Probability of Major Depression Classification Based on the SCID, CIDI and MINI Diagnostic Interviews: A Synthesis of Three Individual Participant Data Meta-Analyses

Yin Wu[1-3]; Brooke Levis[1,3,4]; John P. A. Ioannidis[5]; Andrea Benedetti[3,6,7*]; Brett D. Thombs[1-3,6-9*]; and the DEPRESsion Screening Data (DEPRESSD) Collaboration.

[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] Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK;
[5] Departments of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, Stanford University, Stanford, California, USA;
[6] Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada;
[7] Department of Medicine, McGill University, Montréal, Québec, Canada; Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada;
[8] Department of Psychology, McGill University, Montréal, Québec, Canada;
[9] Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada.
* Co-senior authors

Short title: Comparison of Diagnostic Interviews for Major Depression

Corresponding authors:
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

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

Keywords:
depressive disorders, diagnostic interviews, individual participant data meta-analysis, major depression, classification

ABSTRACT

Objective: To compare odds of major depression classification based on the Structured Clinical Interview for DSM (SCID), the Composite International Diagnostic Interview (CIDI), and the Mini International Neuropsychiatric Interview (MINI).

Methods: We included and standardized data from three individual participant data meta-analysis (IPDMA) databases, which included primary studies with depressive symptom scores from the Patient Health Questionnaire-9, Edinburgh Postnatal Depression Scale, or Hospital Anxiety and Depression Scale – Depression subscale plus diagnostic interview-based major depression status. For each IPDMA, separately, we fit binomial generalized linear mixed models to compare adjusted odds ratios (aORs) of (1) major depression classification, controlling for depression symptom severity and participant characteristics, and (2) the interaction between interview and symptom severity. Next, we synthesized results using DerSimonian-Laird random-effects meta-analysis.

Results: In total, 69,405 participants (7,574 [11%] with major depression) from 212 studies were included. Controlling for symptom severity and participant characteristics, the MINI (74 studies; 25,749 participants) classified major depression more often than the SCID (108 studies; 21,953 participants; aOR [95% CI] = 1.46 [1.11-1.92]). Classification odds for the CIDI (30 studies; 21,703 participants) and SCID did not differ overall (aOR [95% CI] =1.19 [0.79, 1.75]), but as screening scores increased, aOR increased less for the CIDI than the SCID (interaction aOR [95% CI] = 0.64 [0.52-0.80]).

Conclusions: Compared to the SCID, the MINI classified major depression more often. Odds of depression classification with the CIDI increased less as symptom levels increased. Interpretation of research that uses diagnostic interviews to classify depression should consider interview characteristics.

INTRODUCTION

In mental health research, diagnostic interviews are used to classify disorders in a manner consistent with standard classification systems and replicable across studies [1-4]. There are important differences, however, in the designs of commonly used interviews. Semi-structured interviews are designed for administration by trained professionals with diagnostic experience; evaluators can interject queries and use their clinical judgment to determine whether symptoms are present and significant [1-3]. The Structured Clinical Interview for DSM (SCID) [4] is the most commonly used semi-structured interview in depression research [5-7]. Fully structured interviews, in contrast, are designed for lay interviewer administration to reduce the cost of clinician-administered interviews. They are completely scripted, and evaluators cannot provide additional explanations or rephrase questions; minimal judgment is involved. They are intended to maximize reliability but may reduce validity [8]. The Composite International Diagnostic Interview (CIDI) [8] is the most commonly used fully structured interview for depression research [5-7]. The Mini International Neuropsychiatric Interview (MINI) [9,10], also common in depression research, is a very brief fully structured interview, originally described by its developers as a screening interview and intended to be over-inclusive [10].

Despite their differences, semi-structured interviews, fully structured interviews of conventional length, and abbreviated alternatives such as the MINI are usually treated as equivalent. For instance, meta-analyses of depression screening tool accuracy typically pool primary study results without consideration of reference standards [11-17]. Until recently, however, only several small studies, each with 61 depression cases or fewer, compared classification by different diagnostic interviews [2,18-23]. Recently, three individual participant data meta-analyses (IPDMA) compared odds of major depression classification between different diagnostic interviews, controlling for depression symptom severity scores and participant characteristics [5-7]. Those included an IPDMA with 17,158 participants from 57 primary studies that used the Patient Health Questionnaire-9 (PHQ-9) to control for depression symptom severity [5], 12,759 women

in pregnancy or postpartum from 46 studies that used the Edinburgh Postnatal Depression Scale (EPDS) [6], and 15,856 participants from 73 studies that used the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) [7]. Results suggested that, compared to semi-structured interviews (e.g., SCID) [4], the CIDI may classify more people with relatively low-level symptoms as depressed but fewer people with higher symptom levels. The MINI appeared to classify major depression in more people across the symptom spectrum. There was important imprecision in results, however, including wide confidence intervals (CIs) around estimates.

Our objective was to synthesize results from three separate IPDMAs datasets to and compare the most commonly used diagnostic interviews for major depression, the SCID, CIDI, and MINI to determine (1) if odds ratios for major depression classification using the CIDI and MINI differ from the SCID, controlling for depression symptom severity and participant characteristics, and (2) if there is an interaction between the interview and depressive symptom level that would suggest that differences in classification odds are associated with symptom levels.

MATERIALS AND METHODS

We conducted a two-stage evidence synthesis. We first conducted IPDMAs in the PHQ-9, EPDS, and HADS datasets, separately, by fitting models with and without interaction terms for depressive symptom severity in each dataset, separately. Second, we pooled estimates from the results of the three IPDMAs.

Inclusion Criteria for the Included Datasets

For the PHQ-9, EPDS, and HADS-D IPDMAs, datasets from articles in any language were eligible for inclusion if (1) they included diagnostic classification for current Major Depressive Disorder or Major Depressive Episode using Diagnostic and Statistical Manual of Mental Disorders [24-27] or International Classification of Diseases [28] criteria based on a validated semi-structured or fully structured interview; (2) they included PHQ-9, EPDS, or HADS-D scores; (3) the diagnostic interview and depression screening test were administered within two weeks of each other; and (4) participants were ≥ 18 years, not recruited from

youth or college settings, and not recruited from psychiatric settings or because a screening test identified them as having symptoms of depression [29-31]. For the EPDS, participants were women in pregnancy or within 12 months postpartum [30]. In each IPDMA, datasets where not all participants were eligible were included if primary data allowed selection of eligible participants [29-31]. Over 90% of all included studies in the IPDMA databases used the SCID, CIDI, or MINI diagnostic interviews. Thus, for the present study, as we did in the published IPDMAs of the EDPS [6] and HADS-D [7], we restricted analyses to studies that used SCID, CIDI, or MINI.

Search Strategy, Study Selection, Data Acquisition, and Data Extraction

For more details on the search and selection processes, as well as data contribution, extraction, and synthesis, please see Supplementary Method 1. For information on how the IPDMA datasets and the analyses conducted in the present study deviated from our previous published IPDMAs on diagnostic interview performance using the PHQ-9 [5], EPDS [6], and HADS-D [7] IPDMA databases, please see Supplementary Method 2, Supplementary Method 3, and Supplementary Figure 1.

Statistical Analysis

IPDMAs of PHQ-9, EPDS, and HADS-D Datasets:

We initially standardized symptom severity scores in each dataset. To do this, for each measure, we converted raw screening tool scores to standardized scores by Z-transformation (subtracting the mean and dividing by the standard deviation of raw scores). We then meta-analyzed the PHQ-9, EPDS, and HADS datasets, separately. In each dataset, we fit binomial generalized linear mixed models with a logit link function to compare the adjusted odds ratio (aOR) of major depression classification for the CIDI versus the SCID, the MINI versus the SCID, and, as a supplementary analysis, the MINI versus the CIDI, controlling for depressive symptom levels and other participant characteristics. We adjusted for different covariates in the models for each dataset, based on relevant measures. For the PHQ-9 and HADS-D datasets, as in the previously published IPDMAs [5,7], we controlled for depressive symptom severity (continuous

standardized scores), age, sex, country Human Development Index (very high, high, or low-medium) [32], and patient care setting (PHQ-9: primary care, outpatient specialty care, inpatient specialty care, non-medical care [33]; HADS-D: outpatient care, inpatient care, non-medical care, mixed inpatient and outpatient [7]). For the EPDS, we did not control for sex or patient care settings but controlled for pregnancy versus postpartum status [6]. To account for the correlation between subjects within primary studies in each dataset, a random intercept was fit. Fixed slopes were estimated for all covariates in each model. We also fit additional models in each dataset, where we added an interaction term between interview and depressive symptom severity (continuous PHQ-9, EPDS, and HADS-D standardized scores), to evaluate whether any differences in aOR of major depression classification were associated with depression symptom severity.

Synthesis of IPDMA Results:

To synthesize results from the three IPDMAs, we pooled estimates of the aOR for each comparison (CIDI versus SCID, MINI versus SCID, MINI versus CIDI) and the aOR for the interaction of interview and depression symptom severity in each comparison, along with 95% CIs. We used DerSimonian-Laird random effects meta-analysis to pool the aORs [34]. Heterogeneity was examined using the I_2 statistic based on log aORs [35]. Because some studies were included in both the PHQ-9 and HADS-D IPDMAs, as a sensitivity analysis, we re-analyzed results after removing those studies.

All analyses were conducted in R (R version R 3.5.1 and R Studio version 1.1.463) [36,37] using the glmer function within the lme4 package [38] and the rma function within the metafor package [39].

RESULTS

In total, 69,405 participants (7,574 [11%] with major depression) were included in the three individual IPDMAs (Table 1). Of the 212 included primary studies, the SCID was used in 108 studies (21,953 participants, 14% major depression), the CIDI in 30 studies (21,703 participants, 7% major depression), and the MINI in 74 studies (25,749 participants, 12% major depression). Mean (standard deviation) of raw

screening tool scores, prior to standardization, were 4.99 (5.26) for the PHQ-9, 6.98 (5.58) for the EPDS, and 5.16 (4.07) for the HADS-D. Characteristics of individual primary studies are available in Supplementary Table 1 with details for PHQ-9 update in Supplementary Method 1. There were 13 studies that were included in both the PHQ-9 and HADS-D datasets, including 2,383 (6%) participants in the PHQ-9 IPDMA and 2,349 participants (15%) in the HADS-D IPDMA. There was no overlap between the EPDS and the PHQ-9 or HADS-D IPDMAs.

Estimates of aORs of major depression classification by diagnostic interview, controlling for depressive symptom severity and other participant characteristics, individually and pooled, are reported in Table 2. Overall odds of major depression classification did not differ for the CIDI versus the SCID (aOR 1.19, 95% CI = 0.79 to 1.75) in the full model that included the interaction term, but there was a significant interaction between the CIDI and depressive symptom severity; as screening tool scores increased, odds of major depression classification increased less for the CIDI than for the SCID (interaction aOR = 0.64, 95% CI = 0.52 to 0.80). As shown in Figure 1, participants with lower depressive symptom severity were more likely to be classified with major depression with the CIDI compared to the SCID, but the opposite was true with greater symptom severity. Compared to the SCID, the MINI classified major depression more often (aOR 1.45; 95% CI = 1.08 to 1.93), controlling for depressive symptom severity and participant characteristics. There was no apparent interaction between symptom levels and odds of classification (interaction aOR = 0.95, 95% CI = 0.78 to 1.15). See Figure 2.

Trends of the probability of major depression classification by reference standards for individual IPDMAs are presented in Supplementary Figures 2-4. There was minimal between-IPDMA heterogeneity in overall aORs for the comparison of the CIDI versus the SCID and the MINI versus the SCID in models without the interaction term (I^2 = 11% and 0%, respectively) and including the interaction term (I^2 = 0% and 0%, respectively). However, there was substantial between-IPDMA heterogeneity of interaction aORs for both comparisons (I^2 = 82% and 82%). See Table 2.

In the comparison of the MINI versus the CIDI, the MINI was more likely to classify participants as having major depression than the CIDI (aOR = 2.05; 95% CI = 1.36 to 2.10), controlling for depressive symptom levels and other participant characteristics. As screening tool scores increased, the odds of major depression classification increased more for the MINI than for the CIDI (interaction aOR = 1.48, 95% CI = 1.36 to 1.60). Heterogeneity was low for aORs with and without the interaction term, and interaction aORs (I^2 = 0%, 0%, and 0%).

In the individual IPDMAs, some results from the EPDS dataset appeared to diverge from those generated in the PHQ-9 and HADS-D datasets. However, the number of studies and cases included in the EPDS dataset for the CIDI and MINI were smaller than any other combination of screening tool and diagnostic interview. See Table 1.

As a sensitivity analysis, we removed the 13 datasets that were included in both the PHQ-9 and HADS-D IPDMAs and re-ran all analyses. Results were similar (see Supplementary Table 2).

DISCUSSION

There were two main findings. First, overall odds of major depression classification did not differ between the fully structured CIDI and the semi-structured SCID. However, adjusting for depressive symptom levels and participant characteristics, odds of major depression classification with the CIDI increased significantly less than for the SCID as depressive symptom levels increased. This suggests that, compared to the SCID, the CIDI is relatively more likely to classify individuals with subthreshold or mild depressive symptoms and relatively less likely to classify people with more severe symptoms. Second, participants evaluated with the MINI were significantly more likely to be classified as having major depression compared to those assessed with the SCID, independent of symptom severity. Between-study heterogeneity was low for models without the interaction term, but higher for models with interaction terms. Estimates from the EPDS IPDMA appeared to diverge somewhat from the PHQ-9 and HADS-D IPDMAs.

This may have been related to the small numbers of studies and major depression cases for the CIDI and MINI among studies that used the EPDS.

Our findings appear to be consistent with characteristics of the different types of diagnostic interviews. The MINI was designed as a screening interview and described by its developers as over-inclusive in classifying psychiatric disorders [10]. For the CIDI, the lack of sensitivity to different levels of depressive symptoms severity may be because the CIDI assesses symptoms in the last 12 months and over the lifetime, then probes to determine if those symptoms are currently present using only a single question. In contrast, the SCID and the MINI specifically assess symptoms in the past two weeks. In addition, the CIDI is much more complicated than the MINI or the SCID. It includes complex branches and is scored using algorithms subject to calibration, which may influence how well diagnoses map onto DSM criteria. This could lead to error at all symptom levels, which would result in more people classified at lower symptom severity levels and fewer at higher levels.

Results were generally consistent with limited evidence from small studies that previously directly compared depression classification by administering semi- and fully structured diagnostic interviews to the same participants. In two studies that examined general population samples with low prevalence, fully structured interviews classified major depression substantially more frequently than semi-structured interviews [2,20]. On the other hand, in a study of participants in inpatient alcohol treatment, where symptom severity would be expected to be higher, depression classification likelihood was similar with semi-structured and fully structured interviews [22].

Our findings have important implications for research, including clinical trials, prognostic and risk factor studies, diagnostic accuracy studies, and prevalence studies. Concerns have been raised about the degree to which antidepressant trials are generalizable to real-world clinical practice [40]. Based on our findings, the method used to classify depression status is also an important consideration. If used to determine trial eligibility, the CIDI may not identify some participants who would be eligible based on the

SCID, whereas both CIDI and MINI may include some participants who would not be eligible based on the SCID, which could reduce the ability to detect treatment effects and further limit applicability to participants in practice who meet diagnostic criteria. Differences in classifying participants could similarly reduce the ability to identify potential associations between risk factors and depression. In diagnostic test accuracy studies, depression screening tool accuracy has been shown to differ across reference standards [33,41,42]. In studies of major depression prevalence, the MINI will overestimate compared to the SCID, whereas with the CIDI, relative prevalence will depend on the underlying distribution of depressive symptoms.

Our findings, which are contrary to the common belief that different reference standards can be treated equivalently in mental health research, provide evidence that different approaches are needed [43]. Ideally, researchers would use semi-structured interviews, such as the SCID, which are designed to replicate diagnostic procedures as closely as possible, to establish diagnostic status. However, this is not always feasible due to the resources required, including highly trained staff. Future studies are needed to develop models to calibrate weights of major depression classification based on different reference standards that could facilitate synthesis of results using different diagnostic interviews. Meanwhile, in selecting a diagnostic interview for use in research, investigators should consider advantages and disadvantages of different interviews, including performance characteristics and resources required. In published studies, authors should comment on potential implications of the type of diagnostic interview that was used. Users of research, including clinicians, should be aware that results from studies that use the CIDI or MINI may differ from what would be found using semi-structured interviews, which are designed to replicate diagnostic procedures as closely as possible. It is also important to underline that from a clinimetric perspective [44-46], assessment of diagnostic status alone is not sufficient, but that rating tools and self-report questionnaires are needed to characterize symptom severity and the specific nature of experienced symptoms.

A strength of the present study was the inclusion of 69,405 participants with 7,574 (11%) major depression cases from 212 studies. This allowed us to overcome limitations of previous IPDMAs and generate more precise estimates. A second strength was that data within each included dataset were standardized in terms of definitions of major depression classification, eligibility criteria, and variables. A limitation to consider is that for included IPDMAs, we could not obtain primary data for 28 of 117 eligible PHQ-9 studies (24% of eligible studies, 17% of eligible participants), 19 of 64 EPDS studies (30% of eligible studies, 30% of eligible participants), and 47 of 116 HADS-D studies (41% of eligible studies, 29% of eligible participants). A second is that we used standardized scores instead of raw depression symptom scores, which required making the assumption that a standard deviation change in scores was equivalent across different screening tools. Third, because only three estimates were pooled, our ability to estimate heterogeneity and explore possible causes was limited. Fourth, some studies were included in both the PHQ-9 and HADS-D IPDMAs. However, a sensitivity analysis showed that results were similar when these studies were removed. Fifth, we examined the SCID, CIDI, and MINI, because we did not have access to enough studies to include other diagnostic interviews. It is unclear to what degree our findings would generalize to other diagnostic interviews. Finally, our study did not include a head-to-head comparison of interviews from a randomized controlled trial or by administering different interviews to all participants. It is unlikely, however, that such as study would be feasible with a large enough sample to draw conclusions with confidence. Our study design, despite its limitations, overcame this barrier.

To conclude, the semi-structured SCID was designed to replicate diagnostic standards and procedures as closely as possible. By synthesizing results from three large IPDMAs, we found that the most commonly used fully structured diagnostic interviews to classify major depression, the CIDI and MINI, did not perform equivalently to the SCID. The CIDI is not as responsive as the SCID to different levels of reported depressive symptoms, and the MINI identifies more cases across the spectrum of depressive symptom levels. Researchers should carefully consider the advantages and disadvantages of using these diagnostic

interviews, and findings from studies based on the CIDI or the MINI should be interpreted considering how their performance deviates from that of the SCID.

Acknowledgements: Not applicable

Author Contributions: YW, BL, JPAI, AB, and BDT were responsible for the study conception and design. BDT contributed a primary dataset that was included in this study. YW, BL, and BDT contributed to data extraction and coding for the meta-analysis. YW, BL, AB, JPAI, and BDT contributed to data analysis and interpretation. YW, AB and BDT contributed to drafting the manuscript. All authors provided a critical review and approved the final manuscript. AB and BDT are the guarantors; they had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analyses.

Group Author Contributions: The DEPRESSD Collaboration includes collaborators who contributed:

To data extraction, coding, and synthesis: Ying Sun, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Chen He, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Ankur Krishnan, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Parash Mani Bhandari, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Dipika Neupane, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Zelalem Negeri, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Mahrukh Imran, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Danielle B. Rice, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Kira E. Riehm, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Nazanin Saadat, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Marleine Azar, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Alexander W. Levis, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Tatiana A. Sanchez, Lady Davis Institute for Medical

Research, Jewish General Hospital, Montréal, Québec, Canada; Matthew J. Chiovitti, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Xin Wei Yan, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada.

Via the design and conduct of database searches: Jill Boruff, McGill University, Montréal, Québec, Canada; Lorie A. Kloda, Concordia University, Montréal, Québec, Canada.

As members of the DEPRESSD Steering Committee, including conception and oversight of collaboration: Pim Cuijpers, Vrije Universiteit, Amsterdam, the Netherlands; Simon Gilbody, University of York, Heslington, York, UK; Dean McMillan, University of York, Heslington, York, UK; Scott B. Patten, University of Calgary, Calgary, Alberta, Canada; Ian Shrier, McGill University, Montréal, Québec, Canada; Roy C. Ziegelstein, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

As knowledge user consultants: Liane Comeau, International Union for Health Promotion and Health Education, École de santé publique de l'Université de Montréal, Montréal, Québec, Canada; Nicholas D. Mitchell, Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada; Marcello Tonelli, Department of Medicine, University of Calgary, Calgary, Alberta, Canada; Simone N. Vigod, Women's College Hospital and Research Institute, University of Toronto, Toronto, Ontario, Canada; Melissa Henry, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Zahinoor Ismail, Hotchkiss Brain Institute & O'Brien Institute for Public Health, Calgary, Alberta, Canada; Carmen G. Loiselle, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada.

By contributing included datasets: Dickens H. Akena, Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda; Samir Al-Adawi, Department of Behavioural Medicine, College of Medicine & Health Sciences, Sultan Qaboos University, Muscat, Oman; Sultan H. Alamri, King Abdulaziz University, Abdullah Sulayman, Jeddah, Makkah, Saudi Arabia; Rubén Alvarado, School of Public Health, Faculty of Medicine, Universidad de Chile, Santiago, Chile; Cosme Alvarado-

Esquivel, Laboratorio de Investigación Biomédica, Facultad de Medicina y Nutrición, Avenida Universidad, Dgo, Mexico; Dagmar Amtmann, Department of Rehabilitation Medicine, University of Washington, Seattle, Washington, USA; Bruce Arroll, Department of General Practice and Primary Health Care, University of Auckland, New Zealand; Liat Ayalon, Louis and Gabi Weisfeld School of Social Work, Bar Ilan University, Ramat Gan, Israel; Muideen O. Bakare, Child and Adolescent Unit, Federal Neuropsychiatric Hospital, Enugu, Nigeria; Hamid R. Baradaran, Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran; Jacqueline Barnes, Department of Psychological Sciences, Birkbeck, University of London, UK; Amar D. Bavle, Department of Psychiatry, Rajarajeswari Medical College and Hospital, Bengaluru, Karnataka, India; Cheryl Tatano Beck, University of Connecticut School of Nursing, Mansfield, Connecticut, USA; Anna Beraldi, Kbo-Lech-Mangfall-Klinik Garmisch-Partenkirchen, Klinik für Psychiatrie, Psychotherapie & Psychosomatik, Lehrkrankenhaus der Technischen Universität München, Munich, Germany; Charles N. Bernstein , University of Manitoba IBD Clinical and Research Centre, Winnipeg, Manitoba, Canada; Arvin Bhana, Centre for Rural Health, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa; Carola Bindt, Department of Child and Adolescent Psychiatry, University Medical Center Hamburg-Eppendorf, Germany; Charles H. Bombardier, Department of Rehabilitation Medicine, University of Washington, Seattle, Washington, USA; Philip M. Boyce, Discipline of Psychiatry, Westmead Clinical School, Sydney Medical School, University of Sydney, Sydney, Australia; Natalie Büel-Drabe, Department of Psychiatry and Psychotherapy, University Hospital Zürich, Zürich Switzerland; Ryna Imma Buji, Department of Psychiatry, Hospital Mesra Bukit Padang, Sabah, Malaysia; Adomas Bunevicius, Neuroscience Institute, Lithuanian University of Health Sciences, Kaunas ,Lithuania; Dr. Jurate Butnoriene, PhD, who did the data collection and analysis as part of her PhD thesis for the primary study by Butnoriene et al., passed away and was unable to participate in this project; Dr. Robertas Bunevicius, MD, PhD (1958-2016) was the Principal Investigator of the primary studies by

Butnoriene et al. and Bunevicius et al, but passed away and was unable to participate in this project; Peter Butterworth, Centre for Research on Ageing, Health and Wellbeing, Research School of Population Health, The Australian National University, Canberra, Australia; Gregory Carter, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; Marcos H. Chagas, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Juliana C. N. Chan, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China; Lai Fong Chan, Department of Psychiatry, National University of Malaysia, Kuala Lumpur, Malaysia; Linda H. Chaudron, Departments of Psychiatry, Pediatrics, Obstetrics and Gynecology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA; Chih-Ken Chen, Community Medicine Research Center, Keelung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Keelung, Taiwan; Rushina Cholera, Department of Pediatrics, Duke University, Durham, North Carolina, USA; Kerrie Clover, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; Ronán M. Conroy, Royal College of Surgeons in Ireland Division of Population Health Sciences, Dublin, Ireland; Aaron Conway, Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, Canada; Yeates Conwell, Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA; Humberto Correa, Faculty of Medicine, Universidade Federal de Minas Gerais. Belo Horizonte, MG, Brazil; Tiago Castro e Couto, Federal University of Uberlândia, Brazil; Daniel Cukor, Rogosin Institute, New York, New York, USA; Eli Dabscheck, The Alfred Hospital, Prahran, VIC, Australia; Federico M. Daray, Institute of Pharmacology, School of Medicine, University of Buenos Aires, Argentina; Felipe Pinheiro de Figueiredo, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, Brazil; Janneke M. de Man-van Ginkel, Julius Center for Health Sciences and Primary Care, Department of Nursing Science, University Medical Center Utrecht – University Utrecht, Utrecht, the Netherlands; Crisanto Diez-Quevedo, Servei de Psiquiatria, Hospital Germans Trias i Pujol, Badalona,

Spain; Elles Douven, Alzheimer Center Limburg and School for Mental Health and Neuroscience (MHeNs), Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, the Netherlands; Marina G. Downing, School of Psychological Sciences, Monash University, Melbourne VIC, Australia; Valsamma Eapen, School of Psychiatry, University of New South Wales, Kensington, Australia; Jesse R. Fann, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; Anthony Feinstein, University of Toronto, Toronto, Ontario, Canada; Panagiotis P. Ferentinos, National and Kapodistrian University of Athens, 2nd Department of Psychiatry, Attikon General Hospital, Athens, Greece; Michelle Fernandes, Faculty of Medicine, Department of Paediatrics, University of Southampton, Southampton and Nuffield Department of Women's & Reproductive Health, University of Oxford, UK; Sally Field, Perinatal Mental Health Project, Alan J Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa; Barbara Figueiredo, School of Psychology, University of Minho, Portugal; Felix H. Fischer, Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Germany; Jane R. W. Fisher, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Alastair J. Flint, University Health Network, Toronto, Ontario, Canada; Maiko Fujimori, Section of Psychological Science, Division of Health Care Research, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; Daniel S. S. Fung, Department of Developmental Psychiatry, Institute of Mental Health, Singapore; Pamela Gallagher, School of Psychology, Dublin City University, Dublin, Ireland; Milena Gandy, The Department of Psychology, Macquarie University, Sydney, Australia; Lluïsa Garcia-Esteve, Perinatal Mental Health Unit CLINIC-BCN. Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Emily C. Garman, Alan J Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town; Bizu Gelaye, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; Leila Gholizadeh, Faculty of Health, University of Technology Sydney, Sydney, Australia; Lisa Giardinelli,

Psychiatry Unit, Department of Health Sciences, University of Florence, Firenze, Italy; Lorna J. Gibson, Tropical Epidemiology Group, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; Felicity Goodyear-Smith, Department of General Practice and Primary Health Care, University of Auckland, New Zealand; Luigi Grassi, Institute of Psychiatry, Department of Biomedical and Specialty Surgical Sciences, University of Ferrara, Ferrara, Italy; Eric P. Green, Duke Global Health Institute, Duke University, Durham, North Carolina, USA; Catherine G. Greeno, School of Social Work, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; Brian J. Hall, Global and Community Mental Health Research Group, Department of Psychology, Faculty of Social Sciences, University of Macau, Macau Special Administrative Region, China; Liisa Hantsoo, Department of Psychiatry & Behavioral Sciences , The Johns Hopkins University School of Medicine, Baltimore, Maryland; Emily E. Haroz, Center For American Indian Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health; Martin Härter, Department of Medical Psychology, University of Hamburg, Hamburg, Germany; Ulrich Hegerl, Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe-Universität Frankfurt, Germany; Nadine Helle, Department of Child and Adolescent Psychiatry, University Medical Center Hamburg-Eppendorf, Germany; Leanne Hides, School of Psychology, University of Queensland, Brisbane, Queensland, Australia; Stevan E. Hobfoll, STAR-Stress, Anxiety and Resilience Consultants, Chicago, Illinois, USA; Simone Honikman, Perinatal Mental Health Project, Alan J Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town; Louise M. Howard, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; Marie Hudson, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Thomas Hyphantis, Department of Psychiatry, Faculty of Medicine, School of Health Sciences, University of Ioannina, Greece; Masatoshi Inagaki, Department of Psychiatry, Faculty of Medicine, Shimane University, Shimane, Japan; Josef Jenewein, Clinic Zugersee, Center for Psychiatry and Psychotherapie, Oberwil-Zug, Switzerland; Hong Jin Jeon, Department of Psychiatry,

Depression Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Suwon, South Korea; Nathalie Jetté, Departments of Neurology and Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, New York, USA; Monika Keller, Division of Psycho-oncology, Department of General Internal Medicine and Psychosomatics, University Hospital Heidelberg, Germany; Dina Sami Khalifa, Faculty of Health Sciences, Ahfad University for Women, Omdurman, Sudan; Mohammad E. Khamseh, Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran; Kim M. Kiely, School of Psychology, The University of New South Wales, and Neuroscience Research Australia (NeuRA), Sydney, Australia; Sung-Wan Kim, Department of Psychiatry, Chonnam National University Medical School, Republic of Korea; Marie Kjærgaard, Endocrinology Research Group, Medical Clinic, University Hospital of North Norway, Norway; Sebastian Köhler, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands; Jane Kohlhoff, School of Psychiatry, University of New South Wales, Kensington, Australia; Brandon A. Kohrt, Department of Psychiatry and Behavioral Sciences, The George Washington University, Washington, DC, USA; Zoltán Kozinszky, Department of Obstetrics and Gynecology, Danderyd Hospital, Stockholm, Sweden; Laima Kusminskas, Private Practice, Hamburg, Germany; Yunxin Kwan, Department of Psychological Medicine, Tan Tock Seng Hospital, Singapore; Femke Lamers, Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, the Netherlands; Maria Asunción Lara, Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz. San Lorenzo Huipulco, Tlalpan, México D. F. Mexico; Lorenzo Lelli, Psychiatry Unit, Department of Health Sciences, University of Florence, Firenze, Italy; Angeliki A. Leonardou, First Department of Psychiatry, Women's Mental Health Clinic, Athens University Medical School, Athens, Greece; Holly F. Levin-Aspenson, Department of Psychology, University of Notre Dame, Notre Dame, Indiana, USA; Manote Lotrakul, Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; Sonia R. Loureiro, Department of Neurosciences and

Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Bernd Löwe, Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Nagendra P. Luitel, Research Department, TPO Nepal, Kathmandu, Nepal; Crick Lund , Alan J Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town; Michael Maes, Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; Ruth Ann Marrie, Departments of Medicine and Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada; Laura Marsh, Baylor College of Medicine, Houston and Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas, USA; Rocio Martin-Santos, Department of Psychiatry and Psychology, Hospital Clinic, IDIBAPS, CIBERSAM, Barcelona, Spain; Brian P. Marx, National Center for PTSD at VA Boston Healthcare System, Boston, MA, USA; Loreto Massardo, Centro de Biología Celular y Biomedicina, Facultad de Medicina y Ciencia, Universidad San Sebastián. Santiago, Chile; Yutaka Matsuoka, Division of Health Care Research, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; Anja Mehner, Department of Medical Psychology and Medical Sociology, University of Leipzig, Germany; Valentina Meuti, Department of Neurology and Psychiatry, Sapienza University of Rome, Rome, Italy; Ioannis Michopoulos, 2nd Department of Psychiatry, Attikon General Hospital, National and Kapodistrian University of Athens, Athens, Greece; Laurent Misery, Department of Dermatology, University Hospital of Brest, Brest, France; Sherina Mohd Sidik, Cancer Resource & Education Centre, and Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia; Tiago N. Munhoz, Post-graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS, Brazil; Kumiko Muramatsu, Department of Clinical Psychology, Graduate School of Niigata Seiryo University, Niigata, Japan; Sandra Nakić Radoš, Department of Psychology, Catholic University of Croatia, Zagreb, Croatia; Juliet E. M. Nakku, Butabika National Referral Teaching Hospital; Laura Navarrete, Department of Epidemiology and Psychosocial Research,

Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Ciudad de México, México; Purificación Navarro García, Perinatal Mental Health Unit CLINIC-BCN. Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Ricard Navines, Department of Psychiatry and Psychology, Hospital Clinic, IDIBAPS, CIBERSAM, Barcelona, Spain; Daisuke Nishi, Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Japan; Meaghan L. O'Donnell, Phoenix Australia, Carlton VIC, Australia; Daniel Okitundu Luwa E-Andjafono, Unité de Neuropsychologie, Département de Neurologie, Centre Neuro-psycho-pathologique, Faculté de Médecine, Université de Kinshasa, République Démocratique du Congo; Flávia L. Osório, Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; Ahmet Öztürk, Bezmialem Vakif University, Istanbul, Turkey; Jurate Peceliuniene, Vilnius University Faculty of Medicine, Clinic of Internal Diseases, Family Medicine and Oncology, Vilnius, Lithuania; Brian W. Pence, Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; Philippe Persoons, Department of Adult Psychiatry, University Hospitals Leuven, Leuven, Belgium; Angelo Picardi, Centre for Behavioural Sciences and Mental Health, Italian National Institute of Health, Rome, Italy; Luis Pintor, Consultation Liaison Psychiatry Unit. Hospital Clínico de Barcelona, Barcelona, Spain; Jennie L. Ponsford, School of Psychological Sciences, Monash University, Melbourne VIC, Australia; Stephanie L. Pugh, NRG Oncology Statistics and Data Management Center, Philadelphia, PA, USA; Terence J. Quinn, Institute of Cardiovascular & Medical Sciences, University of Glasgow, Glasgow, Scotland; Elmars Rancans, Department of Psychiatry and Narcology, Riga Stradins University, Latvia; Sujit D. Rathod, Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom; Silje E. Reme, Department of psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway; Katrin Reuter, Group Practice for Psychotherapy and Psycho-oncology, Freiburg, Germany; Emma Robertson-Blackmore, Halifax Health, Graduate Medical Education, Daytona Beach, FL. USA; Tamsen J. Rochat, Developmental Pathways to

Health Research Unit, Faculty of Health Sciences, University of Witwatersrand, South Africa; Alasdair G. Rooney, Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, Scotland; Heather J. Rowe, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Roberto Sánchez-González, Department of Psychiatry. Institut de Neuropsiquiatria i Addiccions, Centre Emili Mira, Parc de Salut Mar. Barcelona. Spain; Iná S. Santos, Post-graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS, Brazil; Miranda T. Schram, Department of Internal Medicine, Maastricht University Medical Center, Maastricht, The Netherlands; Marcelo L. Schwarzbold, Department of Internal Medicine, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil; Vesile Senturk Cankorur, Ankara University Faculty of Medicine Psychiatry Department, Ankara, Turkey; Juwita Shaaban, Department of Family Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia; Louise Sharpe, School of Psychology, The University of Sydney, Sydney NSW, Australia; Eileen H. Shinn, Department of Behavioral Science, University of Texas M. D. Anderson Cancer Center, Houston, Texas, USA; Abbey Sidebottom, Allina Health, Minneapolis, Minnesota, USA; Sébastien Simard, Département des sciences de la santé, Université du Québec à Chicoutimi (UQAC), Québec, Canada; Adam Simning, Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA; Susanne Singer, University Medical Centre Mainz, Institute of Medical Biostatistics, Epidemiology and Informatics, Mainz, Germany; Bonnie W. M. Siu, Department of Psychiatry, Castle Peak Hospital, Hong Kong SAR, China; Alkistis Skalkidou , Department of Women’s and Children's Health, Uppsala University, Uppsala, Sweden; Lena Spangenberg, Department of Medical Psychology and Medical Sociology, University of Leipzig, Germany; Lesley Stafford, Centre for Women's Mental Health, Royal Women's Hospital, Parkville, Australia; Alan Stein, Department of Psychiatry, University of Oxford, Oxford, UK; Robert C. Stewart, Division of Psychiatry, University of Edinburgh, Edinburgh, UK; Jon Stone, University of Edinburgh, Edinburgh, UK; Kuan-Pin Su, Tainan Municipal An-Nan Hospital & College of Medicine, China Medical University, Taichung, Taiwan; Serge Sultan,

Université de Montréal, Québec, Canada; Inger Sundström-Poromaa, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; Sharon C. Sung, Department of Child & Adolescent Psychiatry, Institute of Mental Health, Singapore; Keiko Suzuki, Department of General Medicine, Asahikawa University Hospital, Asahikawa, Hokkaido, Japan; Meri Tadinac, Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia; Pei Lin Lynnette Tan, Department of Psychological Medicine, Tan Tock Seng Hospital, Singapore; S. Darius Tandon, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; Martin Taylor-Rowan, Institute of Cardiovascular and Medical Science, University of Glasgow, Glasgow, Scotland; Antonio L. Teixeira, University of Texas Health Science Center at Houston, Houston, Texas, USA; Iva Tendais, School of Psychology, University of Minho, Portugal; Pavaani Thiagayson, Institute of Mental Health, Singapore; Istvan Tiringer, Pécs University, Medical School, Institute of Behavioral Sciences, Pécs, Hungary; Annamária Töreki, Department of Emergency, University of Szeged, Hungary; Anna Torres-Giménez, Perinatal Mental Health Unit CLINIC-BCN. Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Thach D. Tran, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Kylee Trevillion, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; Ka-Yee Tung, Kwai Chung Hospital, Hong Kong SAR, China; Alyna Turner, School of Medicine and Public Health, University of Newcastle, New South Wales, Newcastle, Australia; Katherine Turner, Epilepsy Center-Child Neuropsychiatry Unit, ASST Santi Paolo Carlo, San Paolo Hospital, Milan, Italy; Christina M. van der Feltz-Cornelis, Department of Health Sciences, HYMS, University of York, York, UK; Thandi van Heyningen , Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa ; Henk C. van Weert, Department General Practice, Institute Public Health, Amsterdam Universities Medical Centers, Amsterdam, the Netherlands; Johann M. Vega-Dienstmaier, Facultad de Medicina Alberto Hurtado, Universidad Peruana Cayetano Heredia. Lima, Perú; Paul A. Vöhringer, Department of Psychiatry

and Mental Health, Clinical Hospital, Universidad de Chile, Santiago, Chile; Lynne I. Wagner, Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA; Mark Walterfang, Neuropsychiatry Unit, Royal Melbourne Hospital, Melbourne, Australia; Jian Li Wang, University of Ottawa Institute of Mental Health Research, Ottawa, Canada; Wenzheng Wang, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China; Liang-Jen Wang, Department of Child and Adolescent Psychiatry, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan; Jennifer White, Department of Physiotherapy, School of Primary and Allied Health Care, Monash University, Melbourne, Australia; Dana K. Wong, School of Psychology & Public Health, La Trobe University, Melbourne, Australia; Karen Wynter, School of Nursing and Midwifery, Deakin University, Melbourne, Australia; Mitsuhiko Yamada, Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawa-Higashi, Kodaira, Tokyo, Japan; Kimberly A. Yonkers, Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut, USA; Qing Zhi Zeng, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China; Yuying Zhang, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China.

Funding Sources: This study was funded by the Canadian Institutes of Health Research (KRS-134297, PCG-155468, PJT-162206, KRS-140994, KRS-144045). Drs. Wu and Levis were supported by Fonds de recherche du Québec – Santé (FRQS) Postdoctoral Training Fellowships. Dr. Benedetti was supported by a FRQS researcher salary award.

Disclosure Statement: All authors have completed the ICJME uniform disclosure form and declare no support from any organization for the submitted work; no financial relationships with any organizations that

might have an interest in the submitted work in the previous three years. 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.

Statement of Ethics: The authors have no ethical conflicts to disclose.

Transparency Declaration: The manuscript's guarantor affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Data Sharing: Requests to access data should be made to the corresponding authors.

REFERENCES

1. Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychol Med. 1999;29:1013-1020.
2. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE: A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Psychol Med. 2001;31:1001-1013.
3. Nosen E, Woody SR: Chapter 8: Diagnostic Assessment in Research: In Handbook of research methods in abnormal and clinical psychology. Edited by McKay D. Thousand Oaks, CA, Sage, 2008, pp 109-124.
4. First MB: Structured Clinical Interview for the DSM (SCID). John Wiley & Sons, Inc. 1995.
5. Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured vs. fully structured diagnostic interviews. Br J Psychiatry. 2018;212:377-385
6. 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:e1803.
7. 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.
8. Robins LN, Wing J, Wittchen HU, et al: The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry. 1988;45:1069-1077.

9. Lecrubier Y, Sheehan DV, Weiller E, et al: The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatry. 1997;12:224-231.
10. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. Eur Psychiatry. 1997;12:232-241.
11. Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. CMAJ. 2008;178:997-1003.
12. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry. 2015;37:567-576.
13. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ. 2012;184:E191-E196.
14. Manea L, Gilbody S, Hewitt C, et al: Identifying depression with the PHQ-2: a diagnostic meta-analysis. J Affect Disord. 2016;203:382-395.
15. Hewitt CE, Gilbody SM, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess. 2009;13:1-145,147-230.
16. Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. J Affect Disord. 2010;126;335-348.
17. Rice DB, Kloda LA, Shrier I, Thombs BD: Reporting completeness and transparency of meta-analyses of depression screening tool accuracy: a comparison of meta-analyses published before and after the PRISMA statement. J Psychosom Res. 2016;87:57-69.
18. Lu J, Huang YQ, Liu ZR, Cao XL. Validity of Chinese version of the composite international diagnostic interview-3.0 in psychiatric settings. Chin Med J. 2015;128:2462-2466.

19. Gelaye B, Williams MA, Lemma S, et al. Diagnostic validity of the composite international diagnostic interview (CIDI) depression module in an East African population. Int J Psychiat Med. 2013;46:387-405.
20. Anthony JC, Folstein M, Romanoski AJ, et al. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in eastern Baltimore. Arch Gen Psychiatry. 1985;42:667-675.
21. Booth BM, Kirchner JA, Hamiltonc G, Harrell R, Smith GR. Diagnosing depression in the medically ill: validity of a lay-administered structured diagnostic interview. J Psychiatr Res. 1998;32:353-360.
22. Hesselbrock V, Stabenau J, Hesselbrock M, Mirkin P, Meyer R. A comparison of two interview schedules: the Schedule for Affective Disorders and Schizophrenia-Lifetime and the National Institute for Mental Health Diagnostic Interview Schedule. Arch Gen Psychiatry. 1982;39:674-677.
23. Jordanova V, Wickramesinghe C, Gerada C, Prince M. Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. Psychol Med. 2004;34:1013-1024.
24. Diagnostic and statistical manual of mental disorders: DSM-III 3rd ed, revised. Washington, DC: American Psychiatric Association 1987.
25. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed. Washington, DC: American Psychiatric Association 1994.
26. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed, text revised. Washington, DC: American Psychiatric Association 2000.
27. Diagnostic and statistical manual of mental disorders: DSM-V 5th ed. Washington, DC: American Psychiatric Association 2013.
28. The ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization 1992.

29. 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 (PHQ9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. Syst Rev. 2014;3:124.
30. Thombs BD, Benedetti A, Kloda LA, et al: Diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: protocol for a systematic review and individual patient data meta-analyses. BMJ Open. 2015;5:e009742.
31. Thombs BD, Benedetti A, Kloda LA, et al: Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. BMJ Open. 2016;6:e011913.
32. United Nations Development Programme: Human Development Report 2019. http://hdr.undp.org/sites/default/files/hdr2019.pdf. Accessed January 14, 2020.
33. Levis B, Benedetti A, Thombs BD, and the DEPRESsion Screening Data (DEPRESSD) Collaboration: Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ. 2019;365:l1476.
34. Deeks JJ, Higgins JP, Altman DG: Chapter 9: Analysing data and undertaking meta-analyses. In Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Edited by Higgins JP, Green S, editors. The Cochrane Collaboration, 2011.
35. Higgins JP, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-560.
36. Team RC: R: a language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, 2018.
37. Team R: RStudio: integrated development for R. Boston, MA, RStudio, Inc., 2015, pp 639-640.

38. Bates D, Machler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2014;67:1-48.
39. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1-48.
40. Zimmerman M, Balling C, Chelminski I, Dalrymple K. Have treatment studies of depression become even less generalizable? Applying the inclusion and exclusion criteria in placebo-controlled antidepressant efficacy trials published over 20 years to a clinical sample. Psychother Psychosom. 2019;88:165-70.
41. Wu Y, Levis B, Riehm KE, et al: Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. Psychol Med 2019;12:1-13.
42. He C, Levis B, Riehm KE, et al. The accuracy of the Patient Health Questionnaire-9 algorithm for screening to detect major depression: An individual participant data meta-analysis. Psychother Psychosom. 2020;89:25-37.
43. Fava GA. The decline of pluralism in medicine: dissent is welcome. Psychother Psychosom. 2020;89:1-5.
44. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. Int J Clin Pract. 2012;66:11-15.
45. Fava GA, Carrozzino D, Lindberg L, Tomba E. The clinimetric approach to psychological assessment: a tribute to Per Bech, MD (1942-2018). Psychother Psychosom. 2018;87:321-326.
46. Tomba E, Bech P. Clinimetrics and clinical psychometrics: macro- and micro-analysis. Psychother Psychosom. 2012;81:333-343.

FIGURE LEGENDS

Figure 1. Comparison of major depression classification odds of the Composite International Diagnostic Interview (CIDI) versus the Structured Clinical Interview for DSM (SCID)

The figure presents the aOR of major depression classification for the CIDI compared to the SCID for primary studies based on the PHQ-9, EPDS, and HADS-D and pooled estimates at standardized scores of -1, 0, 1, 2 and 3. The standardized scores of -1, 0, 1, 2 and 3 are approximately equal to scores of 0, 5, 10, 16 and 21 on the PHQ-9 (SD = 5.26); 1, 7, 13, 18 and 24 on the EPDS (SD = 5.58); and 1, 5, 9, 13 and 17 on the HADS-D (SD = 4.07). We present standardized scores from -1 to 3, because raw scores corresponding to standardized scores below -1 or above 3 would be negative or beyond the maximum scores of the included screening tools.

Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; HADS-D: Depression subscale of Hospital Anxiety and Depression Scale; META: Pooled estimates from the synthesis meta-analysis. PHQ-9: Patient Health Questionnaire-9.

719 **Figure 2.** Comparison of major depression classification odds of the Mini International Neuropsychiatric
720 Interview (MINI) vs. the SCID considering the interaction between depressive symptom severity and the
721 MINI

722

723 The figure presents the aOR of major depression classification for the MINI compared to the SCID for
724 primary studies based on the PHQ-9, EPDS, and HADS-D and pooled estimates at standardized scores of -
725 1, 0, 1, 2 and 3. The standardized scores of -1, 0, 1, 2 and 3 are approximately equal to scores of 0, 5, 10, 16
726 and 21 on the PHQ-9 (SD = 5.26); 1, 7, 13, 18 and 24 on the EPDS (SD = 5.58); and 1, 5, 9, 13 and 17 on
727 the HADS-D (SD = 4.07). We present standardized scores from -1 to 3, because raw scores corresponding
728 to standardized scores below -1 or above 3 would be negative or beyond the maximum scores of the
729 included screening tools.

730

731 Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; HADS-D: Depression subscale of Hospital
732 Anxiety and Depression Scale; META: Pooled estimates from the synthesis meta-analysis. PHQ-9: Patient
733 Health Questionnaire-9.

734