Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis

Brooke Levis, PhD[1,2]; Andrea Benedetti, PhD[2-4]; John P. A. Ioannidis, MD[5]; Ying Sun, MPH[1]; Zelalem Negeri, PhD[1,2]; Chen He, MScPH[1]; Yin Wu, PhD[1,2,6]; Ankur Krishnan, MSc[1]; Parash Mani Bhandari, BPH[1,2]; Dipika Neupane, BPH[1,2]; Mahrukh Imran, MScPH[1]; Danielle B. Rice, MSc[1,7]; Kira E. Riehm, MSc[1,8]; Nazanin Saadat, MSc[1]; Marleine Azar, MSc[1,2]; Jill Boruff, MLIS[9]; Pim Cuijpers, PhD[10]; Simon Gilbody, PhD[11]; Lorie A. Kloda, PhD[12]; Dean McMillan, PhD[11]; Scott B. Patten, MD[13,14]; Ian Shrier, MD[1,2,15]; Roy C. Ziegelstein, MD[16]; Sultan H. Alamri, MD[17]; Dagmar Amtmann, PhD[18]; Liat Ayalon, PhD[19]; Hamid R. Baradaran, MD[20,21]; Anna Beraldi, PhD[22]; Charles N. Bernstein, MD[23,24]; Arvin Bhana, PhD[25,26]; Charles H. Bombardier, PhD[18]; Gregory Carter, FRANZCP[27]; Marcos H. Chagas, MD[28]; Dixon Chibanda, PhD[29]; Kerrie Clover, PhD[27]; Yeates Conwell, MD[30]; Crisanto Diez-Quevedo, PhD[31,32]; Jesse R. Fann, MD[33]; Felix H. Fischer, PhD[6,34]; Leila Gholizadeh, PhD[35]; Lorna J. Gibson, MPhil[36]; Eric P. Green, PhD[37]; Catherine G. Greeno, PhD[38]; Brian J. Hall, PhD[39,40]; Emily E. Haroz, PhD[41]; Khalida Ismail, MD[42]; Nathalie Jetté, MD[13,14,43]; Mohammad E. Khamseh, MD[20]; Yunxin Kwan, Mmed[44]; Maria Asunción Lara, PhD[45]; Shen-Ing Liu, MD[46-49]; Sonia R. Loureiro, PhD[28]; Bernd Löwe, MD[50]; Ruth Ann Marrie, MD[51]; Laura Marsh, MD[52]; Anthony McGuire, PhD[53]; Kumiko Muramatsu, MD[54]; Laura Navarrete, PhD[55]; Flávia L. Osório, PhD[28,56]; Inge Petersen, PhD[57]; Angelo Picardi, MD[58]; Stephanie L. Pugh, PhD[59,60]; Terence J. Quinn, MD[61]; Alasdair G. Rooney, MD[62]; Eileen H. Shinn, PhD[63]; Abbey Sidebottom, PhD[64]; Lena Spangenberg, PhD[65]; Pei Lin Lynnette Tan, MMed[44]; Martin Taylor-Rowan, PhD[66]; Alyna Turner, PhD[67,68]; Henk C.

van Weert, MD[69]; Paul A. Vöhringer, MD[70-72]; Lynne I. Wagner, PhD[73,74]; Jennifer White, PhD[75]; Kirsty Winkley, PhD[76]; Brett D. Thombs, PhD[1,2,4,6,7,77,78]

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; [2]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; [3]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada; [4]Department of Medicine, McGill University, Montréal, Québec, Canada; [5]Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA; [6]Department of Psychiatry, McGill University, Montréal, Québec, Canada; [7]Department of Psychology, McGill University, Montréal, Québec, Canada; [8]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA; [9]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Québec, Canada; [10]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, the Netherlands; [11]Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; [12]Library, Concordia University, Montréal, Québec, Canada; [13]Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada; [14]Hotchkiss Brain Institute and O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada; [15]Department of Family Medicine, McGill University, Montréal, Québec, Canada; [16]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [17]Faculty of Medicine, King Abdulaziz University, Jeddah, Makkah, Saudi Arabia; [18]Department of Rehabilitation Medicine, University of

Washington, Seattle, Washington, USA; [19]Louis and Gabi Weisfeld School of Social Work, Bar Ilan University, Ramat Gan, Israel; [20]Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran; [21]Ageing Clinical & Experimental Research Team, Institute of Applied Health Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, Scotland, UK; [22]Kbo-Lech-Mangfall-Klinik Garmisch-Partenkirchen, Klinik für Psychiatrie, Psychotherapie & Psychosomatik, Lehrkrankenhaus der Technischen Universität München, Munich, Germany; [23]University of Manitoba IBD Clinical and Research Centre, Winnipeg, Manitoba, Canada; [24]Department of Internal Medicine, Max rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada; [25]Centre for Rural Health, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa; [26]Health Systems Research Unit, South African Medical Research Council, South Africa; [27]Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; [28]Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; [29]Department of Community Medicine, University of Zimbabwe, Harare, Zimbabwe; [30]Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA; [31]Servei de Psiquiatria, Hospital Germans Trias i Pujol, Badalona, Spain; [32]Departament de Psiquiatria i Medicina Legal, Universitat Autònoma de Barcelona, Badalona, Spain; [33]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; [34]Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Germany; [35]Faculty of Health, University of Technology Sydney, Sydney, Australia; [36]Tropical Epidemiology Group, Faculty of Epidemiology and Population

Health, London School of Hygiene and Tropical Medicine, London, UK; [37]Duke Global Health Institute, Duke University, Durham, North Carolina, USA; [38]School of Social Work, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; [39]Global and Community Mental Health Research Group, Department of Psychology, Faculty of Social Sciences, University of Macau, Macau Special Administrative Region, China; [40]Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [41]Center For American Indian Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States; [42]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neurosciences, King's College London Weston Education Centre, London, UK; [43]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [44]Department of Psychological Medicine, Tan Tock Seng Hospital, Singapore; [45]Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz. San Lorenzo Huipulco, Tlalpan, México D. F. Mexico; [46]Programme in Health Services & Systems Research, Duke-NUS Medical School, Singapore; [47]Department of Psychiatry, Mackay Memorial Hospital, Taipei, Taiwan; [48]Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan; [49]Department of Medicine, Mackay Medical College, Taipei, Taiwan; [50]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [51]Departments of Medicine and Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada; [52]Baylor College of Medicine, Houston and Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas, USA; [53]Department of Nursing, St. Joseph's College, Standish, Maine, USA; [54]Department of Clinical Psychology, Graduate School of Niigata Seiryo University, Niigata, Japan; [55]Department of Epidemiology and Psychosocial Research, Instituto

Nacional de Psiquiatría Ramón de la Fuente Muñiz, Ciudad de México, México; [56]National Institute of Science and Technology, Translational Medicine, Ribeirão Preto, Brazil; [57]Centre for Rural Health, School of Nursing and Public Health, University of KwaZulu-Natal, South Africa; [58]Centre for Behavioural Sciences and Mental Health, Italian National Institute of Health, Rome, Italy; [59]NRG Oncology Statistics and Data Management Center, Philadelphia, PA, USA; [60]American College of Radiology, Philadelphia, PA, USA; [61]Institute of Cardiovascular & Medical Sciences, University of Glasgow, Glasgow, Scotland; [62]Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburg, Edinburgh, Scotland, UK; [63]Department of Behavioral Science, University of Texas M. D. Anderson Cancer Center, Houston, Texas, USA; [64]Allina Health, Minneapolis, Minnesota, USA; [65]Department of Medical Psychology and Medical Sociology, University of Leipzig, Germany; [66]Institute of Cardiovascular and Medical Science, University of Glasgow, Glasgow, Scotland; [67]School of Medicine and Public Health, University of Newcastle, New South Wales, Newcastle, Australia; [68]Deakin University, IMPACT Strategic Research Centre, School of Medicine, Barwon Health, Geelong, Victoria, Australia; [69]Department of General Practice, Amsterdam Institute for General Practice and Public Health, Amsterdam University Medical Centers, Location AMC, Amsterdam, the Netherlands; [70]Department of Psychiatry and Mental Health, Clinical Hospital, Universidad de Chile, Santiago, Chile; [71]Millennium Institute for Depression and Personality Research (MIDAP), Ministry of Economy, Macul, Santiago, Chile; [72]Psychiatry Department, Tufts Medical Center, Tufts University, Boston, USA; [73]Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA; [74]Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, North Carolina, USA; [75]Department of Physiotherapy, School of Primary and Allied Health Care, Monash University, Melbourne, Australia; [76]Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, UK; [77]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada; [78]Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada.

Correspondence Author: Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Ste Catherine Road; Montreal, Quebec, Canada. H3T 1E4. Tel: (514) 340-8222 ext. 25112; Email: brett.thombs@mcgill.ca. ORCID: 0000-0002-5644-8432

Word count: 3,000

ABSTRACT

Objective: Depression symptom questionnaires are not for diagnostic classification. Patient Health Questionnaire-9 (PHQ-9) scores ≥ 10 are nonetheless often used to estimate depression prevalence. We compared PHQ-9 ≥ 10 prevalence to Structured Clinical Interview for DSM (SCID) major depression prevalence and assessed whether an alternative PHQ-9 cutoff could more accurately estimate prevalence.

Study design and setting: Individual participant data meta-analysis of datasets comparing PHQ-9 scores to SCID major depression status.

Results: 9,242 participants (1,389 SCID major depression cases) from 44 primary studies were included. Pooled PHQ-9 ≥ 10 prevalence was 24.6% (95% CI: 20.8%, 28.9%); pooled SCID major depression prevalence was 12.1% (95% CI: 9.6%, 15.2%); pooled difference was 11.9% (95% CI: 9.3%, 14.6%). Mean study-level PHQ-9 ≥ 10 to SCID-based prevalence ratio was 2.5 times. PHQ-9 ≥ 14 and the PHQ-9 diagnostic algorithm provided prevalence closest to SCID major depression prevalence, but study-level prevalence differed from SCID-based prevalence by an average absolute difference of 4.8% for PHQ-9 ≥ 14 (95% prediction interval: -13.6%, 14.5%) and 5.6 % for the PHQ-9 diagnostic algorithm (95% prediction interval: -16.4%, 15.0%).

Conclusion: PHQ-9 ≥ 10 substantially overestimates depression prevalence. There is too much heterogeneity to correct statistically in individual studies.

Key words: depression prevalence, PHQ-9, SCID, individual participant data meta-analysis

Running title: Depression prevalence based on PHQ-9 vs. SCID

HIGHLIGHTS

- We compared Patient Health Questionnaire-9 (PHQ-9) ≥ 10 prevalence to Structured Clinical Interview for DSM (SCID) major depression prevalence in 44 primary studies (9,242 participants, 1,389 SCID major depression cases) that administered the PHQ-9 and SCID.
- We also examined whether an alternative PHQ-9 cutoff could more accurately estimate prevalence.
- Pooled PHQ-9 ≥ 10 prevalence (25%) was double pooled SCID major depression prevalence (12%); pooled difference from each study was 12%.
- PHQ-9 ≥ 14 and PHQ-9 diagnostic algorithm prevalence most closely matched SCID major depression prevalence, but study-level PHQ-9 ≥ 14 and PHQ-9 diagnostic algorithm prevalence differed from SCID major depression prevalence with 95% prediction intervals of -14% to 15% and -16% to 15%, respectively.
- Estimates of depression prevalence should be based on validated diagnostic interviews designed for determining case status; users should evaluate published reports of depression prevalence to ensure that they are based on methods intended to classify major depression.

1. INTRODUCTION

Disease prevalence estimates have important implications for interpreting medical research, understanding disease burden, and making decisions about healthcare resource utilization.[1] In mental health research, major depression classification requires using validated diagnostic interviews.[2,3] Administering diagnostic interviews in large enough samples to estimate prevalence, however, is resource intensive. Thus, researchers sometimes use self-report depression symptom questionnaires, or screening tools, instead, and label the percentage of participants scoring above a screening cutoff as depression prevalence.[4,5] A 2018 study identified 19 primary studies listed in PubMed in a 3-month period whose titles indicated that they assessed prevalence of depression or depressive disorders and found that 89% were based on screening questionnaires only.[4]

Some self-report questionnaires include the same symptoms evaluated in validated diagnostic interviews. None, however, include all components of diagnostic interviews, such as assessment of functional impairment or investigation of non-psychiatric medical conditions that can cause similar symptoms.[4] Using depression symptom questionnaires and cutoffs intended for screening to assess depression prevalence may overestimate prevalence. This is because screening attempts to identify previously unrecognized cases; cutoffs are set to cast a wide net and identify many more patients who may have depression than meet diagnostic criteria.

A recent review examined meta-analyses of depression prevalence published in 2008-2017.[5] Of 81 prevalence estimates reported in abstracts of 69 meta-analyses, 10% were based on diagnostic interviews, 44% were based on screening or rating tools, and 46% combined results from diagnostic interviews and screening or rating tools. Mean reported prevalence was 31% among meta-analyses based on screening or rating tools compared to 17% with diagnostic interviews.[5] The degree to which screening tools exaggerate prevalence, however, depends on the screening tool and cutoff used.[4,5]

We do not know of any studies that have evaluated the degree to which specific screening tool and cutoff combinations overestimate depression prevalence.[4,5] The Patient Health Questionnaire-9 (PHQ-9)[6-8] is the most commonly used depression screening tool in primary care.[9] Its nine items align with the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for major depressive episode.[10-12] The standard cutoff, ≥ 10, is well-established for screening to detect major depression and maximized combined sensitivity and specificity in a recent individual participant data meta-analysis (IPDMA).[6-8,13] PHQ-9 ≥ 10 has been used to estimate depression prevalence in primary research studies and via synthesis in meta-analyses, including in very high-impact journals.[14-16] It is also sometimes used to diagnose depression and make treatment decisions for individual patients.[6,17-19]

Our objective was to use an IPDMA approach to (1) compare PHQ-9 ≥ 10 prevalence to major depression prevalence based on a well-validated semi-structured diagnostic interview, the Structured Clinical Interview for DSM (SCID);[20] and (2) use a prevalence matching approach[4,21] to determine if a PHQ-9 cutoff could be set to match SCID-based prevalence with sufficiently low heterogeneity to accurately estimate prevalence in individual studies.

2. METHODS

This study used a subset of data accrued for an IPDMA of the accuracy of the PHQ-9 for screening to detect major depression.[13] Detailed methods were registered in PROSPERO (CRD42014010673), and a protocol was published.[22] This analysis was not part of the original IPDMA protocol.

2.1 Study Selection

In the main IPDMA, datasets from articles in any language were eligible for inclusion if (1) they included PHQ-9 scores; (2) they included diagnostic classifications for current Major Depressive Episode (MDE) or Major Depressive Disorder (MDD) based on DSM[10-12] or International Classification of Diseases[23] criteria, using a validated semi-structured or fully structured interview; (3) the PHQ-9 and diagnostic interview were administered within two weeks of each other; (4) participants were ≥18 years and not recruited from youth or school-based settings; and (5) participants were not recruited from psychiatric settings or because they were identified as having depressive symptoms. Datasets where not all participants were eligible were included if primary data allowed selection of eligible participants.

For the present study, we included primary studies that based diagnoses on the SCID.[20] The SCID is a semi-structured diagnostic interview intended to be conducted by an experienced diagnostician; it requires clinical judgment and allows rephrasing questions and probes. The reason for including only SCID studies is that in analyses using large IPDMA databases,[24-26] we found that, compared to semi-structured interviews, fully structured interviews, which are designed for administration by lay interviewers, identify more participants with low-level symptoms as depressed but fewer participants with high-level symptoms. These results were consistent with the idea that semi-structured interviews most closely replicate clinical interviews done by trained professionals, whereas fully structured interviews are less resource-intensive options that can be administered by research staff without diagnostic skills but may misclassify major depression in many participants. In our PHQ-9 IPDMA database, 44 of 47 studies that used semi-structured interviews used the SCID. Thus, to reduce heterogeneity, we only included these 44 studies in main analyses.

In sensitivity analyses, we also included the three studies that used other semi-structured interviews. We considered also incorporating published results from eligible studies that did not contribute data to the IPDMA. However, only 3 of 14 such studies[27-29] (970 participants, 77 major depression cases) reported sufficient information to compare PHQ-9 ≥ 10 and SCID-based prevalence, and these studies did not report information necessary to be included in all prevalence matching analyses.

2.2 Data Sources and Searches

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations via Ovid, PsycINFO, and Web of Science from January 1 2000-May 9 2018, using a peer-reviewed[30] search strategy (Supplementary Material: Appendix Methods). We also reviewed reference lists of relevant reviews and queried contributing authors about non-published studies.

Two investigators independently reviewed titles and abstracts for eligibility. If either deemed a study potentially eligible, the full-text was reviewed by two investigators, independently, with disagreements resolved by consensus, consulting a third investigator when necessary.

2.3 Data Contribution and Synthesis

Authors of eligible datasets were invited to contribute de-identified primary data, including PHQ-9 scores and major depression classification status. We emailed corresponding authors of eligible studies at least three times, as necessary. If no response, we emailed co-authors and attempted phone contact.

Prior to integrating individual datasets into our synthesized dataset, we compared published participant characteristics and diagnostic accuracy results with results from raw datasets and resolved discrepancies with the original investigators. When datasets included statistical weights to reflect sampling procedures, we used provided weights. For studies where sampling procedures merited weighting, but the original study did not weight, we constructed weights using inverse selection probabilities.

2.4 Data Analysis

Comparison of PHQ-9 ≥ 10 Prevalence and SCID Major Depression Prevalence

For each primary study, we estimated the percentage of participants who scored ≥ 10 on the PHQ-9, the percentage of participants classified as having major depression based on the SCID, the difference of these percentages, and the ratio. Then, across studies, we pooled prevalence for PHQ-9 ≥ 10, prevalence for the SCID, and differences in prevalence.

Prevalence Matching

To identify which PHQ-9 scoring approach best matched SCID-based prevalence, we estimated pooled differences in prevalence for each possible PHQ-9 cutoff and the PHQ-9 diagnostic algorithm compared to SCID. The scoring approach with the smallest pooled difference was chosen to be the “prevalence match scoring approach.” Then, for each included study, we estimated the difference and ratio in prevalence for the prevalence match scoring approach versus SCID. We determined the mean and median absolute difference and range of differences across all studies. To illustrate the range of difference values that would be expected if a new study were to compare prevalence based on the prevalence match scoring approach to prevalence based on SCID, we estimated 95% prediction intervals for the differences. For the diagnostic algorithm, which requires five or more items with scores of ≥ 2 points, with at least one being depressed mood or anhedonia,[8] three studies[31-33] (524 participants) and 88 additional participants from other studies (612 participants total, 7%) were excluded, as they did not provide PHQ-9 item scores, which are necessary to determine diagnostic algorithm criteria. In sensitivity analyses, we evaluated if results differed if the 612 participants were excluded from all analyses rather than just those involving the diagnostic algorithm.

All meta-analyses incorporated sampling weights and were conducted in R (R version 3.4.1; R Studio version 1.0.143) using the lme4 package. To estimate pooled prevalence values, generalized linear mixed-effects models with a logit link function were fit using the glmer function. To estimate pooled difference values, linear mixed-effects models were fit using the lmer function. To account for correlation between subjects within the same primary study, random intercepts were fit for each primary study. To quantify heterogeneity, we reported the estimated between-study variance (τ^2) for each analysis.

In post-hoc analyses, we investigated whether differences in prevalence for the PHQ-9 prevalence match scoring approach and SCID were associated with study and participant characteristics. To do this, we fit additional linear mixed-effects models for pooled prevalence difference, including age, sex, country human development index ("very high", "high", or "low-medium", based on the United Nation's 2018 Human Development Index) and recruitment setting category (primary care, nonmedical care, inpatient specialty care, or outpatient specialty care) as fixed-effect covariates. For these analyses, we excluded 56 participants (<1%) missing age or sex data.

3. RESULTS

3.1 Search Results and Inclusion of Primary Study Datasets

Of 9,674 unique titles and abstracts identified from the database search for the main IPDMA, 9,198 were excluded after title and abstract review and 297 after full-text review, leaving 179 eligible articles with data from 123 unique participant samples, of which 95 (77.2%) contributed datasets. Authors of included studies contributed data from five unpublished studies, for a total of 100 datasets. Of these, for the present study's main analyses, we excluded 56 studies that classified major depression using a diagnostic interview other than the SCID (Figure 1). Thus, the main analyses of the present study included 9,242 participants (1,389 major depression cases) from 44 primary studies.[31-72] Among the 28 eligible primary studies that did not provide datasets for the main IPDMA, 14 used the SCID (4,408 participants). Thus, the main analyses included 75.9% of eligible studies that used the SCID (44 of 58) and 67.7% of eligible participants (9,242 of 13,650). Table 1 shows the characteristics of each included study.

In sensitivity analyses, we included data from three additional studies (1,992 participants; 139 major depression cases) that provided individual participant data but administered a semi-structured interview other than the SCID (Table 1)[73-75].

3.2 Comparison of PHQ-9 ≥ 10 Prevalence and SCID Major Depression Prevalence

The percentage of participants with PHQ-9 ≥ 10 in each of the 44 SCID studies ranged from 5.3% to 64.8%; pooled prevalence was 24.6% (95% confidence interval [CI]: 20.8%, 28.9%; τ^2: 0.505). The percentage of participants with SCID major depression ranged from 0.6% to 56.4%; pooled prevalence was 12.1% (95% CI: 9.6%, 15.2%; τ^2: 0.703).

Differences in prevalence (PHQ-9 ≥ 10 minus SCID) ranged from -6.0% to 46.9%. The pooled difference was 11.9% (95% CI: 9.3%, 14.6%; τ^2: 0.007).

The ratio of PHQ-9 ≥ 10 prevalence to SCID-based prevalence ranged from 0.7 to 10.0 times (mean: 2.5; median: 1.9). The mean ratio was 3.8 times for the 17 studies with SCID-based prevalence < 10% (mean difference: 13.3%), 2.0 times for the 16 studies with SCID-based prevalence between 10% and 20% (mean difference: 12.7%), and 1.3 times for the 11 studies with SCID-based prevalence of ≥ 20% (mean difference: 8.9%).

3.3 Prevalence Matching

PHQ-9 ≥ 14 (pooled difference in prevalence: 0.5%, 95% CI: -1.7%, 2.6%, τ^2: 0.005) and the PHQ-9 diagnostic algorithm (pooled difference in prevalence: -0.7%, 95% CI: -3.2%, 1.8%; τ^2: 0.006) provided prevalence closest to SCID-based prevalence. Pooled differences in prevalence for PHQ-9 ≥ 13 and ≥ 15 compared to SCID were 2.6% and -2.0%.

In the 44 individual SCID studies, differences between the percentage of participants with PHQ-9 ≥ 14 and SCID major depression ranged from -18.7% to 29.7% (mean absolute difference: 4.8%). Of 44 prevalence estimates based on PHQ-9 ≥ 14, 24 (54.5%) were ≤ 0.75 times or ≥ 1.25 times the SCID-based prevalence. The 95% prediction interval for the difference in prevalence was -13.6% to 14.5%. For the PHQ-9 diagnostic algorithm, study-level differences in prevalence ranged from -20.1% to 27.1% (mean absolute difference: 5.6%). Of 41 prevalence estimates based on the PHQ-9 diagnostic algorithm, 28 (68.3%) were ≤ 0.75 times or ≥ 1.25 times the SCID-based prevalence. The 95% prediction interval for the difference in prevalence was -16.4% to 15.0%. No study or participant characteristics were significantly associated with differences in prevalence for either of the PHQ-9 prevalence match scoring approaches compared to SCID.

3.4 Sensitivity Analyses

Results for all analyses were similar when data from the three studies with semi-structured interviews other than the SCID were added or when the 612 participants without data to determine PHQ-9 diagnostic algorithm classification were excluded.

4. DISCUSSION

Primary studies and meta-analyses that describe their results as reflecting prevalence of depression or depressive disorders are frequently based on depression screening tools, which are not designed for this purpose, rather than validated diagnostic interviews.[4,5] The PHQ-9 is often used to generate what are described by researchers as depression prevalence estimates. The present study found that using PHQ-9 ≥ 10 to assess depression prevalence, which is commonly done, overestimated depression prevalence compared to prevalence based on actual diagnostic criteria by 11.9% (mean ratio: 2.5 times).

These results are consistent with what was predicted in a previous analysis that used hypothetical estimates of sensitivity and specificity to demonstrate how depression screening tools would be expected to inflate prevalence.[4] Results are also consistent with the findings of a meta-research review of prevalence estimates from 69 meta-analyses that found higher mean depression prevalence based on screening or rating tools than based on diagnostic interviews.[5] Thus, if a screening tool, such as the PHQ-9 ≥ 10, is used to estimate prevalence, prevalence will appear to be substantial in virtually all populations, even when true prevalence is very low. This could have important ramifications in terms of policies, service planning and healthcare budgets.

Identifying a PHQ-9 cutoff that could be used to match true prevalence based on a diagnostic interview would allow researchers to use inexpensive questionnaires instead of more costly interview methods for prevalence estimation. We tested a prevalence matching approach and found that PHQ-9 ≥ 14 and the PHQ-9 diagnostic algorithm provided the smallest differences in prevalence compared to SCID major depression, but heterogeneity was high and not associated with study or participant characteristics. The mean absolute difference between prevalence based on PHQ-9 versus SCID in individual studies was 4.8% for PHQ-9 ≥ 14 and 5.6% for the PHQ-9 diagnostic algorithm, reflecting both overestimation and underestimation. For more than half of the studies examined, PHQ-9 ≥ 14 prevalence was less than 75% or more than 125% of SCID-based prevalence; for the PHQ-9 diagnostic algorithm the fraction was over two-thirds. The 95% prediction interval for the difference between PHQ-9 ≥ 14 and SCID-based prevalence ranged from 14% below to 15% above SCID-based prevalence; for the PHQ-9 diagnostic algorithm it was from 16% below to 15% above SCID-based prevalence.

Researchers sometimes report prevalence estimates based on cutoffs from questionnaires, including the PHQ-9, as prevalence of “clinically significant” symptoms or “symptoms” of depression, rather than “depression”.[14,76,77] However, screening tool cutoffs do not reflect a meaningful divide between impairment and non-impairment. Patients scoring at or above virtually any cutoff would be expected to have greater impairment than patients scoring below the cutoff, but no evidence has established any single cutoff for establishing an impairment threshold or that would support clinical decision-making for individual patients without a validated clinical assessment.[4]

Research on screening using the PHQ-9 would be expected to report the proportion of patients who score at or above screening cutoffs because this provides information on the number of patients who would need resources for further mental health assessment. Reporting this percentage as depression prevalence, however, would be akin, for example, to reporting the proportion of women with positive mammogram screens as the prevalence of breast cancer and, as shown in the present study, would dramatically overestimate prevalence.

This is the first study to estimate the degree to which using PHQ-9 ≥ 10 to estimate depression prevalence, a common practice, leads to overestimation of prevalence. Strengths of the study are that we incorporated data from 44 primary studies and that we directly compared PHQ-9 ≥ 10 prevalence estimates to those based on the SCID, a rigorous semi-structured interview intended to facilitate the standardized application of actual diagnostic criteria by trained diagnosticians.[10-12] This study had some limitations. First, we were unable to include data from 14 of 58 published eligible datasets (24%). Second, included datasets were almost exclusively from patients in healthcare settings where the presence of transdiagnostic somatic symptoms and adjustment to illness or injury may have contributed to error variance.[75] Third, included datasets were from a wide range of study settings, which may account for some of the observed heterogeneity. Fourth, overestimation of prevalence when screening tools are used is expected to be greater with lower true prevalence. This is because false positives are disproportionately high in low-prevalence populations and only minimally offset by false negative screens, which occur when true cases are missed by the screening test. However, we were unable to assess this because of the small number of heterogeneous datasets included. Fifth, not all SCID studies described interviewer qualifications; untrained interviewers may have reduced the ability to detect differences across inteviews. Sixth, we only examined one depression screening tool, the PHQ-9, although we expect that other tools would similarly exaggerate depression prevalence.[4,5]

In summary, we found that using PHQ-9 ≥ 10 to estimate depression prevalence results in estimates that are, on average, 12% greater than what would be obtained using validated semi-structured diagnostic interviews. Substantial heterogeneity presents a barrier to using statistical methods to estimate major depression prevalence based on PHQ-9 ≥ 10 or based on any other PHQ-9 cutoff. Researchers should not report results from the PHQ-9 as prevalence of major depression. Users of evidence should evaluate reports of prevalence with caution and ensure that they are based on methods intended to classify major depression.

5. ACKNOWLEDGEMENTS

5.1 Funding

This study was funded by the Canadian Institutes of Health Research (CIHR; KRS-134297, PCG-155468, PJT-162206). Dr. Levis and Dr. Wu were supported by Fonds de recherche du Québec - Santé (FRQS) Postdoctoral Training Fellowships. Drs. Benedetti and Thombs were supported by a FRQS researcher salary award. Mr. Bhandari was supported by a studentship from the Research Institute of the McGill University Health Centre. Ms. Neupane was supported by G.R. Caverhill Fellowship from the Faculty of Medicine, McGill University. Ms. Rice was supported by a Vanier Canada Graduate Scholarship. Ms. Riehm and Ms. Saadat were supported by CIHR Frederick Banting and Charles Best Canada Graduate Scholarship master's awards.

The primary studies by Fiest et al., Patten et al., Amoozegar et al. and Prisnie et al. were supported by the Cumming School of Medicine, University of Calgary, and Alberta Health Services through the Calgary Health Trust, as well as the Hotchkiss Brain Institute. Dr. Patten was supported by a Senior Health Scholar award from Alberta Innovates Health Solutions. Dr. Jetté was supported by a Canada Research Chair in Neurological Health Services Research and an AIHS Population Health Investigator Award. The primary study by Amtmann et al. was supported by a grant from the Department of Education (NIDRR grant number H133B080025) and by the National Multiple Sclerosis Society (MB 0008). 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 studies by Marrie et al. and Bernstein et al. were supported by CIHR (THC-135234) and Crohn's and Colitis Canada. Dr. Bernstein was supported in part by the Bingham Chair in Gastroenterology. The primary study by Bhana et al. was output of the PRogramme for

Improving Mental health carE (PRIME) and was supported by the UK Department for International Development (201446). The views expressed do not necessarily reflect the UK Government's official policies. 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). The primary study by Chibanda et al. was supported by a grant from Grand Challenges Canada (0087-04). Dr. Conwell received support from NIMH (R24MH071604) and the Centers for Disease Control and Prevention (R49 CE002093). The primary study by Martin-Subero et al. was supported in part by a grant from the Spanish Ministry of Health's Health Research Fund (Fondo de Investigaciones Sanitarias, project 97/1184). Collection of data for the primary study by Fann et al. was supported by grant RO1 HD39415 from the US National Center for Medical Rehabilitation Research. The primary study by Fischer et al. was funded by the German Federal Ministry of Education and Research (01GY1150). 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 Gholizadeh et al. was supported by University of Technology Sydney under UTS Research Reestablishment Grants. The primary study by Green et al. (2018) was supported by a grant from the Duke Global Health Institute (453-0751). The primary study by Eack et al. was funded by the NIMH (R24 MH56858). The primary study by Garabiles et al. was supported by the Macao (SAR) Government, through the University of Macau RSKTO grants: MYRG-2014-111. The primary study by Haroz et al. was supported by the United States Agency for International Development Victims of Torture Fund: AID-DFD A-00-08-00308. Dr. Haroz was supported by a NIMH T32 predoctoral training grant

(MH014592-38) and postdoctoral training grant (MH103210) during the conduct of primary study. The primary study by Twist et al. was funded by the UK National Institute for Health Research under its Programme Grants for Applied Research Programme (grant reference number RP-PG-0606-1142). The primary study by Lara et al. was supported by the Consejo Nacional de Ciencia y Tecnología/National Council for Science and Technology (CB-2009-133923-H). The primary study by Liu et al. (2011) was funded by a grant from the National Health Research Institute, Republic of China (NHRI-EX97-9706PI). The primary studies by Osório et al. (2012) were funded by Reitoria de Pesquisa da Universidade de São Paulo (grant number 09.1.01689.17.7) and Banco Santander (grant number 10.1.01232.17.9). 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. Dr. Marrie was supported by the Waugh Family Chair in Multiple Sclerosis and the Research Manitoba Chair, and CIHR grants, during the conduct of the study. Collection of data for the primary study by Williams et al. was supported by a NIMH grant to Dr. Marsh (RO1-MH069666). The primary study by Muramatsu et al. (2018) was supported by grants from Niigata Seiryo University. Dr. Osório was supported by Productivity Grants (PQ-CNPq-2 -number 301321/2016-7). The primary study by Picardi et al. was supported by funds for current research from the Italian Ministry of Health. The primary study by Wagner et al. was supported by grants U10CA21661, U10CA180868, U10CA180822, and U10CA37422 from the National Cancer Institute. The study was also funded in part by a grant from the Pennsylvania Department of Health. The department specifically declaims responsibility for any analyses, interpretations, or conclusions of the primary study. The primary study by Rooney et al. was funded by the United Kingdom National Health Service Lothian Neuro-Oncology Endowment Fund. The primary study by Shinn et al.

was supported by grant NCI K07 CA 093512 and the Lance Armstrong Foundation. The primary study by Sidebottom et al. was funded by a grant from the United States Department of Health and Human Services, Health Resources and Services Administration (grant number R40MC07840). Simning et al.'s research was supported in part by grants from the NIH (T32 GM07356), Agency for Healthcare Research and Quality (R36 HS018246), NIMH (R24 MH071604), and the National Center for Research Resources (TL1 RR024135). The primary study by Spangenberg et al. was supported by a junior research grant from the medical faculty, University of Leipzig. Collection of data for the studies by Turner et al. (2012) were funded by a bequest from Jennie Thomas through the Hunter Medical Research Institute. The study by 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. Dr Vöhringer was supported by the Fund for Innovation and Competitiveness of the Chilean Ministry of Economy, Development and Tourism, through the Millennium Scientific Initiative (grant number IS130005).No other authors reported funding for primary studies or for their work on the present study. No sponsor or funder was involved in the study design; in the collection, analysis and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

5.2 Declaration of Competing Interests

All authors have completed the ICJME uniform disclosure form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years with the following exceptions: Dr. Bernstein declares that he receives grants and personal fees from Abbvie, Janssen, Pfizer, and Takeda; grants from Shire Canada, Celgene, Boeringher Ingelheim, and

Roche; and personal fees from Mylan Pharmaceuticals; outside the submitted work. Dr. Ismail declares that she has received honorarium for speaker fees for educational lectures for Sanofi, Sunovion, Janssen and Novo Nordisk. Dr. Pugh declares that she received salary support from Pfizer-Astella and Millennium, outside the submitted work. Dr. Wagner declares that she receives personal fees from Celgene, outside 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.

5.3 Author Contributions

BLevis, ABenedetti, JPAI, JB, PC, SG, LAK, DM, SBP, IS, RCZ, and BDT were responsible for the study conception and design. JB and LAK designed and conducted database searches to identify eligible studies. SBP, SHA, DA, LA, HRB, ABeraldi, CNB, ABhana, CHB, GC, MHC, DC, KC, YC, CDQ, JRF, FHF, LG, LJG, EPG, CGG, BJH, EEH, KI, NJ, MEK, YK, MAL, SIL, SRL, BLöwe, RAM, LM, AM, KM, LN, FLO, IP, AP, SLP, TJQ, AGR, EHS, AS, LS, PLLT, MTR, AT, HCvW, PAV, LIW, JW and KW contributed primary datasets that were included in this study. BLevis, YS, ZN, CH, YW, AK, PMB, DN, MI, DBR, KER, NS, MA and BDT contributed to data extraction and coding for the individual participant data meta-analysis. BLevis, ABenedetti and BDT conducted analyses and interpreted results. BLevis and BDT drafted the manuscript. All authors provided a critical review and approved the final manuscript. BDT is the guarantor.

5.4 Data Sharing

Statistical codes and dataset used in the individual patient data meta-analysis can be requested from the corresponding author, Dr. Brett D. Thombs.

REFERENCES

1. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. Am J Epidemiol. 1978;107(1):71-76.
2. Wittchen H-U. Reliability and validity studies of the WHO-Composite International Diagnostic Intervivew (CIDI): a critical review. J Psychiatr Res. 1994;28(1):57-84.
3. Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID) – I: History, rationale, and description. Arch Gen Psychiatry. 1992;49(8):624-629.
4. Thombs BD, Kwakkenbos L, Levis AW, Benedetti A. Addressing overestimation of the prevalence of depression prevalence based on self-report screening questionnaires. CMAJ. 2018;190:E44-49.
5. Levis B, Yan XW, He C, Sun Y, Benedetti A, Thombs BD. A comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. BMC Med. 2019;17:65.
6. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-613.
7. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann. 2002;32(9):1-7.
8. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999;282(18):1737-1744.
9. Maurer DM, Raymond TJ, Davis BN. Depression: screening and diagnosis. Am Fam Physician. 2018;98(8):508-515.

10. Diagnostic and statistical manual of mental disorders: DSM-III 3rd ed, revised. Washington, DC: American Psychiatric Association 1987.

11. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed. Washington, DC: American Psychiatric Association 1994.

12. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed, text revised. Washington, DC: American Psychiatric Association 2000.

13. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. The diagnostic accuracy of the Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: an individual participant data meta-analysis. BMJ. 2019;365:I1476.

14. Mata DA, Ramos MA, Bansal N, et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. JAMA 2015;314(22):2373-2383.

15. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA 2016;316(21):2214-2236.

16. Qato DM, Ozenberger K, Olfson M. Prevalence of prescription medications with depression as a potential adverse effect among adults in the United States. JAMA. 2018;319(22):2289-2298.

17. Dejesus RS, Vickers KS, Melin GJ, Williams MD. A system-based approach to depression management in primary care using the Patient Health Questionnaire-9. Mayo Clin Proc. 2007;82(11):1395-1402.

18. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2. Medical Care. 2003;41(11):1284-1292.

19. Whooley MA. Depression and cardiovascular disease: healing the broken-hearted. JAMA. 2006;295:2874-2881.

20. First MB. Structured clinical interview for the DSM (SCID). John Wiley & Sons, Inc. 1995.

21. Kelly MJ, Dunstan FD, Lloyd K, Fone DL. Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods. BMC Psychiatry. 2008;8:10.

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

23. The ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines Geneva: World Health Organization 1992.

24. Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. Br J Psychiatry. 2018;212(6):377-385.

25. Levis B, McMillan D, Sun Y, et al. Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: An individual participant data meta-analysis. Int J Methods Psychiatr Res. 2019;28(4):e1803.

26. Wu Y, Levis B, Sun Y, et al. Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic interviews controlling for Hospital Anxiety and

Depression Scale – Depression subscale scores: an individual participant data meta-analysis of 73 primary studies. J Psychosom Res. 2020;129:109892.

27. Phelan E, Williams B, Meeker K, et al. A study of the diagnostic accuracy of the PHQ-9 in primary care elderly. BMC Fam Pract. 2010;11:63.
28. Watnick S, Wang PL, Demadura T, Ganzini L. Validation of 2 depression screening tools in dialysis patients. Am J Kidney Dis. 2005;46:919-924.
29. Liu ZW, Yu Y, Hu M, Liu HM, Zhou L, Xiao SY. PHQ-9 and PHQ-2 for screening depression in Chinese rural elderly. PLoS One. 2016;11:e0151042.
30. PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). Ottawa: CADTH; 2016.
31. Alamri SH, Bari AI, Ali AT. Depression and associated factors in hospitalized elderly: a cross-sectional study in a Saudi teaching hospital. Ann Saudi Med. 2017;37:122-129.
32. Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehabil. 2005;20:501-511.
33. Vöhringer PA, Jimenez MI, Igor MA, et al. Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. J Med Screen. 2013;20:118-124.
34. Amoozegar F, Patten SB, Becker WJ, et al. The prevalence of depression and the accuracy of depression screening tools in migraine patients. Gen Hosp Psychiatry. 2017;48:25-31.
35. Amtmann D, Bamer AM, Johnson KL, et al. A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis. J Psychosom Res. 2015;79:550-557.

36. Ayalon L, Goldfracht M, Bech P. '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. Int J Geriatr Psychiatry. 2010;25:497-502.

37. Beraldi A, Baklayan A, Hoster E, Hiddemann W, Heussner P. Which questionnaire is most suitable for the detection of depressive disorders in haemato-oncological patients? Comparison between HADS, CES-D and PHQ-9. Oncol Res Treat. 2014;37:108-109.

38. Bernstein CN, Zhang L, Lix LM, et al. The validity and reliability of screening measures for depression and anxiety disorders in inflammatory bowel disease. Inflamm Bowel Dis. 2018;24:1867-1875.

39. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. BMC psychiatry. 2015;15:118.

40. Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR. Validity of the Patient Health Questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. Arch Phys Med. 2012;93:1838-1845.

41. Chagas MH, Tumas V, Rodrigues GR, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. Age Ageing. 2013;42:645-649.

42. Chibanda D, Verhey R, Gibson LJ, et al. Validation of screening tools for depression and anxiety disorders in a primary care population with high HIV prevalence in Zimbabwe. J Affect Disord. 2016;198:50-55.

43. Eack SM, Greeno CG, Lee BJ. Limitations of the Patient Health Questionnaire in identifying anxiety and depression in community mental health: many cases are undetected. Res Soc Work Pract. 2006;16:625-631.

44. Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N. Validating screening tools for depression in epilepsy. Epilepsia. 2014;55:1642-1650.

45. Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. Qual Life Res. 2014;23:1609-1618.

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

47. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ). Results from the German validation study. Diagnostica. 2004;50:171-181.

48. Green JD, Annunziata A, Kleiman SE, et al. Examining the diagnostic utility of the DSM-5 PTSD symptoms among male and female returning veterans. Depress Anxiety. 2017;34:752-760.

49. Green EP, Tuli H, Kwobah E, Menya D, Chesire I, Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: a mixed methods study. J Affect Disord. 2018;228:49-59.

50. Haroz EE, Bass J, Lee C, et al. Development and cross-cultural testing of the International Depression Symptom Scale (IDSS): a measurement instrument designed to represent global presentations of depression. GMH. 2017;4.

51. Hitchon CA, Zhang L, Peschken CA, et al. The validity and reliability of screening measures for depression and anxiety disorders in rheumatoid arthritis. Arthrit Care Res. 2019.

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

53. Kwan Y, Tham WY, Ang A. Validity of the Patient Health Questionnaire-9 (PHQ-9) in the screening of post-stroke depression in a multi-ethnic population. Biol Psychiatry. 2012;71:141S-141S.

54. Lambert SD, Clover K, Pallant JF, et al. Making sense of variations in prevalence estimates of depression in cancer: a co-calibration of commonly used depression scales using Rasch analysis. J Natl Compr Canc Netw. 2015;13:1203-1211.

55. Lara MA, Navarrete L, Nieto L, Martín JP, Navarro JL, Lara-Tapia H. Prevalence and incidence of perinatal depression and depressive symptoms among Mexican women. J Affect Disord. 2015;175:18-24.

56. Marrie RA, Zhang L, Lix LM, et al. The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis. Mult Scler Relat Dis. 2018;20:9-15.

57. Martin-Subero M, Kroenke K, Diez-Quevedo C, et al. Depression as measured by PHQ-9 versus clinical diagnosis as an independent predictor of long-term mortality in a prospective cohort of medical inpatients. Psychosom Med. 2017;79:273-282.

58. Osório FL, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. Perspect Psychiatr Care. 2009;45:216-227.

59. Osório FL, Carvalho AC, Fracalossi TA, Crippa JA, Loureiro ES. Are two items sufficient to screen for depression within the hospital context? Int J Psychiatry Med. 2012;44:141-148.

60. Patten SB, Burton JM, Fiest KM, et al. Validity of four screening scales for major depression in MS. Mult Scler. 2015;21:1064-1071.

61. Picardi A, Adler DA, Abeni D, et al. Screening for depressive disorders in patients with skin diseases: a comparison of three screeners. Acta Derm Venereol. 2005;85:414-419.

62. Prisnie JC, Fiest KM, Coutts SB, et al. Validating screening tools for depression in stroke and transient ischemic attack patients. Int J Psychiatry Med. 2016;51:262-277.

63. Richardson TM, He H, Podgorski C, Tu X, Conwell Y. Screening depression aging services clients. Am J Geriatr Psychiatry. 2010;18:1116-1123.

64. Rooney AG, McNamara S, Mackinnon M, et al. Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. Neuro-oncology. 2013;15:122-129.

65. Shinn EH, Valentine A, Baum G, et al. Comparison of four brief depression screening instruments in ovarian cancer patients: diagnostic accuracy using traditional versus alternative cutpoints. Gynecol Oncol. 2017;145:562-568.

66. Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Arch Womens Ment Health. 2012;15:367-374.

67. Simning A, van Wijngaarden E, Fisher SG, Richardson TM, Conwell Y. Mental healthcare need and service utilization in older adults living in public housing. Am J Geriatr Psychiatry. 2012;20:441-451.

68. Spangenberg L, Glaesmer H, Boecker M, Forkmann T. Differences in Patient Health Questionnaire and Aachen Depression Item Bank scores between tablet versus paper-and-pencil administration. Qual Life Res. 2015;24:3023-3032.

69. Turner A, Hambridge J, White J, et al. Depression screening in stroke: a comparison of alternative measures with the structured diagnostic interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (major depressive episode) as criterion standard. Stroke. 2012;43:1000-1005.

70. Wagner LI, Pugh SL, Small Jr W, et al. Screening for depression in cancer patients receiving radiotherapy: feasibility and identification of effective tools in the NRG Oncology RTOG 0841 trial. Cancer. 2017;123:485-93.

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

72. Wittkampf K, van Ravesteijn H, Baas K, et al. The accuracy of Patient Health Questionnaire- 9 in detecting depression and measuring depression severity in high-risk groups in primary care. Gen Hosp Psychiatry. 2009;31:451-459.

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

74. McGuire AW, Eastwood JA, Macabasco-O'Connell A, Hays RD, Doering LV. Depression screening: utility of the Patient Health Questionnaire in patients with acute coronary syndrome. Am J Crit Care. 2013;22:12-19.

75. Twist K, Stahl D, Amiel SA, Thomas S, Winkley K, Ismail K. Comparison of depressive symptoms in type 2 diabetes using a two-stage survey design. Psychosom Med. 2013;75:791-797.

76. Scott JE, Mathias JL, Kneebone AC. Depression and anxiety after total hip replacement among older adults; a meta-analysis. Aging Ment Health. 2016;20(12):1243-1254.