Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: A systematic review and individual participant data meta-analysis

Yin Wu, PhD[1-3]; Brooke Levis, MSc[1,3]; Kira E. Riehm, MSc[1]; Nazanin Saadat, MSc[1]; Alexander W. Levis, MSc[1]; Marleine Azar, MSc[1]; Danielle B. Rice, MSc[1,4]; Jill Boruff, MLIS[5]; Pim Cuijpers, PhD[6]; Simon Gilbody, PhD[7]; John P.A. Ioannidis, MD[8]; Lorie A. Kloda, PhD[9]; Dean McMillan, PhD[7]; Scott B. Patten, MD[10,11]; Ian Shrier, MD[1,3]; Roy C. Ziegelstein, MD[12]; Dickens H. Akena, PhD[13]; Bruce Arroll, MBChB[14]; Liat Ayalon, PhD[15]; Hamid R. Baradaran, MD[16,17]; Murray Baron, MD[1,18]; Charles H. Bombardier, PhD[19]; Peter Butterworth, PhD[20,21]; Gregory Carter, FRANZCP[22]; Marcos H. Chagas, MD[23]; Juliana C. N. Chan, MD[24-26]; Rushina Cholera, MD[27]; Yeates Conwell, MD[28]; Janneke M. de Man-van Ginkel, PhD[29]; Jesse R. Fann, MD[30]; Felix H. Fischer, PhD[31]; Daniel Fung, MD[32-35]; Bizu Gelaye, PhD[36]; Felicity Goodyear-Smith, MD[14]; Catherine G. Greeno, PhD[37]; Brian J. Hall, PhD[38,39]; Patricia A. Harrison, PhD[40]; Martin Härter, MD, PhD Dipl Psych[41]; Ulrich Hegerl, MD[42]; Leanne Hides, PhD(Clin)[43]; Stevan E. Hobfoll, PhD[44]; Marie Hudson, MD[1,18]; Thomas Hyphantis, MD, PhD[45]; Masatoshi Inagaki, MD[46]; Nathalie Jetté, MD[10,11,47]; Mohammad E. Khamseh, MD[16]; Kim M. Kiely, PhD[48,49]; Yunxin Kwan, MMed (Psychiatry)[50]; Femke Lamers, PhD[51]; Shen-Ing Liu, MD[35,52-54]; Manote Lotrakul, MD[55]; Sonia R. Loureiro, PhD[23]; Bernd Löwe, MD[56]; Anthony McGuire, PhD[57]; Sherina Mohd-Sidik, PhD[58]; Tiago N. Munhoz, PhD[59]; Kumiko Muramatsu, MD[60]; Flávia L. Osório, PhD[23,61]; Vikram Patel, MD[62,63]; Brian W. Pence, PhD[64]; Philippe Persoons, MD[65,66]; Angelo Picardi, MD[67]; Katrin Reuter, PhD[68]; Alasdair G. Rooney, MD[69]; Iná S. Santos, MD[59]; Juwita Shaaban, MMed (Fam. Med)[70]; Abbey Sidebottom, PhD[71]; Adam Simning, MD, PhD[28]; Lesley Stafford, PhD[72,73]; Sharon Sung, PhD[32,35]; Pei Lin Lynnette Tan, MMed (Psychiatry)[50];

Alyna Turner, PhD[74,75]; Henk C. van Weert, MD[76]; Jennifer White, PhD[77]; Mary A. Whooley, MD[78-80]; Kirsty Winkley, PhD[81]; Mitsuhiko Yamada, MD[82]; Andrea Benedetti, PhD[3,18,83]; Brett D. Thombs, PhD[1-4,18,84].

Affiliations:

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; [2]Department of Psychiatry, McGill University, Montréal, Québec, Canada; [3]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; [4]Department of Psychology, McGill University, Montréal, Québec, Canada; [5]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Quebec, Canada; [6]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, the Netherlands; [7]Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; [8]Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA; [9]Library, Concordia University, Montréal, Québec, Canada; [10]Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada; [11]Hotchkiss Brain Institute and O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada; [12]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [13]Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda; [14]Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand; [15]Louis and Gabi Weisfeld School of Social Work, Bar Ilan University, Ramat Gan, Israel; [16]Endocrine Research Center, Institute of Endocrinology and Metabolism,

Iran University of Medical Sciences, Tehran, Iran; [17]Ageing Clinical & Experimental Research Team, Institute of Applied Health Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, Scotland, UK; [18]Department of Medicine, McGill University, Montréal, Québec, Canada; [19]Department of Rehabilitation Medicine, University of Washington, Seattle, Washington, USA; [20]Centre for Research on Ageing, Health and Wellbeing, Research School of Population Health, The Australian National University, Canberra, Australia; [21]Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Melbourne, Australia; [22]Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; [23]Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; [24]Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong Special Administrative Region (SAR), China; [25]Asia Diabetes Foundation, Prince of Wales Hospital, Hong Kong SAR, China; [26]Hong Kong Institute of Diabetes and Obesity, Hong Kong SAR, China; [27]Department of Pediatrics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA; [28]Department of Psychiatry, University of Rochester Medical Center, New York, USA; [29]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; [30]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; [31]Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Germany; [32]Department of Child & Adolescent Psychiatry, Institute of Mental Health, Singapore; [33]Yong Loo Lin School of Medicine, National University of Singapore, Singapore; [34]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore; [35]Programme in Health Services & Systems Research, Duke-NUS

Medical School, Singapore; [36]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; [37]School of Social Work, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; [38]Global and Community Mental Health Research Group, Department of Psychology, Faculty of Social Sciences, University of Macau, Macau Special Administrative Region, China; [39]Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [40]City of Minneapolis Health Department, Minneapolis, Minnesota, USA; [41]Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [42]Depression Research Center of the German Depression Foundation and Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, Germany; [43]School of Psychology, University of Queensland, Brisbane, Queensland, Australia; [44]STAR-Stress, Anxiety & Resilience Consultants, Chicago, Illinois, USA; [45]Department of Psychiatry, University of Ioannina, Ioannina, Greece; [46]Department of Psychiatry, Faculty of Medicine, Shimane University, Shimane, Japan; [47]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [48]School of Psychology, University of New South Wales, Sydney, Australia; [49]Neuroscience Research Australia, Sydney, Australia; [50]Department of Psychological Medicine, Tan Tock Seng Hospital, Singapore; [51]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands; [52]Department of Psychiatry, Mackay Memorial Hospital, Taipei, Taiwan; [53]Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan; [54]Department of Medicine, Mackay Medical College, Taipei, Taiwan; [55]Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; [56]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg,

Germany; [57]Department of Nursing, St. Joseph's College, Standish, Maine, USA; [58]Cancer Resource & Education Centre, and Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia; [59]Post-graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS, Brazil; [60]Department of Clinical Psychology, Graduate School of Niigata Seiryo University, Niigata, Japan; [61]National Institute of Science and Technology, Translational Medicine, Ribeirão Preto, Brazil; [62]Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts, USA; [63]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [64]Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; [65]Department of Adult Psychiatry, University Hospitals Leuven, Leuven, Belgium; [66]Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium; [67]Centre for Behavioural Sciences and Mental Health, Italian National Institute of Health, Rome, Italy; [68]Department of Psychiatry and Psychotherapy, University Medical Center Freiburg, Freiburg, Germany; [69]Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburg, Edinburgh, Scotland, UK; [70]Department of Family Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia; [71]Allina Health, Minneapolis, Minnesota, USA; [72]Centre for Women's Mental Health, Royal Women's Hospital, Parkville, Melbourne, Australia; [73]Melbourne School of Psychological Sciences, University of Melbourne, Australia; [74]School of Medicine and Public Health, University of Newcastle, New South Wales, Newcastle, Australia; [75]IMPACT Strategic Research Centre, School of Medicine, Deakin University, Geelong, Victoria, Australia; [76]Department of General Practice, Amsterdam Institute for General Practice and Public Health, Amsterdam University Medical Centers, location AMC; [77]Monash University, Melbourne,

Australia; [78]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA; [79]Department of Medicine, Veterans Affairs Medical Center, San Francisco, California, USA; [80]Department of Medicine, University of California San Francisco, San Francisco, California, USA; [81]Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, UK, [82]Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawa-Higashi, Kodaira, Tokyo, Japan; [83]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada; [84]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada.

Full Addresses:

Yin Wu

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Brooke Levis

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Kira E. Riehm

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Nazanin Saadat

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Alexander W. Levis

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Marleine Azar

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Danielle B. Rice

Lady Davis Institute for Medical Research, Jewish General Hospital, 4333 Chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 1E4, Canada

Jill Boruff

Macdonald-Stewart Library Building, 809 Sherbrooke Street West, Montreal, Quebec H3A 0C1, Canada

Pim Cuijpers

Department of Clinical, Neuro and Developmental Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

Simon Gilbody

Mental Health and Addiction Research Group, Department of Health Sciences and Hull York Medical School, University of York, Heslington YO10 5DD, United Kingdom

John P.A. Ioannidis

Stanford University, 1265 Welch Road, MSOB X306, Stanford, CA, 94305, USA

Lorie A. Kloda

Concordia University, 1455, boul. de Maisonneuve Ouest, LB-331, Montréal, QC, H3G 1M8, Canada

Dean McMillan

Mental Health and Addiction Research Group, Department of Health Sciences and Hull York Medical School, University of York, Heslington YO10 5DD, United Kingdom

Scott B. Patten

Department of Community Health Sciences, 3rd Floor, TRW Building, University of Calgary, 3280 Hospital Drive NW, Calgary, AB, T2N 4Z6, Canada

Ian Shrier

Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Cote Ste-Catherine Rd, Montréal, QC, H3T 1E2, Canada

Roy C. Ziegelstein

Johns Hopkins University School of Medicine, Miller Research Building, 733 N. Broadway, Suite 115, Baltimore, MD, 21205, USA

Dickens H. Akena

Department of Psychiatry, Makerere University College of Health Sciences, P.O.Box 7062 Kampala, Uganda

Bruce Arroll

Department of General Practice and Primary Health Care, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

Liat Ayalon

Louis and Gabi Weisfeld School of Social Work, Ramat Gan, Bar Ilan University, 52900, Israel

Hamid R. Baradaran

Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran 15937-16615, Iran

Murray Baron

Jewish General Hospital, Suite A 725, 3755 Cote St Catherine Rd, Montréal, QC, H3T 1E2, Canada

Charles H. Bombardier

Division of Clinical and Neuropsychology, Department of Rehabilitation Medicine, University of Washington, Box 359612, Harborview Medical Center, 325 9th Avenue, Seattle, WA, 98104, USA

Peter Butterworth

Research School of Population Health, Florey Building (54), Australian National University, Canberra, ACT, 2600, Australia

Gregory Carter

Locked Bag #7, Hunter Region Mail Centre, NSW 2310, Australia

Marcos H. Chagas

University of São Paulo, Av. Bandeirantes, 3900, 14048-900-Ribeirão Preto, SP, Brazil

Juliana C. N. Chan

Department of Medicine and Therapeutics, The Chinese University of Hong Kong, 9/F Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital, Shatin, Hong Kong

Rushina Cholera

UNC Department of Pediatric, 260 MacNider Building, CB# 7220, 321 S. Columbia Street, UNC School of Medicine, Chapel Hill, NC 27599-7220, USA

Yeates Conwell

University of Rochester Medical Center, 300 Crittenden Blvd., Rochester, NY, 14642, USA

Janneke M. de Man-van Ginkel

University Medical Center Utrecht, Internal mail no Str. 6.131, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands

Jesse R. Fann

Department of Psychiatry & Behavioral Sciences, Univerisity of Washington, Box 356560, Seattle, WA 98195

Felix H. Fischer

Medizinische Klinik mit Schwerpunkt Psychosomatik, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10098 Berlin, Germany

Daniel Fung

Institute of Mental Health, 10 Buangkok View, 539747, Singapore

Bizu Gelaye

Department of Epidemiology, 677 Huntington Ave, Room 505F, Boston, MA, 02115, USA

Felicity Goodyear-Smith

Department of General Practice and Primary Health Care, University of Auckland, PB 92019, Auckland, 1142, New Zealand

Catherine G. Greeno

2204 Cathedral of Learning, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh, PA, 15260, USA

Brian J. Hall

Department of Psychology, Faculty of Social Sciences, Humanities and Social Sciences Building E21-3040, University of Macau, E21 Avenida da Universidade, Taipa, Macau, China

Patricia A. Harrison

City of Minneapolis Health Department, 250 S. Fourth St., Room 510, Minneapolis, MN 55415, USA

Martin Härter

University Medical Center Hamburg-Eppendorf, Department of Medical Psychology, Martinistraße 52, 20246 Hamburg, Germany

Ulrich Hegerl

Depression Research Center of the German Depression Foundation and Department pf Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt 60323 Frankfurt, Germany

Leanne Hides

School of Psychology, University of Queensland, St Lucia, Brisbane, Queensland, 4072, Australia

Stevan E. Hobfoll

STAR-Stress, Anxiety & Resilience Consultants, 1000 N Lake Shore Plaza 9B, Chicago, IL, 60611

Marie Hudson

Jewish General Hospital and Lady Davis Research Institute, 3755 Côte Ste-Catherine Rd, Room A725, Montréal, QC, H3T 1E2, Canada

Thomas Hyphantis

Department of Psychiatry, Faculty of Medicine, School of Health Sciences, University of Ioannina, Ioannina 451 10, Greece

Masatoshi Inagaki

Department of Psychiatry, Faculty of Medicine, Shimane University, 89-1 Enya-cho, Izumo, Shimane, Japan

Nathalie Jetté

Department of Neurology, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY 10029, USA

Mohammad E. Khamseh

Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran 15937-16615, Iran

Kim M. Kiely

Neuroscience Research Australia, Margarete Ainsworth Building, Barker Street, NSW 2031, Sydney, Australia

Yunxin Kwan

Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, 308433, Singapore

Femke Lamers

Amsterdam UMC, Vrije Universiteit, Department Psychiatry, Oldenaller 1, 1081 HJ Amsterdam, The Netherlands

Shen-Ing Liu

Department of Psychiatry, Mackay Memorial Hospital, No. 92, Section 2, Chung-Shan North Rd, Taipei, Taiwan

Manote Lotrakul

Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand

Sonia R. Loureiro

Pós-graduação em Saúde Mental, Depto. de Neurociências e Ciências do Comportamento da Faculdade de Medicina de Ribeirão Preto, USP, Rua Tenente Catão Roxo, 2650, CEP 14051-140, Ribeirão Preto, SP, Brazil

Bernd Löwe

Universitätsklinikum Hamburg-Eppendorf, Institut und Poliklinik für Psychosomatische Medizin und Psychotherapie, Martinistr. 52, Gebäude O25, 20246 Hamburg, Germany

Anthony McGuire

Department of Nursing, St. Joseph's College, 278 Whites Bridge Rd., Standish, ME, 04084, USA

Sherina Mohd-Sidik

Cancer Resource & Education Centre/ Department of Psychiatry, Faculty of Medicine & Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Tiago N. Munhoz

Depto Medicina Social, Programa Pós-graduação Epidemiologia, Universidade Federal de Pelotas, Rua Marechal Deodoro 1160, 3º piso, 96020-220 - Pelotas, RS, Brasil

Kumiko Muramatsu

Department of Clinical Psychology, Graduate School of Niigata Seiryo University, 1-5939, Suidocho, Chuo-ku, Niigata 951-8121, Japan.

Flávia L. Osório

Department of Neurosciences and Behaviour, Medical School of RibeirãoPreto, University of São Paulo, Avenida dos Bandeirantes 3900, CEP 14048-900, Brazil

Vikram Patel

Department of Global Health and Social Medicine, Harvard Medical School, Boston, USA 02119, USA

Brian W. Pence

Department of Epidemiology, UNC-Chapel Hill, McGavran-Greenberg 2103C, CB#7435, 135 Dauer Dr, Chapel Hill NC 27599-7435, USA

Philippe Persoons

Katholieke Universiteit Leuven, Department of Neurosciences, Research Group Psychiatry, University Psychiatric Center KU Leuven, Herestraat 49, 3000 Leuven, Belgium

Angelo Picardi

Italian National Institute of Health, Centre for Behavioural Sciences and Mental Health, Viale Regina Elena 299, 00161 Rome, Italy

Katrin Reuter

Outpatient Service for Psychotherapy and Psychooncology, Stadtstrasse 11, 79104 Freiburg, Germany

Alasdair G. Rooney

Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital Edinburgh, EH10 5HF, Scotland

Iná S. Santos

Depto Medicina Social, Programa Pós-graduação Epidemiologia, Universidade Federal de Pelotas, Rua Marechal Deodoro 1160, 3º piso 96020-220 - Pelotas, RS, Brasil

Juwita Shaaban

School of Medical Science, Health Campus Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

Abbey Sidebottom

Allina Health, 800 E 28th Street, MR 15521, Minneapolis, MN 55407-3799, USA

Adam Simning

University of Rochester Medical Center, School of Medicine and Dentistry, 601 Elmwood Ave, Box PSYCH, Rochester, NY 14642, USA

Lesley Stafford

Centre for Women's Mental Health, The Royal Women's Hospital, Locked Bag 300, Parkville Victoria 3052, Australia

Sharon Sung

Programme in Health Services & Systems Research, Duke-NUS Medical School, 20 College Road, Level 6, 169856, Singapore

Pei Lin Lynnette Tan

Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, 308433, Singapore

Alyna Turner

IMPACT SRC, Deaking University, HERB Building Level 3, PO Box 281 Geelong 3220, Australia

Henk C. van Weert

Department of General Practice, Amsterdam Institute for General Practice and Public Health, Amsterdam University Medical Centers, location AMC, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Jennifer White

Department of Physiotherapy, School of Primary and Allied Health Care, Monash University, Building G, Level 3, McMahons Road, Frankston Victoria 3199 Australia

Mary A. Whooley

Department of Veterans Affairs Medical Center, 4150 Clement Street (111A1), San Francisco, CA 94121, USA

Kirsty Winkley

Faculty of Nursing, Midwifery & Palliative Care, King's College London, Strand, London WC2R 2LS, UK

Mitsuhiko Yamada

National Center of Neurology and Psychiatry, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187-8553, Japan

Andrea Benedetti

Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve, Montréal, QC, H4A 3S5, Canada

Brett D. Thombs

Jewish General Hospital; 4333 Cote Ste Catherine Road; Montreal, Quebec H3T 1E4, Canada

Corresponding authors:

Andrea Benedetti, PhD; Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve, Montréal, QC, 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 H3T 1E4; Tel (514) 340-8222 ext. 25112; E-mail: brett.thombs@mcgill.ca

388 **Word count:** 4,258

ABSTRACT

Background: Item 9 of the Patient Health Questionnaire-9 (PHQ-9) queries about thoughts of death and self-harm, but not suicidality. Although it is sometimes used to assess suicide risk, most positive responses are not associated with suicidality. The PHQ-8, which omits Item 9, is thus increasingly used in research. We assessed equivalency of total score correlations and the diagnostic accuracy to detect major depression of the PHQ-8 and PHQ-9.

Methods: We conducted an individual patient data meta-analysis. We fit bivariate random-effects models to assess diagnostic accuracy.

Results: 16,742 participants (2,097 major depression cases) from 54 studies were included. The correlation between PHQ-8 and PHQ-9 scores was 0·996 (95% confidence interval 0·996 to 0·996). The standard cutoff score of 10 for the PHQ-9 maximized sensitivity + specificity for the PHQ-8 among studies that used a semi-structured diagnostic interview reference standard (N = 27). At cutoff 10, the PHQ-8 was less sensitive by 0·02 (-0·06 to 0·00) and more specific by 0·01 (0·00 to 0·01) among those studies (N = 27), with similar results for studies that used other types of interviews (N = 27). For all 54 primary studies combined, across all cutoffs, the PHQ-8 was less sensitive than the PHQ-9 by 0·00 to 0·05 (0·03 at cutoff 10), and specificity was within 0·01 for all cutoffs (0·00 to 0·01).

Conclusions: PHQ-8 and PHQ-9 total scores were similar. Sensitivity may be minimally reduced with the PHQ-8, but specificity is similar.

INTRODUCTION

The 9-item Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer & Williams, 2001) is a self-report questionnaire that is commonly used for identifying people who may have depression based on matching symptoms to diagnostic criteria or, more commonly, on a standard cutoff of a score of 10 or greater (Moriarty et al. 2015; Levis et al. 2019). It is also used as a continuous measure to assess depressive symptom severity in research and clinical care (Kroenke, Spitzer & Williams, 2001). The nine items of the PHQ-9 are designed to capture the nine Diagnostic and Statistical Manual of Mental Disorders (DSM) symptom criteria for a major depressive episode (American Psychiatric Association 2013). Response options on the items range from "not at all" (0 points) to "nearly every day" (3 points). Per the DSM-5, the ninth criterion for major depression requires "Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide" (American Psychiatric Association 2013). Item 9 of the PHQ-9 taps into this criterion but also assesses self-harm, which is not part of the DSM criterion, or passive thoughts of death within the last two weeks: "…how often have you been bothered by…thoughts that you would be better off dead or of hurting yourself in some way?" It does not query specifically about suicidality, and positive responses may be due to thoughts about death or to thoughts about self-harm.

Item 9 is sometimes used as an indicator of suicide risk, and it may be useful as a component of modelling approaches for stratifying suicide risk among participants in psychiatric settings (Simon et al. 2016; Simon et al. 2013). However, responses to the item may not accurately reflect whether or not suicide risk is present, particularly among patients with serious medical conditions for whom thoughts of death may not reflect suicidal ideation, and it appears to perform poorly in identifying individuals at risk in these settings. Four studies in non-

psychiatric settings have compared positive responses on Item 9 to responses to questions that explicitly assess suicidal thoughts or intentionality. In these studies, which included US military veterans in primary care (Corson, Gerrity & Dobscha, 2004), patients with coronary artery disease (Razykov et al. 2012; Suarez et al. 2015), and cancer patients (Walker et al. 2011), 7% to 21% of all study participants had positive responses on Item 9, but of those, only 18% to 35% had thoughts of suicide based on questions designed specifically to address suicide risk, and only 3% to 20% had a plan (Corson, Gerrity & Dobscha, 2004; Razykov et al. 2012; Suarez et al. 2015; Walker et al. 2011). Thus, concerns have been raised that using Item 9 to identify individuals at risk would result in a high rate of false indications, compared to questions designed specifically to assess suicidal thoughts or intentionality (Razykov et al. 2012; Suarez et al. 2015; Walker et al. 2011).

The PHQ-8 omits Item 9 from the PHQ-9. Many research studies use the PHQ-8 as a depression screening tool or to assess depressive symptom severity in order to avoid the high risk of inaccurate indications of suicide risk based on Item 9 (Corson, Gerrity & Dobscha, 2004; Razykov et al. 2012; Kroenke et al. 2009; Wells et al. 2013; Barrera et al. 2017). This is a particularly important consideration in studies that are not focused on depression or psychiatric disorders, but would need to allocate substantial resources to follow-up on responses to Item 9 of the PHQ-9. Similarly, many large epidemiological studies that include assessments of depressive symptoms are not able to provide adequate assessment and intervention with telephone or internet surveys (Kroenke et al. 2009).

Although differences in performance between the PHQ-8 and PHQ-9 might be expected to be minimal, to the best of our knowledge, only one study has attempted to verify this by comparing diagnostic accuracy between the PHQ-8 and PHQ-9 (Razykov et al. 2012). That

study evaluated the diagnostic testing accuracy of the PHQ-8 versus the PHQ-9 and the correlation between PHQ-8 and PHQ-9 scores in a sample of 1,022 coronary artery disease outpatients (233 major depression cases). Differences between sensitivity and specificity for the PHQ-8 (50%, 91%) and PHQ-9 (54%, 90%) based on a cutoff score of 10 or greater were minimal. In addition, PHQ-8 and PHQ-9 scores were highly correlated (r = 0·997) (Razykov et al. 2012). One additional study reported correlations between continuous PHQ-8 and PHQ-9 scores (Corson, Gerrity & Dobscha, 2004). That study, which included over 1000 patients from a US Department of Veterans Affairs primary care setting, reported a correlation of r = 0·998 (Corson, Gerrity & Dobscha, 2004).

We have synthesized a large database of individual participant data (IPD) from primary studies on the PHQ-9 (Levis et al. 2019; Levis et al. 2018). In the present study we included studies from that database that provided individual item scores (not just total PHQ-9 scores), which allowed for calculation of PHQ-8 scores. The objectives of the present study were (1) to evaluate the equivalency of the correlation between PHQ-8 and PHQ-9 scores for assessing depressive symptom severity; and (2) to assess the equivalency of the diagnostic accuracy of PHQ-8 and PHQ-9 across relevant cutoffs for screening to detect major depression.

METHOD

Data Source

The present study used a subset of participants from an IPD database of PHQ-9 (Levis et al. 2019; Levis et al. 2018). The main PHQ-9 IPD meta-analysis (IPDMA) was registered in PROSPERO (CRD42014010673), and a protocol was published (Thombs et al. 2014). Analyses of the diagnostic accuracy of the PHQ-8 were conducted according to protocol with two exceptions: (1) we stratified results by reference standard categories and (2) we added an

examination of equivalency with the PHQ-9. Results from the main IPDMA of the PHQ-9 are available elsewhere (Levis et al. 2019).

Search Strategy

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and Web of Science from January 1, 2000 through February 7, 2015, using a peer-reviewed search strategy (Canadian Agency for Drugs and Technologies in Health 2016) (SupplementaryMethods1). We limited our search to these databases based on research showing that adding other databases when the Medline search is highly sensitive does not identify additional eligible studies (Rice et al. 2016). The search was limited to the year 2000 forward because the PHQ-9 was originally published in 2001 (Kroenke, Spitzer & Williams, 2001). In addition to the database search, we 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). After de-duplication, unique citations were uploaded into DistillerSR (Evidence Partners, Ottawa, Canada), which was used to store and track search results, conduct screening for eligibility, document correspondence with primary study authors, and extract study characteristics.

Identification of Eligible Studies

Datasets from articles in any language were eligible for inclusion if they included diagnostic classification among participants aged 18 or older for current Major Depressive Disorder (MDD) or Major Depressive Episode (MDE) based on a validated semi-structured or fully structured interview conducted within two weeks of PHQ-9 administration, since diagnostic criteria for major depression are for symptoms in the last two weeks. Datasets where not all participants were at least 18 years of age were included if the primary data allowed us to select

participants who were at least 18 years of age. Datasets where not all participants were administered the PHQ-9 within two weeks of the diagnostic interview were included if the primary data allowed us to select participants who were administered both instruments within two weeks. Data from studies where the PHQ-9 was administered exclusively to individuals with known psychiatric diagnoses or symptoms or who were seeking psychiatric care were excluded, because screening is not indicated for patients already seeking care or managed in psychiatric settings. For defining major depression cases, we considered MDD or MDE based on the DSM or the International Classification of Diseases (ICD). If more than one was reported, we prioritized MDE over MDD and DSM over ICD. Across all studies, there were only 23 discordant diagnoses that depended on classification prioritization (0·1% of participants). For the present study, we only included primary studies that provided individual PHQ-9 item scores and not just PHQ-9 total scores, because only those datasets allowed us to generate PHQ-8 scores and compare the PHQ-8 with the PHQ-9.

Two investigators independently reviewed titles and abstracts for eligibility. If either reviewer deemed a study potentially eligible, full-text article review was done by two investigators, independently. Disagreement between reviewers after full-text review was resolved by consensus, consulting a third investigator when necessary. Translators were consulted to evaluate titles, abstracts and full-text articles for languages other than those for which team members were fluent.

Data Contribution and Synthesis

Authors of eligible datasets were invited to contribute de-identified primary data. Primary study country, clinical setting, language, and diagnostic interview administered were extracted from published reports by two investigators independently, with disagreements resolved by

consensus. Countries were categorized as “very high”, “high”, or “low-medium” development level based on the United Nation’s human development index (Whiting et al. 2011). Recruitment settings were categorized as “non-medical”, “primary care”, “inpatient specialty care”, or “outpatient specialty care.” Participant-level data included age, sex, major depression status, current diagnosis or treatment for a mental health problem, and PHQ-9 item scores. In two primary studies, multiple recruitment settings were included, thus recruitment setting was coded at the participant-level. When primary study datasets included appropriate statistical weighting to reflect sampling procedures, we used the provided weights. For studies where sampling procedures merited weighting, but the original study did not weight, we constructed appropriate weights using inverse selection probabilities. Weighting occurred, for instance, when all participants with positive screens, but only a random subset of participants with negative screens, were administered a diagnostic interview.

Individual participant data were converted to a standard format and entered into a single dataset that also included study-level data. We compared published participant characteristics and diagnostic accuracy results with those obtained using the raw datasets. When primary data and original publications were discrepant, we identified and corrected errors when possible and resolved any outstanding discrepancies in consultation with the original investigators.

Statistical Analyses

To evaluate the equivalence of the PHQ-8 and PHQ-9 scores for assessing depressive symptom severity, a Pearson correlation with a 95% confidence interval (CI) was calculated between the total scores of PHQ-8 (which excluded Item 9) and PHQ-9.

To estimate the diagnostic accuracy of the PHQ-8 and compare with the PHQ-9, we analyzed primary studies separately by the type of diagnostic interview that was used as the

reference standard, as we did in the previously published main PHQ-9 meta-analysis (Levis et al. 2019). This was done because of differences in the performance of the different types of interviews. Semi-structured interviews involve clinical judgement and are designed to be administered by clinically trained professionals; fully structured interviews are completely scripted and designed for lay administration, but the resulting increased standardization and reliability across interviewers may lead to increased misclassification (Brugha, Bebbington & Jenkins, 1999; Nosen & Woody 2008). The Mini International Neuropsychiatric Interview (MINI), which is a fully structured interview, was developed to be administered in a fraction of the time necessary for other fully structured interviews and was described by its developers as designed to be over-inclusive (Robins et al. 1988; Sheehan et al. 1997). In a previous study, we found that semi-structured and fully structured diagnostic interviews are not interchangeable reference standards for major depression and that fully structured interviews may diagnose depression at higher rates than semi-structured interviews at low symptom levels and at lower rates at high symptom levels (Levis et al. 2018). We also found that the MINI classifies approximately twice as many participants as cases compared to the most commonly used fully-structured interview, the Composite International Diagnostic Interview (CIDI) (Levis et al. 2018). In the main PHQ-9 meta-analysis, the diagnostic accuracy of the PHQ-9 differed substantially depending on the reference standard used for the comparison (Levis et al. 2019). Thus, for the present study, we analyzed primary studies separately based on whether they used a semi-structured interview, a fully structured interview (non-MINI), or the MINI.

For each reference standard and for the PHQ-8 and PHQ-9 cutoffs 5-15, separately, bivariate random-effects models were fitted using an adaptive Gauss Hermite quadrature with 1 quadrature point (Riley et al. 2008). This 2-stage meta-analytic approach models sensitivity and

specificity at the same time, taking the inherent correlation between them and the precision of estimates within studies into account. A random-effects model was used as we assumed true values of sensitivity and specificity would likely to vary across primary studies.

In order to examine the equivalence between PHQ-8 and PHQ-9 across reference standards, for each analysis, we used the results of the random-effects meta-analyses at each cutoff to construct separate empirical receiver operating characteristic (ROC) curves based on the pooled estimates. Equivalence tests between PHQ-8 and PHQ-9 sensitivity and specificity were conducted at each cutoff. This allowed us to test whether the sensitivity and specificity of the PHQ-8 was similar to that of the PHQ-9, up to a pre-specified maximum clinically acceptable difference, that is, an equivalence margin (Walker & Nowacki 2011). In the present study, an equivalence margin of $\delta = 0{\cdot}05$ was used, which is the same margin that was used in a previous study that used the same IPD database (Ishihara et al. 2019). CIs for the differences between PHQ-8 and PHQ-9 sensitivity and specificity at each cutoff were constructed via a cluster bootstrap approach (van der Leeden, Busing & Meijer, 1997; van der Leeden, Meijer & Busing, 2008), with resampling at the study and subject level. For each comparison, we ran 1000 iterations of the bootstrap. For each bootstrap iteration, the bivariate random-effects model was fitted to the PHQ-8 and PHQ-9 data, and the pooled sensitivities and specificities were computed separately, as described above, for each cutoff score. Equivalence tests were done by comparing the CIs around the pooled sensitivity and specificity differences to the equivalence margin of $\delta = 0{\cdot}05$. If the entire CI was included within the interval of +/- 0·05, then we rejected the hypothesis that there were differences large enough to be important and concluded that equivalence was present. If the entire CI was outside of the interval, then we failed to reject the hypothesis that the

PHQ-8 and PHQ-9 were not equivalent. When the CIs crossed the +/- 0·05 threshold, findings were deemed equivocal, and the equivalence was indeterminate.

Although we previously found that sensitivity and specificity of the PHQ-9 differs by type of reference standard, we did not believe that the differences in sensitivity and specificity between the PHQ-8 and PHQ-9 would vary depending on the reference standard. This is because for each included study the PHQ-8 and PHQ-9 were compared to the same reference standard. Thus, we reported pooled sensitivity and specificity only stratified by reference standards, but we investigated equivalence both stratified by reference standards and pooled across all studies. To investigate heterogeneity across studies, by reference standard and overall, we generated forest plots for the differences in sensitivity and specificity estimates between PHQ-8 and PHQ-9 for the standard cutoff 10 for each study. We also quantified heterogeneity at cutoff 10, by reporting the estimated variances of the random effects for the differences in PHQ-8 and PHQ-9 sensitivity and specificity (τ^2) (Fagerland, Lydersen & Laake, 2014; Higgins & Thompson 2002).

All analyses were run in R (R version R 3.5.0 and R Studio version 1.1.423) using the lme4 package.

RESULTS

Search Results and Inclusion of Primary Data

For the main IPDMA, of 5,248 unique titles and abstracts identified from the database search, 5,039 were excluded after title and abstract review and 113 after full-text (SupplementaryList1), leaving 96 eligible articles with data from 69 unique participant samples (SupplementaryFigure1). Of the 69 unique samples, 55 contributed data (80%). In addition, authors of included studies contributed data from three unpublished studies, for a total of 58 PHQ-9 datasets contributed to our IPDMA. Four studies without PHQ-9 individual item scores

were excluded from the present study (see SupplementaryTable1b). Thus, 16,742 participants (2,097 major depression cases) from 54 studies were analyzed (78% of 21,572 participants from the 69 eligible published studies and 3 eligible unpublished studies). Included study characteristics are shown in SupplementaryTable1a. Characteristics of eligible studies that did not provide data, including the 4 studies excluded because they only provided PHQ-9 total scores, are shown in SupplementaryTable1b.

There were 27 included primary studies that used semi-structured interviews to assess major depression (6,362 participants), 13 that used fully structured interviews other than the MINI (7,596 participants), and 14 that used the MINI (2,784 participants). The Structured Clinical Interview for DSM Disorders (SCID) was the most commonly used semi-structured interview (24 studies, 4,378 participants), and the CIDI was the most commonly used fully structured interview (10 studies, 6,291 participants). The average study sample size and number of major depression cases was 236 and 29 for studies that used a semi-structured interview; 584 and 61 for studies that used a fully structured interview; and 199 and 37 for studies that used the MINI.

Participant characteristics are shown in Table 1.

PHQ-8 and PHQ-9 Scores

Among all participants in all studies, the mean (standard deviation [SD]) PHQ-8 score (range = 0–24) was 5·3 (5·2), and the mean (SD) PHQ-9 score (range 0–27) was 5·4 (5·4). Overall, 11·8% of participants had a non-zero score on Item 9 (score of 1-3). As shown in Table 2, this included 1·9% among participants with PHQ-8 scores 0-4 and increased to 64·7% among those with scores 20-24. The correlation (95% CI) between PHQ-8 and PHQ-9 scores was 0·996

(0·996, 0·996). The correlation of the score of Item 9 with PHQ-8 scores was 0·480 (0·469, 0·492).

Diagnostic Accuracy of the PHQ-8 and PHQ-9

ROC curves comparing sensitivity and specificity estimates for cutoffs 5-15 between the PHQ-8 and PHQ-9 for the three reference standard categories, separately, are shown in Figure 1. The curves for the PHQ-8 and PHQ-9 were highly overlapping for each reference standard, and the area under the curve for the PHQ-8 and PHQ-9 were similar for semi-structured interviews (0·930 versus 0·933), fully structured interviews (excluding the MINI; 0·852 versus 0·856), and the MINI (0·894 versus 0·899).

Comparisons of sensitivity and specificity estimates between PHQ-8 and PHQ-9 cutoffs 5-15 across the three reference standard categories are shown in Table 3. Cutoff 10 maximized combined sensitivity and specificity for PHQ-8 (sensitivity [95% CI] = 0·86 [0·80, 0·90], specificity [95% CI] = 0·86 [0·83, 0·89]) and PHQ-9 (sensitivity [95% CI] = 0·88 [0·82, 0·92], specificity [95% CI] = 0·86 [0·82, 0·88]) among studies using a semi-structured interview as the reference standard; cutoff 8 for PHQ-8 (sensitivity [95% CI] = 0·77 [0·66, 0·85], specificity [95% CI] = 0·78 [0·71, 0·84]) and PHQ-9 (sensitivity [95% CI] = 0·79 [0·68, 0·86], specificity [95% CI] = 0·77 [0·70, 0·83]) among studies using a fully structured interview; and cutoff 8 for PHQ-8 (sensitivity [95% CI] = 0·83 [0·75, 0·89], specificity [95% CI] = 0·80 [0·75, 0·84]) and PHQ-9 (sensitivity [95% CI] = 0·86 [0·77, 0·91], specificity [95% CI] = 0·79 [0·74, 0·83]) among studies using the MINI.

In comparisons stratified by reference standard, for sensitivity, results of equivalence tests showed that for semi-structured diagnostic interviews, estimates were equivalent from cutoffs 5 through 9 and indeterminate from cutoffs 10 through 15; for fully structured interviews

(excluding the MINI), they were equivalent on cutoffs 5 and 7 and indeterminate at cutoffs 6 and 8 through 15; and for the MINI, they were equivalent from cutoffs 5 through 7 and indeterminate from cutoffs 8 through 15. Estimates of specificity were equivalent in all analyses, regardless of reference standards and cutoffs. See Table 3.

Overall, including all 54 primary studies, as shown in Table 4, across cutoffs, sensitivity was between 0·00 and 0·05 percentage points lower for the PHQ-8 compared to the PHQ-9. At cutoff 10, the difference (95% CI) was -0·03 (-0·06, -0·02). For specificity, the PHQ-8 and PHQ-9 were within 0·01 for all cutoffs. For sensitivity, estimates were equivalent for cutoffs 5 to 8 and indeterminate for cutoffs 9 to 15. For specificity, estimates were equivalent for all cutoffs.

A forest plot of the difference of sensitivity and specificity estimates for cutoff 10 between PHQ-8 and PHQ-9 for all studies is shown in Figure 2. At the commonly used cutoff of 10 or greater, there was low heterogeneity in the differences across the 54 studies with estimated inter-study heterogeneity (τ^2) <0·01 for sensitivity and <0·01 for specificity. Forest plots of the differences of sensitivity and specificity estimates for cutoff 10 between PHQ-8 and PHQ-9 among studies by reference standard category are shown in SupplementaryFigure2.

DISCUSSION

In the present study, we assessed the correlation of continuous PHQ-8 and PHQ-9 scores for assessing depression severity in research and clinical practice, and we compared the diagnostic accuracy of the PHQ-8 and PHQ-9 across all cutoffs for detecting major depression. There were two main findings. First, the correlation of continuous PHQ-8 and PHQ-9 scores was high (0·996). Second, to screen for major depression, the PHQ-8 at different possible cutoffs including the standard cutoff of 10 or greater, was similarly accurate compared to the PHQ-9 overall and across all three types of reference standards. The cutoffs that maximized combined

683 sensitivity and specificity were the same for the PHQ-8 and PHQ-9 across reference standard
684 categories.

685 Overall, for all 54 primary studies combined, across all cutoffs, the PHQ-8 was slightly
686 less sensitive than the PHQ-9 by 0·00 to 0·05 (0·03 at cutoff 10). For specificity, the differences
687 between PHQ-8 and PHQ-9 were within 0·01 for all cutoffs. Although the CIs for the difference
688 in sensitivity for cutoff 10 did not fit the study definition of equivalency, the reduction in
689 sensitivity if the PHQ-8 is used is small, and specificity is equivalent.

690 Previous studies have shown that Item 9 of the PHQ-9 does not accurately assess suicide
691 risk and identifies far more patients or study participants as at risk than would be identified with
692 items designed to assess suicide risk (Corson, Gerrity & Dobscha, 2004; Razykov et al. 2012;
693 Walker et al. 2011). Thus, unintended consequences of using Item 9 could include substantial
694 additional costs for research, as well as possible harms or inconvenience to patients. Research
695 ethics boards sometimes require follow-up for all patients with positive responses to Item 9.
696 Using the PHQ-8, which is minimally different from PHQ-9 in terms of diagnostic accuracy
697 characteristics, would reduce unintended consequences of false signals of suicide risk without
698 substantive changes to continuous measurement properties or diagnostic accuracy for major
699 depression.

700 It is possible that use of the PHQ-8 could result in not identifying a small subset of
701 people with suicidal thoughts, although, if the case, based on our findings, this number would be
702 small. Furthermore, there is no evidence that using questionnaires to screen for suicide in general
703 medical settings, above and beyond screening for depression, would reduce risk of suicide
704 (Allaby 2010; Crawford et al. 2011; Siu & the US Preventive Services Task Force 2016). Tools
705 are available to screen patients or stratify by risk for suicidality. However, a review concluded

706 that they are not accurate enough at this point for use in practice and that alternative methods are
707 more appropriate (Carter & Spittal 2018). Indeed, in mental health settings or when there is
708 reason to suspect possible suicidality, standards of care indicate that engagement with patients is
709 needed to assess suicide risk and determine the best management plan, as appropriate (Carter &
710 Spittal 2018).

711 To our knowledge, this is the first meta-analysis and also the first study using a large IPD
712 database to compare diagnostic accuracy characteristics of PHQ-8 and PHQ-9. Strengths of this
713 study included the large overall sample size, the ability to compare results for PHQ-8 and PHQ-9
714 from all cutoffs from all studies (rather than just published cutoff results), and the ability to
715 assess diagnostic accuracy separately in studies that used semi- and fully structured diagnostic
716 interviews as the reference standard. There are also limitations to consider. First, for the full
717 IPDMA data, we were unable to include primary data from 14 of 69 published eligible datasets
718 (20% datasets; 17% of eligible participants), and we restricted our analyses to those with
719 complete data for all individual PHQ-9 item scores (95% of available data). Nonetheless, this
720 sample was much larger than the few previous primary studies that have compared the PHQ-8
721 and PHQ-9. Second, we categorized studies based on the diagnostic interview that was used, but
722 adaptations to interviews are sometimes made and, thus, all studies may have not used the
723 diagnostic interviews in the way that they were originally designed. However, when we analyzed
724 data from all studies, regardless of reference standard, heterogeneity was minimal, suggesting
725 that findings can be applied across reference standards.

726 **CONCLUSIONS**

727 In summary, although the PHQ-9 was designed to reflect the 9 symptoms included in DSM
728 criteria for major depression, the item assessing suicide risk also assesses self-harm. This study

used a large IPD dataset and found that the PHQ-8 performs similarly to the PHQ-9 in terms of the correlation of continuous scores and the diagnostic accuracy across all cutoffs for detecting major depression. Removing Item 9 and using the PHQ-8 instead of the PHQ-9 has minimal influence on performance of the measure and will likely reduce the number of false positive signals from people who endorse this item but would not be considered to be at risk for suicide based on measures intended to assess suicide risk.

Acknowledgements

This study was funded by the Canadian Institutes of Health Research (KRS-134297; PCG-155468). Dr. Wu was supported by an Utting Postdoctoral Fellowship from the Jewish General Hospital, Montreal, Quebec. Drs. Benedetti and Thombs were supported by Fonds de recherche du Québec - Santé (FRQS) researcher salary awards. Ms. Levis was supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award. Ms. Riehm and Ms. Saadat were supported by CIHR Frederick Banting and Charles Best Canadian Graduate Scholarships – Master's Awards. Mr. Levis and Ms. Azar were supported by FRQS Masters Training Awards. Ms. Rice was supported by a Vanier Canada Graduate Scholarship. Collection of data for the study by Arroll et al. was supported by a project grant from the Health Research Council of New Zealand. Data collection for the study by Ayalon et al. was supported from a grant from Lundbeck International. The primary study by Khamseh et al. was supported by a grant (M-288) from Tehran University of Medical Sciences. 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). Dr. Butterworth was supported by Australian Research Council Future Fellowship FT130101444. 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 studies by Amoozegar and by Fiest et al. were funded by the Alberta Health Services, the University of Calgary Faculty of Medicine, and the Hotchkiss Brain Institute. The primary study by Fischer et al. was funded by the German Federal Ministry of Education and Research (01GY1150). Dr. Fischler was supported by a grant from the Belgian Ministry of Public Health and Social Affairs and a restricted grant from Pfizer Belgium. Data for the primary study by Gelaye et al. was supported by grant from the NIH (T37 MD001449). Collection of data for the primary study by Gjerdingen et al. was supported by grants from the NIMH (R34 MH072925, K02 MH65919, P30 DK50456). The primary study by Eack et al. was funded by the NIMH (R24 MH56858). Collection of data provided by Drs. Härter and Reuter was supported by the Federal Ministry of Education and Research (grants No. 01 GD 9802/4 and 01 GD 0101) and by the Federation of German Pension Insurance Institute. Collection of data for the primary study by Hobfoll et al. was made possible in part from grants from NIMH (RO1 MH073687) and the Ohio Board of Regents. Dr. Hall received support from a grant awarded by the Research and Development Administration Office, University of Macau (MYRG2015-00109-FSS). The primary study by Hides et al. was funded by the Perpetual Trustees, Flora and Frank Leith Charitable Trust, Jack Brockhoff Foundation, Grosvenor Settlement, Sunshine Foundation and Danks Trust. The primary study by Henkel et al. was funded by the German Ministry of Research and Education. Data for the study by Razykov et al. was collected by the Canadian Scleroderma Research Group, which was funded by the CIHR (FRN 83518), the Scleroderma Society of Canada, the Scleroderma Society of Ontario, the Scleroderma Society of Saskatchewan, Sclérodermie Québec, the Cure Scleroderma Foundation, Inova Diagnostics Inc., Euroimmun, FRQS, the Canadian Arthritis Network, and the Lady Davis

Institute of Medical Research of the Jewish General Hospital, Montreal, QC. Dr. Hudson was supported by a FRQS Senior Investigator Award. Collection of data for the primary study by Hyphantis et al. was supported by grant from the National Strategic Reference Framework, European Union, and the Greek Ministry of Education, Lifelong Learning and Religious Affairs (ARISTEIA-ABREVIATE, 1259). The primary study by Inagaki et al. was supported by the Ministry of Health, Labour and Welfare, Japan. Dr. Jetté was supported by a Canada Research Chair in Neurological Health Services Research. Collection of data for the primary study by Kiely et al. was supported by National Health and Medical Research Council (grant number 1002160) and Safe Work Australia. Dr. Kiely was supported by funding from an Australian National Health and Medical Research Council fellowship (grant number 1088313). The primary study by Lamers et al. was funded by the Netherlands Organisation for Health Research and development (grant number 945-03-047). The primary study by Liu et al. 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 Santos et al. was funded by the National Program for Centers of Excellence (PRONEX/FAPERGS/CNPq, Brazil). The primary study by Muramatsu et al. was supported by an educational grant from Pfizer US Pharmaceutical Inc. Collection of primary data for the study by Dr. Pence was provided by NIMH (R34MH084673).

806 The primary studies by Osório et al. were funded by Reitoria de Pesquisa da Universidade de
807 São Paulo (grant number 09.1.01689.17.7) and Banco Santander (grant number
808 10.1.01232.17.9). Dr. Osório was supported by Productivity Grants (PQ-CNPq-2 -number
809 301321/2016-7). The primary study by Picardi et al. was supported by funds for current research
810 from the Italian Ministry of Health. Dr. Persoons was supported by a grant from the Belgian
811 Ministry of Public Health and Social Affairs and a restricted grant from Pfizer Belgium. Dr.
812 Shaaban was supported by funding from Universiti Sains Malaysia. The primary study by
813 Rooney et al. was funded by the United Kingdom National Health Service Lothian Neuro-
814 Oncology Endowment Fund. The primary study by Sidebottom et al. was funded by a grant from
815 the United States Department of Health and Human Services, Health Resources and Services
816 Administration (grant number R40MC07840). Simning et al.'s research was supported in part by
817 grants from the NIH (T32 GM07356), Agency for Healthcare Research and Quality (R36
818 HS018246), NIMH (R24 MH071604), and the National Center for Research Resources (TL1
819 RR024135). Dr. Stafford received PhD scholarship funding from the University of Melbourne.
820 Collection of data for the studies by Turner et al were funded by a bequest from Jennie Thomas
821 through the Hunter Medical Research Institute. Collection of data for the primary study by
822 Williams et al. was supported by a NIMH grant to Dr. Marsh (RO1-MH069666). The primary
823 study by Thombs et al. was done with data from the Heart and Soul Study (PI Mary Whooley).
824 The Heart and Soul Study was funded by the Department of Veterans Epidemiology Merit
825 Review Program, the Department of Veterans Affairs Health Services Research and
826 Development service, the National Heart Lung and Blood Institute (R01 HL079235), the
827 American Federation for Aging Research, the Robert Wood Johnson Foundation, and the
828 Ischemia Research and Education Foundation. Dr. Thombs was supported by an Investigator

Award from the Arthritis Society. 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 study by Wittkampf et al. was funded by The Netherlands Organization for Health Research and Development (ZonMw) Mental Health Program (nos. 100.003.005 and 100.002.021) and the Academic Medical Center/University of Amsterdam. Collection of data for the primary study by Zhang et al. was supported by the European Foundation for Study of Diabetes, the Chinese Diabetes Society, Lilly Foundation, Asia Diabetes Foundation and Liao Wun Yuk Diabetes Memorial Fund. No other authors reported funding for primary studies or for their work on the present study.

Conflict of Interest.

Drs. Jetté and Patten declare that they received a grant, outside the submitted work, from the University of Calgary Hotchkiss Brain Institute, which was jointly funded by the Institute and Pfizer. Pfizer was the original sponsor of the development of the PHQ-9, which is now in the public domain. Dr. Chan is a steering committee member or consultant of Astra Zeneca, Bayer, Lilly, MSD and Pfizer. She has received sponsorships and honorarium for giving lectures and providing consultancy and her affiliated institution has received research grants from these companies. Dr. Hegerl declares that within the last three years, he was an advisory board member for Lundbeck and Servier; a consultant for Bayer Pharma; a speaker for Roche Pharma and Servier; and received personal fees from Janssen, all outside the submitted work. 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. 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.

Author Contributions.

YW, BLevis, JB, PC, SG, JPAI, LAK, DM, SBP, IS, RCZ, AB, and BDT were responsible for the study conception and design. JB and LAK designed and conducted database searches to identify eligible studies. DHA, BA, LA, HRB, MB, CHB, PB, GC, MHC, JCNC, RC, YC, JMG, JRF, FHF, DF, BG, FGS, CGG, BJH, PAH, MHärter, UH, LH, SEH, MHudson, TH, MI, NJ, MEK, KMK, YK, FL, SL, ML, SRL, BLöwe, AM, SM, TNM, KM, FLO, VP, BWP, PP, AP, KR, AGR, ISS, JS, ASidebottom, ASimning, LS, SS, PLLT, AT, HCvW, JW, MAW, KW, MY, and BDT were responsible for collection of primary data included in this study. BLevis, KER, NS, AWL, MA, DBR, and BDT contributed to data extraction and coding for the meta-analysis. YW, BLevis, AB, and BDT contributed to the data analysis and interpretation. YW, BLevis, AB, and BDT contributed to drafting the manuscript. All authors provided a critical review and approved the final manuscript. AB and BDT are guarantors.

REFERENCES

Allaby M. (2010). *Screening for depression: A report for the UK National Screening Committee* (Revised report). UK National Screening Committee: London, United Kingdom.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Barrera TL, Cummings JP, Armento M, Cully JA, Bush Amspoker A, Wilson NL, Mallen MJ, Shrestha S, Kunik ME, Stanley MA (2017). Telephone-delivered cognitive-behavioral therapy for older, rural veterans with depression and anxiety in home-based primary care. *Clinical gerontologist* **40**, 114-123.

Brugha TS, Bebbington PE, Jenkins R (1999). A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychological Medicine* **29**, 1013-1020.

Canadian Agency for Drugs and Technologies in Health. (2016). *PRESS – Peer review of electronic search strategies: 2015 Guideline explanation and elaboration* (PRESS E&E). CADTH: Ottawa.

Carter G, Spittal MJ (2018). Suicide risk assessment: Risk stratification is not accurate enough to be clinically useful and alternative approaches are needed. *Crisis* **39**, 229.

Corson K, Gerrity MS, Dobscha SK (2004). Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *The American Journal of Managed Care* **10**, 839-845.

Crawford MJ, Thana L, Methuen C, Ghosh P, Stanley SV, Ross J, Gordon F, Blair G, Bajaj P (2011). Impact of screening for risk of suicide: randomised controlled trial. *British Journal of Psychiatry* **198**, 379-384.

Fagerland MW, Lydersen S, Laake P (2014). Recommended tests and confidence intervals for paired binomial proportions. *Statistics in Medicine* **33**, 2850-2875.

Higgins JP, Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539-1558.

Ishihara M, Harel D, Levis B, Levis AW, Riehm KE, Saadat N, Azar M, Rice DB, Sanchez TA, Chiovitti MJ, Cuijpers P (2019). Shortening self-report mental health symptom measures through optimal test assembly methods: Development and validation of the Patient Health Questionnaire-Depression-4. *Depression and anxiety* **36**, 82-92.

Kroenke K, Spitzer RL, Williams JBW (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606-613.

Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders* **114**, 163-173.

Levis B, Benedetti A, Riehm KE, Saadat N, Levis AW, Azar M, Rice DB, Chiovitti MJ, Sanchez TA, Cuijpers P, Gilbody S (2018). Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *British Journal of Psychiatry* **212**, 377-385.

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

Moriarty AS, Gilbody S, McMillan D, Manea L (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *General Hospital Psychiatry* **37**, 567-576.

Nosen E, Woody SR (2008). Chapter 8: Diagnostic assessment in research. In *Handbook of Research Methods in Abnormal and Clinical Psychology* (ed. D. McKay), pp. 109-124. Sage: Thousand Oaks.

Razykov I, Ziegelstein R, Whooley M, Thombs BD (2012). The PHQ-9 versus the PHQ-8 – Is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the heart and Soul study. *Journal of Psychosomatic Research* **73**, 163-168.

Rice DB, Kloda LA, Levis B, Qi B, Kingsland E, Thombs BD (2016). Are MEDLINE searches sufficient for systematic reviews and meta-analyses of the diagnostic accuracy of depression screening tools? A review of meta-analyses. *Journal of Psychosomatic Research* **87**, 7-13.

Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR (2008). Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Statistics in Medicine* **27**, 6111-6136.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* **45**, 1069-1077.

Sheehan DV, Lecrubier Y, Sheehan KH, Janavs J, Weiller E, Keskiner A, Schinka J, Knapp E, Sheehan MF, Dunbar GC (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry* **12**, 232-241.

Simon GE, Coleman KJ, Rossom RC, Beck A, Oliver M, Johnson E, Whiteside U, Operskalski B, Penfold RB, Shortreed SM, Rutter C (2016). Risk of suicide attempt and suicide death following completion of the Patient Health Questionnaire depression module in community practice. *The Journal of Clinical Psychiatry* **77**, 221-227.

Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, Ludman EJ (2013). Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death?. *Psychiatric Services* **64**, 1195-1202.

Suarez L, Beach SR, Moore SV, Mastromauro CA, Januzzi JL, Celano CM, Chang TE, Huffman JC (2015). Use of the Patient Health Questionnaire-9 and a detailed suicide evaluation in determining imminent suicidality in distressed patients with cardiac disease. *Psychosomatics* **56**,181-189.

Siu AL, and the US Preventive Services Task Force (USPSTF) (2016). Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA: The Journal of the American Medical Association* **315**, 380-387.

Thombs BD, Benedetti A, Kloda LA, Levis B, Nicolau I, Cuijpers P, Gilbody S, Ioannidis JP, McMillan D, Patten SB, Shrier I (2014). The diagnostic accuracy of the Patient Health Questionnaire-2 (PHQ-2), Patient Health Questionnaire-8 (PHQ-8), and Patient Health Questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *Systematic Reviews* **3**, 124.

Walker J, Hansen CH, Butcher I, Sharma N, Wall L, Murray G, Sharpe M (2011). Thoughts of death and suicide reported by cancer patients who endorsed the "suicidal thoughts" item of the PHQ-9 during routine screening for depression. *Psychosomatics* **52**, 424-427.

Walker E, Nowacki AS (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine* **26**, 192-196.

Wells TS, Horton JL, LeardMann CA, Jacobson IG, Boyko EJ (2013). A comparison of the PRIME-MD PHQ-9 and PHQ-8 in a large military prospective study, the Millennium Cohort Study. *Journal of Affective Disorders* **148**, 77-83.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* **155**, 529-536.

van der Leeden R, Busing FMTA, Meijer E. (1997). *Bootstrap methods for two-level models. Technical report PRM 97-04.* Leiden University, Department of Psychology: Leiden, The Netherlands.

van der Leeden R, Meijer E, Busing FMTA (2008). Chapter 11: Resampling multilevel models. In *Handbook of Multilevel Analysis* (ed. J. Leeuw, E. Meijer), pp. 401-433. Springer: New York.

Table 1. Participant characteristics by subgroup

Participant Subgroup	N Participants	N (%) Major Depression
All participants	16,742	2,097 (13)
Type of diagnostic interview		
Semi-structured diagnostic interview	6,362	790 (12)
Fully structured diagnostic interview	7,596	790 (10)
Mini International Neuropsychiatric Interview	2,784	517 (19)
Age[a]		
< 60	11,144	1,402 (13)
≥ 60	5,552	692 (12)
Sex[a]		
Women	9,552	1,259 (13)
Men	7,180	835 (12)
Care setting		
Non-medical care	1,832	252 (14)
Primary care	7,846	760 (10)
Inpatient specialty care	1,245	136 (11)
Outpatient specialty care	5,819	949 (16)
Country human development index		
Very high	13,297	1,577 (12)
High	1,337	276 (21)
Low-medium	2,108	244 (12)

[a] Due to missing participant data, total participant numbers for these variables were <16,742.

Table 2. Characteristics of participants who rated Item 9 as present for several days, more than half the days, or nearly every day (i.e., scores 1-3) in last two weeks by total PHQ-8 score

Total PHQ-8 Score	N of participants[a]	% with non-zero Item 9	Item 9 Mean (SD)
0-4	11,034	1·9%	0·02 (0·16)
5-9	5,071	13·2%	0·16 (0·45)
10-14	2,231	31·3%	0·44 (0·74)
15-19	1,044	48·3%	0·78 (0·97)
20-24	380	64·7%	1·39 (1·25)

[a] Numbers of participants add up to >16,742 as they were weighted by sampling weights.

Table 3a. Comparison of sensitivity and specificity estimates between PHQ-8 and PHQ-9 among studies that used a semi-structured reference standard

	PHQ-8[a]				PHQ-9				PHQ-8 – PHQ-9[b]			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0·98	(0·95, 0·99)	0·55	(0·50, 0·60)	0·98	(0·95, 0·99)	0·55	(0·50, 0·60)	0·00	(-0·01, 0·00)	0·00	(0·00, 0·01)
6	0·98	(0·95, 0·99)	0·63	(0·58, 0·68)	0·98	(0·95, 0·99)	0·63	(0·58, 0·67)	0·00	(-0·00, 0·00)	0·00	(0·00, 0·01)
7	0·97	(0·93, 0·99)	0·70	(0·66, 0·74)	0·98	(0·93, 0·99)	0·70	(0·65, 0·74)	-0·01	(-0·02, 0·00)	0·00	(0·00, 0·01)
8	0·94	(0·89, 0·96)	0·76	(0·72, 0·79)	0·95	(0·90, 0·97)	0·75	(0·71, 0·79)	-0·01	(-0·03, 0·00)	0·01	(0·00, 0·01)
9	0·89	(0·84, 0·92)	0·81	(0·78, 0·84)	0·91	(0·87, 0·95)	0·80	(0·77, 0·83)	-0·02	(-0·06, -0·00)	0·01	(0·00, 0·01)
10[b]	0·86	(0·80, 0·90)	0·86	(0·83, 0·89)	0·88	(0·82, 0·92)	0·86	(0·82, 0·88)	-0·02	(-0·06, -0·00)	0·00	(0·00, 0·02)
11	0·81	(0·75, 0·87)	0·90	(0·87, 0·92)	0·84	(0·84, 0·84)	0·89	(0·89, 0·89)	-0·03	(-0·06, -0·00)	0·01	(0·00, 0·02)
12	0·74	(0·68, 0·79)	0·92	(0·89, 0·93)	0·78	(0·71, 0·83)	0·91	(0·89, 0·93)	-0·04	(-0·09, -0·01)	0·01	(0·00, 0·01)
13	0·67	(0·60, 0·73)	0·94	(0·92, 0·95)	0·69	(0·63, 0·75)	0·93	(0·91, 0·95)	-0·02	(-0·07, -0·00)	0·01	(0·00, 0·01)
14	0·59	(0·53, 0·65)	0·96	(0·94, 0·97)	0·64	(0·57, 0·70)	0·95	(0·93, 0·96)	-0·05	(-0·09, -0·01)	0·01	(0·00, 0·01)
15	0·51	(0·44, 0·57)	0·97	(0·95, 0·98)	0·55	(0·48, 0·62)	0·96	(0·94, 0·97)	-0·04	(-0·09, -0·02)	0·01	(0·00, 0·01)

[a] N Studies = 27; N Participants = 6,362; N major depression = 790

[b] For PHQ-8 cutoff 10, among studies that used semi-structured interviews as the reference standard, the default optimizer in glmer failed to converge, thus bobyqa was used instead.

CI: confidence interval

Table 3b. Comparison of sensitivity and specificity estimates between PHQ-8 and PHQ-9 among studies that used a fully structured reference standard (MINI excluded)

	PHQ-8[a]				PHQ-9				PHQ-8 – PHQ-9			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0·92	(0·85, 0·96)	0·57	(0·49, 0·66)	0·93	(0·85, 0·96)	0·57	(0·48, 0·65)	-0·01	(-0·03, 0·00)	0·00	(0·00, 0·02)
6	0·88	(0·79, 0·93)	0·65	(0·57, 0·73)	0·90	(0·81, 0·95)	0·65	(0·56, 0·72)	-0·02	(-0·07, 0·00)	0·00	(0·00, 0·02)
7	0·83	(0·73, 0·90)	0·72	(0·64, 0·79)	0·84	(0·73, 0·90)	0·71	(0·63, 0·78)	-0·01	(-0·01, 0·00)	0·01	(0·00, 0·02)
8	0·77	(0·66, 0·85)	0·78	(0·71, 0·84)	0·79	(0·68, 0·86)	0·77	(0·70, 0·83)	-0·02	(-0·07, -0·00)	0·01	(0·00, 0·01)
9	0·69	(0·59, 0·77)	0·83	(0·76, 0·87)	0·71	(0·62, 0·80)	0·81	(0·75, 0·86)	-0·02	(-0·07, -0·00)	0·02	(0·01, 0·02)
10	0·63	(0·52, 0·72)	0·86	(0·81, 0·90)	0·67	(0·57, 0·76)	0·85	(0·80, 0·90)	-0·04	(-0·09, -0·01)	0·01	(0·00, 0·02)
11	0·57	(0·45, 0·67)	0·89	(0·85, 0·93)	0·59	(0·49, 0·69)	0·88	(0·84, 0·92)	-0·02	(-0·07, -0·01)	0·01	(0·00, 0·02)
12	0·51	(0·38, 0·64)	0·92	(0·88, 0·94)	0·54	(0·43, 0·65)	0·90	(0·86, 0·93)	-0·03	(-0·16, -0·01)	0·02	(0·01, 0·02)
13	0·43	(0·32, 0·55)	0·94	(0·91, 0·96)	0·47	(0·36, 0·58)	0·93	(0·89, 0·95)	-0·04	(-0·12, -0·01)	0·01	(0·00, 0·01)
14	0·36	(0·26, 0·47)	0·95	(0·93, 0·97)	0·41	(0·31, 0·53)	0·95	(0·92, 0·96)	-0·05	(-0·14, -0·01)	0·00	(0·00, 0·01)
15	0·30	(0·22, 0·39)	0·96	(0·95, 0·98)	0·33	(0·24, 0·42)	0·96	(0·94, 0·97)	-0·03	(-0·07, -0·00)	0·00	(0·00, 0·00)

[a] N Studies = 13; N Participants = 7,596; N major depression = 790

CI: confidence interval

Table 3c. Comparison of sensitivity and specificity estimates between PHQ-8 and PHQ-9 among studies that used the MINI reference standard

	PHQ-8[a]				PHQ-9				PHQ-8 – PHQ-9			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0·96	(0·93, 0·98)	0·58	(0·50, 0·65)	0·97	(0·93, 0·98)	0·57	(0·49, 0·65)	-0·01	(-0·01, 0·00)	0·01	(0·00, 0·02)
6	0·92	(0·85, 0·96)	0·67	(0·59, 0·74)	0·93	(0·86, 0·97)	0·66	(0·59, 0·73)	-0·01	(-0·03, 0·00)	0·01	(0·00, 0·02)
7	0·89	(0·81, 0·94)	0·73	(0·67, 0·79)	0·89	(0·81, 0·94)	0·73	(0·66, 0·79)	0·00	(-0·03, 0·00)	0·01	(0·00, 0·01)
8	0·83	(0·75, 0·89)	0·80	(0·75, 0·84)	0·86	(0·77, 0·91)	0·79	(0·74, 0·83)	-0·03	(-0·06, 0·00)	0·01	(0·00, 0·02)
9	0·78	(0·69, 0·85)	0·85	(0·81, 0·89)	0·81	(0·71, 0·88)	0·84	(0·80, 0·88)	-0·03	(-0·07, -0·00)	0·01	(0·00, 0·02)
10	0·72	(0·63, 0·79)	0·88	(0·84, 0·91)	0·75	(0·66, 0·82)	0·88	(0·84, 0·91)	-0·03	(-0·08, -0·01)	0·01	(0·00, 0·02)
11	0·65	(0·57, 0·73)	0·91	(0·88, 0·94)	0·69	(0·61, 0·77)	0·90	(0·87, 0·93)	-0·04	(-0·08, -0·01)	0·01	(0·00, 0·02)
12	0·59	(0·51, 0·66)	0·93	(0·91, 0·95)	0·65	(0·56, 0·73)	0·92	(0·90, 0·94)	-0·06	(-0·11, -0·02)	0·01	(0·00, 0·02)
13	0·53	(0·44, 0·62)	0·95	(0·93, 0·97)	0·57	(0·49, 0·66)	0·94	(0·92, 0·96)	-0·04	(-0·09, -0·01)	0·01	(0·00, 0·02)
14	0·43	(0·35, 0·51)	0·97	(0·95, 0·98)	0·49	(0·49, 0·49)	0·96	(0·96, 0·96)	-0·06	(-0·11, -0·02)	0·01	(0·00, 0·02)
15	0·37	(0·29, 0·45)	0·98	(0·96, 0·99)	0·42	(0·42, 0·42)	0·97	(0·97, 0·97)	-0·05	(-0·10, -0·02)	0·01	(0·00, 0·01)

[a] N Studies = 14; N Participants = 2,784; N major depression = 517

CI: confidence interval; MINI: Mini International Neuropsychiatric Interview

Table 4. Comparison of sensitivity and specificity estimates between PHQ-8 and PHQ-9 across cutoffs 5-15 for all studies

PHQ-8 – PHQ-9				
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	-0·01	(-0·01, 0·00)	0·00	(0·00, 0·01)
6	0·00	(-0·01, 0·00)	0·01	(0·00, 0·01)
7	-0·01	(-0·02, 0·00)	0·00	(0·00, 0·01)
8	-0·01	(-0·04, -0·01)	0·00	(0·01, 0·01)
9	-0·03	(-0·06, -0·01)	0·01	(0·01, 0·01)
10	-0·03	(-0·06, -0·02)	0·01	(0·00, 0·01)
11	-0·03	(-0·06, -0·01)	0·01	(0·01, 0·01)
12	-0·05	(-0·08, -0·03)	0·01	(0·00, 0·01)
13	-0·04	(-0·06, -0·02)	0·00	(0·00, 0·01)
14	-0·05	(-0·08, -0·03)	0·01	(0·00, 0·01)
15	-0·04	(-0·07, -0·03)	0·01	(0·00, 0·01)

[a] N Studies = 54; N Participants = 16,742; N major depression = 2,097

CI: confidence interval

Supplementary Material

SupplementaryMethods1. Search strategies

SupplementaryFigure1. Flow diagram of study selection process

SupplementaryFigure2. Forest plots of the difference in sensitivity and specificity estimates at cutoff 10 between PHQ-8 and PHQ-9 for each reference standard category

SupplementaryTable1. Characteristics of included primary studies as well as eligible primary studies not included in the present study

SupplementaryList1. Studies excluded at full-text review level

SupplementaryMethods1. Search strategies

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)

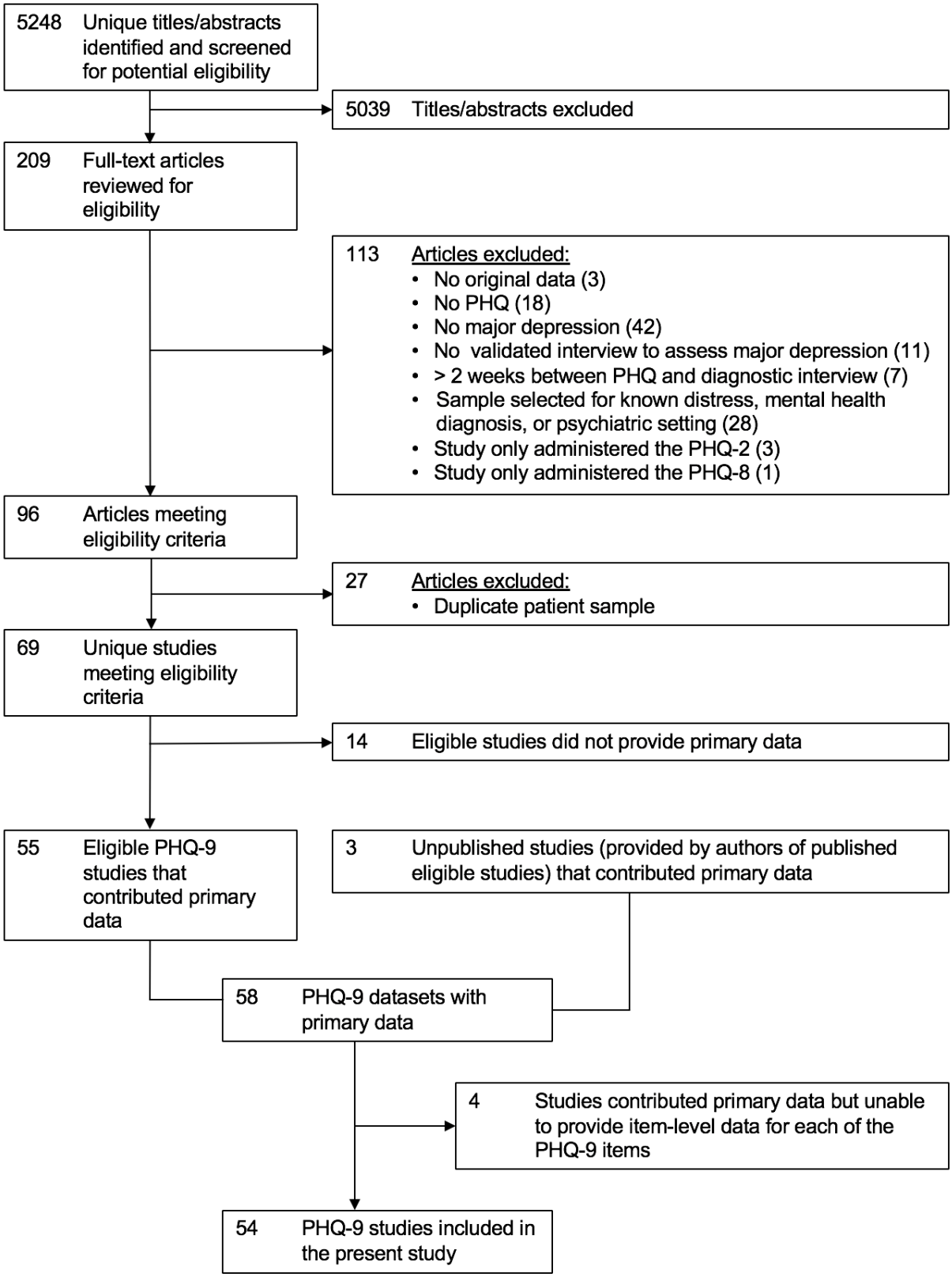
1. PHQ*.af.
2. patient health questionnaire*.af.
3. 1 or 2
4. Diagnosis/
5. Medical Diagnosis/
6. Psychodiagnosis/
7. Misdiagnosis/
8. Screening/
9. Health Screening/
10. Screening Tests/
11. Prediction/
12. Cutting Scores/
13. Psychometrics/
14. Test Validity/
15. screen*.af.
16. predictive value*.af.
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)

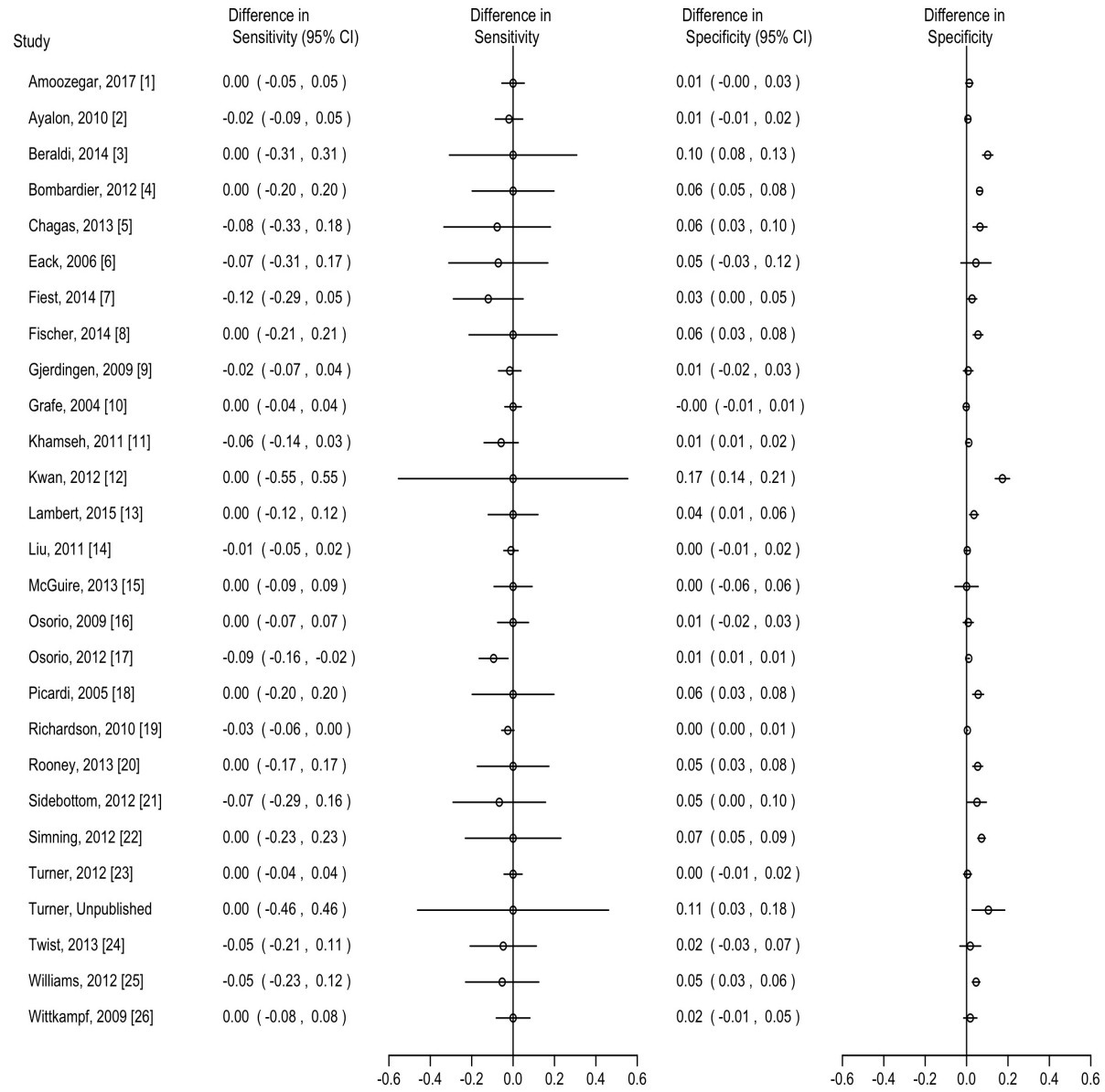
#1: TS=(PHQ* OR "Patient Health Questionnaire*")

#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=2000-2015

SupplementaryFigure1. Flow diagram of study selection process

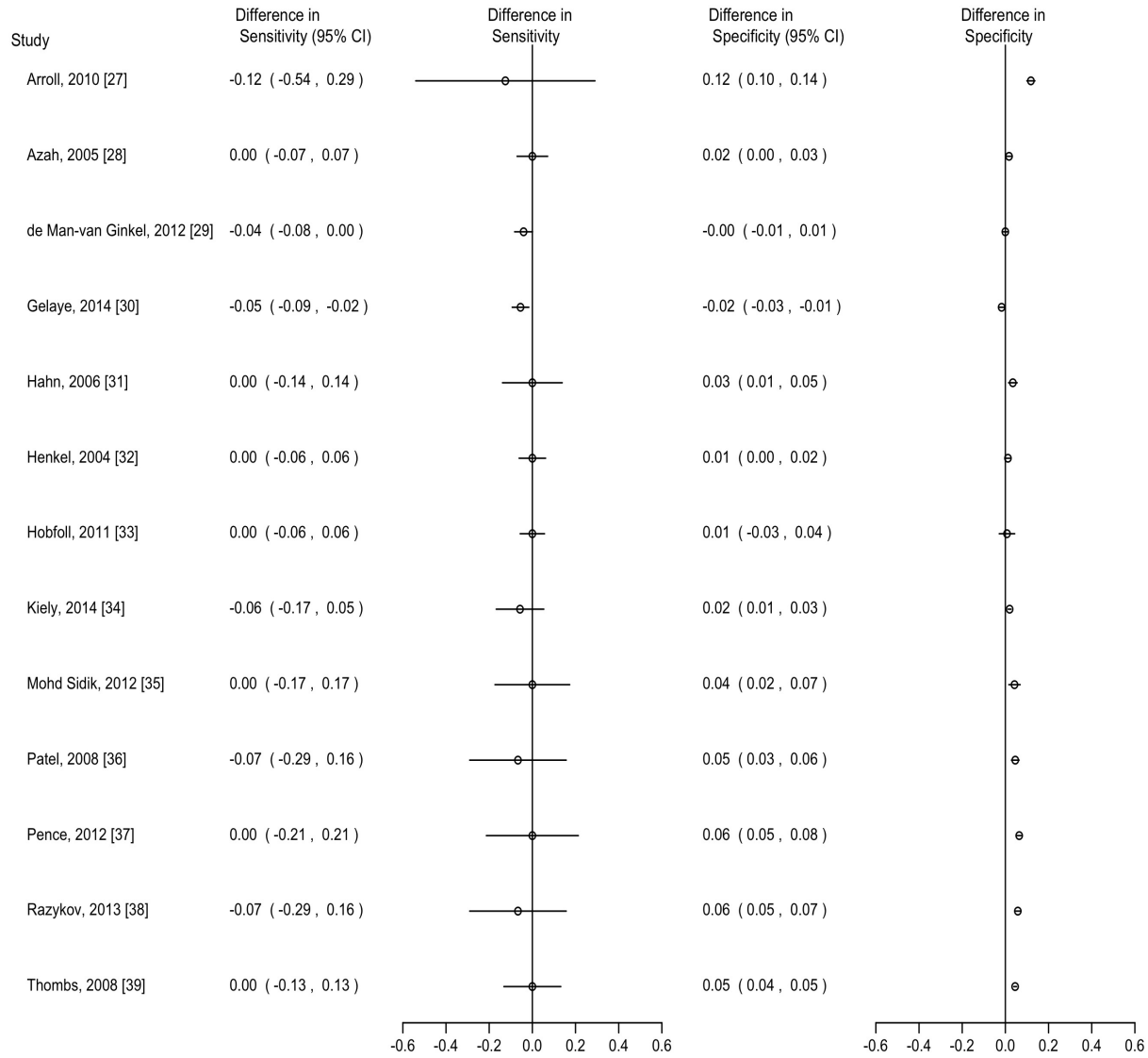


SupplementaryFigure2a. Forest plots of the difference in sensitivity and specificity estimates at cutoff 10 between PHQ-8 and PHQ-9 among studies that used a semi-structured diagnostic interview as the reference standard[a] (N Studies = 27; N Participants = 6,362; N major depression = 790)



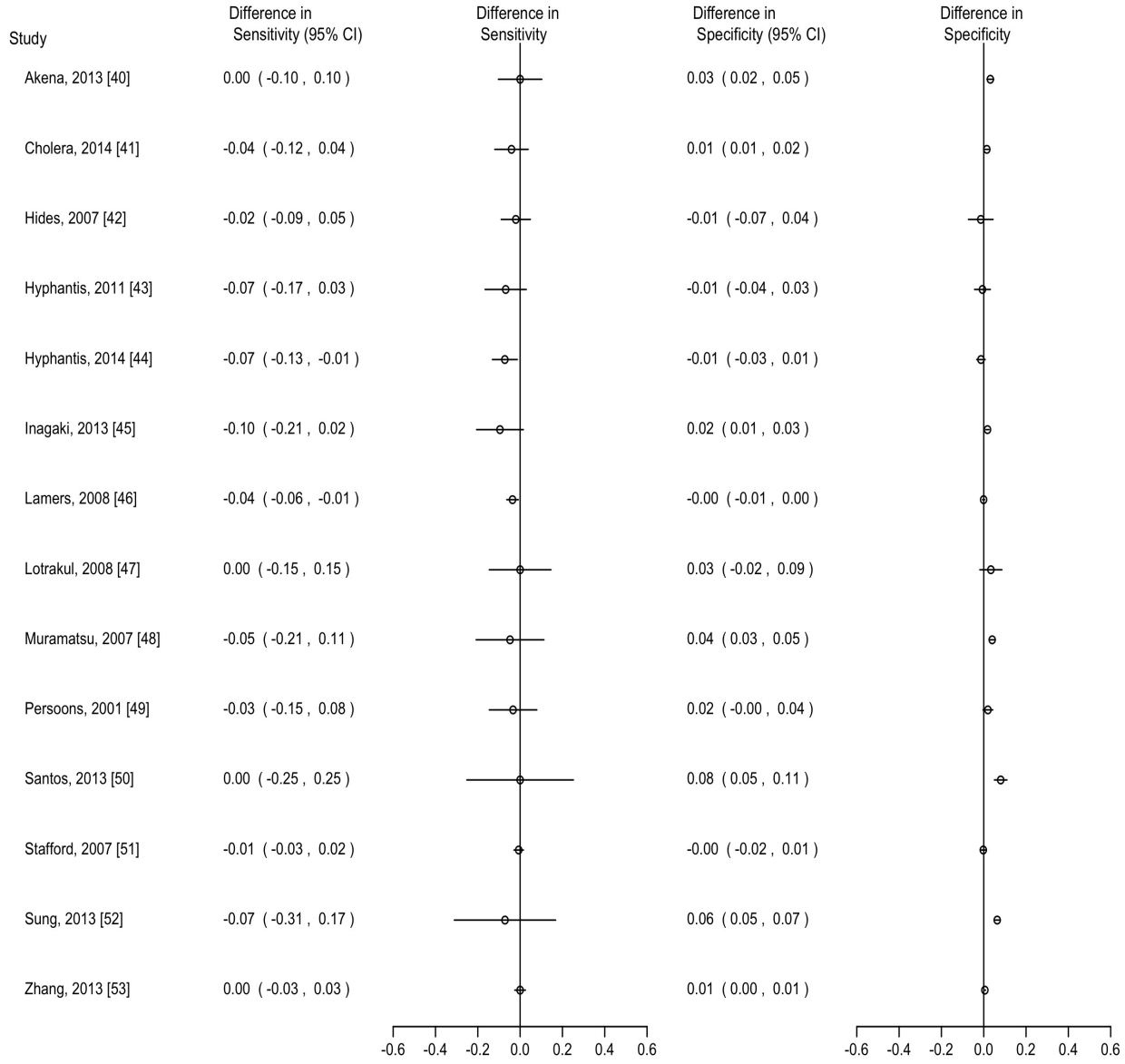
[a] τ^2 for the differences of sensitivity and specificity were both <0·001.

SupplementaryFigure2b. Forest plots of the difference in sensitivity and specificity estimates at cutoff 10 between PHQ-8 and PHQ-9 among studies that used a fully structured diagnostic interview as the reference standard[a] (N Studies = 13; N Participants = 7,596; N major depression = 790)



[a] τ^2 for the differences of sensitivity and specificity were both <0·001.

SupplementaryFigure2c. Forest plots of the difference in sensitivity and specificity estimates at cutoff 10 between PHQ-8 and PHQ-9 among studies that used the MINI as the reference standard[a] (N Studies = 14; N Participants = 2,784; N major depression = 517)



[a] τ^2 for the differences of sensitivity and specificity were both <0·001.

SupplementaryTable1a. Characteristics of included primary studies

First Author, Year	Country	Recruited Population	Diagnostic Interview	Classification System	Total N	Major Depression N (%)
Semi-structured Interviews						
Amoozegar, 2017[1a]	Canada	Migraine patients	SCID	DSM-IV	203	49 (24)
Ayalon, 2010[2]	Israel	Elderly primary care patients	SCID	DSM-IV	151	6 (4)
Beraldi, 2014[3]	Germany	Cancer inpatients	SCID	DSM-IV	116	7 (6)
Bombardier, 2012[4]	USA	Inpatients with spinal cord injuries	SCID	DSM-IV	160	14 (9)
Chagas, 2013[5]	Brazil	Outpatients with Parkinson's Disease	SCID	DSM-IV	84	19 (23)
Eack, 2006[6]	USA	Women seeking psychiatric services for their children at two mental health centers	SCID	DSM-IV	48	12 (25)
Fiest, 2014[7]	Canada	Epilepsy outpatients	SCID	DSM-IV	169	23 (14)
Fischer, 2014[8]	Germany	Heart failure patients	SCID	DSM-IV	194	11 (5)
Gjerdingen, 2009[9]	USA	Mothers registering their newborns for well-child visits at medical or pediatric clinics	SCID	DSM-IV	419	19 (5)
Gräfe, 2004[10]	Germany	Medical and psychosomatic outpatients	SCID	DSM-IV	494	67 (14)
Khamseh, 2011[11]	Iran	Type 2 diabetes patients	SCID	DSM-IV	122	47 (39)
Kwan, 2012[12]	Singapore	Post-stroke inpatients undergoing rehabilitation	SCID	DSM-IV-TR	113	3 (3)
Lambert, 2015[13a]	Australia	Cancer patients	SCID	DSM-IV	147	21 (14)
Liu, 2011[14]	Taiwan	Primary care patients	SCAN	DSM-IV	1532	50 (3)
McGuire, 2013[15]	USA	Acute coronary syndrome inpatients	DISH	DSM-IV	100	9 (9)

Osório, 2009[16]	Brazil	Women in primary care	SCID	DSM-IV	177	60 (34)
Osório, 2012[17]	Brazil	Inpatients from various clinical wards	SCID	DSM-IV	86	28 (33)
Picardi, 2005[18]	Italy	Inpatients with skin diseases	SCID	DSM-IV	138	12 (9)
Richardson, 2010[19]	USA	Older adults undergoing in-home aging services care management assessment	SCID	DSM-IV	377	95 (25)
Rooney, 2013[20]	UK	Adults with cerebral glioma	SCID	DSM-IV	126	14 (11)
Sidebottom, 2012[21]	USA	Pregnant women	SCID	DSM-IV	246	12 (5)
Simning, 2012[22]	USA	Older adults living in public housing	SCID	DSM-IV	190	10 (5)
Turner, Unpublished	Australia	Cardiac rehabilitation patients	SCID	DSM-IV	51	4 (8)
Turner, 2012[23b]	Australia	Stroke patients	SCID	DSM-IV	72	13 (18)
Twist, 2013[24d]	UK	Type 2 diabetes outpatients	SCAN	DSM-IV	352	79 (22)
Williams, 2012[25]	USA	Parkinson's Disease patients	SCID	DSM-IV	235	61 (26)
Wittkampf, 2009[26c]	The Netherlands	Primary care patients at risk for depression	SCID	DSM-IV	260	45 (17)
Fully Structured Interviews						
Arroll, 2010[27]	New Zealand	Primary care patients	CIDI	DSM-IV	2,528	156 (6)
Azah, 2005[28c]	Malaysia	Adults attending family medicine clinics	CIDI	ICD-10	180	30 (17)
de Man-van Ginkel, 2012[29]	The Netherlands	Stroke patients	CIDI	DSM-IV	164	17 (10)
Gelaye, 2014[30]	Ethiopia	Outpatients at a general hospital	CIDI	DSM-IV	923	162 (18)
Hahn, 2006[31]	Germany	Patients with chronic illnesses from rehabilitation centers	CIDI	DSM-IV	211	18 (9)
Henkel, 2004[32]	Germany	Primary care patients	CIDI	ICD-10	430	43 (10)

Hobfoll, 2011[33]	Israel	Jewish and Palestinian residents of Jerusalem exposed to war	CIDI	DSM-IV	144	42 (29)
Kiely, 2014[34]	Australia	Community sample of adults	CIDI	ICD-10	822	33 (4)
Mohd-Sidik, 2012[35d]	Malaysia	Primary care patients	CIDI	DSM-IV	146	31 (21)
Patel, 2008[36]	India	Primary care patients	CIS-R	ICD-10	299	13 (4)
Pence, 2012[37]	Cameroon	HIV-infected patients	CIDI	DSM-IV	398	11 (3)
Razykov, 2013[38]	Canada	Patients with systemic sclerosis	CIDI	DSM-IV	345	13 (4)
Thombs, 2008[39]	USA	Outpatients with coronary artery disease	C-DIS	DSM-IV	1,006	221 (22)
Mini International Neuropsychiatric Interviews (MINI)						
Akena, 2013[40]	Uganda	HIV/AIDS patients	MINI	DSM-IV	91	11 (12)
Cholera, 2014[41]	South Africa	Patients undergoing routine HIV counseling and testing at a primary health care clinic	MINI	DSM-IV	397	47 (12)
Hides, 2007[42]	Australia	Injection drug users accessing a needle and syringe program	MINI	DSM-IV	103	47 (46)
Hyphantis, 2011[43c]	Greece	Patients with various rheumatologic disorders	MINI	DSM-IV	213	69 (32)
Hyphantis, 2014[44]	Greece	Patients with chronic illnesses presenting at the emergency department	MINI	DSM-IV	349	95 (27)
Inagaki, 2013[45d]	Japan	Internal medicine outpatients	MINI	DSM-III-R	104	21 (20)
Lamers, 2008[46c]	The Netherlands	Elderly primary care patients with diabetes mellitus or chronic obstructive pulmonary	MINI	DSM-IV	104	59 (57)

		disease				
Lotrakul, 2008[47]	Thailand	Outpatients	MINI	DSM-IV	278	19 (7)
Muramatsu, 2007[48]	Japan	Primary care patients	MINI	DSM-IV	116	32 (28)
Persoons, 2001[49b]	Belgium	Inpatients and patients at gastroenterological and hepatology wards	MINI	DSM-IV	173	28 (16)
Santos, 2013[50]	Brazil	General population	MINI	DSM-IV	196	25 (13)
Stafford, 2007[51]	Australia	Inpatients with coronary artery disease who had undergone surgery	MINI	DSM-IV	193	35 (18)
Sung, 2013[52]	Singapore	Primary care patients	MINI	DSM-IV	399	12 (3)
Zhang, 2013[53]	China	Type 2 diabetes patients	MINI	DSM-IV	68	17 (25)

Abbreviations: C-DIS: Computerized Diagnostic Interview Schedule; 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; 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 the electronic database search.

[b]Multiple recruitment settings were included, thus recruitment setting was coded at the participant-level.

[c]Appropriate statistical weighting to reflect sampling procedures was included in the primary study dataset. The provided weights were used for analysis.

[d]Sampling procedures merited weighting, but the original dataset did not weight. Appropriate weights were constructed using inverse selection probabilities.

SupplementaryTable1b. Characteristics of eligible primary studies not included in the present study

First Author, Year	Country	Recruited Population	Diagnostic Interview	Classification System	Total N	Major Depression N (%)
Semi-structured Interviews						
Becker, 2002[54]	Saudi Arabia	Primary care patients	SCID	DSM-III-R	173	NR
Chen, 2013[55]	China	Primary care populations	SCID	DSM-IV	280	NR[a]
Chen, 2012[56]	China	Adults over 60 in primary care	SCID	DSM-IV	262	97 (37)
Fann, 2005[57b]	USA	Inpatients with traumatic brain injury	SCID	DSM-IV	134	45 (34)
Lai, 2010[58]	Hong Kong	Men with postpartum wives	SCID	DSM-IV	551	8 (1)
Navinés, 2012[59]	Spain	Chronic hepatitis C patients	SCID	DSM-IV	104	21 (20)
Phelan, 2010[60]	USA	Elderly primary care patients	SCID	DSM-IV	69	8 (12)
Thompson, 2011[61]	USA	Parkinson's patients	SCID	DSM-IV	214	30 (14)
Vöhringer, 2013[62b]	Chile	Primary care patients	SCID	DSM-IV	190	59 (31)
Watnick, 2005[63]	USA	Long term dialysis patients	SCID	DSM-IV	62	12 (19)
Fully Structured Interviews						
Al-Ghafri, 2014[64]	Oman	Medical trainees	CIDI	NR	131	NR[a]
Delgadillo, 2011[65b]	UK	Outpatients in drug addiction treatment	CIS-R	ICD-10	103	51 (50)
Haddad, 2013[66]	UK	Coronary heart disease patients	CIS-R	ICD-10	730	32 (4)

Mini International Neuropsychiatric Interviews (MINI)						
Persoons, 2003[67]	Belgium	Otorhinolaryngology outpatients	MINI	DSM-IV	97	16 (16)
Rathore, 2014[68]	USA	Adults with epilepsy	MINI	DSM-IV	172	33 (19)
Scott, 2011[69]	USA	Chronic hepatitis C patients	MINI	DSM-IV and ICD-10	30	NR[a]
van Steenbergen-Weijenburg, 2010[70b]	The Netherlands	Diabetes patients	MINI	DSM-IV	172	33 (19)
Wang, 2014[71]	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; 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.

[b]Contributed primary data to the main individual participant data meta-analysis of PHQ-9, but were excluded from the present study because PHQ-9 individual item scores were not available.

SupplementaryList1. Studies excluded at full-text review level

Studies excluded because no original data was used:

1. Shoukri MM, Donner A. Bivariate modeling of interobserver agreement coefficients. *Stat Med.* 2009;**28**:430-440.
2. Priyanka P, Boyle LL, Tu XM, Conwel Yl. 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. *Am J Geriatr Psychiatr*. 2010;**18**:S113-S114.
3. Lowe B, Grafe K, Quenter A, Buchholz C, Wild B, Zipfel S, Herzog W. 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. *Psychother Psychosom Med Psychol*. 2001;**51**:109-109.

Studies excluded because there was no administration of the PHQ:

4. Fine TH, Contractor AA, Tamburrino M, Elhai JD, Prescott MR, Cohen GH, Shirley E, Chan PK, Goto T, Slembarski R, Liberzon I, Galea S, Calabrese JR. Validation of the telephone-administered PHQ-9 against the in-person administered SCID-I major depression module. *J Affect Disord*. 2013;**150**:1001-1007.
5. 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. *Eur Psychiatry.* 2013;**28**:442-447.
6. Ryan DA, Gallagher P, Wright S, Cassidy EM. Sensitivity and specificity of the Distress Thermometer and a two-item depression screen (Patient Health Questionnaire-2) with a 'help' question for psychological distress and psychiatric morbidity in patients with advanced cancer.. *Psychooncology*. 2012;**21**:1275-1284.

7. Saliba D, DiFilippo S, Edelen MO, Kroenke K, Buchanan J, Streim J. Testing the PHQ-9 interview and observational versions (PHQ-9 OV) for MDS 3.0. *J Am Med Dir Assoc*. 2012;**13**:618-625.
8. 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. *Gen Hosp Psychiatry*. 2012;**34**:121-126.
9. Morina N, von Lersner U, Prigerson HG. War and bereavement: consequences for mental and physical distress. *PLoS ONE*. 2011;**6**:e22140.
10. Watson LC, Zimmerman S, Cohen LW, Dominik R. Practical depression screening in residential care/assisted living: five methods compared with gold standard diagnoses. *Am J Geriatr Psychiatry*. 2009;**17**:556-564.
11. 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. *Psychol Med*. 2009;**39**:1107-1116.
12. 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. *J Affect Disord*. 2007;**97**:261-264.
13. Husain N, Gater R, Tomenson B, Creed F. Comparison of the Personal Health Questionnaire and the Self Reporting Questionnaire in rural Pakistan. *JPMA J Pak Med Assoc*. 2006;**56**:366-370.
14. Lowe B, Grafe K, Zipfel S, Spitzer RL, Herrmann-Lingen C, Witte S, Herzog W. 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. *J Psychosom Res*. 2003;**55**:515-519.

15. 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. *Psychol Med*. 2000;**30**:831-840.
16. Husain N, Creed F, Tomenson B. Depression and social stress in Pakistan. *Psychol Med*. 2000;**30**:395-402.
17. Tschudi-Madsen H, Kjeldsberg M, Natvig B, Ihlebaek C, Dalen I, Straand J, Bruusgaard D. Multiple symptoms and medically unexplained symptoms-Closely related concepts in general practitioners' evaluations. A linked doctor-patient study. *J Psychosom Res*. 2013;**74**:186-190.
18. Creed F. The relationship between somatic symptoms, health anxiety, and outcome in medical out-patients. *Psychiatr Clin North Am*. 2011;**34**:545-564.
19. Allgaier AK, Pietsch K, Fruhe B, Prast E, Sigl-Glockner, Schulte-Korne G. Depression in pediatric care: Is the WHO-Five Well-Being Index a valid screening instrument for children and adolescents?. *Gen Hosp Psychiatry*. 2012;**34**:234-241.
20. Gellis ZD. Depression screening in medically ill homecare elderly. *Best Pract Ment Health*. 2010;**6**:1-16.
21. Lowe B, Grafe K, Kroenke K, Zipfel S, Quenter A, Wild B, Fiehn C, Herzog W. Predictors of Psychiatric Comorbidity in Medical Outpatients. *Psychosom Med*. 2003;**65**:764-770.

Studies excluded because there was no clinical interview to diagnose current major depression:

22. Muller KW, Beutel ME, Wolfling K. A contribution to the clinical characterization of Internet addiction in a sample of treatment seekers: validity of assessment, severity of psychopathology and type of co-morbidity. *Compr Psychiatry*. 2014;**55**:770-777.
23. Ringoir L, Pedersen SS, Widdershoven JW, Pop VJ. Prevalence of psychological distress in elderly hypertension patients in primary care. *Neth Heart J.* 2014;**22**:71-76.
24. Gigantesco A, Mirante N, Granchelli C, Diodati G, Cofini V, Mancini C, Carbonelli A, Tarolla E, Minardi V, Salmaso S, D'Argenio P. Psychopathological chronic sequelae of the 2009 earthquake in L'Aquila, Italy. *J Affect Disord*. 2013;**148**:265-271.
25. 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. *Confl Health*. 2012;**6**:12.
26. Tabb KM, Gavin AR, Guo Y, Huang H, Debiec K, Katon W. 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.
27. Gold KJ, Spangenberg K, Wobil P, Schwenk TL. Depression and risk factors for depression among mothers of sick infants in Kumasi, Ghana. *Int J Gynaecol Obstet*. 2013;**120**:228-231.
28. Kamphuis MH, Stegenga BT, Zuithoff NP, King M, Nazareth I, de Wit NJ, Geerlings MI. Does recognition of depression in primary care affect outcome? The PREDICT-NL study. *Fam Pract.* 2012;**29**:16-23.

29. 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. *Soc Psychiatry Psychiatr Epidemiol*. 2012;**47**:87-95.

30. Zuithoff NP, Vergouwe Y, King M, Nazareth I, van Wezep MJ, Moons KG, Geerlings MI. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study. *BMC Fam Pract*. 2010;**11**:98.

31. Osorio Fde L, de Carvalho AC, Crippa JA, Loureiro SR. Screening for smoking in a general hospital: scale validation, indicators of prevalence, and comorbidity. *Perspect Psychiatr Care*. 2013;**49**:5-12.

32. 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.. *J Nerv Ment Dis*. 2013;**201**:286-291.

33. Hauser W, Glaesmer H, Schmutzer G, Brahler E. Widespread pain in older Germans is associated with posttraumatic stress disorder and lifetime employment status--results of a cross-sectional survey with a representative population sample. *Pain*. 2012;**153**:2466-2472.

34. Maneeton B, Maneeton N, Mahathep P. Prevalence of depression and its correlations: a cross-sectional study in Thai cancer patients. *Asian Pac J Cancer Prev.* 2012;**13**:2039-2043.

35. Hauffa R, Rief W, Brahler E, Martin A,R Mewes R, Glaesmer H. Lifetime traumatic experiences and posttraumatic stress disorder in the German population: results of a representative population survey. *J Nerv Ment Dis*. 2011;**199**:934-939.

36. Hausteiner-Wiehle C, Sokollu F. Magical thinking in somatoform disorders: an exploratory study among patients with suspected allergies. *Psychopathology*. 2011;**44**:283-288.

37. Schmitz-Hubsch T, Coudert M, Tezenas du Montcel S, Giunti P, Labrum R, Durr A, Ribai P, Charles P, Linnemann C, Schols L, Rakowicz M, Rola R, Zdzienicka E, Fancellu R, Mariotti C, Baliko L, Melegh B, Filla A, Salvatore E, van de Warrenburg BP, Szymanski S, Infante J, Timmann D, Boesch S, Depondt C, Kang JS, Schulz JB, Klopstock T, Lossnitzer N, Lowe B, Frick C, Rottlander D, Schlaepfer TE, Klockgether T. Depression comorbidity in spinocerebellar ataxia. *Mov Disord*. 2011;**26**:870-876.

38. Lowe B, Kroenke K, Spitzer RL, Williams JB, Mussell M, Rose M, Wingenfeld K, Sauer N, Spitzer C. Trauma exposure and posttraumatic stress disorder in primary care patients: cross-sectional criterion standard study. *J Clin Psychiatry*. 2011;**72**:304-312.

39. Poutanen O, Koivisto AM, Salokangas RK. Applicability of the DEPS Depression Scale: assessing format and individual items in subgroups of patients. *Nord J Psychiatry*. 2010;**64**:384-390.

40. Arroll B, Goodyear-Smith F, Kerse N, Hwang M, Crengle S, Gunn J, Fishman T, Hatcher S, Pradhan S, Sidhu K. The prevalence of depression among Maori patients in Auckland general practice. *J Prim Health Care*. 2009;**1**:26-29.

41. Mussell M, Kroenke K, Spitzer RL, Williams JB, Herzog W, Lowe B. Gastrointestinal symptoms in primary care: prevalence and association with depression and anxiety. *J Psychosom Res*. 2008;**64**:605-612.

42. Holzapfel N, Muller-Tasch T, Wild B, Junger J, Zugck C, Remppis A, Herzog W, Lowe B. Depression profile in patients with and without chronic heart failure. *J Affect Disord*. 2008;**105**:53-62.

43. Cannon DS, Tiffany ST, Coon H, Scholand MB, McMahon WM, Leppert MF. The PHQ-9 as a brief assessment of lifetime major depression. *Psychol Assess*. 2007;**19**:247-251.

44. Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *J Am Geriatr Soc*. 2007;**55**:596-602.

45. Kissane DW, Wein S, Love A, Lee XQ, Kee PL, Clarke DM. The Demoralization Scale: a report of its development and preliminary validation. *J Palliat Care*. 2004;**20**:269-276.

46. 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?. *J Diabetes*. 2013;**5**:471-481.

47. Galek A, Erbsloeh-Moeller B, Koellner V, Kuehn-Becker H, Langhorst J, Petermann F, Prothmann U, Winkelmann A, Haeuser W. Mental disorders in patients with fibromyalgia syndrome. Screening in centres of different medical specialties. *Schmerz*. 2013;**27**:296-304.

48. Armstrong G, Nuken A, Samson L, Singh S, Jorm AF, Kermode M. Quality of life, depression, anxiety and suicidal ideation among men who inject drugs in Delhi, India. *BMC Psychiatry*. 2013;**13**:151.

49. Suzuki T, Shiga T, Nishimura K, Ishigooka J, Hagiwara N. PHQ-9 screening for depression in hospitalized patients with heart failure. *Eur J Heart Fail*. 2013;**12**:S242.

50. Whitlow NR, Ryan GL, Stuart SP. The Patient Health Questionnaire (Phq) is a Poor Psychological Screening Tool in in Vitro Fertilization (Ivf) Patients. *Fertil Steril*. 2011;**96**:S11.

51. Karekla M, Pilipenko N, Feldman J. Greek Language Validation of the Patient Health Questionnaire (PHQ). *Ann Behav Med*. 2011;**41**:S20.

52. Pilipenko N, Karekla M, Feldman J. Validation of Patient Health Questionnaire in Greek-Language Sample. *Eur Psychiat*. 2011;**26**:473

53. 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. *J Clin Nurs*. 2010;**19**:88.

54. Ulhaq S, Symeon C, Agius M. Use of the Phq-9 as a Screening Tool for Post-Stroke Depression. *Eur Psychiat*. 2010;**25**:502.

55. 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?. *J Card Fail*. 2009;**15**:S114.

56. Subramanian U, Perkins SM, Kim J, Ding Y, Pressler SJ. Depressive symptoms in heart failure: Validity and reliability of the PHQ-8. *J Gen Intern Med*. 2008;**23**:276.

57. 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. *Ann Behav Med*. 2007;**33**:S56.

58. 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.

59. Smith GC, Macasey P, Trauer T. Screening and monitoring in renal dialysis and transplant patients using the SF36 and Patient Health Questionnaire. *Aust N Z J Psychiatry*. 2000;**34**:A62.

60. Howell EA, Bodnar-Deren S, Balbierz A, Loudon H, Mora PA, Zlotnick C, Wang J, Leventhal H. An intervention to reduce postpartum depressive symptoms: A randomized controlled trial. *Arch Womens Ment Health*. 2014;**17**:57-63.

61. Mulligan L, Fear NT, Jones N, Alvarez H, Hull L, Naumann U, Wessely S, Greenberg N. Postdeployment Battlemind training for the U.K. armed forces: A cluster randomized controlled trial. *J Consult Clin Psychol*. 2012;**80**:331-341.

62. Gothwal VK, Bagga DK, Bharani S, Sumalini R, Reddy SP. The Patient Health Questionnaire-9: Validation among patients with glaucoma.. *PLoS ONE*. 2014;**9**:Art e101295-8.

63. Mautner E, Ashida C, Greimel E, Lang U, Kolman C, Alton D, Inoue W. Are There Differences in the Health Outcomes of Mothers in Europe and East-Asia? A Cross-Cultural Health Survey. *Biomed Res Int*. 2014;**2014**:856543.

Studies excluded because there was no validated diagnostic interview to assess current major depression:

64. 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. *Int J Psychiatry Clin Pract*. 2004;**8**:11-18.

65. 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.

66. Inoue T, Tanaka T, Nakagawa S, Nakato Y, Kameyama R, Boku S, Toda H, Kurita T, Koyama T. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry*. 2012;**12**:73.

67. Orive M, Padierna JA, Quintana JM, Las-Hayas C, Vrotsou K, Aguirre U. Detecting depression in medically ill patients: Comparative accuracy of four screening questionnaires and physicians' diagnoses in Spanish population. *J Psychosom Res*. 2010;**69**:399-406.

68. 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. *Aust N Z J Public Health*. 2008;**32**:317-321.

69. Carballeira Y, Dumont P, Borgacci S, Rentsch D, de Tonnac N, Archinard M, Andreoli A. Criterion validity of the French version of Patient Health Questionnaire (PHQ) in a hospital department of internal medicine. *Psychol Psychother*. 2007;**80**:69-77.

70. Rentsch D, Dumont P, Borgacci S, Carballeira Y, deTonnac N, Archinard M, Andreoli A. Prevalence and treatment of depression in a hospital department of internal medicine. *Gen Hosp Psychiatry.* 2007;**29**:25-31.

71. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;**41**:1284-1292.

72. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;**16**:606-613.

73. 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. *Psychosom Med*. 2001;**63**:679-686.

74. 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.

Studies excluded because the PHQ and diagnostic interview were not administered within two weeks of each other:

75. Choi Y, Mayer TG, Williams MJ, Gatchel RJ. What is the best screening test for depression in chronic spinal pain patients?. *Spine J*. 2014;**14**:1175-1182.

76. Prescott MR, Tamburrino M, Calabrese JR, Liberzon I, Slembarski R, Shirley E, Fine T, Goto T, Wilson K, Ganocy S, Chan P, Derus A, Serrano MB, Sizemore J, Kauffman J, Galea S. Validation of lay-administered mental health assessments in a large Army National Guard cohort. *Int J Methods Psychiatr Res*. 2014;**23**:109-119.

77. Cassin S, Sockalingam S, Hawa R, Wnuk S, Royal S, Taube-Schiff M, Okrainec A. Psychometric properties of the Patient Health Questionnaire (PHQ-9) as a depression screening tool for bariatric surgery candidates. *Psychosomatics*. 2013;**54**:352-358.

78. Liu LT, Chen SL, Jin T, Li L, Fan HN, Yu XE, Fang Y, Ma Y, Xu BH. Natural outcome and risk-prediction model of late-life depression. *Zhejiang da Xue Xue Bao.Yi Xue Ban*. 2012;**41**:653-658.

79. Chen S, Chiu H, Xu B, Ma Y, Jin T, Wu M, Conwell Y. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. *Int J Geriatr Psychiatry*. 2010;**25**:1127-1133.

80. Rief W, Mewes R, Martin A, Glaesmer H, Braehler E. Are psychological features useful in classifying patients with somatic symptoms?. *Psychosom Med*. 2010;**72**:648-655.

81. Yeung A, Fung F, Yu SC, Vorono S, Ly M, Wu S, Fava M. Validation of the Patient Health Questionnaire-9 for depression screening among Chinese Americans. *Compr Psychiatry*. 2008;**49**:211-217.

Studies excluded because their sample selection was based on the presence of distress or depression:

82. 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. *Depress Res Treat*. 2014;**2014**:768978.

83. Thapar A, Hammerton G, Collishaw S, Potter R, Rice F, Harold G, Craddock N, Thapar A, Smith DJ. Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures. *Br J Gen Pract*. 2014;**64**:e31-7.

84. 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. *Prim Care Companion CNS Disord*. 2011;**13**: doi: 10.4088/PCC.10m01027.

85. Vera M, Reyes-Rabanillo ML, Huertas S, Juarbe D, Perez-Pedrogo C, Huertas A, Pena M. Suicide ideation, plans, and attempts among general practice patients with chronic health conditions in Puerto Rico. *Int J Gen Med*. 2011;**4**:197-205.

86. Gibbons RD, Hooker G, Finkelman MD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kupfer DJ. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *J Clin Psychiatry*. 2013;**74**:669-674.

87. Smith AB, Rush R, Wright P, Stark D, Velikova G, Sharpe M. Validation of an item bank for detecting and assessing psychological distress in cancer patients.. *Psychooncology*. 2009;**18**:195-199.

88. Thekkumpurath P, Walker J, Butcher I, Hodges L, Kleiboer A, O'Connor M, Wall L, Murray G, Kroenke K, Sharpe M. Screening for major depression in cancer outpatients: the diagnostic accuracy of the 9-item patient health questionnaire. *Cancer*. 2011;**117**:218-227.

89. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, Kupfer DJ. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry*. 2012;**69**:1104-1112.

90. 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. *Arch Women Ment Health*. 2013;**16**:271-277.

91. Berghofer A, Hartwich A, Bauer M, Unutzer J, Willich SN, Pfennig A. Efficacy of a systematic depression management program in high utilizers of primary care: a randomized trial. *BMC Health Serv Res*. 2012;**12**:298.

92. 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**:1150-1158.

93. 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**:218-224.

94. Mittal D, Fortney JC, Pyne JM, Wetherell JL. Predictors of persistence of comorbid generalized anxiety disorder among veterans with major depressive disorder. *J Clin Psychiatry*. 2011;**72**:1445-1451.

95. Sockalingam S, Blank D, Al Jarad A, Alosaimi F, Hirschfield G, Abbey SE. A comparison of depression screening instruments in hepatitis C and the impact of depression on somatic symptoms. *Psychosomatics*. 2011;**52**:433-440.

96. Lossnitzer N, Muller-Tasch T, Lowe B, Zugck C, Nelles M, Remppis A, Haass M, Rauch B, Junger J, Herzog W, Wild B. Exploring potential associations of suicidal ideation and ideas of self-harm in patients with congestive heart failure. *Depress Anxiety*. 2009;**26**:764-768.

97. Sayers SL, Farrow VA, Ross J, Oslin DW. Family problems among recently returned military veterans referred for a mental health evaluation. *J Clin Psychiatry.* 2009;**70**:163-170.

98. 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.

99. Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *Br J Gen Pract*. 2007;**57**:650-652.

100. Yeung A, Yu SC, Fung F, Vorono S, Fava M. Recognizing and engaging depressed Chinese Americans in treatment in a primary care setting. *Int J Geriatr Psychiatry*. 2006;**21**:819-823.

101. Williams LS, Brizendine EJ, Plue L, Bakas T, Tu W, Hendrie H, Kroenke K. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 2005;**36**:635-638.

102. Buehler B, Kocalevent R, Berger R, Mahler A, Preiss B, Liwowsky I, Carl P, Hegerl U. Treatment situation of long-term unemployed with psychological disorders. *Nervenarzt*. 2013;**84**:603-607.

103. Gawlik S, Waldeier L, Mueller M, Szabo A, Sohn C, Reck C. Subclinical depressive symptoms during pregnancy and birth outcome-a pilot study in a healthy German sample. *Arch Womens Ment Health*. 2013;**16**:93-100.

104. 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.

105. 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. *J Psychosom Res*. 2014;**77**: 109-115

106. 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**:42-49.

107. Grote NK, Katon WJ, Lohr MJ, Carson K, Curran M, Galvin E, Russo JE, Gregory M. Culturally relevant treatment services for perinatal depression in socio-

economically disadvantaged women: The design of the MOMCare study. *Contemp Clin Trials*. 2014;**39**:34-49.

108. 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. *J Pain Symptom Manage*. 2015;**49**:127-135.

109. Olariu E, Castro-Rodriguez JI, Alvarez P, Garnier C, Reinoso M, Martin-Lopez LM, Alonso J, Forero CG. Validation of clinical symptom irt scores for diagnosis and severity assessment of common mental disorders. *Qual Life Res*. 2015;**24**:979-992.

Studies excluded because only the PHQ-2 or PHQ-8 was administrated:

110. 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 Behav*. 2011;**21**:387-390.

111. Lino VT, Portela MC, Camacho LA, Atie S, Lima MJ, Rodrigues NC, Barros MB, Andrade MK. 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-e113778.

112. Park H, Kim J, Hahm B. The Distress Thermometer and The PHQ-2 for Ultra-Brief Screening Depression Of Cancer Patients In Korea. *Psychooncology*. 2013;**22**:303-304.

113. 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?. *Gen Hosp Psychiatry*. 2010;**32**:544-548.

SUPPLEMENTARY MATERIAL REFERENCES

1. **Amoozegar F, Patten SB, Becker WJ, Bulloch AG, Fiest KM, Davenport WJ, Carroll CR, Jette N** (2017). The prevalence of depression and the accuracy of depression screening tools in migraine patients. *General Hospital Psychiatry* **48**, 25-31.
2. **Ayalon L, Goldfracht M, Bech P** (2010). 'Do you think you suffer from depression?' Re-evaluating the use of a single item question for the screening of depression in older primary care patients. *International Journal of Geriatric Psychiatry* **25**, 497-502.
3. **Beraldi A, Baklayan A, Hoster E, Hiddemann W, Heussner P** (2014). Which questionnaire is most suitable for the detection of depressive disorders in haemato-oncological patients? Comparison between HADS, CES-D and PHQ-9. *Oncology Research and Treatment* **37**, 108-109.
4. **Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR** (2012). Validity of the Patient Health Questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. *Archives of Physical Medicine & Rehabilitation* **93**, 1838-1845.
5. **Chagas MH, Tumas V, Rodrigues GR, Machado-de-Sousa JP, Filho AS, Hallak JE, Crippa JA** (2013). Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age & Ageing* **42**, 645–649.
6. **Eack SM, Greeno CG, Lee BJ** (2006). Limitations of the Patient Health Questionnaire in identifying anxiety and depression in community mental health: Many cases are undetected. *Research on Social Work Practice* **16**, 625-631.
7. **Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N** (2014). Validating screening tools for depression in epilepsy. *Epilepsia* **55**, 1642-1650.

8. **Fischer HF, Klug C, Roeper K, Blozik E, Edelmann F, Eisele M, Stork S, Wachter R, Scherer M, Rose M, Herrmann-Lingen M** (2014). Screening for mental disorders in heart failure patients using computer-adaptive tests. *Quality of Life Research* **23**, 1609-1618.
9. **Gjerdingen D, Crow S, McGovern P, Miner M, Center B** (2009). Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Annals of Family Medicine* **7**, 63-70.
10. **Gräfe K, Zipfel S, Herzog W, Löwe B** (2004). Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ). Results from the German validation study. *Diagnostica* **50**, 171-181.
11. **Khamseh ME, Baradaran HR, Javanbakht A, Mirghorbani M, Yadollahi Z, Malek M** (2011). Comparison of the CES-D and PHQ-9 depression scales in people with type 2 diabetes in Tehran, Iran. *BMC Psychiatry* **11**, 61.
12. **Kwan Y, Tham WY, Ang A** (2012). Validity of the Patient Health Questionnaire-9 (PHQ-9) in the screening of post-stroke depression in a multi-ethnic population. *Biological psychiatry* **71**, 141S.
13. **Lambert SD, Clover K, Pallant JF, Britton B, King MT, Mitchell AJ, Carter G** (2015). Making sense of variations in prevalence estimates of depression in cancer: A co-calibration of commonly used depression scales using Rasch analysis. *Journal of the National Comprehensive Cancer Network* **13**, 1203-1211.
14. **Liu SI, Yeh ZT, Huang HC, Sun J, Tjung JJ, Hwang LC, Shih YH, Yeh AW** (2011). Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Comprehensive Psychiatry* **52**, 96–101.

15. **McGuire AW, Eastwood JA, Macabasco-O'Connell A, Hays RD, Doering LV** (2013). Depression screening: utility of the Patient Health Questionnaire in patients with acute coronary syndrome. *American Journal of Critical Care* **22**, 12-19.

16. **Osório FL, Vilela Mendes A, Crippa JA, Loureiro SR** (2009). 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. *Perspectives in Psychiatric Care* **45**, 216-227.

17. **Osório FL, Carvalho AC, Fracalossi TA, Crippa JA, Loureiro ES** (2012). Are two items sufficient to screen for depression within the hospital context? *International Journal of Psychiatry in Medicine* **44**, 141-148.

18. **Picardi A, Adler DA, Abeni D, Chang H, Pasquini P, Rogers WH, Bungay KM** (2005). Screening for depressive disorders in patients with skin diseases: a comparison of three screeners. *Acta Dermato-Venereologica* **85**, 414-419.

19. **Richardson TM, He H, Podgorski C, Tu X, Conwell Y** (2010). Screening depression aging services clients. *American Journal of Geriatric Psychiatry* **18**, 1116-1123.

20. **Rooney AG, McNamara S, Mackinnon M, Fraser M, Rampling R, Carson A, Grant R** (2013). Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. *Neuro-oncology*. 2013;**15**:122–129.

21. **Sidebottom AC, Harrison PA, Godecker A, Kim H** (2012). Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. *Archives of Women's Mental Health* **15**, 367-374.

22. **Simning A, van Wijngaarden E, Fisher SG, Richardson TM, Conwea Y** (2012). Mental healthcare need and service utilization in older adults living in public housing. *American Journal of Geriatric Psychiatry* **20**, 441-451.

23. **Turner A, Hambridge J, White J, Carter G, Clover K, Nelson L, Hackett M** (2012). 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* **43**, 1000-1005.

24. **Twist K, Stahl D, Amiel SA, Thomas S, Winkley K, Ismail K** (2013). Comparison of depressive symptoms in type 2 diabetes using a two-stage survey design. *Psychosomatic Medicine* **75**, 791-797.

25. **Williams JR, Hirsch ES, Anderson K, Bush AL, Goldstein SR, Grill S, Lehmann S, Little JT, Margolis RL, Palanci J, Pontone G** (2012). A comparison of nine scales to detect depression in Parkinson disease: which scale to use? *Neurology* **78**, 998-1006.

26. **Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, Lucassen P, van de Lisdonk P, van Weert H** (2009). The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *General Hospital Psychiatry* **31**, 451-459.

27. **Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S** (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine* **8**, 348-353.

28. **Azah MN, Shah ME, Shaaban J, Bahri IS, Rushidi WM, Jamil YM** (2005). Validation of the Malay version brief Patient Health Questionnaire (PHQ-9) among adult attending family medicine clinics. *International Medical Journal* **12**, 259-263.

29. **De Man-van Ginkel JM, Hafsteinsdóttir T, Lindeman E, Burger H, Grobbee D, Schuurmans M** (2012). An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke* **43**, 854-856.

30. **Gelaye B, Tadesse MG, Williams MA, Fann JR, Vander Stoep A, Zhou XH** (2014). Assessing validity of a depression screening instrument in the absence of a gold standard. *Annals of Epidemiology* **24**, 527-531.

31. **Hahn D, Reuter K, Harter M** (2006). Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *GMS Psycho-Social-Medicine* 3.

32. **Henkel V, Mergl R, Kohnen R, Allgaier AK, Moller HJ, Hegerl U** (2004)**.** Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. *General Hospital Psychiatry* **26**, 190-198.

33. **Hobfoll SE, Canetti D, Hall BJ, Brom D, Palmieri PA, Johnson RJ, Pat-Horenczyk R, Galea S** (2011). Are community studies of psychological trauma's impact accurate? A study among Jews and Palestinians. *Psychological Assessment* 23, 599-605.

34. **Kiely KM, Butterworth P** (2014). Validation of four measures of mental health against depression and generalized anxiety in a community based sample. *Psychiatry Research* **225**, 291-298.

35. **Mohd Sidik S, Arroll B, Goodyear-Smith F** (2012). Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. *Medical Journal of Malaysia* **67**, 309-315.

36. **Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA** (2008). Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychological Medicine* 38, 221-228.

37. **Pence BW, Gaynes BN, Atashili J, O'Donnell JK, Tayong G, Kats D, Whetten R, Whetten K, Njamnshi AK, Ndumbe PM** (2012). Validity of an interviewer-

administered Patient Health Questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *Journal of Affective Disorders* **143**, 208-213.

38. **Razykov I, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research Group** (2013). Utility of the Patient Health Questionnaire-9 to assess suicide risk in patients with systemic sclerosis. *Arthritis Care & Research* **65**, 753-758.
39. **Thombs BD, Ziegelstein RC, Whooley MA** (2008). Optimizing detection of major depression among patients with coronary artery disease using the Patient Health Questionnaire: data from the heart and soul study. *Journal of General Internal Medicine* **23**, 2014-2017.
40. **Akena D, Joska J, Obuku EA, Stein DJ** (2013). Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care* **25**, 1245-1252.
41. **Cholera R, Gaynes BN, Pence BW, Bassett J, Qangule N, Macphail C, Bernhardt S, Pettifor A, Miller WC** (2014). Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *Journal of Affective Disorders* **167**, 160-166.
42. **Hides L, Lubman DI, Devlin H, Cotton S, Aitken C, Gibbie T, Hellard M** (2007). Reliability and validity of the Kessler 10 and Patient Health Questionnaire among injecting drug users. *Australian & New Zealand Journal of Psychiatry* **41**, 166-168.
43. **Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA** (2011). 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 & Research* **63**, 1313-1321.

44. **Hyphantis T, Kroenke K, Papatheodorou E, Paika V, Theocharopoulos N, Ninou A, Tomenson B, Carvalho AF, Guthrie E,ARISTEIA-ABREVIATE Study Grp Member** (2014). 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. *Comprehensive Psychiatry* **55**, 1950-1959.

45. **Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, Kurosawa M, Muramatsu K, Furukawa TA, Yamada M** (2013). 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. *General Hospital Psychiatry* **35**, 592-597.

46. **Lamers F, Jonkers CC, Bosma H, Penninx BW, Knottnerus JA, van Eijk JT** (2008). Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *Journal of Clinical Epidemiology* **61**, 679-687.

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

48. **Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Yoshida M, Otsubo T, Gejyo F** (2007). The Patient Health Questionnaire, Japanese version: validity according to the Mini-International Neuropsychiatric Interview-Plus. *Psychological Reports* **101**, 952-960.

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

50. **Santos IS, Tavares BF, Munhoz TN, Almeida LS, Silva NT, Tams BD, Patella AM, Matijasevich A** (2013). Sensitivity and specificity of the Patient Health Questionnaire-9 (PHQ-9) among adults from the general population. *Cadernos de Saude Publica* **29**, 1533-1543.

51. **Stafford L, Berk M, Jackson HJ** (2007). Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *General Hospital Psychiatry* **29**, 417-424.

52. **Sung SC, Low CC, Fung DS, Chan YH** (2013). 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-Pacific psychiatry : Official Journal of the Pacific Rim College of Psychiatrists* **5**, 249-258.

53. **Zhang Y, Ting R, Lam M, Lam J, Nan H, Yeung R, Yang W, Ji L, Weng J, Wing YK, Sartorius N, Chan JC** (2013). Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. *Journal of Affective Disorders* **151**, 660-666.

54. **Becker S, Al Zaid K, Al Faris E** (2002). Screening for somatization and depression in Saudi Arabia: a validation study of the PHQ in primary care. *The International Journal of Psychiatry in Medicine* **32**, 271-283.

55. **Chen S, Fang Y, Chiu H, Fan H, Jin T, Conwell Y** (2013). Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. *Asia-Pacific Psychiatry* **5**, 61-68.

56. **Chen S, Conwell Y, Vanorden K, Lu N, Fang Y, Ma Y, Fan H, Jin T, Chiu H** (2012). Prevalence and natural course of late-life depression in China primary care: a population based study from an urban community. *Journal of Affective Disorders* **141**, 86-93.

57. **Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, Rau H, Temkin N** (2005). Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *Journal of Head Trauma Rehabilitation* **20**, 501-511.

58. **Lai BP, Tang AK, Lee DT, Yip AS, Chung TK** (2010). Detecting postnatal depression in Chinese men: a comparison of three instruments. *Psychiatry Research* **180**, 80-85.

59. **Navines R, Castellvi P, Moreno-Espana J, Gimenez D, Udina M, Canizares S, Diez-Quevedo C, Valdés M, Solà R, Martín-Santos R** (2012). Depressive and anxiety disorders in chronic hepatitis C patients: reliability and validity of the Patient Health Questionnaire. *Journal of Affective Disorders* **138**, 343-351.

60. **Phelan E, Williams B, Meeker K, Bonn K, Frederick J, LoGerfo J, Snowden M** (2010). A study of the diagnostic accuracy of the PHQ-9 in primary care elderly. *BMC Family Practice* **11**, 63.

61. **Thompson AW, Liu H, Hays RD, Katon WJ, Rausch R, Diaz N, Jacob EL, Vassar SD, Vickrey BG** (2011). Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. *Parkinsonism & Related Disorders* **17**, 40-45.

62. **Vohringer PA, Jimenez MI, Igor MA, Fores GA, Correa MO, Sullivan MC, Holtzman NS, Whitham EA, Barroilhet SA, Alvear K, Logvinenko T** (2013). Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. *Journal of Medical Screening* **20**, 118-124.

63. **Watnick S, Wang PL, Demadura T, Ganzini L** (2005). Validation of 2 depression screening tools in dialysis patients. *American Journal of Kidney Diseases* **46**, 919-924.

64. **Al-Ghafri G, Al-Sinawi H, Al-Muniri A, Dorvlo AS, Al-Farsi YM, Armstrong K, Al-Adawi S** (2014). Prevalence of depressive symptoms as elicited by Patient Health Questionnaire (PHQ-9) among medical trainees in Oman. *Asian Journal of Psychiatry* **8**, 59-62.

65. **Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, Dale V** (2011). How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *Journal of Affective Disorders* **134**, 266-271.

66. **Haddad M, Walters P, Phillips R, Tsakok J, Williams P, Mann A, Tylee A** (2013). 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* **8**, e78493.

67. **Persoons P, Luyckx K, Desloovere C, Vandenberghe J, Fischler B** (2003). Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. *General Hospital Psychiatry* **25**, 316-323.

68. **Rathore JS, Jehi LE, Fan Y, Patel SI, Foldvary-Schaefer N, Ramirez MJ, Busch RM, Obuchowski NA, Tesar GE** (2014). Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy. *Epilepsy & Behavior* **37**, 215-220.

69. **Scott JD, Wang CC, Coppel E, Lau A, Veitengruber J, Roy-Byrne P** (2011). Diagnosis of depression in former injection drug users with chronic hepatitis C. *Journal of Clinical Gastroenterology* **45**, 462-467.

70. **van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF, Hakkaart-van Roijen L, Rutten FF, Beekman AT, van der Feltz-Cornelis CM** (2010). Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Services Research* **10**, 235.

71. **Wang W, Bian Q, Zhao Y, Li X, Wang W, Du J, Zhang G, Zhou Q, Zhao M** (2014). Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry* **36**, 539-544.