Article:

Probability of major depression diagnostic classification using semi-structured vs. fully structured diagnostic interviews

Brooke Levis, Andrea Benedetti, Kira E. Riehm, Nazanin Saadat, Alexander W. Levis, Marleine Azar, Danielle B. Rice, Matthew J. Chiovitti, Tatiana A. Sanchez, Pim Cuijpers, Simon Gilbody, John P.A. Ioannidis, Lorie A. Kloda, Dean McMillan, Scott B. Patten, Ian Shrier, Russell J. Steele, Roy C. Ziegelstein, Dickens H. Akena, Bruce Arroll, Liat Ayalon, Hamid R. Baradaran, Murray Baron, Anna Beraldi, Charles H. Bombardier, Peter Butterworth, Gregory Carter, Marcos H. Chagas, Juliana C. N. Chan, Rushina Cholera, Neerja Chowdhary, Kerrie Clover, Yeates Conwell, Janneke M. de Man-van Ginkel, Jaime Delgadillo, Jesse R. Fann, Felix H. Fischer, Benjamin Fischler, Daniel Fung, Bizu Gelaye, Felicity Goodyear-Smith, Catherine G. Greeno, Brian J. Hall, John Hambridge, Patricia A. Harrison, Ulrich Hegerl, Leanne Hides, Stevan E. Hobfoll, Marie Hudson, Thomas Hyphantis, Masatoshi Inagaki, Khalida Ismail, Nathalie Jetté, Mohammad E. Khamseh, Kim M. Kiely, Femke Lamers, Shen-Ing Liu, Manote Lotrakul, Sonia R. Loureiro, Bernd Löwe, Laura Marsh, Anthony McGuire, Sherina Mohd Sidik, Tiago N. Munhoz, Kumiko Muramatsu, Flávia L. Osório, Vikram Patel, Brian W. Pence, Philippe Persoons, Angelo Picardi, Alasdair G. Rooney, Iná S. Santos, Juwita Shaaban, Abbey Sidebottom, Adam Simning, Lesley Stafford, Sharon Sung, Pei Lin Lynnette Tan, Alyna Turner, Christina M. van der Feltz-Cornelis, Henk C. van Weert, Paul A. Vöhringer, Jennifer White, Mary A. Whooley, Kirsty Winkley, Mitsuhiko Yamada, Yuying Zhang, and Brett D. Thombs.

Corresponding author:

Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Ste Catherine Road; Montreal,

Quebec H3T 1E4; Tel (514) 340-8222 ext. 5112; E-mail: brett.thombs@mcgill.ca

Word count: 3,884

ABSTRACT

Background: Different diagnostic interviews are used as reference standards for major depression classification in research. Semi-structured interviews involve clinical judgement, whereas fully structured interviews are completely scripted. The Mini International Neuropsychiatric Interview (MINI), a brief fully structured interview, is also sometimes used. It is not known whether interview method is associated with probability of major depression classification. **Aims:** To evaluate the association between interview method and odds of major depression classification, controlling for depressive symptom scores and participant characteristics. **Method**: Data collected for an individual participant data meta-analysis of Patient Health Questionnaire-9 (PHQ-9) diagnostic accuracy were analyzed. Binomial Generalized Linear Mixed Models were fit. **Results**: 17,158 participants (2,287 major depression cases) from 57 primary studies were analyzed. Among fully structured interviews, odds of major depression were higher for the MINI compared to the Composite International Diagnostic Interview (CIDI) [OR (95% CI) = 2.10 (1.15-3.87)]. Compared to semi-structured interviews, fully structured interviews (MINI excluded) were non-significantly more likely to classify participants with low-level depressive symptoms (PHQ-9 scores ≤6) as having major depression [OR (95% CI) = 3.13 (0.98-10.00)], similarly likely for moderate-level symptoms (PHQ-9 scores 7-15) [OR (95% CI) = 0.96 (0.56-1.66)], and significantly less likely for high-level symptoms (PHQ-9 scores ≥16) [OR (95% CI) = 0.50 (0.26-0.97)]. **Conclusions**: The MINI may identify more depressed cases than the CIDI, and semi- and fully structured interviews may not be interchangeable methods, but these results should be replicated. **Declaration of Interest**: This study was funded by the Canadian Institutes of Health Research (KRS-134297).

DECLARATION OF INTEREST

Conflict of Interest Disclosures:

INTRODUCTION

Historically, major depression classification in research was done by clinical judgement or unstructured interviews. Lack of agreement between interviewers led to the development of standardized diagnostic interviews, including semi-structured interviews, designed to be administered by clinicians, and fully structured interviews, which can be administered by lay interviewers.[1,2] Semi-structured interviews are akin to a guided diagnostic conversation. Standardized questions are asked, but interviewers may insert additional queries and use clinical judgement to decide whether symptoms are present.[2,3] Examples include the Structured Clinical Interview for DSM (SCID) and Schedules for Clinical Assessment in Neuropsychiatry (SCAN).[4,5] In contrast, fully structured interviews typically involve fully scripted, standardized questions that are read verbatim, without additional probes.[2,3] They are designed to be less subjective and provide greater standardization, but with less flexibility and without incorporating clinical judgment.[2,3,6] Examples include the Composite International Diagnostic Interview (CIDI) and the Diagnostic Interview Schedule (DIS).[7,8] The Mini International Neuropsychiatric Interview (MINI) is also a fully structured interview, but it differs from the CIDI and DIS in that it was described by its authors as designed to be able to be administered in a fraction of the time at the cost of being over-inclusive and generating a higher rate of false-positive diagnoses.[9,10]

Although fully structured interviews are sometimes referred to as imperfect reference standards compared to semi-structured interviews,[11] both are considered appropriate reference standards for major depression classification in research.[2] Consistent with this, existing meta-analyses on depression screening tool accuracy have treated both interview types as equivalent reference standards.[12] For different interviews to be treated as equivalent diagnostic standards, the probability of being classified as meeting diagnostic criteria should not depend on the

interview administered. Different interview formats, however, may lead to different diagnostic

patterns. For instance, it is possible that the greater standardization and reliability across

interviews gained in fully structured interviews, compared to clinician-administered semi-

structured interviews, could increase misclassification.

Five studies have administered validated semi- and fully structured interviews to the same

set of participants in non-psychiatric settings within a 2-week period to assess current major

depression (SupplementaryTable1).[11,13–16] Most included small numbers of participants and

major depression cases. Nonetheless, in the three studies with ≥100 participants, prevalence of

major depression was more than twice as high when assessed with fully structured interviews

compared to semi-structured interviews. No studies have randomized participants to receive

either a fully or semi-structured interview and compared major depression prevalence.

The high cost and burden of administering multiple diagnostic interviews to large numbers

of participants or, alternatively, randomizing large numbers of participants to receive semi- or

fully structured interviews, presents a substantial barrier to testing for differences between

interview types. An alternative would be to compare the probability of being classified as having

major depression using different interview types, controlling for depression symptom severity

and other factors potentially related to classification. Individual participant data (IPD) meta-

analysis, in which participant-level data from many studies are synthesized, offers a way to

examine the association between diagnostic method and probability of major depression

classification across a large number of participants, controlling for factors potentially associated

with classification, including depressive symptom severity.

The objective of this study was to examine the association between diagnostic interview

method and major depression classification. First, we compared the odds of major depression

classification using different diagnostic interviews, first among semi-structured interviews and then separately among fully structured interviews, in each case controlling for depressive symptom severity and study- and participant-level characteristics. Second, we compared the odds of major depression classification between the semi- and fully structured interviews, including a focus on the interviews with the largest numbers of patients, the SCID and the CIDI, and controlling for depressive symptom severity and study and participant-level characteristics. Third, we tested whether differences in the odds of classification across interview types were associated with depressive symptom severity.

METHOD

This study used data accrued for an IPD meta-analysis on the diagnostic accuracy of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool to detect major depression. Detailed methods were registered in PROSPERO (CRD42014010673), and a protocol was published.[17] As an initial step, we assessed the comparability of diagnostic classifications generated by different diagnostic interviews.

Search Strategy

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and Web of Science from January 2000 - December 2014 on February 7, 2015, using a search strategy (SupplementaryMethods1), which was peer-reviewed using PRESS.[18] We limited our search to these databases based on research showing that adding other databases when the Medline search is highly sensitive does not identify additional eligible studies.[19] The search was limited to the year 2000 forward because the PHQ-9 was published in 2001.[20] We reviewed reference lists of relevant reviews and queried contributing authors about non-published studies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda,

MD, USA). After de-duplication, unique citations were uploaded into DistillerSR (Evidence Partners, Ottawa, Canada), which was used to store and track search results and track the review process.

Identification of Eligible Studies

Datasets from articles in any language were eligible for inclusion if they included diagnostic classification for current Major Depressive Disorder (MDD) or Major Depressive Episode (MDE) based on a validated semi- or fully structured interview conducted within two weeks of PHQ-9 administration, since diagnostic criteria are for symptoms in the last two weeks. Datasets where not all participants were administered the PHQ-9 within two weeks of the diagnostic interview were included if the primary data allowed us to select participants administered the diagnostic interview and PHQ-9 within two weeks. Data from studies where the PHQ-9 was administered exclusively to patients known to have psychiatric diagnoses or symptoms were excluded, since screening is not done with patients already managed in psychiatric settings.[21] For defining major depression, we considered MDD or MDE based on the Diagnostic and Statistical Manual of Mental Disorders (DSM), or MDE based on the International Classification of Diseases (ICD). If more than one was reported, we prioritized DSM over ICD, and DSM MDE over DSM MDD. We prioritized MDE over MDD because screening tests are intended to identify symptoms of depression and not rule out due to bipolar disorder. We prioritized DSM over ICD because DSM is more commonly used in existing studies. However, across all studies, there were only 23 discordant diagnoses that depended on classification prioritization (0.1% of participants).

Two investigators independently reviewed titles and abstracts for eligibility. If either reviewer deemed a study potentially eligible, a full-text article review was completed, also by

13

two investigators independently. Seven members of the research team participated in the review process; however, each title and abstract and each full text was reviewed independently by only two of the seven investigators. Disagreement between reviewers after full-text review was resolved by consensus, including a third investigator (either BL or BDT) when necessary. Titles, abstracts and full-text articles in languages other than English were translated by members of the research team or by advanced research trainees who were native speakers of the language and familiar with the topic. They were not paid for their translation services.

Data Contribution and Synthesis

Authors of eligible datasets were invited to contribute de-identified primary data. Primary study country, clinical setting, language, and diagnostic interview administered were extracted from published reports by two investigators independently, with disagreements resolved by consensus. Countries were categorized as "very high", "high", or "low-medium" development level based on the United Nation's human development index.[22] Recruitment settings were categorized as "non-medical", "primary care", "inpatient specialty care", or "outpatient specialty care". Participant-level data included age, sex, major depression status, and PHQ-9 scores. In three primary studies, multiple settings were included, thus setting was coded at the participant-level.

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

Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[23] See

SupplementaryMethods2 for QUADAS-2 coding rules. Discrepancies in data extraction and risk

of bias assessment were resolved by consensus.

Statistical Analyses

To isolate the association between diagnostic assessment method and major

depression classification, we estimated binomial Generalized Linear Mixed Models

(GLMMs) with a logit link function. In all analyses, the outcome was major depression

classification. The predictor of interest was either the specific diagnostic interview or

interview category, depending on the analysis. Covariates were depressive symptom

severity (PHQ-9 score), age, sex, country human development index, and clinical

setting. The PHQ-9 has been shown in many studies, across diverse populations in both

medical and non-medical settings, to be a valid measure of depressive symptom severity

with good convergent validity and a one-dimensional factor structure.[20,24–27] Other

covariates were chosen due to their potential influence on major depression

classification and their availability across primary studies. To account for correlation

between subjects within the same primary study, a random intercept was fit for each

primary study. Fixed slopes were estimated for PHQ-9 score, assessment method, age,

sex, human development index, and clinical setting.

First, we estimated a GLMM among studies that used semi-structured interviews

(SCID, SCAN, Depression Interview and Structured Hamilton [DISH]). Then, we

estimated a GLMM among studies that used fully structured interviews (CIDI, Clinical

Interview Schedule-Revised [CIS-R], Diagnostic Interview Schedule [DIS], MINI). For

each model, we used the interview with the greatest number of participants as the reference category.

Second, because the MINI was intentionally designed to be a brief, but overly inclusive, tool,[9,10] and based on results from the first analyses, which were consistent with this, we compared fully structured diagnostic interviews, without the MINI, to semi-structured interviews. To do this, we estimated a GLMM to compare odds of major depression classification between the remaining semi- and fully structured interviews, (reference = semi-structured). As a sensitivity analysis, we further restricted our analysis to studies using either the CIDI or SCID (reference = SCID), as these interviews were used substantially more often than other included interviews.

Third, we investigated a possible interaction between interview assessment method and depressive symptom severity based on categorical PHQ-9 score classifications. To do this, we separated PHQ-9 scores into 3 categories: low (scores 0-6; reference group), medium (scores 7-15), and high (scores 16-27). Score ranges were chosen because recent meta-analyses of the PHQ-9 have evaluated cutoff scores from 7 to 15, suggesting a mid-level range.[28] To compare models with and without the interaction term, a likelihood ratio test was used. We then replicated the model comparing semi- and fully structured interviews in each PHQ-9 category separately to obtain stratum-specific classification odds ratios for fully versus semi-structured interviews. Additionally, we conducted a separate interaction analysis between continuous PHQ-9 score and diagnostic interview method. As a sensitivity analysis, we further restricted our interaction analyses to studies using the CIDI or SCID.

In another set of sensitivity analyses, we reran all of our models adding domain

scores for QUADAS-2. All analyses were run in R using the lme4 package.

Funding and ethics

The study sponsors had no role in study design; in the collection, analysis, and

interpretation of data; in the writing of the report; or in the decision to submit the paper

for publication. BDT had full access to all data in the study and had final responsibility

for the decision to submit for publication. As this study involved secondary analysis of

anonymized previously collected data, the Research Ethics Committee of the Jewish

General Hospital declared that this project did not require research ethics approval.

However, for each included dataset, we confirmed that the original study received

ethics approval and that all patients provided informed consent.

RESULTS

Search Results and Inclusion of Primary Data

Of 5,248 unique titles and abstracts identified from the database search, 5,039 were

excluded after title and abstract review and 113 after full-text review, leaving 96 eligible articles

with data from 69 unique participant samples (SupplentaryFigure1). Of the 69 unique samples,

55 contributed data (80%). In addition, authors of included studies contributed data from three

unpublished studies, for a total of 58 datasets. However, one primary dataset did not include data

for key covariates included in analyses and was excluded, leaving 57 primary datasets. In total,

17,158 participants (2,287 major depression cases) were included. Included study characteristics

are shown in SupplementaryTable2a. Characteristics of eligible studies that did not provide data

for the present study are shown in SupplementaryTable2b. Of the 21,171 participants in 69

eligible published datasets, 16,757 were in the 54 published studies with data included in the
present study (79%).

Of the 57 total included studies, 29 used semi-structured interviews, and 28 used fully
structured interviews (Table 1). The SCID was the most commonly used semi-structured
interview (26 studies, 4,732 participants), and the CIDI (11 studies, 6,271 participants) and MINI
(14 studies, 2,756 participants) were the most commonly used fully structured interviews.

Association of Diagnostic Interview and Major Depression Classification

Semi-structured Interviews

Among semi-structured interviews, compared to the SCID, odds of major depression
were not significantly different for the SCAN (adjusted odds ratio [aOR] = 0.56, 95% confidence
interval [95% CI] = 0.18 to 1.78) or DISH (aOR = 1.13, 95% CI = 0.19 to 6.80). However, only
two studies used the SCAN, and only one used the DISH.

Fully Structured Interviews

Among fully structured interviews, compared to the CIDI, odds of major depression were
higher, but not significantly different for the DIS (aOR = 4.32, 95% CI = 0.95 to 19.62) or CIS-R
(aOR = 1.53, 95% CI = 0.48 to 4.91), although these estimates were based on one and two
studies, respectively. Participants interviewed with the MINI were substantially and statistically
significantly more likely to be classified as having major depression (aOR = 2.10, 95% CI = 1.15
to 3.87).

Semi-structured versus Fully Structured Interviews

Excluding the MINI, odds of major depression were similar using fully versus semi-
structured interviews (aOR = 0.90, 95% CI = 0.51 to 1.57). In a sensitivity analysis restricted to
studies that used the SCID or CIDI, odds of major depression were lower for the CIDI compared

to the SCID, but this was not statistically significantly different (aOR = 0.57, 95% CI = 0.32 to 1.02).

Interaction between PHQ-9 Scores and Diagnostic Interview Method

The proportion of participants classified as having major depression at each PHQ-9 score for semi-structured interviews, fully structured interviews (MINI excluded), and the MINI are shown in Figure 1a, with differences in proportions across interview types shown in Figure 1b. As shown in Figure 1 and SupplementaryTable3, compared to semi-structured interviews, fully structured interviews resulted in a somewhat higher probability of major depression classification for PHQ-9 scores from 0 to 10, but lower probability for PHQ-9 scores of 11 to 27. Consistent with this, there was a significant interaction between assessment method and PHQ-9 score category (Table 2), and the likelihood ratio test comparing models with and without the interaction term was statistically significant (p < 0.001). The interaction was also statistically significant when tested using the PHQ-9 as a continuous variable. The aOR for the interaction between PHQ-9 score and fully structured interview was 0.90 (95% CI = 0.88 to 0.92), which suggested a 10% dilution in the slope of the odds of a major depression diagnosis across PHQ-9 scores for fully structured interviews compared to semi-structured interviews.

When stratified based on PHQ-9 score category, participants with low PHQ-9 scores (0-6) were more likely to receive a major depression classification with a fully structured interview compared to a semi-structured interview (aOR = 3.13, 95% CI = 0.98 to 10.00), although this was not statistically significant. Semi- and fully structured interviews performed similarly among participants in the medium PHQ-9 group (scores 7-15: aOR = 0.96, 95% CI = 0.56 to 1.66). Among participants with high PHQ-9 scores (16-27), participants were significantly less likely to be classified with major depression using fully structured interviews (aOR = 0.50, 95% CI = 0.26

to 0.97, Table 3). These odds ratios corresponded to a crude prevalence of 3.2% among those administered a fully structured interview vs. 1.2% among those administered a semi-structured interview in the low range PHQ-9 group, 21.4% vs. 20.8% in the medium range group, and 54.2% vs. 72.5% in the high range group, not adjusting for PHQ-9 scores or participant characteristics.

In sensitivity analyses restricted to studies that used the SCID or CIDI, results for interaction models were similar.

Risk of Bias Sensitivity Analyses

See SupplentaryTable4 for QUADAS-2 ratings for each included primary study. In sensitivity analyses with models that included QUADAS-2 domains, no domains were significantly associated with major depression, and the inclusion of the QUADAS-2 domains did not substantially change coefficient estimates for any variables.

DISCUSSION

There were two main findings. First, among fully structured interviews, the adjusted odds of being classified as having major depression were approximately twice as high using the MINI compared to the CIDI. Second, excluding the MINI, there was a statistically significant interaction between fully structured versus semi-structured interview and depression symptom severity based on the PHQ-9. Compared to semi-structured interviews, the likelihood of diagnosis increased significantly less for fully structured interviews as symptom severity increased. Fully structured interviews tended to classify more participants with low-level symptoms as having major depression, although this was not statistically significant; they performed similar to semi-structured interviews for participants with moderate symptoms, and

they classified fewer participants with high-level symptoms as having major depression compared to semi-structured interviews.

The finding that odds of major depression classification were twice as high for the MINI compared to the CIDI is consistent with the interviews' designs. Whereas the CIDI and other fully structured interviews are in-depth interviews,[7,8] the MINI was developed to be able to be administered in a fraction of the time as other interviews and was described by its developers as designed to be over-inclusive.[9,10] Our findings suggest that, consistent with the developers' intent, the MINI may identify substantially higher rates of major depression if used to determine case status than other fully structured interviews. The probability of being classified with major depression was also high based on the DIS and CIS-R, but evidence was too limited to draw conclusions. The formats of these interviews, however, are more similar to the CIDI than the MINI.

By standardizing all questions and probes and removing clinical judgment, fully structured interviews are designed to be as reliable as possible, but this may reduce advantages of semi-structured interviews related to the inclusion of a framework for incorporating clinical judgment. Consistent with this, our findings suggest that compared to semi-structured interviews, the association between symptom levels and probability of being classified as having major depression was lower for fully structured interviews (MINI excluded). Compared to semi-structured interviews, participants with low-level depressive symptoms assessed with fully structured interviews appeared more likely to be classified as having major depression, whereas participants with high-level symptoms appeared less likely. Participants with moderate symptoms were similarly likely to be classified as having major depression when semi- and fully structured interviews were used. This suggests that, in practice, the effect of the diagnostic

21

interview that is selected on the prevalence that is generated likely depends on the underlying distribution of symptom levels in the population.

Existing data from other studies is roughly consistent with this. In general population samples, where depressive symptom levels are generally low, major depression prevalence has been found to be substantially higher when fully structured interviews are used versus semi-structured interviews (SupplementaryTable1).[11,13] On the other hand, in a study of patients from an alcoholic treatment unit, where depressive symptoms would be expected to be much higher, major depression prevalence was similar based on semi- and fully structured interviews.[15]

In research settings, semi- and fully structured interviews are typically used interchangeably as appropriate reference standards in depression screening tool diagnostic accuracy studies, for inclusion and exclusion in treatment trials, and for determining major depression prevalence. Based on the findings of the present study, caution is warranted when deciding which interview to use. Prevalence estimates may be influenced, potentially substantially, by this choice. It is not clear to what degree estimates of screening tool accuracy may be influenced by using a fully versus semi-structured interview, and this should be determined by future studies, including a replication of this study using data from IPD meta-analyses of other depression screening tools.[29,30]

This is the first study to compare fully and semi-structured interviews for major depression using an IPD meta-analysis approach. Strengths of this study include the large overall sample size and the ability to consider both study and participant-level factors in analyses, including participant-specific depressive symptom severity. There are also limitations to consider. First, we were unable to include primary data from 15 of 69 eligible datasets (20% of eligible datasets, 21% of eligible participants), and we restricted our analyses to those with complete data for all

variables in our models (98% of available data). Nonetheless, this was a very large sample, many times the size of existing studies that have attempted to compare fully and semi-structured interviews for major depression. None of those studies included more than 61 cases based on a fully structured interview or 22 cases based on a semi-structured interview. Second, despite the large overall sample size, there was substantial heterogeneity across studies. We were not able to conduct subgroup analyses based on medical comorbidity or cultural aspects such as country or language because comorbidity data were not available for over half of participants, and many countries and languages were represented in few primary studies. However, studies of differential item functioning with the PHQ-9 have shown that it performs equivalently across multiple languages and between people with and without medical disorders.[31–35] Third, it is possible that residual confounding may exist, given that we were only able to consider variables collected in the original investigations, and the included study-level variables may not apply uniformly to all participants in a study. Fourth, although we coded for the qualifications of the interviewer for all semi-structured interviews as part of our QUADAS-2 rating, two studies used interviewers who did not meet typical standards, and approximately half of studies were rated unclear. This may have influenced the quality of the reference standard in some studies. Fifth, particularly for semi-structured interviews, lack of interviewer blinding may have influenced classifications. Although only two studies were coded as having non-blinded interviewers, 11 were coded as unclear. We did not query authors on interviewer characteristics and blinding if information was not published due to concern that recollection, in some cases, after over a decade had passed, may not have been accurate.

CONCLUSIONS

We found that the MINI diagnostic interview was associated with a substantially higher probability of major depression classification than the CIDI, controlling for depression symptom scores on the PHQ-9 and other patient characteristics. We also found that compared to semi-structured interviews, fully structured interviews tend to classify more people with low-level symptoms as depressed, but fewer people with high-level symptoms. This suggests that the choice of using a fully structured diagnostic interview or a semi-structured interviews may influence research findings. This is the first study that has used a large participant sample and IPD meta-analysis to compare diagnostic interview methods, and future research should replicate this study to verify results.

Author Contributions:

BLevis, ABenedetti, PC, SG, JPAI, LAK, DM, SBP, IS, RJS, RCZ and BDT were responsible for the study conception and design. DHA, BA, LA, HRB, MB, ABeraldi, CHB, PB, GC, MHC, JCNC, RC, NC, KC, YC, JMG, JD, JRF, FHF, BF, DF, BG, SG, FGS, CGG, BJH, JH, PAH, UH, LH, SEH, MH, TH, MI, KI, NJ, MEK, KMK, FL, SL, ML, SRL, BLöwe, LM, AM, SMS, TNM, KM, FLO, VP, BWP, PP, AP, AGR, ISS, JS, ASidebottom, ASimning, LS, SS, PLLT, AT, CMvdFC, HCvW, PAV, JW, MAH, KW, MY, YZ, and BDT were responsible for collection of primary data included in this study. BLevis, KER, NS, MA, DBR, MJC, TAS, and BDT contributed to data extraction and coding for the meta-analysis. BLevis, ABenedetti, AWL, and BDT contributed to the data analysis and interpretation. BLevis, ABenedetti, and BDT contributed to drafting the manuscript. All authors provided a critical review and approved the final manuscript. BDT is the guarantor.

Author degrees:

Brooke Levis, MSc; Andrea Benedetti, PhD; Kira E. Riehm, MSc; Nazanin Saadat, BSc; Alexander W. Levis, BSc; Marleine Azar, BSc; Danielle B. Rice, MSc; Matthew J. Chiovitti, MISt; Tatiana A. Sanchez, BSc; Pim Cuijpers, PhD; Simon Gilbody, PhD; John P.A. Ioannidis, MD; Lorie A. Kloda, PhD; Dean McMillan, PhD; Scott B. Patten, MD; Ian Shrier, MD; Russell J. Steele, PhD; Roy C. Ziegelstein, MD; Dickens H. Akena, PhD; Bruce Arroll, MBChB; Liat Ayalon, PhD; Hamid R. Baradaran, MD; Murray Baron, MD; Anna Beraldi, PhD; Charles H. Bombardier, PhD; Peter Butterworth, PhD; Gregory Carter, FRANZCP; Marcos H. Chagas, MD; Juliana C. N. Chan, MD; Rushina Cholera, MD; Neerja Chowdhary, MD; Kerrie Clover, PhD; Yeates Conwell, MD; Janneke M. de Man-van Ginkel, PhD; Jaime Delgadillo, PhD; Jesse R. Fann, MD; Felix H. Fischer, PhD; Benjamin Fischler, MD; Daniel Fung, MD; Bizu Gelaye, PhD; Felicity Goodyear-Smith, MD; Catherine G. Greeno, PhD; Brian J. Hall, PhD; John Hambridge, Dip Clin Psych; Patricia A. Harrison, PhD; Ulrich Hegerl, MD; Leanne Hides, PhD(Clin); Stevan E. Hobfoll, PhD; Marie Hudson, MD; Thomas Hyphantis, MD; Masatoshi Inagaki, MD; Khalida Ismail, MD; Nathalie Jetté, MD; Mohammad E. Khamseh, MD; Kim M. Kiely, PhD; Femke Lamers, PhD; Shen-Ing Liu, MD; Manote Lotrakul, MD; Sonia R. Loureiro, PhD; Bernd Löwe, MD; Laura Marsh, MD; Anthony McGuire, PhD; Sherina Mohd Sidik, PhD; Tiago N. Munhoz, PhD; Kumiko Muramatsu, MD; Flávia L. Osório, PhD; Vikram Patel, MD; Brian W. Pence, PhD; Philippe Persoons, MD; Angelo Picardi, MD; Alasdair G. Rooney, MD; Iná S. Santos, MD; Juwita Shaaban, MMed (Fam. Med); Abbey Sidebottom, PhD; Adam Simning, MD; Lesley Stafford, PhD; Sharon Sung, PhD; Pei Lin Lynnette Tan, MMed (Psychiatry); Alyna Turner, PhD; Christina M. van der Feltz-Cornelis, MD; Henk C. van Weert,

MD; Paul A. Vöhringer, MD; Jennifer White, PhD; Mary A. Whooley, MD; Kirsty Winkley, PhD; Mitsuhiko Yamada, MD; Yuying Zhang, PhD; Brett D. Thombs, PhD.

Author Affiliations:

Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada (Levis B, Riehm, Saadat, Levis A, Azar, Rice, Chiovitti, Sanchez, Shrier, Steele, Baron, Hudson, Thombs); Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada (Levis B, Benedetti, Levis A, Azar, Shrier, Thombs); Department of Medicine, McGill University, Montréal, Québec, Canada (Benedetti, Baron, Hudson, Thombs); Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada (Benedetti); Department of Psychology, McGill University, Montréal, Québec, Canada (Rice, Thombs); Department of Clinical, Neuro and Developmental Psychology, EMGO Institute, VU University, Amsterdam, the Netherlands (Cuijpers); Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK (Gilbody, McMillan); Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA (Ioannidis); Library, Concordia University, Montréal, Québec, Canada (Kloda); Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada (Patten, Jetté); Hotchkiss Brain Institute and O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada (Patten, Jetté); Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada (Steele); Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (Ziegelstein); Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda (Akena); Department of General Practice and Primary Health Care, University

of Auckland, New Zealand (Arroll, Goodyear-Smith); Louis and Gabi Weisfeld School of Social Work, Bar Ilan University, Ramat Gan, Israel (Ayalon); Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran (Baradaran, Khamseh); Kbo-Lech-Mangfall-Klinik Garmisch-Partenkirchen, Klinik für Psychiatrie, Psychotherapie & Psychosomatik, Lehrkrankenhaus der Technischen Universität München, Munich, Germany (Beraldi); Department of Rehabilitation Medicine, University of Washington, Seattle, Washington, USA (Bombardier); Centre for Research on Ageing, Health and Wellbeing, Research School of Population Health, The Australian National University, Canberra, Australia (Butterworth, Kiely); Centre for Mental Health, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia (Butterworth); Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Melbourne, Australia (Butterworth); Centre for Translational Neuroscience and Mental Health, University of Newcastle, New South Wales, Australia (Carter, Clover); Department of Neurosciences and Behavior, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil (Chagas, Osório, Loureiro); Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China (Chan, Zhang); Asia Diabetes Foundation, Prince of Wales Hospital, Hong Kong Special Administrative Region, China (Chan, Zhang); Hong Kong Institute of Diabetes and Obesity, Hong Kong Special Administrative Region, China (Chan); Department of Pediatrics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA (Cholera); The World Health Organization, Geneva, Switzerland (Chowdhary); Psycho-Oncology Service, Calvary Mater Newcastle, New South Wales, Australia (Clover); Department of Psychiatry, University of Rochester Medical Center, New York, USA (Conwell, Simning);

Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands (de Man-van Ginkel); Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK (Delgadillo); Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA (Fann); Institute for Social Medicine, Epidemiology, and Health Economics, Charité - Universitätsmedizin Berlin, Germany (Fischer); Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Germany (Fischer); Private Practice, Brussels, Belgium (Fischler); Department of Child & Adolescent Psychiatry, Institute of Mental Health, Singapore (Fung, Sung); Yong Loo Lin School of Medicine, National University of Singapore, Singapore (Fung); Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore (Fung); Office of Clinical Sciences, Duke-NUS Medical School, Singapore (Fung, Liu, Sung); Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA (Gelaye); School of Social Work, University of Pittsburgh, Pittsburgh, Pennsylvania, USA (Greeno); Global and Community Mental Health Research Group, Department of Psychology, Faculty of Social Sciences, University of Macau, Macau Special Administrative Region, China (Hall); Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA (Hall); Liaison Psychiatry Department, John Hunter Hospital, Newcastle, Australia (Hambridge); City of Minneapolis Health Department, Minneapolis, Minnesota, USA (Harrison); Department of Psychiatry and Psychotherapy, University Hospital Leipzig, Leipzig, Germany (Hegerl); Centre for Children's Health Research, Institute of Health & Biomedical Innovation, School of Psychology, University of Queensland, Brisbane, Queensland, Australia (Hides); Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois, USA (Hobfoll);

Department of Psychiatry, University of Ioannina, Ioannina, Greece (Hyphantis); Department of Neuropsychiatry, Okayama University Hospital, Okayama, Japan (Inagaki); Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neurosciences, King's College London Weston Education Centre, London, UK (Ismail, Winkley); Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada (Jetté); Department of Psychological Medicine, Tan Tock Seng Hospital, Singapore (Tan); Department of Psychiatry, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, the Netherlands (Lamers); Department of Psychiatry, Mackay Memorial Hospital, Taipei, Taiwan (Liu); Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan (Liu); Department of Medicine, Mackay Medical College, Taipei, Taiwan (Liu); Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand (Lotrakul); Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (Löwe); Schön Klinik Hamburg Eilbek, Hamburg, Germany (Löwe); Baylor College of Medicine, Houston and Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas, USA (Marsh); Department of Nursing, St. Joseph's College, Standish, Maine, USA (McGuire); Cancer Resource & Education Centre, and Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia (Mohd Sidik); Post-graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS, Brazil (Munhoz, Santos); Department of Clinical Psychology, Graduate School of Niigata Seiryo University, Niigata, Japan (Muramatsu); National Institute of Science and Technology, Translational Medicine, Ribeirão Preto, Brazil (Osório); Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts, USA (Patel); London School of Hygiene & Tropical Medicine, London, UK (Patel); Centre for

Chronic Conditions and Injuries, Public Health Foundation of India, New Delhi, India (Patel); Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (Pence); Department of Adult Psychiatry, University Hospitals Leuven, Leuven, Belgium (Persoons); Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium (Persoons); Centre for Behavioural Sciences and Mental Health, Italian National Institute of Health, Rome, Italy (Picardi); Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburg, Edinburgh, Scotland, UK (Rooney); Department of Family Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia (Shaaban); Allina Health, Minneapolis, Minnesota, USA (Sidebottom); Centre for Women's Mental Health, Royal Women's Hospital, Parkville, Australia (Stafford); Melbourne School of Psychological Sciences, University of Melbourne, Australia (Stafford); School of Medicine and Public Health, University of Newcastle, New South Wales, Newcastle, Australia (Turner); IMPACT Strategic Research Centre, School of Medicine, Deakin University, Geelong, Victoria, Australia (Turner); Clinical Center of Excellence for Body, Mind and Health, GGz Breburg, Tilburg, the Netherlands (van der Feltz-Cornelis); Tilburg University, Faculty of Social Sciences, Tranzo Department, Tilburg, the Netherlands (van der Feltz-Cornelis); Department of General Practice, Academic Medical Centre Amsterdam, University of Amsterdam, Amsterdam, the Netherlands (van Weert); Department of Psychiatry and Mental Health, Clinical Hospital, Universidad de Chile, Santiago, Chile (Vöhringer); Millennium Institute for Depression and Personality Research (MIDAP), Ministry of Economy, Macul, Santiago, Chile (Vöhringer); Mood Disorders Program, Tufts Medical Center, Tufts University, Boston, USA (Vöhringer); Monash University, Melbourne, Australia (White); Department of Epidemiology and Biostatistics, University of California San

Francisco, San Francisco, California, USA (Whooley); Department of Medicine, Veterans Affairs Medical Center, San Francisco, California, USA (Whooley); Department of Medicine, University of California San Francisco, San Francisco, California, USA (Whooley); Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawa-Higashi, Kodaira, Tokyo, Japan (Yamada); Department of Psychiatry, McGill University, Montréal, Québec, Canada (Thombs); Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada (Thombs).

Full Addresses:

Brooke Levis

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

Andrea Benedetti

Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, 5252 boul de Maisonneuve, Office/Workstation # 3D.59, Montréal, QC, H4A 3S5, Canada

Kira E. Riehm

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

Nazanin Saadat

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

Alexander W. Levis

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

Marleine Azar

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

Danielle B. Rice

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

Matthew J. Chiovitti

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

Tatiana A. Sanchez

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

Pim Cuijpers

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

Simon Gilbody

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

John P. A. Ioannidis

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

Lorie Kloda

Concordia University, 1455, boul. de Maisonneuve Ouest, FB-802, Montréal, QC,

H3G 1M8, Canada

Dean McMillan

Mental Health and Addiction Research Group, Department of Health Sciences and

Hull York Medical School, University of York, Heslington YO10 5DD, United

Kingdom

Scott B. Patten

Department of Community Health Sciences, 3rd Floor, TRW Building, University

of Calgary, 3280 Hospital Drive NW, Calgary, AB, T2N 4Z6, Canada

Ian Shrier

Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research,

Jewish General Hospital, 3755 Cote Ste-Catherine Rd, Montréal, QC, H3T 1E2,

Canada

Russell J. Steele

Department of Mathematics and Statistics, McGill University, 805 Rue Sherbrooke

O., Montreal, QC, H3A 0B9, Canada

Roy C. Ziegelstein

Johns Hopkins University School of Medicine, Miller Research Building, 733 N.

Broadway, Suite 115, Baltimore, MD, 21205, USA

Dickens H. Akena

Department of Psychiatry, Makerere University College of Health Sciences,

P.O.Box 7062 Kampala, Uganda

Bruce Arroll

Department of General Practice and Primary Health Care, University of Auckland,

Private Bag 92019, Auckland 1142, New Zealand

Liat Ayalon

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

52900, Israel

Hamid R. Baradaran

Endocrinology and Metabolism Research Institute, Shariati Hospital Complex,

North Kargar Avenue, Tehran, 14114-13137, Iran

Murray Baron

Jewish General Hospital, Suite A 725, 3755 Cote St Catherine Rd, Montréal, QC,

H3T 1E2, Canada

Anna Beraldi

Auenstraße 6, D-82467 Garmisch-Partenkirchen, Germany

Charles H. Bombardier

Division of Clinical and Neuropsychology, Department of Rehabilitation Medicine,

University of Washington, Box 359612, Harborview Medical Center, 325 9th

Avenue, Seattle, WA, 98104, USA

Peter Butterworth

Centre for Mental Health, Level 4, 207 Bouverie St, The University of Melbourne,

Victoria 3010, Australia

Gregory Carter

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

Marcos H. Chagas

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

Brazil

Juliana C. N. Chan

Department of Medicine and Therapeutics, The Chinese University of Hong Kong,

9/F Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital, Shatin,

Hong Kong

Rushina Cholera

UNC School of Medicine, Department of Pediatrics, CB# 7593, Chapel Hill, NC,

27599-7593, USA

Neerja Chowdhary

World Health Organization. Avenue Appia 20, 1211 Genève 27, Switzerland

Kerrie Clover

Psycho-oncology #51, Calvary Mater Newcastle, Platt St, Waratah NSW 2301,

Australia

Yeates Conwell

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

14642, USA

Janneke M. de Man-van Ginkel

University Medical Center Utrecht, Internal mail no Str. 6.131, P.O. Box 85500,

3508 GA, Utrecht, The Netherlands

Jaime Delgadillo

Clinical Psychology Unit, University of Sheffield, Cathedral Court, Floor F, 1 Vicar

Lane, Sheffield, S1 1HD, United Kingdom

Jesse R. Fann

Department of Psychiatry & Behavioral Sciences, Univerisity of Washington, Box

356560, Seattle, WA 98195

Felix H. Fischer

Medizinische Klinik mit Schwerpunkt Psychosomatik, Charité -

Universitätsmedizin Berlin, Charitéplatz 1, 10098 Berlin, Germany

Benjamin Fischler

Private practice Brussels rue du Pepin 4, Belgium

Daniel Fung

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

Bizu Gelaye

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

02115, USA

Felicity Goodyear-Smith

Department of General Practice & Primary Health Care, University of Auckland,

PB 92019, Auckland, 1142, New Zealand

Catherine G. Greeno

2204 Cathedral of Learning, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh,

PA, 15260, USA

Brian J. Hall

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

John Hambridge

University of Newcastle, NSW 2310, Newcastle, Australia

Patricia A. Harrison

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

Ulrich Hegerl

University of Leipzig, Department of Psychiatry and Psychotherapy, Semmelweisstrasse 10, 04103 Leipzig, Germany

Leanne Hides

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

Stevan E. Hobfoll

1645 W. Jackson Blvd, Suite 400, Dept of Behavioral Sciences, Rush University Medical Center, Chicago, IL, 60614, USA

Marie Hudson

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

Thomas Hyphantis

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

Masatoshi Inagaki

Department of Neuropsychiatry, Okayama University Hospital,

2-5-1, Shikata-cho, Kita-ku, Okayama, 700-8558, Japan

Khalida Isamail

Institute of Psychiatry, Psychology and Neuroscience, King's College London, 10

Cutcombe Road, London, SE5 9AF, United Kingdom

Nathalie Jetté

Department of Clinical Neurosciences, University of Calgary, 1403 29th Street NW,

T2N 2T9, Canada

Mohammad Khamseh

Endocrinology and Metabolism Research Institute, Shariati Hospital Complex,

North Kargar Avenue, Tehran, 14114-13137, Iran

Kim E. Kiely

Centre for Research on Ageing, Health and Wellbeing, The Australian National

University, Building 54 Mills Road, ACT 2601, Australia

Femke Lamers

VU University Medical Center, Department Psychiatry, A.J. Ernststraat 1187, room

D2.14, 1081 HL Amsterdam, The Netherlands

Shen-Ing Liu

Department of Psychiatry, Mackay Memorial Hospital, No. 92, Section 2, Chung-

Shan North Rd, Taipei, Taiwan

Manote Lotrakul

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

Sonia R. Loureiro

Rua Tenente Catão Roxo, 2650, CEP 14051-140, Ribeirão Preto, SP, Brazil

Bernd Löwe

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

Laura Marsh

Mental Health Care Line, Michael E DeBakey VA Medical Center, Departments of Psychiatry and Neurology, Baylor College of Medicine, 2002 Holcombe Blvd, Houston, TX, 77030, USA

Anthony McGuire

278 Whites Bridge Rd., Standish, ME, 04084, USA

Sherina Mohd Sidik

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

Tiago N. Munhoz

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

Kumiko Muramatsu

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

Flávia L. Osório

Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - USP. Avenida dos Bandeirantes 3900- 3 andar- alaC. Ribeirão Preto - São Paulo - Brasil - CEP 14049-900

Vikram Patel

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

Brian W. Pence

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

Philippe Persoons

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

Angelo Picardi

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

Alasdair G. Rooney

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

Iná S. Santos

Depto Medicina Social, Programa Pós-graduação Epidemiologia, Universidade

Federal de Pelotas, Rua Marechal Deodoro 1160, 3º piso

96020-220 - Pelotas, RS, Brasil

Juwita Shaaban

School of Medical Science, Health Campus Universiti Sains Malaysia, 16150

Kubang Kerian, Kelantan, Malaysia

Abbey Sidebottom

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

Adam Simning

Strong Behavioral Health, 300 Crittenden Blvd, Rochester, NY, 14642, USA

Lesley Stafford

Centre for Women's Mental Health, The Royal Women's Hospital, Locked Bag

300, Parkville Victoria 3052, Australia

Sharon Sung

Office of Clinical Sciences, Duke-NUS Medical School, 20 College Road, Level 6,

169856, Singapore

Pei Lin Lynnette Tan

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

Alyna Turner

School of Medicine and Public Health, Faculty of Health, University of Newcastle,

University Drive, Callaghan, NSW, 2308, Australia

Christina M. van der Feltz-Cornelis

Tranzo Academic Centre 'Geestdrift', Tilburg University, PO Box 90153, 5000 LE
Tilburg, The Netherlands

Henk C. van Weert

Dpt. General Practice, Academic Medical Centre Amsterdam, Meibergdree 9, 1105
AZ Amsterdam, The Netherlands

Paul A. Vöhringer

Department of Psychiatry, 800 Washington St, Boston, MA 02111, USA

Jennifer White

School of Primary and Allied Health Care, Faculty of Medicine, Nursing & Health
Sciences, Monash University, Kingston Centre, 400 Warrigal Rd, Cheltenham
Victoria 3192, Australia

Mary A. Whooley

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

Kirsty Winkley

King's College London & Institute of Psychiatry, Psychology & Neuroscience,
Weston Education Centre, London SE5 9RS, UK

Mitsuhiko Yamada

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

Yuying Zhang

Department of Medicine and Therapeutics, The Chinese University of Hong Kong,

9/F Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital, Shatin,

Hong Kong

Brett D. Thombs

Room 302, Institute of Community and Family Psychiatry, Jewish General

Hospital, 4333 Cote Ste Catherine Road, Montréal, QC, H3T 1E4, Canada

REFERENCES

1. Jones KD. The Unstructured Clinical Interview. J Couns Dev. 2010;**88**:220–226.

2. 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.

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

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

5. World Health Organization. Schedules for clinical assessment in neuropsychiatry: manual. Amer Psychiatric Pub Inc. 1994.

6. Kurdyak PA, Gnam WH. Small signal, big noise: performance of the CIDI depression module. Can J Psychiatry. 2005;**50**:851–856.

7. 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.

8. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. Arch Gen Psychiatry. 1981;**38**:381–389.

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. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Psychol Medicine. 2001;**31**:1001–1013.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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;**27**:3:124.

18. PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). Ottawa: CADTH; 2016 Jan.

19. Sampson M, Barrowman NJ, Moher D, et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol. 2003;**56**:943–955.

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

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

22. United Nations. International Human Development Indicators. Available: http://hdr.undp.org/en/countries (accessed 2017 April 26).

23. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;**155**:529–536.

24. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med. 2006;**21**:547–552.

25. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry. 2006;**28**:71–77.

26. Adewuya AO, Ola BA, Afolabi OO. Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. J Affect Disord. 2006;**96**:89–93.

27. Milette K, Hudson M, Baron M, Thombs BD. Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates. Rheumatology. 2010;**49**:789–796.

28. 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.

29. 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.

30. 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.

31. Arthurs E, Steele RJ, Hudson M, Baron M, Thombs BD; Canadian Scleroderma Research Group. Are scores on English and French versions of the PHQ-9 comparable? An assessment of differential item functioning. PLoS One. 2012;**7**:e52028.

32. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med. 2006;**21**:547–552.

33. Chung H, Kim J, Askew RL, Jones SMW, Cook KF, Amtmann D. Assessing measurement invariance of three depression scales between neurologic samples and community samples. Qual Life Res. 2015;**24**;1829–1834.

34. Cook KF, Kallen MA, Bombardier C, et al. Do measures of depressive symptoms function differently in people with spinal cord injury versus primary care patients: the CES-D, PHQ-9, and PROMIS-D. Qual Lif Res. 2017;**26**:139–148.

35. Leavens A, Patten SB, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research Group. Influence of somatic symptoms on Patient Health Questionnaire-9 depression scores among patients with systemic sclerosis compared to a healthy general population sample. Arthritis Care Res. 2012;**64**:1195–1201.

FIGURES

Figure 1a. Probability of Major Depression Classification by PHQ-9 Score for Semi-

structured Interviews, Fully structured Interviews (Excluding MINI), and MINI.
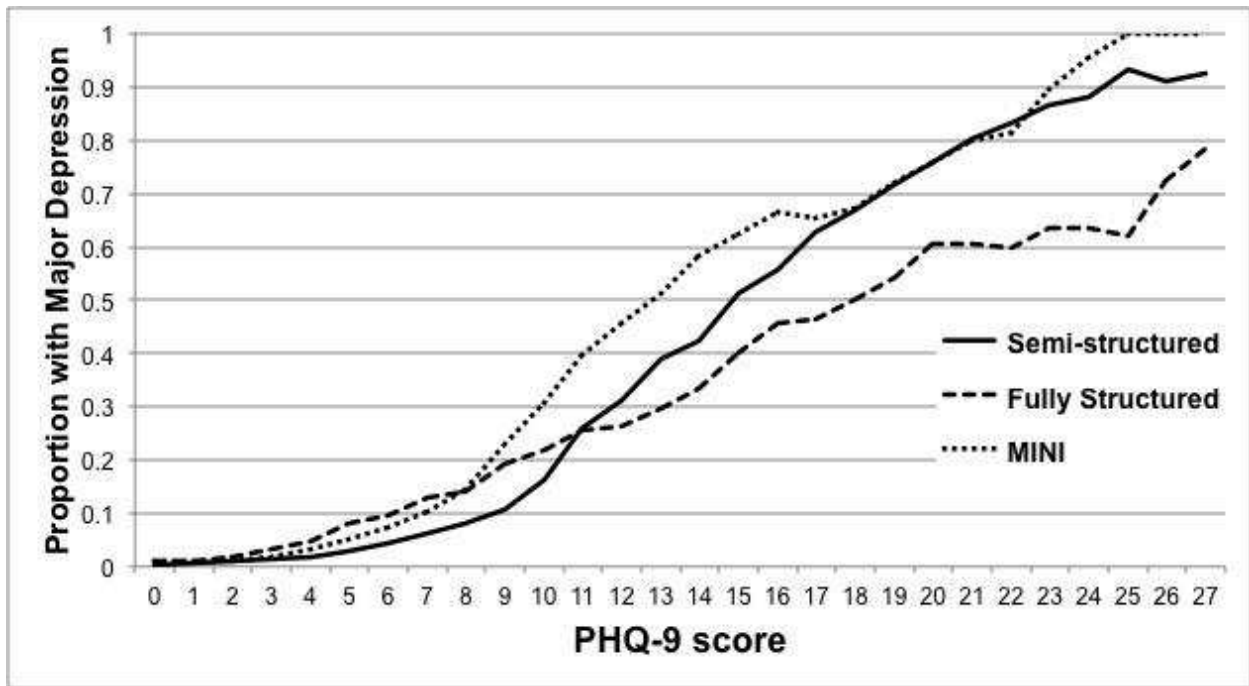
Proportions are plotted as 3-point moving averages (e.g., the proportions at the PHQ-9 score of 10 are

averages of the proportions at PHQ-9 scores of 9, 10, and 11).

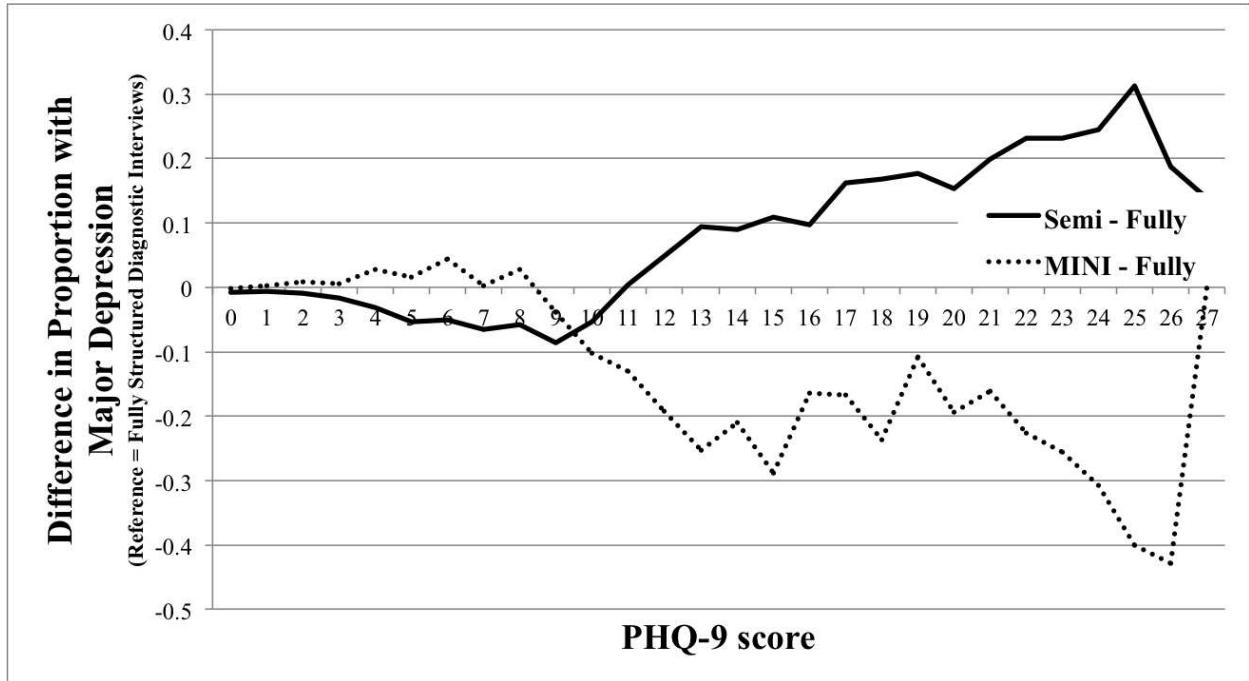**Figure 1b**. Difference in Probability of Major Depression Classification by PHQ-9

Score for Semi-structured Interviews and MINI compared to Fully structured Interviews

(Excluding MINI).

Differences in proportions are plotted as 3-point moving averages (e.g., the differences in proportions at

the PHQ-9 score of 10 are averages of the differences in proportions at PHQ-9 scores of 9, 10, and 11).

Table 1. Participant data by diagnostic interview

Diagnostic Interview	N Studies	N Participants	Major Depression	
			N	%
Semi-structured				
SCID	26	4,732	785	17
SCAN	2	1,891	130	7
DISH	1	100	9	9
Fully structured				
CIDI	11	6,271	554	9
DIS	1	1,006	221	22
MINI	14	2,756	524	19
CIS-R	2	402	64	16
Total	57	17,158	2,287	13

Abbreviations: CIDI: Composite International Diagnostic Interview, CIS-R: Clinical Interview Schedule-Revised, DIS: Diagnostic Interview Schedule, DISH: Depression Interview and Structured Hamilton, MINI: Mini International Neuropsychiatric Interview, SCAN: Schedules for Clinical Assessment in Neuropsychiatry, SCID: Structured Clinical Interview for DSM Disorders

Table 2. Model summary of fixed effects generalized linear mixed model

considering a potential interaction between PHQ-9 score category and

assessment method[a,b]

Variable	Odds ratio (OR)	
	OR	95% CI
Fully structured assessment method	1.49	0.82-2.72
PHQ-9 total score	1.37	1.35-1.40
Age (years)	1.00	0.99-1.00
Male sex	0.89	0.77-1.03
Clinical setting	--	--
Non-medical (reference)	--	--
Primary care	0.67	0.27-1.64
Specialty care: Inpatient	0.33	0.13-0.85
Specialty care: Outpatient	0.64	0.26-1.54
Human development index	--	--
Very high (reference)	--	--
High	2.27	1.11-4.61
Low to medium	0.78	0.27-2.24
PHQ-9 score category * fully structured assessment method[c]	--	--
Low PHQ-9 (0-6) (reference)	--	--
Medium PHQ-9 (7-15)	0.73	0.57-0.92
High PHQ-9 (16-27)	0.26	0.18-0.37

[a]Excluding the MINI.

[b]Estimate of random intercept variance = 0.58.

[c]$p < 0.001$ in likelihood ratio test comparing models with and without interaction term.

Table 3. Generalized linear mixed model summaries for each PHQ-9 score category

PHQ-9 score category	Low PHQ scores (0-6) N = 9,339		Medium PHQ scores (7-15) N = 3,970		High PHQ scores (16-27) N = 1,093	
	OR	95% CI	OR	95% CI	OR	95% CI
OR[a] (95% CI) for fully structured assessment method	3.13	0.98-10·00	0.96	0.56-1.66	0.50	0.26-0.97
N receiving fully structured interview	5,228		1,999		452	
	N	%	N	%	N	%
N (%) with major depression	167	3.2	427	21.4	245	54.2
N receiving semi-structured interview	4,111		1,971		641	
	N	%	N	%	N	%
N (%) with major depression	50	1.2	409	20.8	465	72.5

[a]Excluding the MINI and adjusted for PHQ-9 score, age, sex, clinical setting and human development index.