry.* 2013;**74**:669-674.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Archives of General Psychiatry.* 2012;**69**:1104-1112.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gigantesco A, Mirante N, Granchelli C, et al. Psychopathological chronic sequelae of the 2009 earthquake in L'Aquila, Italy. *Journal of Affective disorders*. 2013;**148**:265-271.	Major depression not assessed
Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *British Journal of General Practice.* 2007;**57**:650-652.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gold KJ, Spangenberg K, Wobil P, Schwenk TL. Depression and risk factors for depression among mothers of sick infants in Kumasi, Ghana. *International Journal of Gynaecology & Obstetrics*. 2013;**120**:228-231.	Major depression not assessed
Gothwal VK, Bagga DK, Bharani S, Sumalini R, Reddy SP. The Patient Health Questionnaire-9: Validation among patients with glaucoma. *PLoS ONE*. 2014;**9** e101295.	Major depression not assessed
Grote NK, Katon WJ, Lohr MJ, et al. Culturally relevant treatment services for perinatal depression in socio-economically disadvantaged women: The design of the MOMCare study. *Contemporary Clinical Trials*. 2014;**39**:34-49.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Hanwella R, Ekanayake S, de Silva VA. The validity and reliability of the Sinhala translation of the Patient Health Questionnaire (PHQ-9) and PHQ-2 screener. *Depression Research and Treatment.* 2014;**2014**:768978.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Hauffa R, Rief W, Brahler E, et al. Lifetime traumatic experiences and posttraumatic stress disorder in the German population: results of a representative population survey. *Journal of Nervous & Mental Disease*. 2011;**199**:934-939.	Major depression not assessed
Hauser W, Glaesmer H, Schmutzer G, Brahler E. Widespread pain in older Germans is associated with posttraumatic stress disorder and lifetime	Major depression not assessed

employment status--results of a cross-sectional survey with a representative population sample. *Pain*. 2012;**153**:2466-2472.	
Hausteiner-Wiehle C, Sokollu F. Magical thinking in somatoform disorders: an exploratory study among patients with suspected allergies. *Psychopathology*. 2011;**44**:283-288.	Major depression not assessed
Holzapfel N, Muller-Tasch T, Wild B, et al. Depression profile in patients with and without chronic heart failure. *Journal of Affective Disorders*. 2008;**105**:53-62.	Major depression not assessed
Howell EA, Bodnar-Deren S, Balbierz A, et al. An intervention to reduce postpartum depressive symptoms: A randomized controlled trial. *Archives of Women's Mental Health*. 2014;**17**:57-63.	Major depression not assessed
Husain N, Creed F, Tomenson B. Depression and social stress in Pakistan. *Psychological Medicine*. 2000;**30**:395-402.	PHQ not administered
Husain N, Gater R, Tomenson B, Creed F. Comparison of the Personal Health Questionnaire and the Self Reporting Questionnaire in rural Pakistan. *JPMA - Journal of the Pakistan Medical Association.* 2006;**56**:366-370.	PHQ not administered
Husain N, Waheed W, Tomenson B, Creed F. The validation of personal health questionnaire amongst people of Pakistani family origin living in the United Kingdom. *Journal of Affective Disorders*. 2007;**97**:261-264.	PHQ not administered
Inoue T, Tanaka T, Nakagawa S. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry*. 2012;**12**:73.	No validated interview to assess major depression
Jacobs SR, Jacobsen PB, Donovan K, Booth-Jones M. Utility of the Patient Health Questionnaire-9 (Phq-9) in identifying depression among hematopoietic stem cell transplant (HSCT) patients. *Annals of Behavioral Medicine*. 2007;**33**:S56-S56.	Major depression not assessed
Jeon HJ, Park JH, Shim EJ. Permissive attitude toward suicide and future intent in individuals with and without depression: results from a nationwide survey in Korea. *Journal of Nervous & Mental Disease*. 2013;**201**:286-291.	Major depression not assessed
Kamphuis MH, Stegenga BT, Zuithoff NP, et al. Does recognition of depression in primary care affect outcome? The PREDICT-NL study. *Family Practice.* 2012;**29**:16-23.	Major depression not assessed
Karekla M, Pilipenko N, Feldman J. Greek language validation of the Patient Health Questionnaire (PHQ). *Annals of Behavioral Medicine*. 2011;**41**:S20-S20.	Major depression not assessed
Kissane DW, Wein S, Love A, et al. The Demoralization Scale: a report of its development and preliminary validation. *Journal of Palliative Care*. 2004;**20**:269-276.	Major depression not assessed
Krause S, Rydall A, Hales S, Rodin G, Lo C. Initial validation of the Death and Dying Distress Scale for the assessment of death anxiety in patients with advanced cancer. *Journal of Pain and Symptom Management*. 2015;**49**:127-135.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care*. 2003;**41**:1284-1292.	No validated interview to assess major depression
Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001;**16**:606-613.	No validated interview to assess major depression
Lewis BA, Gjerdingen DK, Avery MD, et al. Examination of a telephone-based exercise intervention for the prevention of postpartum depression: design, methodology, and baseline data from The Healthy Mom study. *Contemporary Clinical Trials.* 2012;**33**:1150-1158.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Lewis BA, Gjerdingen DK, Avery MD, et al. A randomized trial examining a physical activity intervention for the prevention of postpartum depression: The healthy mom trial. *Mental Health and Physical Activity.* 2014;**7**:42-49.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *Journal of the American Geriatrics Society*. 2007;**55**:596-602.	Major depression not assessed

Lino VT, Portela MC, Camacho LA, et al. Screening for depression in low-income elderly patients at the primary care level: use of the Patient Health Questionnaire-2. *PLoS One*. 2014;**9**:e113778.	Study only administered the PHQ-2
Liu LT, Chen SL, Jin T, et al. Natural outcome and risk-prediction model of late-life depression. *Zhejiang da Xue Xue Bao Yi Xue Ban/Journal of Zhejiang University Medical Sciences.* 2012;**41**:653-658.	> 2 weeks between PHQ and diagnostic interview
Londono A, Romero P, Casas G. The association between armed conflict, violence and mental health: a cross sectional study comparing two populations in Cundinamarca department, Colombia. *Conflict & Health*. 2012;**6**:12.	Major depression not assessed
Lossnitzer N, Muller-Tasch T, Lowe B, et al. Exploring potential associations of suicidal ideation and ideas of self-harm in patients with congestive heart failure. *Depression & Anxiety*. 2009;**26**:764-768.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Lowe B, Grafe K, Kroenke K, et al. Predictors of psychiatric comorbidity in medical outpatients. *Psychosomatic Medicine*. 2003;**65**:764-770.	PHQ not administered
Lowe B, Grafe K, Quenter A, et al. The Patient Health Questionnaire D as a self-rating instrument for screening mental disorders in internal medicine and in general medicine - Preliminary validation results with 1000 outpatients. *Psychotherapie Psychosomatik Medizinische Psychologie*. 2001;**51**:109-109.	No original data
Lowe B, Grafe K, Zipfel S, et al. Detecting panic disorder in medical and psychosomatic outpatients: comparative validation of the Hospital Anxiety and Depression Scale, the Patient Health Questionnaire, a screening question, and physicians' diagnosis. *Journal of Psychosomatic Research*. 2003;**55**:515-519.	PHQ not administered
Lowe B, Kroenke K, Spitzer RL, et al. Trauma exposure and posttraumatic stress disorder in primary care patients: cross-sectional criterion standard study. *Journal of Clinical Psychiatry*. 2011;**72**:304-312.	Major depression not assessed
Mahajan S, Avasthi A, Grover S, Chawla YK. Role of baseline depressive symptoms in the development of depressive episode in patients receiving antiviral therapy for hepatitis C infection. *Journal of Psychosomatic Research*. 2014.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Maneeton B, Maneeton N, Mahathep P. Prevalence of depression and its correlations: a cross-sectional study in Thai cancer patients. *Asian Pacific Journal of Cancer Prevention: APJCP.* 2012;**13**:2039-2043.	Major depression not assessed
Mao HJ, Li HJ, Chiu H, Chan WC, Chen SL. Effectiveness of antenatal emotional self-management training program in prevention of postnatal depression in Chinese women. *Perspectives in Psychiatric Care*. 2012;**48**:218-224.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Margrove K, Mensah S, Thapar A, Kerr M. Depression screening for patients with epilepsy in a primary care setting using the Patient Health Questionnaire-2 and the Neurological Disorders Depression Inventory for Epilepsy. *Epilepsy & Behavior*. 2011;**21**:387-390.	Study only administered the PHQ-2
Mautner E, Ashida C, Greimel E, et al. Are there differences in the health outcomes of mothers in Europe and East-Asia? A cross-cultural health Survey. *Biomed Research International*. 2014;856543.	Major depression not assessed
Mitchell AJ, McGlinchey JB, Young D, Chelminski I, Zimmerman M. Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. *Psychological Medicine*. 2009;**39**:1107-1116.	PHQ not administered
Mittal D, Fortney JC, Pyne JM, Wetherell JL. Predictors of persistence of comorbid generalized anxiety disorder among veterans with major depressive disorder. *Journal of Clinical Psychiatry.* 2011;**72**:1445-1451.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Morina N, von Lersner U, Prigerson HG. War and bereavement: consequences for mental and physical distress. *PLoS ONE*. 2011;**6**:e22140.	PHQ not administered
Muller KW, Beutel ME, Wolfling K. A contribution to the clinical characterization of Internet addiction in a sample of treatment seekers: validity	Major depression not assessed

of assessment, severity of psychopathology and type of co-morbidity. *Comprehensive Psychiatry*. 2014;**55**:770-777.	
Mulligan L, Fear NT, Jones N, et al. Postdeployment Battlemind training for the U.K. armed forces: A cluster randomized controlled trial. *Journal of Consulting and Clinical Psychology*. 2012;**80**:331-341.	Major depression not assessed
Mussell M, Kroenke K, Spitzer RL, et al. Gastrointestinal symptoms in primary care: prevalence and association with depression and anxiety. *Journal of Psychosomatic Research*. 2008;**64**:605-612.	Major depression not assessed
Olariu E, Castro-Rodriguez JI, Alvarez P, et al. Validation of clinical symptom irt scores for diagnosis and severity assessment of common mental disorders. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*. 2014.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Orive M, Padierna JA, Quintana JM, et al. Detecting depression in medically ill patients: Comparative accuracy of four screening questionnaires and physicians' diagnoses in Spanish population. *Journal of Psychosomatic Research*. 2010;**69**:399-406.	No validated interview to assess major depression
Osorio FL, de Carvalho AC, Crippa JA, Loureiro SR. Screening for smoking in a general hospital: scale validation, indicators of prevalence, and comorbidity. *Perspectives in Psychiatric Care*. 2013;**49**:5-12.	Major depression not assessed
Park H, Kim J, Hahm B. The Distress Thermometer and the PHQ-2 for ultra-brief screening depression of cancer patients In Korea. *Psycho-oncology.* 2013;**22**:303-304.	Study only administered the PHQ-2
Pibernik-Okanovic M, Grgurevic M, Ajdukovic D, Novak B, Begic D, Metelko Z. Screening performance of a short versus long version of the Patient Health Questionnaire-depression in outpatients with diabetes. *Diabetologia*. 2009;**52**:S392-S393.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Pilipenko N, Karekla M, Feldman J. Validation of Patient Health Questionnaire in Greek-language sample. *European Psychiatry*. 2011;**26**.	Major depression not assessed
Poutanen O, Koivisto AM, Salokangas RK. Applicability of the DEPS Depression Scale: assessing format and individual items in subgroups of patients. *Nordic Journal of Psychiatry*. 2010;**64**:384-390.	Major depression not assessed
Prescott MR, Tamburrino M, Calabrese JR, et al. Validation of lay-administered mental health assessments in a large Army National Guard cohort. *International Journal of Methods in Psychiatric Research.* 2014;**23**:109-119.	> 2 weeks between PHQ and diagnostic interview
Priyanka P, Boyle LL, Tu XM, Conwell Y. Inter-rater reliability and validity of the PHQ-9 and GAD-7 to identify depression and anxiety in older adults receiving aging services care management. *American Journal of Geriatric Psychiatry*. 2010;**18**:S113-S114.	No original data
Reck C, Stehle E, Reinig K, Mundt C. Maternity blues as a predictor of DSM-IV depression and anxiety disorders in the first three months postpartum. *Journal of Affective Disorders*. 2009;**113**:77-87.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Rentsch D, Dumont P, Borgacci S, et al. Prevalence and treatment of depression in a hospital department of internal medicine. *General Hospital Psychiatry*. 2007;**29**:25-31.	No validated interview to assess major depression
Rief W, Mewes R, Martin A, Glaesmer H, Braehler E. Are psychological features useful in classifying patients with somatic symptoms? *Psychosomatic Medicine.* 2010;**72**:648-655.	> 2 weeks between PHQ and diagnostic interview
Ringoir L, Pedersen SS, Widdershoven JW, Pop VJ. Prevalence of psychological distress in elderly hypertension patients in primary care. *Netherlands Heart Journal.* 2014;**22**:71-76.	Major depression not assessed
Rizzo R, Piccinelli M, Mazzi MA, Bellantuono C, Tansella M. The Personal Health Questionnaire: a new screening instrument for detection of ICD-10 depressive disorders in primary care. *Psychological Medicine*. 2000;**30**:831-840.	PHQ not administered
Ryan DA, Gallagher P, Wright S, Cassidy EM. Sensitivity and specificity of the Distress Thermometer and a two-item depression screen (Patient Health	PHQ not administered

Questionnaire-2) with a 'help' question for psychological distress and psychiatric morbidity in patients with advanced cancer. *Psycho-oncology*. 2012;**21**:1275-1284.	
Saliba D, DiFilippo S, Edelen MO, et al. Testing the PHQ-9 interview and observational versions (PHQ-9 OV) for MDS 3.0. *Journal of the American Medical Directors Association*. 2012;**13**:618-625.	PHQ not administered
Salve H, Goswami K, Nongkynrih B, Sagar R, Sreenivas V. Prevalence of psychiatric morbidity at Mobile Health Clinic in an urban community in North India. *General Hospital Psychiatry*. 2012;**34**:121-126.	PHQ not administered
Sayers SL, Farrow VA, Ross J, Oslin DW. Family problems among recently returned military veterans referred for a mental health evaluation. *Journal of Clinical Psychiatry*. 2009;**70**:163-170.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Schmitz-Hubsch T, Coudert M, Tezenas du Montcel S, et al. Depression comorbidity in spinocerebellar ataxia. *Movement Disorders*. 2011;**26**:870-876.	Major depression not assessed
Shen Q, Bergquist-Beringer S. Relationship between major depression and insulin resistance: Does it vary by gender or race/ethnicity among young adults aged 20-39 years? *Journal of Diabetes*. 2013;**5**:471-481.	Major depression not assessed
Shoukri MM, Donner A. Bivariate modeling of interobserver agreement coefficients. *Statistics in medicine.* 2009;**28**:430-440.	No original data
Smith AB, Rush R, Wright P, et al. Validation of an item bank for detecting and assessing psychological distress in cancer patients. *Psycho-oncology.* 2009;**18**:195-199.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Smith GC, McAsey P, Trauer T. Screening and monitoring in renal dialysis and transplant patients using the SF36 and Patient Health Questionnaire. *Australian and New Zealand Journal of Psychiatry*. 2000;**34**:A62-A62.	Major depression not assessed
Smith GC, McAsey P, Trauer T. Screening and monitoring in renal analysis and transplant patients using the SF36 and Patient Health Questionnaire. *Psychosomatics.* 2001;**42**:182-183.	Major depression not assessed
Smith GC, Trauer T, Kerr PG, Chadban SJ. Prospective psychosocial monitoring of living kidney donors using the Short Form-36 Health Survey: Results at 12 months. *Transplantation.* 2004;**78**:1384-1389.	No validated interview to assess major depression
Smith MV, Gotman N, Lin H, Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *General Hospital Psychiatry*. 2010;**32**:544-548.	Study only administered the PHQ-8
Sockalingam S, Blank D, Al Jarad A, et al. A comparison of depression screening instruments in hepatitis C and the impact of depression on somatic symptoms. *Psychosomatics*. 2011;**52**:433-440.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Stegenga BT, Kamphuis MH, King M, Nazareth I, Geerlings MI. The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study. *Social Psychiatry & Psychiatric Epidemiology*. 2012;**47**:87-95.	Major depression not assessed
Subramanian U, Perkins SM, Kim J, Ding Y, Pressler SJ. Depressive symptoms in heart failure: Validity and reliability of the PHQ-8. *Journal of General Internal Medicine*. 2008;**23**:276-276.	Major depression not assessed
Suzuki T, Shiga T, Nishimura K, Ishigooka J, Hagiwara N. PHQ-9 screening for depression in hospitalized patients with heart failure. *European Journal of Heart Failure*. 2013;S242.	Major depression not assessed
Tabb KM, Gavin AR, Guo Y, et al. Views and experiences of suicidal ideation during pregnancy and the postpartum: findings from interviews with maternal care clinic patients. *Women & Health*. 2013;**53**:519-535.	Major depression not assessed
Tavakkoli M, Ferrando SJ, Rabkin J, Marks K, Talal AH. Depression and fatigue in chronic hepatitis C patients with and without HIV co-infection. *Psychosomatics*. 2013;**54**:466-471.	No validated interview to assess major depression

Thapar A, Hammerton G, Collishaw S, et al. Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures. *British Journal of General Practice.* 2014;**64**:e31-7.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Thekkumpurath P, Walker J, Butcher I, et al. Screening for major depression in cancer outpatients: the diagnostic accuracy of the 9-item Patient Health Questionnaire. *Cancer*. 2011;**117**:218-227.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Tilli V, Suominen K, Karlsson H. The Autonomic Nervous System Questionnaire and the Brief Patient Health Questionnaire as screening instruments for panic disorder in Finnish primary care. *European Psychiatry: the Journal of the Association of European Psychiatrists.* 2013;**28**:442-447.	PHQ not administered
Tschudi-Madsen H, Kjeldsberg M, Natvig B, et al. Multiple symptoms and medically unexplained symptoms-Closely related concepts in general practitioners' evaluations. A linked doctor-patient study. *Journal of Psychosomatic Research*. 2013;**74**:186-190.	PHQ not administered
Uebelacker LA, German NM, Gaudiano BA, Miller IW. Patient Health Questionnaire depression scale as a suicide screening instrument in depressed primary care patients: a cross-sectional study. *The Primary Care Companion to CNS Disorders*. 2011;**13.**	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Ulhaq S, Symeon C, Agius M. Use of the PHQ-9 as a screening tool for post-stroke depression. *European Psychiatry*. 2010;**25**.	Major depression not assessed
Vera M, Reyes-Rabanillo ML, Huertas S, et al. Suicide ideation, plans, and attempts among general practice patients with chronic health conditions in Puerto Rico. *International Journal of General Medicine*. 2011;**4**:197-205.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Watson LC, Zimmerman S, Cohen LW, Dominik R. Practical depression screening in residential care/assisted living: five methods compared with gold standard diagnoses. *American Journal of Geriatric Psychiatry*. 2009;**17**:556-564.	PHQ not administered
Whitlow NR, Ryan GL, Stuart SP. The Patient Health Questionnaire (PHQ) is a poor psychological screening tool in in vitro fertilization (IVF) Patients. *Fertility and Sterility.* 2011;**96**:S11-S11.	Major depression not assessed
Williams LS, Brizendine EJ, Plue L, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 2005;**36**:635-638.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Yeung A, Fung F, Yu SC, et al. Validation of the Patient Health Questionnaire-9 for depression screening among Chinese Americans. *Comprehensive Psychiatry*. 2008;**49**:211-217.	> 2 weeks between PHQ and diagnostic interview
Yeung A, Yu SC, Fung F, Vorono S, Fava M. Recognizing and engaging depressed Chinese Americans in treatment in a primary care setting. *International Journal of Geriatric Psychiatry.* 2006;**21**:819-823.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Zuithoff NP, Vergouwe Y, King M, et al. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study. *BMC Family Practice*. 2010;**11**:98.	Major depression not assessed

Supplementary Table 1b. Reasons for exclusion of all articles excluded at the full-text level for the main IPDMA of the EPDS (N=213)

Reference	Reason for exclusion
Abiodun OA. Postnatal depression in primary care populations in Nigeria. Gen Hosp Psychiatry. 2006;**28**:133-6.	Could not determine eligibility[a]
Abou-Saleh MT, Ghubash R, Karim L, Krymski M, Bhai I. Hormonal aspects of postpartum depression. Psychoneuroendocrinology. 1998;**23**:465-75.	> 2 weeks between EPDS and diagnostic interview
Aceti F, Baglioni V, Ciolli P, De Bei F, Di Lorenzo F, Ferracuti S, et al. Maternal attachment patterns and personality in post partum depression. Riv Psichiatr. 2012;**47**:214-20.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Adewuya AO, Eegunranti AB, Lawal AM. Prevalence of postnatal depression in Western Nigerian women: a controlled study. Int J Psychiatry Clin Pract. 2005;**9**:60-4.	Could not determine eligibility[a]
Adewuya AO. Early postpartum mood as a risk factor for postnatal depression in Nigerian women. Am J Psychiatry. 2006;**163**:1435-7.	No validated interview to assess major depression
Ahn S, Corwin EJ. The association between breastfeeding, the stress response, inflammation, and postpartum depression during the postpartum period: Prospective cohort study. Int J Nurs Stud. 2015;**52**:1582-90.	Major depression not assessed
Alami KM, Kadri N, Berrada S. Prevalence and psychosocial correlates of depressed mood during pregnancy and after childbirth in a Moroccan sample. Arch Womens Ment HealthArch Womens Ment Health. 2006;**9**:343-6.	Could not determine eligibility[a]
Albacar G, Sans T, MartinSantos R, GarciaEsteve L, Guillamat R, Sanjuan J, et al. Thyroid function 48 h after delivery as a marker for subsequent postpartum depression. Psychoneuroendocrinology. 2010;**35**:738-42.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Albacar G, Sans T, MartinSantos R, GarciaEsteve L, Guillamat R, Sanjuan J, et al. An association between plasma ferritin concentrations measured 48h after delivery and postpartum depression. J Affect DisordJ Affect Disord. 2011;**131**:136-42.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Alexander S, Palmer C, Stone PC. Evaluation of screening instruments for depression and anxiety in breast cancer survivors. Breast Cancer Res Treat. 2010;**122**:573-8.	No pregnant or postpartum women
Algul A, Semiz UB, Dundar O, Ates MA, Basoglu C, Ebrinc S, et al. Psychosocial and hormone related risk factors for early postnatal depressive symptoms in Turkish women. Neurol Psychiat Br. 2008;**15**:117-22.	Major depression not assessed
Al-Modayfer O, Alatiq Y, Khair O, Abdelkawi S. Postpartum depression and related risk factors among Saudi females. Int J Cult Ment Health. 2015;**8**:316-24.	No validated interview to assess major depression
Alvarado-Esquivel C, Sifuentes-Alvarez A, Estrada-Martínez S, Salas-Martínez C, Hernndez-Alvarado AB, Ortiz-Rocha SG, et al. Prevalence of postnatal depression in women attending public hospitals in Durango, Mexico. Gac Med Mex. 2010;**146**:1-9.	No validated interview to assess major depression
Alvarado-Esquivel C, Sifuentes-Alvarez A, Salas-Martinez C. Unhappiness with the Fetal Gender is associated with Depression in Adult Pregnant	Sample selected for known distress, mental health

Women Attending Prenatal Care in a Public Hospital in Durango, Mexico. Int J Biomed Sci. 2016;**12**:36-41.	diagnosis, or psychiatric setting
Areias ME, Kumar R, Barros H, Figueiredo E. Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers. Br J Psychiatry. 1996;**169**:30-5.	No validated interview to assess major depression
Areias ME, Kumar R, Barros H, Figueiredo E. Correlates of postnatal depression in mothers and fathers. Br J Psychiatry. 1996;**169**:36-41.	No validated interview to assess major depression
Austin MP, Dudley M, Launders C, Dixon C, Macartney-Bourne F. Description and evaluation of a domiciliary perinatal mental health service focussing on early intervention. Arch Womens Ment Health. 1999;**2**:169-73.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Austin MP, Frilingos M, Lumley J, Hadzi-Pavlovic D, Roncolato W, Acland S, et al. Brief antenatal cognitive behaviour therapy group intervention for the prevention of postnatal depression and anxiety: a randomised controlled trial. J Affect Disord. 2008;**105**:35-44.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Austin MP, Hadzi-Pavlovic D, Priest SR, Reilly N, Wilhelm K, Saint K, Parker G. Depressive and anxiety disorders in the postpartum period: how prevalent are they and can we improve their detection? Arch Womens Ment Health. 2010;**13**:395-401.	Major depression not assessed
Austin MP, Hadzi-Pavlovic D, Saint K, Parker G. Antenatal screening for the prediction of postnatal depression: validation of a psychosocial Pregnancy Risk Questionnaire. Acta Psychiatr Scand. 2005;**112**:310-7.	Major depression not assessed
Azar R, Paquette D, Zoccolillo M, Baltzer F, Tremblay RE. The association of major depression, conduct disorder, and maternal overcontrol with a failure to show a cortisol buffered response in 4-month-old infants of teenage mothers. Biological Psychiatry. 2007;**62**:573-9.	Not a sample of adults
Bågedahl-Strindlund M, Monsen Börjesson K. Postnatal depression: a hidden illness. Acta Psychiatr Scand. 1998;**98**:272-5.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Bergant AM, Heim K, Ulmer H, Illmensee K. Early postnatal depressive mood: associations with obstetric and psychosocial factors. J Psychosom Res. 1999;**46**:391-4.	Major depression not assessed
Bergant AM, Nguyen T, Heim K, Ulmer H, Dapunt O. German language version and validation of the Edinburgh postnatal depression scale. Dtsch Med Wochenschr. 1998;**123**:35-40.	No validated interview to assess major depression
Bick DE, MacArthur C, Lancashire RJ. What influences the uptake and early cessation of breast feeding? Midwifery. 1998;**14**:242-7.	Major depression not assessed
Bloch M, Rotenberg N, Koren D, Klein E. Risk factors associated with the development of postpartum mood disorders. J Affect Disord. 2005;**88**:9-18.	> 2 weeks between EPDS and diagnostic interview
Boath E, Cox J, Lewis M, Jones P, Pryce A. When the cradle falls: the treatment of postnatal depression in a psychiatric day hospital compared with routine primary care. J Affect Disord. 1999;**53**:143-51.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Boyce P, Hickey A. Psychosocial risk factors to major depression after childbirth. Soc Psychiatry Psychiatr Epidemiol. 2005;**40**:605-12.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Boyce P, Stubbs J, Todd A. The Edinburgh Postnatal Depression Scale: validation for an Australian sample. Aust N Z J Psychiatry. 1993;**27**:472-6.	Sample selected for known distress, mental health

	diagnosis, or psychiatric setting
Browne JC, Scott KM, Silvers KM. Fish consumption in pregnancy and omega-3 status after birth are not associated with postnatal depression. J Affect Disord. 2006;**90**:131-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Brugha TS, Wheatley S, Taub NA, Culverwell A, Friedman T, Kirwan P, et al. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. Psychol Med. 2000;**30**:1273-81.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Bunevičius A, Kusminskas L, Bunevičius R. Validation of the Lithuanian version of the Edinburgh Postnatal Depression Scale. Med Lith. 2009;**45**:544.	No validated interview to assess major depression
Bunevicius A, Kusminskas L, Bunevicius R. Validity of the Edinburgh Postnatal Depression Scale. Eur Psychiatry. 2009;**24:**S896.	No validated interview to assess major depression
Burns A, O'Mahen H, Baxter H, Bennert K, Wiles N, Ramchandani P, et al. A pilot randomised controlled trial of cognitive behavioural therapy for antenatal depression. BMC Psychiatry. 2013;**13**:33.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Byatt N, Biebel K, Simas TAM, Sarvet B, Ravech M, Allison J, Straus J. Improving perinatal depression care: The Massachusetts Child Psychiatry Access Project for Moms. Gen Hosp Psychiatry. 2016;**40**:12-7.	Major depression not assessed
Caramlau I, Barlow J, Sembi S, McKenzie-McHarg K, McCabe C. Mums 4 Mums: structured telephone peer-support for women experiencing postnatal depression. Pilot and exploratory RCT of its clinical and cost effectiveness. Trials . 2011;**12**:88.	No original data
Carothers AD, Murray L. Estimating psychiatric morbidity by logistic regression: application to post-natal depression in a community sample. Psychol Med. 1990;**20**:695-702.	No validated interview to assess major depression
Carpiniello B, Pariante CM, Serri F, Costa G, Carta MG. Validation of the Edinburgh Postnatal Depression Scale in Italy. J Psychosom Obstet Gynecol. 1997;**18**:280-5.	No validated interview to assess major depression
Castañón SC, Pinto LJ. Use of the Edinburgh Postnatal Depression Scale to detect postpartum depression. Rev Med Chil. 2008;**136**:851-8.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Chaudron LH, Nirodi N. The obsessive-compulsive spectrum in the perinatal period: a prospective pilot study. Arch Womens Ment Health. 2010;**13**:403-10.	> 2 weeks between EPDS and diagnostic interview
Chee CY, Chong YS, Ng TP, Lee DT, Tan LK, Fones CS. The association between maternal depression and frequent non-routine visits to the infant's doctor--a cohort study. J Affect Disord. 2008;**107**:247-53.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Chee CYI, Lee DTS, Chong YS, Tan LK, Ng TR, Fones CSL. Confinement and other psychosocial factors in perinatal depression: A transcultural study in Singapore. J Affect Disord. 2005;**89**:157-66.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Chen H, Bautista D, Ch'ng YC, Li W, Chan E, Rush AJ. Screening for postnatal depression in Chinese-speaking women using the Hong Kong translated version of the Edinburgh Postnatal Depression Scale. Asia Pac Psychiatry. 2013;**5**:E64-E72.	No validated interview to assess major depression

Chibanda D, Verhey R, Gibson LJ, Munetsi E, Machando D, Rusakaniko S, et al. Validation of screening tools for depression and anxiety disorders in a primary care population with high HIV prevalence in Zimbabwe. J Affect Disord. 2016;**198**:50-55.	EPDS not administered
Clarke PJ. Validation of two postpartum depression screening scales with a sample of First Nations and Metis women. Can J Nurs Res. 2008;**40**:112-25.	Major depression not assessed
Class QA, Verhulst J, Heiman JR. Exploring the heterogeneity in clinical presentation and functional impairment of postpartum depression. J Reprod Infant Psychol. 2013;**31**:183-94.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Clifford C, Day A, Cox J, Werrett J. A cross-cultural analysis of the use of the Edinburgh Post-Natal Depression Scale (EPDS) in health visiting practice. J Adv Nurs. 1999;**30**:655-64.	No validated interview to assess major depression
Coleman R, Morison L, Paine K, Powell RA, Walraven G. Women's reproductive health and depression: a community survey in the Gambia, West Africa. Soc Psychiatry Psychiatr Epidemiol. 2006;**41**:720-7.	No validated interview to assess major depression
Cooper PJ, Murray L, Wilson A, Romaniuk H. Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression. I. Impact on maternal mood. Br J Psychiatry. 2003;**182**:412-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Costas J, Gratacòs M, Escaramís G, Martín-Santos R, de Diego Y, Baca-García E, et al. Association study of 44 candidate genes with depressive and anxiety symptoms in post-partum women. J Psychiatr Res. 2010;**44**:717-24.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J Affect Disord. 1996;**39**:185-9.	No validated interview to assess major depression
Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987;**150**:782-6.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry. 1993;**163**:27-31.	No validated interview to assess major depression
de Souza Ribeiro Martins C, dos Santos Motta JV, Quevedo LA, de Matos MB, Pinheiro KAT, de Mattos Souza LD, et al. Comparison of two instruments to track depression symptoms during pregnancy in a sample of pregnant teenagers in Southern Brazil. J Affect Disord. 2015;**177**:95-100.	Not a sample of adults
Dennis CL, Hodnett E, Kenton L, Weston J, Zupancic J, Stewart DE, Kiss A. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. BMJ. 2009;**338**:a3064.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Ebeigbe PN, Akhigbe KO. Incidence and associated risk factors of postpartum depression in a tertiary hospital in Nigeria. Niger Postgrad Med J. 2008;**15**:15-8.	Major depression not assessed
Eberhard-Gran M, Eskild A, Tambs K, Schei B, Opjordsmoen S. The Edinburgh Postnatal Depression Scale: validation in a Norwegian community sample. Nord J Psychiatry. 2001;**55**:113-7.	No validated interview to assess major depression
Ekeroma AJ, Ikenasio-Thorpe B, Weeks S, Kokaua J, Puniani K, Stone P, Foliaki SA. Validation of the Edinburgh Postnatal Depression Scale	> 2 weeks between EPDS and diagnostic interview

(EPDS) as a screening tool for postnatal depression in Samoan and Tongan women living in New Zealand. N Z M J. 2012;**125**:41-9.	
Ekuklu G, Tokuc B, Eskiocak M, Berberoglu U, Saltik A. Prevalence of postpartum depression in Edirne, Turkey, and related factors. J Reprod Med. 2004;**49**:908-14.	Major depression not assessed
El-Ibiary SY, Hamilton SP, Abel R, Erdman CA, Robertson PA, Finley PR. A pilot study evaluating genetic and environmental factors for postpartum depression. Innov Clin Neurosci. 2013;**10**:15-22.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Elliott SA, Leverton TJ, Sanjack M, Turner H, Cowmeadow P, Hopkins J, Bushnell D. Promoting mental health after childbirth: a controlled trial of primary prevention of postnatal depression. Br J Clin Psychol. 2000;**39**:223-41.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Fairbrother N, Young AH, Janssen P, Antony MM, Tucker E. Depression and anxiety during the perinatal period. BMC Psychiatry. 2015;**15**:206.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Farhat A, Saeidi R, Mohammadzadeh A, Hesari H. Prevalence of Postpartum Depression; a longitudinal study. Iran J Neonatol. 2015;**6**:39-44.	Major depression not assessed
Flynn HA, Sexton M, Ratliff S, Porter K, Zivin K. Comparative performance of the Edinburgh Postnatal Depression Scale and the Patient Health Questionnaire-9 in pregnant and postpartum women seeking Psychiatr Serv. Psychiatry Res. 2011;**187**:130-4.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gallanti AME, Rodríguez CEAM, Rodríguez IM, Sosa MA. Puerperal depression and its association with demographic and social factors, the way of resolution of pregnancy and the newborn clinical evolution. Medula. 2015;**24**:25-34.	Major depression not assessed
Gelabert E, Subira S, Plaza A, Torres A, Navarro P, Imaz ML, et al. The Vulnerable Personality Style Questionnaire: psychometric properties in Spanish postpartum women. Arch Womens Ment Health. 2011;**14**:115-24.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gemmill AW, Leigh B, Ericksen J, Milgrom J. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. BMC Public Health. 2006;**6**:211.	Major depression not assessed
Georgiopoulos AM, Bryan TL, Wollan P, Yawn BP. Routine screening for postpartum depression. J Fam Pract. 2001;**50**:117.	Major depression not assessed
Gerardin P, Wendland J, Bodeau N, Galin A, Bialobos S, Tordjman S, et al. Depression during pregnancy: Is the developmental impact earlier in boys? A prospective case-control study. J Clin Psychiatry. 2011;**72**:378-87.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gerardin P. Characteristics and clinical consequences of prenatal depression. Main results of a prospective case-control study on perinatal depression from pregnancy to one year-old infant. Neuropsychiatr Enfance eAdolesc. 2012;**60**:138-46.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Ghubash R, Abou-Saleh MT, Daradkeh TK. The validity of the Arabic Edinburgh Postnatal Depression Scale. Soc Psychiatry Psychiatr Epidemiol. 1997;**32**:474-6.	> 2 weeks between EPDS and diagnostic interview
Ghubash R, Abou-Saleh MT. Postpartum psychiatric illness in Arab culture: prevalence and psychosocial correlates. Br J Psychiatry. 1997;**171**:65-8.	> 2 weeks between EPDS and diagnostic interview

Ginsburg GS, Barlow A, Goklish N, Hastings R, Baker EV, Mullany B, et al. Postpartum depression prevention for reservation-based American Indians: Results from a Pilot Randomized Controlled Trial. Child Youth Care Forum. 2012;**41**:229-45.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Goeb JL, Férel S, Guetta J, Guibert J, Guedeney A, Coste J, et al. Assisted reproductive techniques when the Man is HIV Seropositive. Psychiatr Enfant. 2009;**52**:63-88.	Major depression not assessed
Goutaudier N, Lopez A, SéjournéN, Denis A, Chabrol H. Premature birth: subjective and psychological experiences in the first weeks following childbirth, a mixed-methods study. J Reprod Infant Psychol. 2011;**29**:364-73.	Major depression not assessed
Goyal D, Park VT, McNiesh S. Postpartum depression among Asian Indian mothers. MCN Am J Matern Child Nurs2015;**40**:256-61.	Major depression not assessed
Grant KA, Bautovich A, McMahon C, Reilly N, Leader L, Austin MP. Parental care and control during childhood: Associations with maternal perinatal mood disturbance and parenting stress. Arch Womens Ment Health. 2012;**15**:297-305.	Could not determine eligibility[a]
Grant KA, McMahon C, Austin MP, Reilly N, Leader L, Ali S. Maternal prenatal anxiety, postnatal caregiving and infants' cortisol responses to the still-face procedure. Dev Psychobiol. 2009;**51**:625-37.	Could not determine eligibility[a]
Grant KA, McMahon C, Reilly N, Austin MP. Maternal sensitivity moderates the impact of prenatal anxiety disorder on infant responses to the still-face procedure. Infant Behav Dev. 2010;**33**:453-62.	Could not determine eligibility[a]
Grigoriadis S, de Camps Meschino D, Barrons E, Bradley L, Eady A, Fishell A, et al. Mood and anxiety disorders in a sample of Canadian perinatal women referred for psychiatric care. Arch Womens Ment Health. 2011;**14**:325-33.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Guedeney N, Fermanian J. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results about use and psychometric properties. Eur Psychiatry. 1998;**13**:83-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gutierrez-Zotes A, Labad J, Martin-Santos R, Garcia-Esteve L, Gelabert E, Jover M, et al. Coping strategies and postpartum depressive symptoms: A structural equation modelling approach. Eur Psychiatry. 2015;**30**:701-8.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gutiérrez -Zotes JA, Farnós A, Vilella E, Labad J. Higher psychoticism as a predictor of thoughts of harming one's infant in postpartum women: a prospective study. Compr Psychiatry. 2013;**54**:1124-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Gutiérrez -Zotes A, Labad J, MartinSantos R, GarciaEsteve L, Gelabert E, Jover M, et al. Coping strategies and postpartum depressive symptoms: A structural equation modelling approach. Eur Psychiatry. 2015;**30**:701-8.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Hamdan A, Tamim H. Psychosocial risk and protective factors for postpartum depression in the United Arab Emirates. Arch Womens Ment Health. 2011;**14**:125-33.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Hamdan A, Tamim H. The relationship between postpartum depression and breastfeeding. Int J Psychiatry Med. 2012;**43**:243-59.	Sample selected for known distress, mental health diagnosis, or psychiatric setting

Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: a comparison of three instruments. J Womens Health. 2008;**17**:585-96.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Harris B, Huckle P, Thomas R, Johns S, Fung H. The use of rating scales to identify post-natal depression. Br J Psychiatry. 1989;**154**:813-7.	Could not determine eligibility[a]
Harris B, Othman S, Davies JA, Weppner GJ, Richards CJ, Newcombe RG, et al. Association between postpartum thyroid dysfunction and thyroid antibodies and depression. BMJ. 1992;**305**:152-6.	No validated interview to assess major depression
Harvey ST, Pun PK. Analysis of positive Edinburgh depression scale referrals to a consultation liaison psychiatry service in a two-year period. Int J Ment Health Nurs. 2007;**16**:161-7.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Hatton DC, HarrisonHohner J, Matarazzo J, Edwards P, Lewy A, Davis L. Missed antenatal depression among high risk women: A secondary analysis. Arch Womens Ment Health. 2007;**10**:121-3.	No validated interview to assess major depression
Henshaw C, Foreman D, Cox J. Postnatal blues: a risk factor for postnatal depression. J Psychosom Obstet Gynecol. 2004;**25**:267-72.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Herz E, Thoma M, Umek W, Gruber K, Linzmayer L, Walcher W, et al. Non-psychotic post-partum depression. Geburtshilfe Frauenheilkd. 1997;**57**:282-8.	Major depression not assessed
Holden JM. Postnatal depression: its nature, effects, and identification using the Edinburgh Postnatal Depression scale. Birth. 1991;**18**:211-21.	No original data
Holt WJ. The detection of postnatal depression in general practice using the Edinburgh postnatal depression scale. N Z M J. 1995;**108**:57.	> 2 weeks between EPDS and diagnostic interview
Howard LM, Flach C, Mehay A, Sharp D, Tylee A. The prevalence of suicidal ideation identified by the Edinburgh Postnatal Depression Scale in postpartum women in primary care: findings from the RESPOND trial. BMC Pregnancy Childbirth. 2011;**11**:57-59.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Huang J, Zhang L, He M, Qiang X, Xiao X, Huang S, et al. Comprehensive evaluation of postpartum depression and correlations between postpartum depression and serum levels of homocysteine in Chinese women. Zhong Nan Da Xue Xue Bao Yi Xue BanZhong Nan Da Xue Xue Bao Yi Xue Ban. 2015;**40**:311-6.	No validated interview to assess major depression
Huang YC, Mathers NJ. Postnatal depression and the experience of South Asian marriage migrant women in Taiwan: survey and semi-structured interview study. Int J Nurs Stud. 2008;**45**:924-31.	Major depression not assessed
Husain N, Cruickshank K, Husain M, Khan S, Tomenson B, Rahman A. Social stress and depression during pregnancy and in the postnatal period in British Pakistani mothers: a cohort study. J Affect Disord. 2012;**140**:268-76.	Could not determine eligibility[a]
Husain N, Kiran T, Sumra A, Naeem Zafar S, Ur Rahman R, Jafri F, et al. Detecting maternal depression in a low-income country: comparison of the self-reporting questionnaire and the Edinburgh Postnatal Depression Scale. J Trop Pediatr. 2014;**60**:129-33.	Could not determine eligibility[a]
Ibanez G, Bernard JY, Rondet C, Peyre H, Forhan A, Kaminski M, et al. Effects of antenatal maternal depression and anxiety on children's early cognitive development: A prospective cohort study. PLOS ONE. 2015;**10**:e0135849.	Major depression not assessed

Ikeda M, Hayashi M, Kamibeppu K. The relationship between attachment style and postpartum depression. Attach Hum Dev. 2014;**16**:557-72.	> 2 weeks between EPDS and diagnostic interview
Inglis AJ, Hippman CL, Carrion PB, Honer WG, Austin JC. Mania and depression in the perinatal period among women with a history of major depressive disorders. Arch Womens Ment Health. 2014;**17**:137-43.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Jadresic E, Araya R, Jara C. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Chilean postpartum women. J Psychosom Obstet Gynecol. 1995;**16**:187-91.	No validated interview to assess major depression
Jaju S, Al Kharusi L, Gowri V. Antenatal prevalence of fear associated with childbirth and depressed mood in primigravid women. Indian J Psychiatry. 2015;**57**:158-61.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Jardri R, Maron M, Pelta J, Thomas P, Codaccioni X, Goudemand M, Delion P. Impact of midwives' training on postnatal depression screening in the first week post delivery: a quality improvement report. Midwifery. 2010;**26**:622-9.	> 2 weeks between EPDS and diagnostic interview
Ji S, Long Q, Newport DJ, Na H, Knight B, Zach EB, et al. Validity of depression rating scales during pregnancy and the postpartum period: impact of trimester and parity. J Psychiatr Res. 2011;**45**:213-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Josefsson A, Larsson C, Sydsjö G, Nylander PO. Temperament and character in women with postpartum depression. Arch Womens Ment Health. 2007;**10**:3-7.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Keshavarzi F, Yazdchi K, Rahimi M, Rezaei M, Farnia V, Davarinejad O, et al. Post partum depression and thyroid function. Iran J Psychiatry. 2011;**6**:117-20.	Major depression not assessed
Kirkan TS, Aydin N, Yazici E, Akcali Aslan P, Acemoglu H, Daloglu AG. The depression in women in pregnancy and postpartum period: A follow-up study. Int J Soc Psychiatry. 2015;**61**:343-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Klier CM, Muzik M, Dervic K, Mossaheb N, Benesch T, Ulm B, Zeller M. The role of estrogen and progesterone in depression after birth. J Psychiatr Res. 2007;**41**:273-9.	> 2 weeks between EPDS and diagnostic interview
Knorring LV. Book review of Depression in women with focus on the postpartum period. Nord J Psychiatry. 2003;**57**:390.	No validated interview to assess major depression
Kohlhoff J, Hickinbotham R, Knox C, Roach V, Barnett Am B. Antenatal psychosocial assessment and depression screening in a private hospital. Aust N Z J Obstet Gynaecol. 2016;**56**:173-8.	Major depression not assessed
Koss J, Bidzan M, Smutek J, Bidzan L. Influence of perinatal depression on labor-associated fear and mmotional attachment to the child in high-risk pregnancies and the first days after delivery. Med Sci Monit. 2016;**22**:1028-37.	Major depression not assessed
Lai BP, Tang AK, Lee DT, Yip AS, Chung TK. Detecting postnatal depression in Chinese men: a comparison of three instruments. Psychiatry Res. 2010;**180**:80-5.	No pregnant or postpartum women
Lau Y, Wang Y, Yin L, Chan KS, Guo X. Validation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale in Chengdu mothers. Int J Nurs Stud. 2010;**47**:1139-51.	Could not determine eligibility[a]

Lawrie TA, Hofmeyr GJ, de Jager M, Berk M. Validation of the Edinburgh Postnatal Depression Scale on a cohort of South African women. S Afr Med J. 1998;**88**:1340-4.	No validated interview to assess major depression
Lee DT, Wong CK, Ungvari GS, Cheung LP, Haines CJ, Chung TK. Screening psychiatric morbidity after miscarriage: application of the 30-item General Health Questionnaire and the Edinburgh Postnatal Depression Scale. Psychosom Med. 1997;**59**:207-10.	No pregnant or postpartum women
Lee DT, Yip AS, Chan SS, Tsui MH, Wong WS, Chung TK. Postdelivery screening for postpartum depression. Psychosom Med. 2003;**65**:357-61.	Major depression not assessed
Lee DT, Yip AS, Chiu HF, Chung TK. Screening for postnatal depression using the double-test strategy. Psychosom Med. 2000;**62**:258-63.	Major depression not assessed
Lee DT, Yip AS, Chiu HF, Leung TY, Chung TK. Screening for postnatal depression: are specific instruments mandatory? J Affect Disord. 2001;**63**:233-8.	Major depression not assessed
Lee DT, Yip SK, Chiu HF, Leung TY, Chan KP, Chau IO, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1998;**172**:433-7.	No validated interview to assess major depression
Leverton TJ, Elliott SA. Is the EPDS a magic wand?: 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. J Reprod Infant Psychol. 2000;**18**:279-96.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Lewis BA, Gjerdingen DK, Avery MD, Guo H, Sirard JR, Bonikowske AR, Marcus BH. Examination of a telephone-based exercise intervention for the prevention of postpartum depression: design, methodology, and baseline data from The Healthy Mom study. Contemp Clin Trials. 2012;**33**:1150-8.	Major depression not assessed
Lewis BA, Gjerdingen DK, Avery MD, Sirard JR, Guo H, Schuver K, Marcus BH. A randomized trial examining a physical activity intervention for the prevention of postpartum depression: the healthy mom trial. Ment Health Phys Act. 2014;7:42-9.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Logsdon MC, Myers JA. Comparative performance of two depression screening instruments in adolescent mothers. J Womens Health. 2010;**19**:1123-8.	Not a sample of adults
Łukasik A, Błaszczyk K, Wojcieszyn M, Belowska A. Characteristic of affective disorders of the first week of puerperium. Ginekol Pol. 2003;**74**:1194-9.	No validated interview to assess major depression
Lundh W, Gyllang C. Use of the Edinburgh Postnatal Depression Scale in some Swedish child health care centres. Scand J Caring Sci. 1993;**7**:149-54.	No validated interview to assess major depression
Lydsdottir LB, Howard LM, Olafsdottir H, Thome M, Tyrfingsson P, Sigurdsson JF. The mental health characteristics of pregnant women with depressive symptoms identified by the Edinburgh Postnatal Depression Scale. J Clin Psychiatry. 2014;**75**:393-8.	> 2 weeks between EPDS and diagnostic interview
Mallett P, Andrew M, Hunter C, Smith J, Richards C, Othman S, et al. Cognitive function, thyroid status and postpartum depression. Acta Psychiatr Scand. 1995;**91**:243-6.	No validated interview to assess major depression
Maloney DM. Postnatal depression: a study of mothers in the metropolitan area of Perth, Western Australia. Aust J Midwifery. 1998;**11**:18-23.	Major depression not assessed
Mao HJ, Li HJ, Chiu H, Chan WC, Chen SL. Effectiveness of antenatal emotional self-management training program in prevention of postnatal depression in Chinese women. Perspect Psychiatr Care. 2012;**48**:218-24.	Sample selected for known distress, mental health

	diagnosis, or psychiatric setting
Martin-Santos R, Gelabert E, Subira S, Gutierrezzotes A, Langorh K, Jover M, et al. Research Letter: Is neuroticism a risk factor for postpartum depression? Psychol Med. 2012;**42**:1559-65.	No original data
Mason L, Poole H. Healthcare professionals' views of screening for postnatal depression. Community Pract. 2008;**81**:30-4.	No pregnant or postpartum women
Matijasevich A, Munhoz TN, Tavares BF, Barbosa AP, da Silva DM, Abitante MS, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) for screening of major depressive episode among adults from the general population. BMC Psychiatry. 2014;**14**:284.	No pregnant or postpartum women
Matthey S, Valenti B, Souter K, Ross-Hamid C. Comparison of four self-report measures and a generic mood question to screen for anxiety during pregnancy in English-speaking women J Affect Disord. 2013;**148**:347-51.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Matthey S. Differentiating between Transient and Enduring distress on the Edinburgh Depression Scale within screening contexts. J Affect Disord. 2016;**196**:252-58.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Matthey S. Using the Edinburgh Postnatal Depression Scale to screen for anxiety disorders. Depress Anxiety. 2008;**25**:926-31.	No pregnant or postpartum women
Mauri M, Banti S, Borri C, Rambelli C, Ramacciotti D, Oppo A, et al. Depressive Symptomatology in Pregnancy Detected with EPDS: the Problem of False Positive. Eur Psychiatry. 2010;**25**:1403.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Mazhari S, Nakhaee N. Validation of the Edinburgh Postnatal Depression Scale in an Iranian sample. Arch Womens Ment Health. 2007;**10**:293-7.	No validated interview to assess major depression
Mazzeo SE, SlofOp't Landt MC, Jones I, Mitchell K, Kendler KS, Neale MC, et al. Associations among postpartum depression, eating disorders, and perfectionism in a population-based sample of adult women. Int J Eat Disord. 2006;**39**:202-11.	Major depression not assessed
McMahon CA, Boivin J, Gibson FL, Hammarberg K, Wynter K, Fisher JR. Older maternal age and major depressive episodes in the first two years after birth: Findings from the Parental Age and Transition to Parenthood Australia (PATPA) study. J Affect Disord. 2015;**175**:454-62.	Major depression not assessed
Meltzer-Brody S, Zerwas S, Leserman J, Von Holle A, Regis T, Bulik C. Eating disorders and trauma history in women with perinatal depression. J Womens Health. 2011;**20**:863-70.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Meuti V, Aceti F, Giacchetti N, Carluccio GM, Zaccagni M, Marini I, et al. Perinatal depression and patterns of attachment: a critical risk factor? Depress Res Treat. 2015;**2015**:105012.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Milgrom J, Gemmill AW, Ericksen J, Burrows G, Buist A, Reece J. Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: A randomised controlled trial. Aust N Z J Psychiatry. 2015;**49**:236-245.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Miller L, Gur M, Shanok A, Weissman M. Interpersonal psychotherapy with pregnant adolescents: two pilot studies. J Child Psychol Psychiatry. 2008;**49**:733-42.	Not a sample of adults

Moayedoddin A, Moser D, Nanzer N. The impact of brief psychotherapy centred on parenthood on the anxio-depressive symptoms of mothers during the perinatal period. Swiss Med Wkly. 2013;**143**:w13769.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Murray D, Cox JL, Chapman G, Jones P. Childbirth: life event or start of a long-term difficulty? Further data from the Stoke-on-Trent controlled study of postnatal depression. Br J Psychiatry. 1995;**166**:595-600.	No validated interview to assess major depression
Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). J Reprod Infant Psychol. 1990;**8**:99-107.	No validated interview to assess major depression
Murray L, Carothers AD. The validation of the Edinburgh Post-natal Depression Scale on a community sample. Br J Psychiatry. 1990;**157**:288-90.	No validated interview to assess major depression
O'Mahen H, Himle JA, Fedock G, Henshaw E, Flynn H. A pilot randomized controlled trial of cognitive behavioral therapy for perinatal depression adapted for women with low incomes. Depress Anxiety. 2013;**30**:679-87.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
O'Neill T. Postnatal depression--aetiological factors. Ir Med J. 1990;**83**:17-18.	> 2 weeks between EPDS and diagnostic interview
Ortiz Collado MA, Saez M, Favrod J, Hatem M. Antenatal psychosomatic programming to reduce postpartum depression risk and improve childbirth outcomes: a randomized controlled trial in Spain and France. BMC Pregnancy & Childbirth. 2014;**14**:22.	Major depression not assessed
Owoeye AO, Aina OF, Morakinyo O. Risk factors of postpartum depression and EPDS scores in a group of Nigerian women. Trop Doct. 2006;**36**:100-3.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Parker G, Hegarty B, Granville-Smith I, Ho J, Paterson A, Gokiert A, Hadzi-Pavlovic D. Is essential fatty acid status in late pregnancy predictive of post-natal depression?. Acta Psychiatr Scand. 2015;**131**:148-56.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Parker GB, Hegarty B, Paterson A, Hadzi-Pavlovic D, Granville-Smith I, Gokiert A. Predictors of post-natal depression are shaped distinctly by the measure of 'depression'. J Affect Disord. 2015;**173**:239-44.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Patton GC, Romaniuk H, Spry E, Coffey C, Olsson C, Doyle LW, et al. Prediction of perinatal depression from adolescence and before conception (VIHCS): 20-year prospective cohort study. Lancet. 2015;**386**:875-83.	Major depression not assessed
Peindl KS, Wisner KL, Hanusa BH. Identifying depression in the first postpartum year: guidelines for office-based screening and referral. J Affect Disord. 2004;**80**:37-44.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Phillips J, Sharpe L, Nemeth D. Maternal psychopathology and outcomes of a residential mother-infant intervention for unsettled infant behaviour. Aust N Z J Psychiatry. 2010;**44**:280-9.	> 2 weeks between EPDS and diagnostic interview
Piacentini D, Leveni D, Primerano G, Cattaneo M, Volpi L, Biffi G, Mirabella F. Prevalence and risk factors of postnatal depression among women attending antenatal courses. Epidemiologia Psichiatr Soc. 2009;**18**:214-20.	> 2 weeks between EPDS and diagnostic interview

Pitanupong J, Liabsuetrakul T, Vittayanont A. Validation of the Thai Edinburgh Postnatal Depression Scale for screening postpartum depression. Psychiatry Res. 2007;**149**:253-9.	No validated interview to assess major depression
Pollock JI, Manaseki-Holland S, Patel V. Detection of depression in women of child-bearing age in non-Western cultures: a comparison of the Edinburgh Postnatal Depression Scale and the Self-Reporting Questionnaire-20 in Mongolia. J Affect Disord. 2006;**92**:267-71.	Not a sample of adults
Reck C, Stehle E, Reinig K, Mundt C. Maternity blues as a predictor of DSM-IV depression and anxiety disorders in the first three months postpartum. J Affect Disord. 2009;**113**:77-87.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Reck C, Struben K, Backenstrass M, Stefenelli U, Reinig K, Fuchs T, et al. Prevalence, onset and comorbidity of postpartum anxiety and depressive disorders. Acta Psychiatr Scand. 2008;**118**:459-68.	> 2 weeks between EPDS and diagnostic interview
Regmi S, Sligl W, Carter D, Grut W, Seear M. A controlled study of postpartum depression among Nepalese women: validation of the Edinburgh Postpartum Depression Scale in Kathmandu. Trop Med Int Health. 2002;**7**:378-82.	Major depression not assessed
Robakis TK, Williams KE, Crowe S, Kenna H, Gannon J, Rasgon NL. Optimistic outlook regarding maternity protects against depressive symptoms postpartum. Arch Womens Ment Health. 2015;**18**:197-208.	No validated interview to assess major depression
Roca A, Imaz ML, Torres A, Plaza A, Subira S, Valdes M, et al. Unplanned pregnancy and discontinuation of SSRIs in pregnant women with previously treated affective disorder. J Affect Disord. 2013;**150**:807-13.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Rojas G, Fritsch R, Solis J, Gonzalez M, Guajardo V, Araya R. Quality of life of women depressed in the post-partum period. Rev Med Chil. 2006;**134**:713-20.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Rubertsson C, Borjesson K, Berglund A, Josefsson A, Sydsjo G. The Swedish validation of Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. Nord J Psychiatry. 2011;**65**:414-8.	No validated interview to assess major depression
Saleh ES, El-Bahei W, El-Hadidy MA, Zayed A. Predictors of postpartum depression in a sample of Egyptian women. Neuropsychiatr Dis Treat. 2012;**9**:15-24.	EPDS not administered
Sanjuan J, MartinSantos R, GarciaEsteve L, Carot JM, Guillamat R, GutierrezZotes A, et al. Mood changes after delivery: Role of the serotonin transporter gene. Br J Psychiatry. 2008;**193**:383-8.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Santos IS, Matijasevich A, Tavares BF, Barros AJ, Botelho IP, Lapolli C, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of mothers from the 2004 Pelotas Birth Cohort Study.Cad Saude Publica. 2007;**23**:2577-88.	No validated interview to assess major depression
Santos IS, Matijasevich A, Tavares BF, da Cruz Lima AC, Riegel RE, Lopes BC. Comparing validity of Edinburgh scale and SRQ20 in screening for post-partum depression. Clin Pract Epidemiol Ment Health. 2007;**3**:18.	No validated interview to assess major depression
Savarimuthu RJ, Ezhilarasu P, Charles H, Antonisamy B, Kurian S, Jacob KS. Post-partum depression in the community: a qualitative study from rural South India. Int J Soc Psychiatry. 2010;**56**:94-102.	Could not determine eligibility[a]

Séjourné N, Alba J, Onorrus M, Goutaudier N, Chabrol H. Intergenerational transmission of postpartum depression. J Reprod Infant Psychol. 2011;**29**:115-24.	No validated interview to assess major depression
Seth S, Lewis AJ, Saffery R, Lappas M, Galbally M. Maternal prenatal mental health and placental 11 beta-HSD2 gene expression: initial findings from the Mercy Pregnancy and Emotional Wellbeing study. Int J Mol Sci. 2015;**16**:27482-96.	Major depression not assessed
Simpson W, Glazer M, Michalski N, Steiner M, Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Can J Psychiatry. 2014;**59**:434-40.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Sit DK, Flint C, Svidergol D, White J, Wimer M, Bish B, Wisner KL. Best practices: an emerging best practice model for perinatal depression care. Psychiatr Serv. 2009;**60**:1429-31.	No validated interview to assess major depression
Slade P, Morrell CJ, Rigby A, Ricci K, Spittlehouse J, Brugha TS. Postnatal women's experiences of management of depressive symptoms: a qualitative study. Br J Gen Pract. 2010;**60**:e440-e448.	Major depression not assessed
Smith-Nielsen J, Steele H, Mehlhase H, Cordes K, Steele M, Harder S, Vaever MS. Links among high EPDS scores, state of mind regarding attachment, and symptoms of personality disorder. J Pers Disord. 2015;**29**:771-93.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Sundaram S, Harman JS, Cook RL. Maternal morbidities and postpartum depression: An analysis using the 2007 and 2008 pregnancy risk assessment monitoring system. Womens Health Issues. 2014;**24**:e381-8.	EPDS not administered
Sutter-Dallay AL, Giaconne-Marcesche V, Glatigny-Dallay E, Verdoux H. Women with anxiety disorders during pregnancy are at increased risk of intense postnatal depressive symptoms: a prospective survey of the MATQUID cohort. Eur Psychiatry. 2004;**19**:459-63.	> 2 weeks between EPDS and diagnostic interview
Tam LW, Newton RP, Dern M, Parry BL. Screening women for postpartum depression at well baby visits: resistance encountered and recommendations. Arch Womens Ment Health. 2002;**5**:79-82.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Tan EC, Chua TE, Lee TMY, Tan HS, Ting JLY, Chen HY. Case-control study of glucocorticoid receptor and corticotrophin-releasing hormone receptor gene variants and risk of perinatal depression. BMC Pregnancy Childbirth. 2015;**15**:283.	Major depression not assessed
Tang Y, Shi S, Lu W, Chen Y, Wang Q, Zhu Y, et al. Prenatal psychological prevention trial on postpartum anxiety and depression. Chin Ment Health J. 2009;**23**:83-89.	Could not determine eligibility[a]
Teng HW, Hsu CS, Shih SM, Lu ML, Pan JJ, Shen WW. Screening postpartum depression with the Taiwanese version of the Edinburgh Postnatal Depression scale. Compr Psychiatry. 2005;**46**:261-65.	Could not determine eligibility[a]
Tesfaye M, Hanlon C, Wondimagegn D, Alem A. Detecting postnatal common mental disorders in Addis Ababa, Ethiopia: validation of the Edinburgh Postnatal Depression Scale and Kessler Scales. J Affect Disord. 2010;**122**:102-8.	No validated interview to assess major depression
Tharner A, Luijk MPCM, van IJzendoorn MH, BakermansKranenburg MJ, Jaddoe VWV, Hofman A, et al. Maternal lifetime history of depression and depressive symptoms in the prenatal and early postnatal period do not	> 2 weeks between EPDS and diagnostic interview

predict infant-mother attachment quality in a large, population-based Dutch cohort study. Attach Hum Dev. 2012;**14**:63-81.	
Thorpe K. A study of the use of the Edinburgh Postnatal Depression Scale with parent groups outside the postpartum period. J Reprod Infant Psychol. 1993;**11**:119-25.	No pregnant or postpartum women
Tietz A, Zietlow AL, Reck C. Maternal bonding in mothers with postpartum anxiety disorder: the crucial role of subclinical depressive symptoms and maternal avoidance behaviour. Arch Womens Ment Health. 2014;**17**:433-42.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Ueda M, Yamashita H, Yoshida K. Impact of infant health problems on postnatal depression: pilot study to evaluate a health visiting system. Psychiatry Clin Neurosci. 2006;**60**:182-9.	> 2 weeks between EPDS and diagnostic interview
Uguz F, Akman C, Sahingoz M, Kaya N, Kucur R. One year follow-up of post-partum-onset depression: the role of depressive symptom severity and personality disorders. J Psychosom Obstet Gynecol. 2009;**30**:141-5.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Uwakwe R, Okonkwo JE. Affective (depressive) morbidity in puerperal Nigerian women: validation of the Edinburgh Postnatal Depression Scale. Acta Psychiatr Scand. 2003;**107**:251-9.	No validated interview to assess major depression
Venkatesh KK, Zlotnick C, Triche EW, Ware C, Phipps MG. Accuracy of brief screening tools for identifying postpartum depression among adolescent mothers. Pediatrics. 2014;**133**:e45-e45.	Not a sample of adults
Venter MD, Smets J, Raes F, Wouters K, Franck E, Hanssens M, et al. Impact of childhood trauma on postpartum depression: A prospective study. Arch Womens Ment Health. 2016;**19**:337-42.	Major depression not assessed
Verkerk GJ, Denollet J, Van Heck GL, Van Son MJ, Pop VJ. Personality factors as determinants of depression in postpartum women: a prospective 1-year follow-up study. Psychosom Med. 2005;**67**:632-7.	No validated interview to assess major depression
Verkerk GJM, Pop VJM, Van Son MJM, Van Heck GL. Prediction of depression in the postpartum period: A longitudinal follow-up study in high-risk and low-risk women. J Affect Disord. 2003;**77**:159-66.	> 2 weeks between EPDS and diagnostic interview
Viktorin A, Meltzer-Brody S, Kuja-Halkola R, Sullivan PF, Landen M, Lichtenstein P, Magnusson PK. Heritability of perinatal depression and genetic overlap with nonperinatal depression. Am J Psychiatry. 2016;**173**:158-65.	EPDS not administered
Wang Y, Guo X, Lau Y, Chan KS, Yin L, Chen J. Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale. Int J Nurs Stud. 2009;**46**:813-23.	Could not determine eligibility[a]
Warner R, Appleby L, Whitton A, Faragher B. Attitudes toward motherhood in postnatal depression: development of the Maternal Attitudes Questionnaire. J Psychosom Res. 1997;**43**:351-8.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Warnock FF, Bakeman R, Shearer K, Misri S, Oberlander T. Caregiving behavior and interactions of prenatally depressed mothers (antidepressant-treated and non-antidepressant-treated) during newborn acute pain. Infant Ment Health J. 2009;**30**:384-406.	Could not determine eligibility[a]
Weobong B, Akpalu B, Doku V, Agyei SO, Hurt L, Kirkwood B, Prince M. The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. J Affect Disord. 2009;**113**:109-17.	No validated interview to assess major depression

Werrett J, Clifford C. Validation of the Punjabi version of the Edinburgh postnatal depression scale (EPDS). Int J Nurs Stud. 2006;**43**:227-36.	Major depression not assessed
Wickberg B, Hwang CP. Counselling of postnatal depression: a controlled study on a population based Swedish sample. J Affect Disord. 1996;**39**:209-16.	No validated interview to assess major depression
Wickberg B, Hwang CP. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. Acta Psychiatr Scand. 1996;**94**:181-84.	No validated interview to assess major depression
Wu M, Li X, Feng B, Wu H, Qiu C, Zhang W. Correlation between sleep quality of third-trimester pregnancy and postpartum depression. Med Sci Monit. 2014;**20**:2740-5.	Could not determine eligibility[a]
Yamashita H, Yoshida K, Nakano H, Tashiro N. Postnatal depression in Japanese women. Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. J Affect Disord. 2000;**58**:145-54.	No validated interview to assess major depression
Yonkers KA, Ramin SM, Rush AJ, Navarrete CA, Carmody T, March D, et al. Onset and persistence of postpartum depression in an inner-city maternal health clinic system. Am J Psychiatry. 2001;**158**:1856-63.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Yoshida K, Yamashita H, Ueda M, Tashiro N. Postnatal depression in Japanese mothers and the reconsideration of 'Satogaeri bunben'. Pediatr Int. 2001;**43**:189-93.	No validated interview to assess major depression
Zammit S, Thomas K, Thompson A, Horwood J, Menezes P, Gunnell D, et al. Maternal tobacco, cannabis and alcohol use during pregnancy and risk of adolescent psychotic symptoms in offspring. Br J Psychiatry. 2009;**195**:294-300.	No pregnant or postpartum women
Zelkowitz P, Milet TH. Postpartum psychiatric disorders: Their relationship to psychological adjustment and marital satisfaction in the spouses. J Abnorm Psychol. 1996;**105**:281-5.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Zlotnick C, Capezza NM, Parker D. An interpersonally based intervention for low-income pregnant women with intimate partner violence: A pilot study. Arch Womens Ment Health. 2011;**14**:55-65.	Sample selected for known distress, mental health diagnosis, or psychiatric setting
Zubaran C, Foresti K, Schumacher MV, Amoretti AL, Thorell MR, Muller LC. The correlation between postpartum depression and health status. Matern & Child Health J. 2010;**14**:751-7.	> 2 weeks between EPDS and diagnostic interview

Supplementary Table 2a. Characteristics of eligible primary studies that did not provide primary data for the main IPDMA of the PHQ-9 (N=14)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Total N	Major depression N (%)
Becker, 2002[1]	Saudi Arabia	Primary care patients	SCID	DSM-III-R	173	NR
Chen, 2013[2]	China	Primary care populations	SCID	DSM-IV	280	NR[a]
Chen, 2012[3]	China	Adults over 60 in primary care	SCID	DSM-IV	262	97 (37)
Lai, 2010[4]	Hong Kong	Men with postpartum wives	SCID	DSM-IV	551	8 (1)
Navinés, 2012[5]	Spain	Chronic hepatitis C patients	SCID	DSM-IV	104	21 (20)
Phelan, 2010[6]	USA	Elderly primary care patients	SCID	DSM-IV	69	8 (12)
Thompson, 2011[7]	USA	Parkinson's patients	SCID	DSM-IV	214	30 (14)
Watnick, 2005[8]	USA	Long term dialysis patients	SCID	DSM-IV	62	12 (19)
Al-Ghafri, 2014[9]	Oman	Medical trainees	CIDI	NR	131	NR[a]
Haddad, 2013[10]	UK	Coronary heart disease patients	CIS-R	ICD-10	730	32 (4)
Persoons, 2003[11]	Belgium	Otorhinolaryngology outpatients	MINI	DSM-IV	97	16 (16)
Rathore, 2014[12]	USA	Adults with epilepsy	MINI	DSM-IV	172	33 (19)
Scott, 2011[13]	USA	Chronic hepatitis C patients	MINI	DSM-IV and ICD-10	30	NR[a]
Wang, 2014[14]	China	General population	MINI	DSM-IV	1045	28 (3)

Abbreviations: CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule Revised; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; IPDMA: Individual Participant Data Meta-Analysis; MINI: Mini International Neuropsychiatric Interview; NR: Not Reported; PHQ-9: Patient

Health Questionnaire-9; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America.
[a]Reported numbers implausible

Supplementary Table 2b. Characteristics of eligible primary studies that did not provide primary data for the main IPDMA of the EPDS (N=24)

First author, year	Country	Recruited population	Diagnostic interview	Total N	N (%) Major depression
Adewuya, 2006[15]	Nigeria	Pregnancy, 32-36 weeks	MINI	86	9 (10)
Adouard, 2005[16]	France	Pregnancy, 28-34 weeks	MINI	60	15 (25)
Agoub, 2005[17]	Morocco	Postpartum, 2-3 weeks	MINI	144	27 (19)
Aydin, 2004[18]	Turkey	Postpartum, within first year.	SCID	341	28 (8)
Banti, 2011[19]	Italy	Pregnancy, 3 months.	SCID	1066	NR
Barnett, 1999[20]	Australia	Postpartum, 6 weeks	DIS	316	21 (7)
Benvenuti, 1999[21]	Italy	Postpartum, 8-12 weeks	MINI	113	18 (16)
Bergink, 2011[22]	The Netherlands	Pregnancy, 12 weeks	CIDI	845	47 (6)
Berle, 2003[23]	Norway	Postpartum, 6-12 weeks	MINI	100	27 (27)
Brodey, 2016[24]	USA	Pregnant/Postpartum mixed sample. Postpartum sample was 0-150 days.	SCID	879	NR
Chibanda, 2010[25]	Zimbabwe	Postpartum, 6 weeks	SCID	210	NR
Christl, 2013[26]	Australia	Postpartum, between 0-12 weeks	MINI	232	13 (6)
Crotty, 2004[27]	Ireland	Postpartum, 6 weeks	SCAN	113	NR

Gausia, 2007[28]	Bangladesh	Postpartum, 6-8 weeks	SCID	100	3 (3)
Gorman, 2004[29]	France, Ireland, Italy, USA, UK, Portugal, Austria, Switzerland	Pregnancy, second or third semester	SCID	289	10 (4)
Li, 2011[30]	China	Postpartum, 2-12 weeks.	SCID	387	24 (6)
Mahmud, 2003[31]	Malaysia	Postpartum, 4-12 weeks	CIDI	64	9 (14)
Matthey, 2001[32]	Australia	Postpartum, 6-7 weeks	DIS	230	11 (5)
Moses-Kolko, 2012[33]	USA	Postpartum, 0-16 weeks	SCID	33	13 (39)
O'Brien, 2004[34]	UK	Postpartum, ≤2 years	CIS-R	216	31 (14)
Pedersen, 2016[35]	USA	Pregnancy, 35-36 weeks	MINI	199	NR
Pinheiro 2013[36]	Brazil	Postpartum, 45-90 days	MINI	207	27 (13)
Priest, 2003[37]	Australia	Postpartum, 2 months	SADS	292	NR
Stuebe, 2013[38]	USA	Pregnancy, 3rd trimester	SCID	47	8 (17)

Abbreviations: CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule – Revised; DIS: Diagnostic Interview Schedule; IPDMA: Individual Participant Data Meta-Analysis; MINI: Mini International Neurospsychiatric Interview; NR: Not Reported; SADS: Schedule for Affective Disorders and Schizophrenia; SCAN: Schedule for Clinical Assessment in Neuropsychiatry; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America

Supplementary Table 3a. Characteristics of primary studies that were excluded for the present study because they were unpublished or did not publish accuracy estimates for any cutoff for PHQ-9 (N=14)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Total N	Major depression N (%)
Study excluded because it was not published						
Turner, Unpublished	Australia	Cardiac rehabilitation patients	SCID	DSM-IV	51	4 (8)
Studies excluded because they did not publish accuracy estimates for any cutoff for PHQ-9						
Ayalon, 2010[39]	Israel	Elderly primary care patients	SCID	DSM-IV	151	6 (4)
Beraldi, 2014[40]	Germany	Cancer inpatients	SCID	DSM-IV	116	7 (6)
Eack, 2006[41]	USA	Women seeking psychiatric services for their children at two mental health centers	SCID	DSM-IV	48	12 (25)
Henkel, 2004[42]	Germany	Primary care patients	CIDI	ICD-10	430	43 (10)
Hides, 2007[43]	Australia	Injection drug users accessing a needle and syringe program	MINI	DSM-IV	103	47 (46)
Hobfoll, 2011[44]	Israel	Jewish and Palestinian residents of Jerusalem exposed to war	CIDI	DSM-IV	144	42 (29)
Hyphantis, 2014[45]	Greece	Patients with chronic illnesses presenting at the emergency department	MINI	DSM-IV	349	95 (27)
Kwan, 2012[46]	Singapore	Post-stroke inpatients undergoing rehabilitation	SCID	DSM-IV-TR	113	3 (3)
Muramatsu, 2007[47]	Japan	Primary care patients	MINI	DSM-IV	116	32 (28)

Persoons, 2001[48]	Belgium	Inpatients and patients at gastroenterological and hepatology wards	MINI	DSM-IV	173	28 (16)
Picardi, 2005[49]	Italy	Inpatients with skin diseases	SCID	DSM-IV	138	12 (9)
Razykov, 2013[50]	Canada	Patients with systemic sclerosis	CIDI	DSM-IV	345	13 (4)
Simning, 2012[51]	USA	Older adults living in public housing	SCID	DSM-IV	190	10 (5)

Abbreviations: CIDI: Composite International Diagnostic Interview; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; MINI: Mini Neurospsychiatric Diagnostic Interview; PHQ-9: Patient Health Questionnaire-9; SCID: Structured Clinical Interview for DSM Disorders; USA: United States of America.

Supplementary Table 3b. Characteristics of primary studies that were excluded for the present study because they did not publish accuracy estimates for any cutoff for EPDS (N=21)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Total N	Major depression N (%)
Aceti, 2012[52]	Italy	Pregnant women in the third trimester	SCID	DSM-IV	44	22 (50)
Alvarado-Esquivel, 2016[53]	Mexico	Pregnant women recruited at a public hospital in Durango City, Mexico	MINI	DSM-IV	184	12 (7)
Barnes, 2009[54]	UK	Socially disadvantaged mothers at 2 months postpartum	SCID	DSM-III-R	347	25 (7)
Bavle, 2016[55]	India	Pregnant women recruited from an outpatient obstetrics department in a tertiary care hospital	SCID	DSM-IV	318	6 (2)
Comasco, 2016[56]	Sweden	Pregnant women	MINI	DSM-IV	419	34 (8)
Eapen, 2013[57]	Australia	Women attending an antenatal clinic in Sydney	MINI	DSM-IV	131	26 (20)
Felice, 2004[58]	Malta	Pregnant women attending an antenatal clinic	CIS-R	ICD-10	443	51 (12)
Giardinelli, 2012[59]	Italy	Women between 28 and 32 weeks pregnant recruited from a obstetric course in Florence	SCID	DSM-IV	588	28 (5)
Helle, 2015[60]	Germany	Mothers with very low birthweight and normal weight infants between 4 and 6 weeks postpartum	SCID	DSM-IV	224	12 (5)
Hickey, 1997[61]	Australia	Postpartum women recruited in the hospital after delivery	SCID	DSM-III-R	72	31 (43)
Howard, 2018[62]	UK	Pregnant women recruited from an inner-city London maternity service	SCID	DSM-IV	527	130 (25)
Imbula, 2012[63]	Democratic Republic of Congo	Women between 1 and 10 months postpartum recruited from 'well-baby' clinics	MINI	DSM-IV-TR	117	29 (25)

Prenoveau, 2013[64]	UK	Postpartum women at 10 months recruited from mixed health centres.	SCID	DSM-IV	579	69 (9)
Robertson-Blackmore, 2013[65]	USA	Women at 18 weeks gestation	SCID	DSM-IV-TR	864	65 (8)
Roomruangwong, 2016[66]	Thailand	Pregnant women at the end of their term	MINI	DSM-IV-TR	348	5 (1)
Rowe, 2008[67]	Australia	English speaking women admitted with their up to 1-year-old infants to private parenting centers	CIDI	DSM-IV	137	25 (18)
Siu, 2012[68]	China	Postpartum women	SCID	DSM-IV	805	126 (16)
Turner, 2009[69]	Italy	Women from a regional epilepsy center in Italy between 5 and 8 weeks postpartum	SCID	DSM-IV-TR	54	5 (9)
Usuda, 2016[70]	Japan	Pregnant women between 12 and 24 weeks of gestation recruited at maternity hospital in Japan	MINI	DSM-IV	177	2 (1)
Yonkers, 2014[71]	USA	Women at 17 weeks gestation	CIDI	DSM-IV	7303	267 (6)
Fisher, 2010[72]	Australia	Postpartum women recruited in Australian maternal and child health centres at 6 months postpartum	CIDI	DSM-IV	192	1 (1)

Abbreviations: CIDI: Composite International Diagnostic Interview; DSM: Diagnostic and Statistical Manual of Mental Disorders; EPDS: Edinburgh Postnatal Depression Scale; ICD: International Classification of Diseases; MINI: Mini Neurospsychiatric Diagnostic Interview; SCID: Structured Clinical Interview for DSM Disorders; USA: United States of America.

[a]The published paper was not identified in our database search, dataset later provided by the author

Supplementary Table 4a. Characteristics of primary studies that were excluded in the present study because the difference in sample size or MD cases between IPDMA dataset and published data was >10% for PHQ-9 and because eligibility could not be determined (N=14)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Total N in IPDMA dataset	Total N published	Percentage difference in N	MD N IPDMA	MD N published	Percentage difference in MD
Fiest, 2014[73]	Canada	Epilepsy outpatients	SCID	DSM-IV	169	185	9	23	27	15
Fischer, 2014[74]	Germany	Heart failure patients	SCID	DSM-IV	194	194	0	11	27	59
Hahn, 2006[75]	Germany	Patients with chronic illnesses from rehabilitation centers	CIDI	DSM-IV	211	204	3	18	35	49
Kiely, 2014[76]	Australia	Community sample of adults	CIDI	ICD-10	822	886	7	33	62	47
Lamers, 2008[77]	The Netherlands	Elderly primary care patients with diabetes mellitus or chronic obstructive pulmonary disease	MINI	DSM-IV	104	620	83	-	-	-
McGuire, 2013[78]	USA	Acute coronary syndrome inpatients	DISH	DSM-IV	100	101	1	9	23	61
Osório, 2012[79]	Brazil	Inpatients from various clinical wards	SCID	DSM-IV	86	100	14	-	-	-
Patel, 2008[80]	India	Primary care patients	CIS-R	ICD-10	299	299	0	13	51[b]	75
Santos, 2013[81]	Brazil	General population	MINI	DSM-IV	196	447	57	-	-	-
Sidebottom, 2012[82]	USA	Pregnant women	SCID	DSM-IV	246	745	67	-	-	-
Williams, 2012[83]	USA	Parkinson's Disease patients	SCID	DSM-IV	235	229	3	61	78	22
Wittkampf, 2009[84]	The Netherlands	Primary care patients at risk for depression	SCID	DSM-IV	260	440	41	-	-	-

Zhang, 2013[85]	China	Type 2 diabetes patients	MINI	DSM-IV	68	99	31	-	-	-
Studies excluded because eligibility could not be determined										
Azah, 2005[86]	Malaysia	Adults attending family medicine clinics	CIDI	ICD-10	180	180	0	30 (17)	NA	NA

Abbreviations: CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule Revised; DISH: Depression Interview and Structured Hamilton; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; ; IPDMA: Individual Participants Data Meta-analysis; MD: Major Depression; MINI: Mini Neurospsychiatric Diagnostic Interview; PHQ-9: Patient Health Questionnaire-9; SCID: Structured Clinical Interview for DSM Disorders; USA: United States of America.

[a]Was unpublished at the time of electronic database search

[b]MD cases back calculated using sample size and diagnostic accuracy information

Supplementary Table 4b. Characteristics of primary studies that were excluded in the present study because the difference in sample size or MD cases between the IPDMA dataset and published dataset was >10% for EPDS (N=9)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Total N in IPDMA dataset	Total N published	Percentage difference in N	MD N IPDMA	MD N published	Percentage difference in MD
Alvarado-Esquivel, 2006[87]	Mexico	Women within 3 months postpartum	MINI	DSM-IV	42 49	49 51	14 4	- 6	- 7	- 14
de Figueiredo, 2015[88]	Brazil	Postpartum women enrolled in prenatal care outpatient services in a Brazilian city	SCID	DSM-IV	241	199	21	-	-	-
Fernandes, 2011[89]	India	Rural women in their third trimester	MINI	DSM-IV	133	194	31	-	-	-
Figueira, 2009[90]	Brazil	Postpartum mothers recruited from hospitalization records	MINI	DSM-IV	239	245	2	18	66	72
Leonardou, 2009[91]	Greece	Postpartum women recruited from private and public maternity wards on their second day postpartum	SCID	DSM-III-R	81	81	0	4	10	60
Navarro, 2007[92]	Spain	Women presenting for postpartum care at 6 weeks	SCID	DSM-IV	401	405	1	84	180	53
Stewart, 2013[93]	Malawi	Pregnant women attending an antenatal clinic in rural Malawi	SCID	DSM-IV	186	92	102	-	-	-
Tendais, 2014[94]	Portugal	Pregnant women recruited in an obstetrics outpatient unit	SCID	DSM-IV	141 94	148 99	5 5	18 9	38 14	53 36
Tran, 2011[95]	Vietnam	Pregnant and postpartum	SCID	DSM-IV	359	364	1	52	109	52

											Vietnamese women recruited from the commune health centre

Abbreviations: DSM: Diagnostic and Statistical Manual of Mental Disorders; EPDS: Edinburgh Postnatal Depression Scale; ICD: International Classification of Diseases; ; IPDMA: Individual Participants Data Meta-analysis; MD: Major Depression; MINI: Mini Neurospsychiatric Diagnostic Interview; SCID: Structured Clinical Interview for DSM Disorders; SS: Sample Size.

Supplementary Table 5a. Characteristics of primary studies for PHQ-9 included in the present study (N=30)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Published cutoffs	Total N in IPDMA	MD N IPDMA (%)	Total N published	MD N published (%)
Akena, 2013[96]	Uganda	HIV/AIDS patients	MINI	DSM-IV	8-13	91	11 (12)	92	11[d] (12)
Amoozegar, 2017[a97]	Canada	Migraine patients	SCID	DSM-IV	10-15	203	49 (24)	208	52 (25)
Arroll, 2010[98]	New Zealand	Primary care patients	CIDI	DSM-IV	8,10,12,15	2528	156 (6)	2642	163 (6)
Bombardier, 2012[99]	USA	Inpatients with spinal cord injuries	SCID	DSM-IV	9-12	134	14 (10)	142	14 (10)
Chagas, 2013[100]	Brazil	Outpatients with Parkinson's Disease	SCID	DSM-IV	8-11	84	19 (23)	84	19[d] (23)
Cholera, 2014[101]	South Africa	Patients undergoing routine HIV counseling and testing at a primary health care clinic	MINI	DSM-IV	8,10,12	397	47 (12)	397	47 (12)
de Man-van Ginkel, 2012[102]	The Netherlands	Stroke patients	CIDI	DSM-IV	10	164	20[b] (12)	164	20 (12)
Delgadillo, 2011[103]	UK	Outpatients in drug addiction treatment	CIS-R	ICD-10	12	103	51 (50)	103	51 (50)
Fann, 2005[104]	USA	Inpatients with traumatic brain injury	SCID	DSM-IV	10,12	135[c]	22[c] (17)	135	23 (17)
Gelaye, 2014[105]	Ethiopia	Outpatients at a general hospital	CIDI	DSM-IV	9,10,11	923	162 (18)	926	162 (17)
Gjerdingen, 2009[106]	USA	Mothers registering their newborns for well-child visits at medical or pediatric clinics	SCID	DSM-IV	10	419	19 (5)	438	20 (5)
Hyphantis, 2011[107]	Greece	Patients with various rheumatologic disorders	MINI	DSM-IV	4-16	213	69 (32)	213	69 (32)

Inagaki, 2013[108]	Japan	Internal medicine outpatients	MINI	DSM-III-R	4-13	104	21 (20)	104	21
Khamseh, 2011[109]	Iran	Type 2 diabetes patients	SCID	DSM-IV	13	184	79 (43)	185	80 (43)
Lambert, 2015[a110]	Australia	Cancer patients	SCID	DSM-IV	5,9,10,15,20	147	21 (14)	148	21[d] (14)
Liu, 2011[111]	Taiwan	Primary care patients	SCAN	DSM-IV	9-11	1532	50 (3)	1532	50 (50)
Lotrakul, 2008[112]	Thailand	Outpatients	MINI	DSM-IV	6-15	278	19 (7)	279	19 (7)
Lowe, 2004[113]	Germany	Medical and psychosomatic outpatients	SCID	DSM-IV	11-13	494	67 (14)	501	66 (13)
Mohd Sidik, 2012[114]	Malaysia	Primary care patients	CIDI	DSM-IV	10	146	31 (21)	146	31 (21)
Osório, 2009[115]	Brazil	Women in primary care	SCID	DSM-IV	10-21	177	60 (34)	177	60 (34)
Pence, 2012[116]	Cameroon	HIV-infected patients	CIDI	DSM-IV	8,10,12	398	11 (3)	398	11 (3)
Richardson, 2010[117]	USA	Older adults undergoing in-home aging services care management assessment	SCID	DSM-IV	7-12	377	95 (25)	378	101 (27)
Rooney, 2013[118]	UK	Adults with cerebral glioma	SCID	DSM-IV	8-11	126	14 (11)	129	15[d] (12)
Stafford, 2007[119]	Australia	Inpatients with coronary artery disease who had undergone surgery	MINI	DSM-IV	5-6	193	35 (18)	193	35 (18)
Sung, 2013[120]	Singapore	Primary care patients	MINI	DSM-IV	6	399	12 (3)	400	12 (3)
Thombs, 2008[121]	USA	Outpatients with coronary artery disease	C-DIS	DSM-IV	1-10	1006	221 (22)	1024	224 (22)
Turner, 2012[122]	Australia	Stroke patients	SCID	DSM-IV	7,9,10	72	13 (18)	72	13 (18)
Twist, 2013[123]	UK	Type 2 diabetes outpatients	SCAN	DSM-IV	10-14	360	80 (22)	368	84 (23)

van Steenbergen-Weijenburg, 2010[124]	The Netherlands	Diabetes patients	MINI	DSM-IV	8-12	196	37 (19)	197	37 (19)
Vöhringer, 2013[125]	Chile	Primary care patients	SCID	DSM-IV	10	190	59 (31)	197	59 (30)

Abbreviations: C-DIS: Computerized Diagnostic Interview Schedule; CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule Revised; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; IPDMA: Individual Participants Data Meta-analysis; MD: Major Depression; MINI: Mini Neurospsychiatric Diagnostic Interview; PHQ-9: Patient Health Questionnaire-9; SCAN: Schedules for Clinical Assessment in Neuropsychiatry; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America.

[a]Was unpublished at the time of electronic database search

[b]MD cases aggregate of those diagnosed with ICD MDE, DSM MDD or both

[c]The value obtained after applying weight

[d]MD cases back calculated using sample size and diagnostic accuracy information

[e]Included only the participants that were administered index test and reference standard within a week

Supplementary Table 5b. Characteristics of primary studies for EPDS included in the present study (N=19)

First author, year	Country	Recruited population	Diagnostic interview	Classification system	Published cutoffs	Total N IPDMA	MD N IPDMA (%)	Total N published	MD N published (%)
Alvarado, 2015[126]	Chile	Pregnant women up to 28 weeks gestation	MINI	DSM-IV	7-16	111	38 (34)	111	38 (34)
Töreki, 2014[127]	Hungary	Women between 6 and 8 weeks postpartum	SCID	DSM-IV	5-16	265	8 (3)	266	8 (3)
Couto, 2015[128]	Brazil	Women in their second trimester of pregnancy recruited at antenatal care in a public hospital	MINI	DSM-IV-TR	8-14	173	36 (21)	188	33 (18)
Bakare, 2014[129]	Nigeria	Postpartum women	MINI	DSM-IV	9	405	62 (15)	408	62 (15)
Rochat, 2013[130]	South Africa	Women recruited from their antenatal appointment at a primary health care clinic between 26 and 34 weeks of pregnancy	SCID	DSM-IV	13	104	50 (48)	109	51 (47)
Töreki, 2013[131]	Hungary	Women at 12 weeks antenatal	SCID	DSM-IV	5-14	219	7 (3)	219	7 (3)
Thiagayson, 2013[132]	Singapore	Inpatient high-risk pregnant women at 23 or more weeks of gestation	MINI	DSM-IV	7-12	200	22 (11)	200	22 (11)
Tandon, 2012[133]	USA	Pregnant and postpartum women enrolled in home visitation programs	SCID	DSM IV	11,13	89	25 (28)	95	27(28)
Chaudron, 2010[134]	USA	Postpartum women recruited from Well-Child Care visits with infants 0-14 months of age	SCID	DSM-IV	9,13	187	70 (37)	198	73 (37)
Bunevicius, 2009[135]	Lithuania	Pregnant women 12 to 16 weeks pregnant attending an obstetric clinic	SCID	DSM-III-R	9-15	230	12 (5)	230	12 (5)
Phillips, 2009[136]	Australia	Postpartum mothers with unsettled infants	SCID	DSM-IV	11,12,13	158	42 (27)	166	42 (25)

Pawlby, 2008[137]	UK	Postpartum women at 3 months	CIS	ICD-9	13	144	31 (22)	147	34 (23)
Su, 2007[138]	Taiwan	Women in their second and third trimesters	MINI	DSM-IV	13	185	23 (12)	185	23 (12)
Garcia-Esteve, 2003[139]	Spain	Women at 6 weeks postpartum	SCID	DSM-III-R	8-15	334	36 (11)	334	36 (11)
Vega-Dienstmaier, 2002[140]	Peru	Women up to 12 months postpartum	SCID	DSM-IV	1-26	306	19 (6)	321	19 (6)
Beck, 2001[141]	USA	Postpartum mothers	SCID	DSM-IV	13	150	18 (12)	150	18 (12)
Nakić Radoš, 2013[142]	Croatia	Women between 6 and 8 weeks postpartum	SCID	DSM-IV-TR	7-14	272	10 (4)	272	10 (4)
Tissot, 2015[143]	Switzerland	Women at 3 months postpartum	DIGS	DSM-IV	9-13	65	4 (6)	65	4 (6)
Khalifa, 2015[144]	Sudan	Women at 3 months postpartum	MINI	ICD-10	1-15	40	18 (45)	40	18 (45)

Abbreviations: CIS: Clinical Interview Schedule; DIGS: Diagnostic Interview of Genetic Studies; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; MINI: Mini Neuropsychiatric Diagnostic Interview; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America

[a]Data from only one time point was included

SUPPLEMENTAY FILE REFERENCES

(1) Becker S, Al Zaid K, Al Faris E. Screening for somatization and depression in Saudi Arabia: a validation study of the PHQ in primary care. *Int J Psychiatry Med* 2002; **32:** 271-83.

(2) Chen S, Fang Y, Chiu H, Fan H, Jin T, Conwell Y. Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. *Asia Pac Psychiatry* 2013; **5:** 61-8.

(3) Chen S, Conwell Y, Vanorden K, et al. Prevalence and natural course of late-life depression in China primary care: a population based study from an urban community. *J Affect Disord* 2012; **141:** 86-93.

(4) Lai BP, Tang AK, Lee DT, Yip AS, Chung TK. Detecting postnatal depression in Chinese men: a comparison of three instruments. *Psychiatry Res* 2010; **180:** 80-5.

(5) Navines R, Castellvi P, Moreno-Espana J, et al. Depressive and anxiety disorders in chronic hepatitis C patients: reliability and validity of the Patient Health Questionnaire. *J Affect Disord* 2012; **138:** 343-51.

(6) Phelan E, Williams B, Meeker K, et al. A study of the diagnostic accuracy of the PHQ-9 in primary care elderly. *BMC Fam Pract* 2010; **11:** 63-2296.

(7) Thompson AW, Liu H, Hays RD, et al. Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. *Parkinsonism Relat Disord* 2011; **17:** 40-5.

(8) Watnick S, Wang PL, Demadura T, Ganzini L. Validation of 2 depression screening tools in dialysis patients. *Am J Kidney Dis* 2005; **46:** 919-24.

(9) Al-Ghafri G, Al-Sinawi H, Al-Muniri A, et al. Prevalence of depressive symptoms as elicited by Patient Health Questionnaire (PHQ-9) among medical trainees in Oman. *Asian J Psychiatr* 2014; **8:** 59-62.

(10) Haddad M, Walters P, Phillips R, et al. Detecting depression in patients with coronary heart disease: a diagnostic evaluation of the PHQ-9 and HADS-D in primary care, findings from the UPBEAT-UK study. *PLoS One* 2013; **8:** e78493.

(11) Persoons P, Luyckx K, Desloovere C, Vandenberghe J, Fischler B. Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. *Gen Hosp Psychiatry* 2003; **25:** 316-23.

(12) Rathore JS, Jehi LE, Fan Y, et al. Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy. *Epilepsy Behav* 2014; **37:** 215-20.

(13) Scott JD, Wang CC, Coppel E, Lau A, Veitengruber J, Roy-Byrne P. Diagnosis of depression in former injection drug users with chronic hepatitis C. *J Clin Gastroenterol* 2011; **45:** 462-7.

(14) Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry* 2014; **36:** 539-44.

(15) Adewuya AO, Ola BA, Dada AO, Fasoto OO. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for depression in late pregnancy among Nigerian women. *J Psychosom Obstet Gynaecol* 2006; **27:** 267-72.

(16) Adouard F, Glangeaud-Freudenthal NM, Golse B. Validation of the Edinburgh postnatal depression scale (EPDS) in a sample of women with high-risk pregnancies in France. *Arch Womens Ment Health* 2005; **8:** 89-95.

(17) Agoub M, Moussaoui D, Battas O. Prevalence of postpartum depression in a Moroccan sample. *Arch Womens Ment Health* 2005; **8:** 37-43.

(18) Aydin N, Inandi T, Yigit A, Hodoglugil NN. Validation of the Turkish version of the Edinburgh Postnatal Depression Scale among women within their first postpartum year. *Soc Psychiatry Psychiatr Epidemiol* 2004; **39:** 483-6.

(19) Banti S, Mauri M, Oppo A, et al. From the third month of pregnancy to 1 year postpartum. Prevalence, incidence, recurrence, and new onset of depression. Results from the perinatal depression-research & screening unit study. *Compr Psychiatry* 2011; **52:** 343-51.

(20) Barnett B, Matthey S, Gyaneshwar R. Screening for postnatal depression in women of non-English speaking background. *Arch Womens Ment Health* 1999; **2:** 67-74.

(21) Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. *J Affect Disord* 1999; **53:** 137-41.

(22) Bergink V, Kooistra L, Lambregtse-van den Berg, M P, et al. Validation of the Edinburgh Depression Scale during pregnancy. *J Psychosom Res* 2011; **70:** 385-9.

(23) Berle JO, Aarre TF, Mykletun A, Dahl AA, Holsten F. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. *J Affect Disord* 2003; **76:** 151-6.

(24) Brodey BB, Goodman SH, Baldasaro RE, et al. Development of the Perinatal Depression Inventory (PDI)-14 using item response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. *Arch Womens Ment Health* 2016; **19:** 307-16.

(25) Chibanda D, Mangezi W, Tshimanga M, et al. Validation of the Edinburgh Postnatal Depression Scale among women in a high HIV prevalence area in urban Zimbabwe. *Arch Womens Ment Health* 2010; **13:** 201-6.

(26) Christl B, Reilly N, Smith M, Sims D, Chavasse F, Austin MP. The mental health of mothers of unsettled infants: is there value in routine psychosocial assessment in this context? *Arch Womens Ment Health* 2013; **16:** 391-9.

(27) Crotty F, Sheehan J. Prevalence and detection of postnatal depression in an Irish community sample. *Ir J Psychol Med* 2004; **21:** 117-21.

(28) Gausia K, Fisher C, Algin S, Oosthuizen J. Validation of the Bangla version of the Edinburgh Postnatal Depression Scale for a Bangladeshi sample. *J Reprod Infant Psychol* 2007; **25:** 308-15.

(29) Gorman LL, O'Hara MW, Figueiredo B, et al. Adaptation of the structured clinical interview for DSM-IV disorders for assessing depression in women during pregnancy and post-partum across countries and cultures. *Br J Psychiatry Suppl* 2004; **46:** s17-23.

(30) Li L, Liu F, Zhang H, Wang L, Chen X. Chinese version of the Postpartum Depression Screening Scale: translation and validation. *Nurs Res* 2011; **60:** 231-9.

(31) Mahmud WM, Awang A, Mohamed MN. Revalidation of the Malay Version of the Edinburgh Postnatal Depression Scale (EPDS) Among Malay Postpartum Women Attending the Bakar Bata Health Center in Alor Setar, Kedah, North West Of Peninsular Malaysia. *Malays J Med Sci* 2003; **10:** 71-5.

(32) Matthey S, Barnett B, Kavanagh DJ, Howie P. Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. *J Affect Disord* 2001; **64:** 175-84.

(33) Moses-Kolko EL, Price JC, Wisner KL, et al. Postpartum and depression status are associated with lower [[^{11}C]raclopride BP(ND) in reproductive-age women. *Neuropsychopharmacology* 2012; **37:** 1422-32.

(34) O'Brien LM, Heycock EG, Hanna M, Jones PW, Cox JL. Postnatal depression and faltering growth: a community study. *Pediatrics* 2004; **113:** 1242-7.

(35) Pedersen C, Leserman J, Garcia N, Stansbury M, Meltzer-Brody S, Johnson J. Late pregnancy thyroid-binding globulin predicts perinatal depression. *Psychoneuroendocrinology* 2016; **65:** 84-93.

(36) Pinheiro RT, Coelho FM, Silva RA, et al. Association of a serotonin transporter gene polymorphism (5-HTTLPR) and stressful life events with postpartum depressive symptoms: a population-based study. *J Psychosom Obstet Gynaecol* 2013; **34:** 29-33.

(37) Priest SR, Henderson J, Evans SF, Hagan R. Stress debriefing after childbirth: a randomised controlled trial. *Med J Aust* 2003; **178:** 542-5.

(38) Stuebe AM, Grewen K, Meltzer-Brody S. Association between maternal mood and oxytocin response to breastfeeding. *J Womens Health (Larchmt)* 2013; **22:** 352-61.

(39) Ayalon L, Goldfracht M, Bech P. 'Do you think you suffer from depression?' Reevaluating the use of a single item question for the screening of depression in older primary care patients. *Int J Geriatr Psychiatry* 2010; **25:** 497-502.

(40) Beraldi A, Baklayan A, Hoster E, Hiddemann W, Heuner P. Which questionnaire is most suitable for the detection of depressive disorders in haemato-oncological patients? Comparison between HADS, CES-D and PHQ-9. *Oncol Res Treat* 2014; **37:** 108-9.

(41) Eack SM, Greeno CG, Lee BJ. Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected. *Res Soc Work Pract* 2006; **16:** 625-31.

(42) Henkel V, Mergl R, Kohnen R, Allgaier AK, Moller HJ, Hegerl U. Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. *Gen Hosp Psychiatry* 2004; **26:** 190-8.

(43) Hides L, Lubman DI, Devlin H, et al. Reliability and validity of the Kessler 10 and Patient Health Questionnaire among injecting drug users. *Aust N Z J Psychiatry* 2007; **41:** 166-8.

(44) Hobfoll SE, Canetti D, Hall BJ, et al. Are community studies of psychological trauma's impact accurate? A study among Jews and Palestinians. *Psychol Assess* 2011; **23:** 599-605.

(45) Hyphantis T, Kroenke K, Papatheodorou E, et al. Validity of the Greek version of the PHQ 15-item Somatic Symptom Severity Scale in patients with chronic medical conditions and correlations with emergency department use and illness perceptions. *Compr Psychiatry* 2014; **55:** 1950-9.

(46) Kwan Y, Tham WY, Ang A. Validity of the Patient Health Questionnaire-9 (PHQ-9) in the Screening of Post-Stroke Depression in a Multi-Ethnic Population. *Biol Psychiatry* 2012; **71:** 141S.

(47) Muramatsu K, Miyaoka H, Kamijima K, et al. The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. *Psychol Rep* 2007; **101:** 952-60.

(48) Persoons P, Luyckx K, Fischler B. Psychiatric diagnoses in gastroenterolgy: Validation of a self-report instrument (PRIME-MD Patient Health Questionnaire), epidemiology and recognition. *Gastroenterology* 2001; **120:** A114.

(49) Picardi A, Adler DA, Abeni D, et al. Screening for depressive disorders in patients with skin diseases: a comparison of three screeners. *Acta Derm Venereol* 2005; **85:** 414-9.

(50) Razykov I, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research Group. Utility of the Patient Health Questionnaire-9 to assess suicide risk in patients with systemic sclerosis. *Arthritis Care Res (Hoboken)* 2013; **65:** 753-8.

(51) Simning A, van Wijngaarden E, Fisher SG, Richardson TM, Conwell Y. Mental healthcare need and service utilization in older adults living in public housing. *Am J Geriatr Psychiatry* 2012; **20:** 441-51.

(52) Aceti F, Aveni F, Baglioni V, et al. Perinatal and postpartum depression: from attachment to personality. A pilot study. *J Psychopathology* 2012; **18:** 328-34.

(53) Alvarado-Esquivel C, Sifuentes-Alvarez A, Salas-Martinez C. Detection of Mental Disorders Other Than Depression with the Edinburgh Postnatal Depression Scale in a Sample of Pregnant Women in Northern Mexico. *Ment Illn* 2016; **8:** 6021.

(54) Barnes J, Senior R, MacPherson K. The utility of volunteer home-visiting support to prevent maternal depression in the first year of life. *Child Care Health Dev* 2009; **35:** 807-16.

(55) Bavle AD, Chandahalli AS, Phatak AS, Rangaiah N, Kuthandahalli SM, Nagendra PN. Antenatal Depression in a Tertiary Care Hospital. *Indian J Psychol Med* 2016; **38:** 31-5.

(56) Comasco E, Gulinello M, Hellgren C, Skalkidou A, Sylven S, Sundstrom-Poromaa I. Sleep duration, depression, and oxytocinergic genotype influence prepulse inhibition of the startle reflex in postpartum women. *Eur Neuropsychopharmacol* 2016; **26:** 767-76.

(57) Eapen V, Johnston D, Apler A, Rees S, Silove DM. Adult separation anxiety during pregnancy and its relationship to depression and anxiety. *J Perinat Med* 2013; **41:** 159-63.

(58) Felice E, Saliba J, Grech V, Cox J. Prevalence rates and psychosocial characteristics associated with depression in pregnancy and postpartum in Maltese women. *J Affect Disord* 2004; **82:** 297-301.

(59) Giardinelli L, Innocenti A, Benni L, et al. Depression and anxiety in perinatal period: prevalence and risk factors in an Italian sample. *Arch Womens Ment Health* 2012; **15:** 21-30.

(60) Helle N, Barkmann C, Bartz-Seel J, et al. Very low birth-weight as a risk factor for postpartum depression four to six weeks postbirth in mothers and fathers: Cross-sectional results from a controlled multicentre cohort study. *J Affect Disord* 2015; **180:** 154-61.

(61) Hickey AR, Boyce PM, Ellwood D, Morris-Yates AD. Early discharge and risk for postnatal depression. *Med J Aust* 1997; **167:** 244-7.

(62) Howard LM, Ryan EG, Trevillion K, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *Br J Psychiatry* 2018; **212:** 50-6.

(63) Imbula Essam B, Okitundu Luwa EA, Mampunza Ma-Miezi S. Postpartum depression in Kinshasa (DR Congo): prevalence and risk factors. *Med Sante Trop* 2012; **22:** 379-84.

(64) Prenoveau J, Craske M, Counsell N, et al. Postpartum GAD is a risk factor for postpartum MDD: the course and longitudinal relationships of postpartum GAD and MDD. *Depress Anxiety* 2013; **30:** 506-14.

(65) Robertson-Blackmore E, Putnam FW, Rubinow DR, et al. Antecedent trauma exposure and risk of depression in the perinatal period. *J Clin Psychiatry* 2013; **74:** e942-8.

(66) Roomruangwong C, Kanchanatawan B, Sirivichayakul S, Maes M. Antenatal depression and hematocrit levels as predictors of postpartum depression and anxiety symptoms. *Psychiatry Res* 2016; **238:** 211-7.

(67) Rowe HJ, Fisher JR, Loh WM. The Edinburgh Postnatal Depression Scale detects but does not distinguish anxiety disorders from depression in mothers of infants. *Arch Womens Ment Health* 2008; **11:** 103-8.

(68) Siu BW, Leung SS, Ip P, Hung SF, O'Hara MW. Antenatal risk factors for postnatal depression: a prospective study of Chinese women at maternal and child health centres. *BMC Psychiatry* 2012; **12:** 22-244X.

(69) Turner K, Piazzini A, Franza A, Marconi AM, Canger R, Canevini MP. Epilepsy and postpartum depression. *Epilepsia* 2009; **50 Suppl 1:** 24-7.

(70) Usuda K, Nishi D, Makino M, et al. Prevalence and related factors of common mental disorders during pregnancy in Japan: a cross-sectional study. *Biopsychosoc Med* 2016; **10:** 17-016.

(71) Yonkers KA, Smith MV, Forray A, et al. Pregnant women with posttraumatic stress disorder and risk of preterm birth. *JAMA Psychiatry* 2014; **71:** 897-904.

(72) Fisher JR, Wynter KH, Rowe HJ. Innovative psycho-educational program to prevent common postpartum mental disorders in primiparous women: a before and after controlled study. *BMC Public Health* 2010; **10:** 432-2458.

(73) Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N. Validating screening tools for depression in epilepsy. *Epilepsia* 2014; **55:** 1642-50.

(74) Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res* 2014; **23:** 1609-18.

(75) Hahn D, Reuter K, Harter M. Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *Psychosoc Med* 2006; **3:** Doc09.

(76) Kiely KM, Butterworth P. Validation of four measures of mental health against depression and generalized anxiety in a community based sample. *Psychiatry Res* 2015; **225:** 291-8.

(77) Lamers F, Jonkers CC, Bosma H, Penninx BW, Knottnerus JA, van Eijk JT. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol* 2008; **61:** 679-87.

(78) McGuire AW, Eastwood JA, Macabasco-O'Connell A, Hays RD, Doering LV. Depression screening: utility of the patient health questionnaire in patients with acute coronary syndrome. *Am J Crit Care* 2013; **22:** 12-9.

(79) Osorio FL, Carvalho AC, Fracalossi TA, Crippa JA, Loureiro ES. Are two items sufficient to screen for depression within the hospital context? *Int J Psychiatry Med* 2012; **44:** 141-8.

(80) Patel V, Araya R, Chowdhary N, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med* 2008; **38:** 221-8.

(81) Santos IS, Tavares BF, Munhoz TN, et al. Sensitivity and specificity of the Patient Health Questionnaire-9 (PHQ-9) among adults from the general population. *Cad Saude Publica* 2013; **29:** 1533-43.

(82) Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. *Arch Womens Ment Health* 2012; **15:** 367-74.

(83) Williams JR, Hirsch ES, Anderson K, et al. A comparison of nine scales to detect depression in Parkinson disease: which scale to use? *Neurology* 2012; **78:** 998-1006.

(84) Wittkampf K, van Ravesteijn H, Baas K, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009; **31:** 451-9.

(85) Zhang Y, Ting R, Lam M, et al. Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. *J Affect Disord* 2013; **151:** 660-6.

(86) N Azah M, M Shah M, Juwita S, S Bahri I, WM Rushidi W, M Jamil Y. Validation of the Malay version brief Patient Health Questionnaire (PHQ-9) among adult attending family medicine clinics. *Int Med J* 2005; **12:** 259.

(87) Alvarado-Esquivel C, Sifuentes-Alvarez A, Salas-Martinez C, Martinez-Garcia S. Validation of the Edinburgh Postpartum Depression Scale in a population of puerperal women in Mexico. *Clin Pract Epidemiol Ment Health* 2006; **2:** 33-0179.

(88) Figueiredo FP, Parada AP, Cardoso VC, et al. Postpartum depression screening by telephone: a good alternative for public health and research. *Arch Womens Ment Health* 2015; **18:** 547-53.

(89) Fernandes MC, Srinivasan K, Stein AL, Menezes G, Sumithra R, Ramchandani PG. Assessing prenatal depression in the rural developing world: a comparison of two screening measures. *Arch Womens Ment Health* 2011; **14:** 209-16.

(90) Figueira P, Correa H, Malloy-Diniz L, Romano-Silva MA. Edinburgh Postnatal Depression Scale for screening in the public health system. *Rev Saude Publica* 2009; **43 Suppl 1:** 79-84.

(91) Leonardou A, Zervas Y, Papageorgiou C, et al. Validation of the Edinburgh Postnatal Depression Scale and prevalence of postnatal depression at two months postpartum in a sample of Greek mothers. *J Reprod Infant Psychol* 2009; **27:** 28-39.

(92) Navarro P, Ascaso C, Garcia-Esteve L, Aguado J, Torres A, Martin-Santos R. Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. *Gen Hosp Psychiatry* 2007; **29:** 1-7.

(93) Stewart RC, Umar E, Tomenson B, Creed F. Validation of screening tools for antenatal depression in Malawi--a comparison of the Edinburgh Postnatal Depression Scale and Self Reporting Questionnaire. *J Affect Disord* 2013; **150:** 1041-7.

(94) Tendais I, Costa R, Conde A, Figueiredo B. Screening for depression and anxiety disorders from pregnancy to postpartum with the EPDS and STAI. *Span J Psychol* 2014; **17:** E7.

(95) Tran TD, Tran T, La B, Lee D, Rosenthal D, Fisher J. Screening for perinatal common mental disorders in women in the north of Vietnam: a comparison of three psychometric instruments. *J Affect Disord* 2011; **133:** 281-93.

(96) Akena D, Joska J, Obuku EA, Stein DJ. Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care* 2013; **25:** 1245-52.

(97) Amoozegar F, Patten SB, Becker WJ, et al. The prevalence of depression and the accuracy of depression screening tools in migraine patients. *Gen Hosp Psychiatry* 2017; **48:** 25-31.

(98) Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 2010; **8:** 348-53.

(99) Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR. Validity of the Patient Health Questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. *Arch Phys Med Rehabil* 2012; **93:** 1838-45.

(100) Chagas MH, Tumas V, Rodrigues GR, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age Ageing* 2013; **42:** 645-9.

(101) Cholera R, Gaynes BN, Pence BW, et al. Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *J Affect Disord* 2014; **167:** 160-6.

(102) de Man-van Ginkel, J M, Hafsteinsdottir T, Lindeman E, Burger H, Grobbee D, Schuurmans M. An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke* 2012; **43:** 854-6.

(103) Delgadillo J, Payne S, Gilbody S, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *J Affect Disord* 2011; **134:** 266-71.

(104) Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil* 2005; **20:** 501-11.

(105) Gelaye B, Tadesse MG, Williams MA, Fann JR, Vander Stoep A, Andrew Zhou XH. Assessing validity of a depression screening instrument in the absence of a gold standard. *Ann Epidemiol* 2014; **24:** 527-31.

(106) Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med* 2009; **7:** 63-70.

(107) Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA. Diagnostic accuracy, internal consistency, and convergent validity of the Greek version of the patient health questionnaire 9 in diagnosing depression in rheumatologic disorders. *Arthritis Care Res (Hoboken)* 2011; **63:** 1313-21.

(108) Inagaki M, Ohtsuki T, Yonemoto N, et al. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: a cross-sectional study. *Gen Hosp Psychiatry* 2013; **35:** 592-7.

(109) Khamseh ME, Baradaran HR, Javanbakht A, Mirghorbani M, Yadollahi Z, Malek M. Comparison of the CES-D and PHQ-9 depression scales in people with type 2 diabetes in Tehran, Iran. *BMC Psychiatry* 2011; **11:** 61-244X.

(110) Lambert SD, Clover K, Pallant JF, et al. Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis. *J Natl Compr Canc Netw* 2015; **13:** 1203-11.

(111) Liu SI, Yeh ZT, Huang HC, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry* 2011; **52:** 96-101.

(112) Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry* 2008; **8:** 46-244X.

(113) Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78:** 131-40.

(114) Sherina MS, Arroll B, Goodyear-Smith F. Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. *Med J Malaysia* 2012; **67:** 309-15.

(115) de Lima Osorio F, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care* 2009; **45:** 216-27.

(116) Pence BW, Gaynes BN, Atashili J, et al. Validity of an interviewer-administered patient health questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *J Affect Disord* 2012; **143:** 208-13.

(117) Richardson TM, He H, Podgorski C, Tu X, Conwell Y. Screening depression aging services clients. *Am J Geriatr Psychiatry* 2010; **18:** 1116-23.

(118) Rooney AG, McNamara S, Mackinnon M, et al. Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. *Neuro Oncol* 2013; **15:** 122-9.

(119) Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry* 2007; **29:** 417-24.

(120) Sung SC, Low CC, Fung DS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: a comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). *Asia Pac Psychiatry* 2013; **5:** 249-58.

(121) Thombs BD, Ziegelstein RC, Whooley MA. Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *J Gen Intern Med* 2008; **23:** 2014-7.

(122) Turner A, Hambridge J, White J, et al. Depression screening in stroke: a comparison of alternative measures with the structured diagnostic interview for the diagnostic and statistical manual of mental disorders, fourth edition (major depressive episode) as criterion standard. *Stroke* 2012; **43:** 1000-5.

(123) Twist K, Stahl D, Amiel SA, Thomas S, Winkley K, Ismail K. Comparison of depressive symptoms in type 2 diabetes using a two-stage survey design. *Psychosom Med* 2013; **75:** 791-7.

(124) van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Serv Res* 2010; **10:** 235-6963.

(125) Vohringer PA, Jimenez MI, Igor MA, et al. Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. *J Med Screen* 2013; **20:** 118-24.

(126) Alvarado R, Jadresic E, Guajardo V, Rojas G. First validation of a Spanish-translated version of the Edinburgh postnatal depression scale (EPDS) for use in pregnant women. A Chilean study. *Arch Womens Ment Health* 2015; **18:** 607-12.

(127) Toreki A, Ando B, Dudas RB, et al. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. *Midwifery* 2014; **30:** 911-8.

(128) Castro E Couto T, Martins Brancaglion MY, Nogueira Cardoso M, et al. What is the best tool for screening antenatal depression? *J Affect Disord* 2015; **178:** 12-7.

(129) Bakare MO, Okoye JO, Obindo JT. Introducing depression and developmental screenings into the national programme on immunization (NPI) in southeast Nigeria: an experimental cross-sectional assessment. *Gen Hosp Psychiatry* 2014; **36:** 105-12.

(130) Rochat TJ, Tomlinson M, Newell ML, Stein A. Detection of antenatal depression in rural HIV-affected populations with short and ultrashort versions of the Edinburgh Postnatal Depression Scale (EPDS). *Arch Womens Ment Health* 2013; **16:** 401-10.

(131) Toreki A, Ando B, Kereszturi A, et al. The Edinburgh Postnatal Depression Scale: translation and antepartum validation for a Hungarian sample. *Midwifery* 2013; **29:** 308-15.

(132) Thiagayson P, Krishnaswamy G, Lim ML, et al. Depression and anxiety in Singaporean high-risk pregnancies - prevalence and screening. *Gen Hosp Psychiatry* 2013; **35:** 112-6.

(133) Tandon SD, Cluxton-Keller F, Leis J, Le HN, Perry DF. A comparison of three screening tools to identify perinatal depression among low-income African American women. *J Affect Disord* 2012; **136:** 155-62.

(134) Chaudron LH, Szilagyi PG, Tang W, et al. Accuracy of depression screening tools for identifying postpartum depression among urban mothers. *Pediatrics* 2010; **125:** e609-17.

(135) Bunevicius A, Kusminskas L, Pop VJ, Pedersen CA, Bunevicius R. Screening for antenatal depression with the Edinburgh Depression Scale. *J Psychosom Obstet Gynaecol* 2009; **30:** 238-43.

(136) Phillips J, Charles M, Sharpe L, Matthey S. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. *J Affect Disord* 2009; **118:** 101-12.

(137) Pawlby S, Sharp D, Hay D, O'Keane V. Postnatal depression and child outcome at 11 years: the importance of accurate diagnosis. *J Affect Disord* 2008; **107:** 241-5.

(138) Su KP, Chiu TH, Huang CL, et al. Different cutoff points for different trimesters? The use of Edinburgh Postnatal Depression Scale and Beck Depression Inventory to screen for depression in pregnant Taiwanese women. *Gen Hosp Psychiatry* 2007; **29:** 436-41.

(139) Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *J Affect Disord* 2003; **75:** 71-6.

(140) Vega-Dienstmaier JM, Mazzotti Suarez G, Campos Sanchez M. Validation of a Spanish version of the Edinburgh Postnatal Depression Scale. *Actas Esp Psiquiatr* 2002; **30:** 106-11.

(141) Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nurs Res* 2001; **50:** 242-50.

(142) Radoš SN, Tadinac M, Herman R. Validation study of the Croatian version of the Edinburgh Postnatal Depression Scale (EPDS). *Suvrem Psihol* 2013; **16:** 203-8.

(143) Tissot H, Favez N, Frascarolo-Moutinot F, Despland J. Assessing postpartum depression: Evidences for the need of multiple methods. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 2015; **65:** 61-6.

(144) Khalifa DS, Glavin K, Bjertness E, Lien L. Postnatal depression among Sudanese women: prevalence and validation of the Edinburgh Postnatal Depression Scale at 3 months postpartum. *Int J Womens Health* 2015; **7:** 677-84.