

This is a repository copy of *Depression prevalence of the Geriatric Depression Scale-15 was compared to Structured Clinical Interview for DSM using individual participant data meta-analysis*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/215909/>

Version: Published Version

Article:

(2024) Depression prevalence of the Geriatric Depression Scale-15 was compared to Structured Clinical Interview for DSM using individual participant data meta-analysis. Scientific reports. 17430. ISSN 2045-2322

<https://doi.org/10.1038/s41598-024-68496-3>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



OPEN

Depression prevalence of the Geriatric Depression Scale-15 was compared to Structured Clinical Interview for DSM using individual participant data meta-analysis

Marc Parsons^{1,2}, Lu Qiu³, Brooke Levis³, Suiqiong Fan³, Ying Sun³, Lara S. N. Amiri³, Daphna Harel⁴, Sarah Markham⁵, Simone N. Vigod⁶, Roy C. Ziegelstein⁷, Yin Wu^{3,8}, Jill T. Boruff⁹, Pim Cuijpers¹⁰, Simon Gilbody¹¹, Scott B. Patten^{12,13,14}, Andrea Benedetti^{1,2,15}, Brett D. Thombs^{1,3,8,15,16,17}✉ & the DEPRESSION Screening Data (DEPRESSD) GDS Group*

Depression questionnaire cutoffs are calibrated for screening accuracy and not to assess prevalence, but the Geriatric Depression Scale (GDS-15) is often used to estimate diagnostic prevalence among older adults, most commonly with scores of ≥ 5 . We conducted an individual participant data meta-analysis to compare depression prevalence based on GDS-15 ≥ 5 to Structured Clinical Interview for Diagnostic and Statistical Manual (SCID) diagnoses and assessed whether an alternative cutoff could be more accurate. We used generalized linear mixed models to estimate prevalence. Data from 14 studies (3602 participants, 434 SCID major depression) were included. Pooled GDS-15 ≥ 5 prevalence was 34.2% (95% confidence interval [CI] 27.5–41.6%), and pooled SCID prevalence was 14.8% (95% CI 10.0–21.5%; difference of 17.6%, 95% CI 11.6–23.6%). GDS-15 ≥ 8 provided the closest estimate to SCID with mean difference of -0.3% (95% prediction interval -17.0 – 16.5%). Prevalence estimate differences were not associated with study or participant characteristics. In sum, GDS-15 ≥ 5 substantially overestimated depression prevalence. A cutoff of ≥ 8 was accurate overall, but heterogeneity was too high for implementation in practice. Validated diagnostic interviews should be used to estimate major depression prevalence among older adults.

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada. ²Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, QC, Canada. ³Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Cote Ste-Catherine, Pavillon H4.83, Montreal, QC H3T 1E2, Canada. ⁴Department of Applied Statistics, Social Science, and Humanities, New York University, New York, NY, USA. ⁵Department of Biostatistics and Health Informatics, King's College London, London, UK. ⁶Women's College Hospital and Research Institute, University of Toronto, Toronto, ON, Canada. ⁷Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁸Department of Psychiatry, McGill University, Montreal, QC, Canada. ⁹Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, QC, Canada. ¹⁰Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ¹¹Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK. ¹²Departments of Community Health Sciences and Psychiatry, University of Calgary, Calgary, AB, Canada. ¹³Mathison Centre for Mental Health Research & Education, University of Calgary, Calgary, AB, Canada. ¹⁴Cuthbertson & Fischer Chair in Pediatric Mental Health, University of Calgary, Calgary, AB, Canada. ¹⁵Department of Medicine, McGill University, Montreal, QC, Canada. ¹⁶Department of Psychology, McGill University, Montreal, QC, Canada. ¹⁷Biomedical Ethics Unit, McGill University, Montreal, QC, Canada. *A list of authors and their affiliations appears at the end of the paper. ✉email: brett.thombs@mcgill.ca

Objectives

Accurate estimates of depression prevalence are important to understand disease burden, interpret research on etiology, and use healthcare resources efficiently¹. In mental health research, validated diagnostic interviews are required to determine whether a disorder is present, including major depression². These interviews, however, are regarded as costly to administer, particularly in large studies, due to the time and trained personnel required. Because of this, self-reported screening tools are often used in research instead of diagnostic interviews, and the percentage of study participants who score above a cutoff threshold is sometimes described as the prevalence of depression^{3,4}. Score cutoffs, however, are typically set for screening and, thus, to cast a wide net to identify people who may have depression. These cutoffs are not intended to be used to classify diagnostic status, and they typically overestimate prevalence of disorders, sometimes substantially^{3–7}.

In older adults, accurately diagnosing depression can be challenging because of the high prevalence of medical comorbidities whose symptoms overlap with those of a depressive disorder. Moreover, as individuals age, polypharmacy and medication adverse reactions become more common and may overlap with symptoms of depression. The Geriatric Depression Scale (GDS) is used as a screening tool to identify possible depression in older adults, including those with a medical illness or with mild to moderate cognitive impairment⁸. To avoid overlap with symptoms of physical illness, the GDS does not include somatic symptoms, such as insomnia, loss of appetite, or fatigue; this may, though, remove key features of depressive disorders that could be evaluated in a diagnostic interview to assess origin. The GDS included 30-item and 15-item versions, as well as briefer 10-, 5- and 4-item versions^{9–11}. For the GDS-30, cutoffs of ≥ 11 for at least mild depression and ≥ 21 for moderate to severe depression have been recommended^{12,13}. For the GDS-15, cutoffs of ≥ 5 for at least mild depression and ≥ 10 for moderate to severe depression are used^{12,13} for the purpose of screening.

The GDS is often used to estimate prevalence. We reviewed recent studies indexed in PubMed (2020–2022) and identified 35 studies that reported one or more estimates of “prevalence” of depression based on a version of the GDS. There were 27 estimates from 25 studies based on the GDS-15, 12 estimates from 10 studies based on the GDS-30, and no estimates based on other GDS versions. For the GDS-15, reported cutoffs used to estimate prevalence ranged from 5 to 12 with the most common being ≥ 5 (11 studies; 44%) and ≥ 6 (6 studies; 24%). See eMethods1 and eTable1.

No studies have compared prevalence estimates based on GDS score cutoffs and validated diagnostic interviews. Three previous studies have used an individual participant data meta-analysis (IPDMA) approach to compare prevalence based on other depression screening tool cutoffs with prevalence based on a validated diagnostic interview. One found that prevalence based on the Patient Health Questionnaire-9 (PHQ-9) and its standard cutoff of ≥ 10 exceeded prevalence based on the Structured Clinical Interview for the DSM (SCID) by a mean of 11.9% among 9,242 participants from 44 primary studies⁵. Among perinatal women (29 studies, 7,315 participants), cutoffs of ≥ 10 (22.2%) and ≥ 13 (11.5%) on the Edinburgh Postnatal Depression Scale (EPDS) overestimated prevalence compared to the SCID (9.0%)⁶. Pooled prevalence using the Hospital Anxiety and Depression Scale – depression subscale (HADS-D; 41 studies, 6,005 participants), was 24.5% for HADS-D ≥ 8 and 10.7% for HADS-D ≥ 11 compared to 11.6% for SCID major depression⁷. In all three IPDMAs, alternative cutoffs were evaluated to attempt to match SCID-based prevalence. However, prediction intervals, which were used to estimate the expected difference between cutoff-based prevalence and SCID-based prevalence in a new study, were -13.6 to 14.5% for PHQ ≥ 14 ⁵, -13.7 to 12.3% for EPDS ≥ 14 ⁶, and -21.1 to 19.5% for the HADS-D⁷, all too imprecise to be useful in the interpretation of individual study results.

The objectives of the present study were to use an IPDMA approach to: (1) compare GDS-15 ≥ 5 prevalence to major depression prevalence based on the most commonly used semi-structured diagnostic interview, the SCID;¹⁴ and (2) use a prevalence matching approach^{3,15} to determine whether any cutoff threshold on the GDS-15 matches SCID major depression prevalence closely and with sufficiently low heterogeneity to be able to estimate major depression prevalence in individual studies. We did not attempt to estimate prevalence with the GDS-30 due to the small number of studies with required data in our dataset.

Methods

This study used a subset of data collected for an IPDMA of the diagnostic accuracy of the GDS for screening to detect major depression in older adults. Detailed methods of this IPDMA were registered in PROSPERO (CRD42018104329), and a protocol was published¹⁶. The PROSPERO registration was updated to include the analyses described in the present protocol as an addendum.

Study selection criteria

In the main IPDMA, datasets from studies in any language were eligible for inclusion if (1) they included GDS scores for any version of the GDS; (2) they included diagnostic classifications for current Major Depressive Episode (MDE) or Major Depressive Disorder (MDD) based on Diagnostic and Statistical Manual or International Classification of Diseases criteria, using a validated semi-structured or fully structured interview; (3) the GDS and diagnostic interview were administered within two weeks of each other, since diagnostic criteria for major depression are for symptoms experienced in the last two weeks; and (4) participants were not recruited from psychiatric settings or because they were identified as having symptoms of depression, since screening is done to identify unrecognized cases. Datasets where not all participants were eligible were included if primary data allowed for the selection of eligible participants.

For the present analysis, we only included studies with GDS-15 scores or individual item scores that we could use to calculate GDS-15 total scores. We also only included primary studies that classified major depression using the SCID. The SCID is a semi-structured diagnostic interview designed to be conducted by an experienced clinician; it requires professional judgment and allows rephrasing questions and probes to follow up responses.

We did not include other semi-structured interviews because they are not commonly used in screening accuracy studies. We did not include studies that evaluated major depression with fully structured interviews. Fully structured interviews are entirely scripted so that they can be administered by a trained lay interviewer. They are intended to increase standardization, but this may come at the cost of reduced validity^{17–20}, and several studies have shown that they perform differently than clinician-administered semi-structured interviews^{21–24}. Interrater reliability for diagnosing major depression by independent raters with the SCID has been shown to be moderate to high (κ :0.66 to 0.93)^{25,26}.

Internationally, major guidelines on depression screening from the UK National Institute for Health and Care Excellence,^{27,28} the United States Preventive Services Task Force²⁹, and the Canadian Task Force on Preventive Health Care³⁰ all address screening for major depression, but do not consider screening for less severe mood indications (e.g., dysthymic disorder or minor depression), for which treatment options and efficacy are much less well delineated. Thus, we similarly compare prevalence estimates based on GDS-15 scores to major depression prevalence.

Data sources, search strategy and study selection

A health sciences librarian searched Medline, Medline In-Process & Other Non-Indexed Citations (Ovid), PsycINFO (Ovid), and Web of Science from inception to July 16, 2018, using a peer-reviewed search strategy³¹ (see supplementary material eMethods 2). We also reviewed reference lists of relevant reviews and queried contributing authors about non-published studies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA). After de-duplication, unique citations were uploaded into DistillerSR (Evidence Partners, Ottawa, Canada) for screening.

Two investigators independently reviewed studies by title and abstract for eligibility. If either investigator deemed a study to be potentially eligible, a full-text review was done by both investigators independently. Any disagreements were resolved by consensus and consulting a third investigator when necessary. For languages other than those in which team members were fluent, translators were consulted.

Data contribution and synthesis

Authors of eligible datasets were invited to contribute de-identified primary data, including GDS scores and major depression classification statuses. We emailed corresponding authors of eligible primary studies at least three times, as necessary, with at least two weeks between each email. If we did not receive a response, we emailed co-authors and attempted to contact corresponding authors by phone.

Before integrating individual datasets into our synthesized dataset, we compared published participant characteristics and diagnostic accuracy results with results from raw datasets and resolved any discrepancies in consultation with the original investigators.

Data analysis

Comparison of GDS-15 ≥ 5 prevalence and SCID major depression prevalence

First, for each primary study, we estimated four values: (1) the percentage of participants who scored ≥ 5 on the GDS-15; (2) the percentage of participants classified as having major depression based on the SCID; (3) the difference of these percentages, and (4) the ratio of these percentages. Then, across all studies, we pooled prevalence for GDS-15 ≥ 5 , prevalence of major depression based on the SCID, and the difference in prevalence from each study.

To estimate pooled prevalence values, we fit generalized linear mixed-effects models with a logit link function using the `glmer` function in R³². All analyses were conducted in R (version 4.2.1) using RStudio (version 2022.07.2)^{33,34}. We fit fifteen models for the indicator of having a GDS score greater than or equal to each GDS cutoff (one model for each possible cutoff of 1 to 15). We fit another model for the indicator of classification with SCID major depression. We also fit linear mixed-effects models using the `lmer` function in R to estimate pooled prevalence differences³³. The outcome for each cutoff was the participant-level difference in binary outcomes that reflected scoring at or above each GDS cutoff and major depression classification using the SCID. To account for between-subject correlation within the same primary study, we included random intercepts. As an estimate of heterogeneity, we calculated τ^2 , estimating between-study variance, and I^2 , estimating the proportion of variability due to heterogeneity. τ^2 can range from 0 (no heterogeneity) to 1 (extreme heterogeneity) and I^2 from 0 to 100%. To describe the range of expected prevalence differences if a new study were included in our analysis, we estimated a 95% prediction interval for the difference.

Prevalence matching

To identify which GDS-15 cutoff best matched SCID-based prevalence, we estimated the pooled differences in prevalence for each possible GDS-15 cutoff compared to the SCID. The GDS-15 cutoff with the smallest pooled difference was identified as the “prevalence-match cutoff.” Then, for each included study, in addition to the already estimated difference in prevalence based on the cutoff versus SCID major depression, we also estimated the ratio of prevalence based on the prevalence-match cutoff to that of the SCID. We produced a scatter plot of study-level differences by sample size.

In a sensitivity analysis, we fit a meta-regression generalised linear mixed model using the `lmerTest` package³⁴ to assess whether differences in prevalence for the prevalence-matched cutoff and SCID were associated with study (country human development index category, recruitment setting category, and sample size) or participant (age, sex,) characteristics. We excluded participants if they were missing data on any characteristic. In a second sensitivity analysis, we compared GDS-15 ≥ 5 and SCID prevalence from eligible published studies that did not provide individual participant data. See eMethods 3.

Results

Search results and inclusion of primary study datasets

For the main IPDMA, of the 9,318 unique titles and abstracts identified from the search, 8,914 were excluded after title and abstract review and 317 after full-text review. The 87 remaining articles comprised data from 75 unique samples, of which 21 (28%) contributed individual participant data. Authors of included studies also contributed data from two unpublished studies, resulting in a total of 23 datasets.

For the present study analyses, we excluded nine studies that classified major depression using a diagnostic interview other than the SCID. In total, our analyses included 3,062 participants (434 major depression cases; prevalence 14.8%) from 14 primary studies (see Fig. 1). Table 1 shows characteristics of each included study. Of the 14 included studies, two included GDS-30 data.

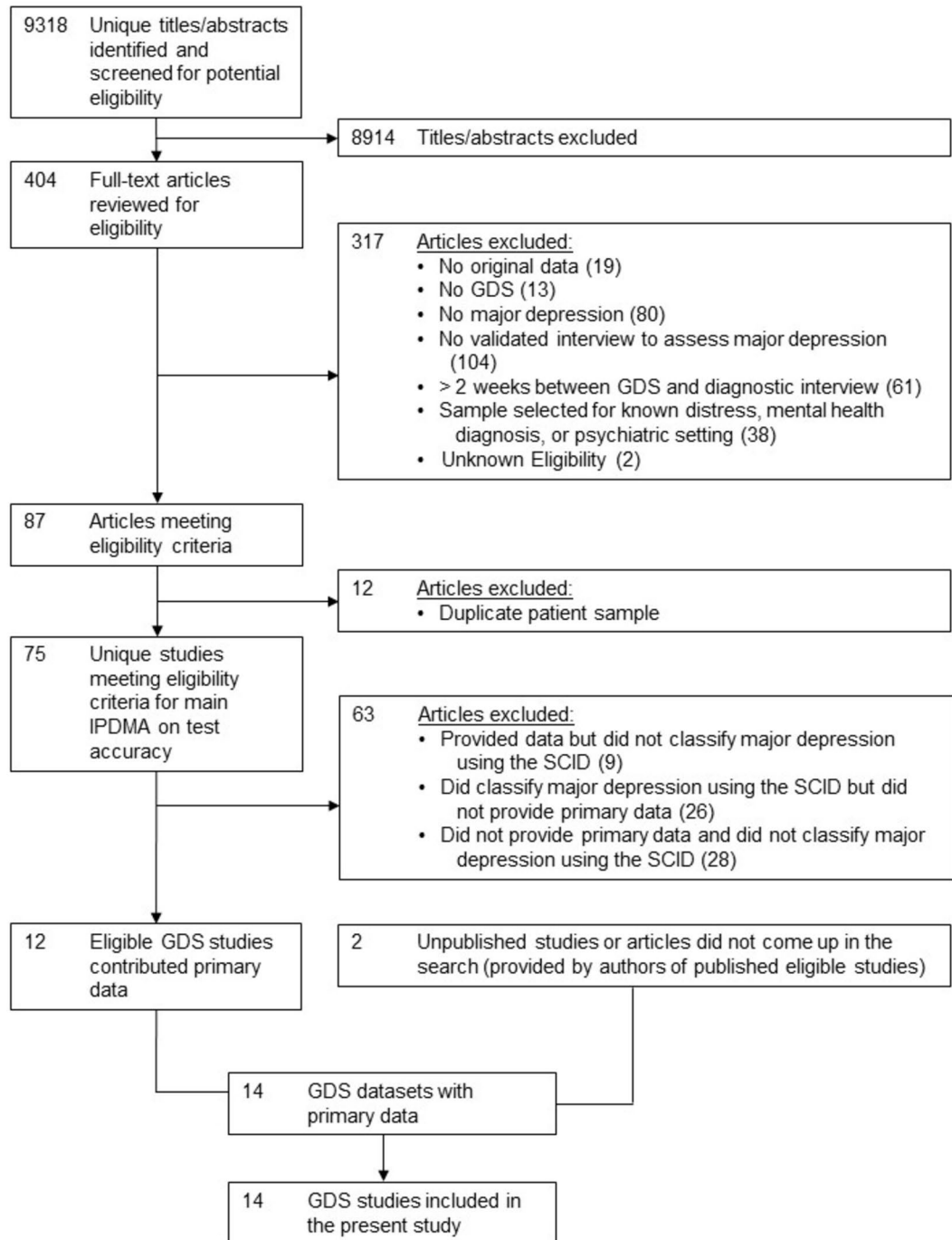


Figure 1. Flow diagram of the study selection process. GDS, Geriatric Depression Scale; IPDMA, individual participant data meta-analysis; SCID, Structured Clinical Interview for DSM.

Author, year	Country	Population	N total	N (%) major depression	Mean age (years)	% female	N (%) GDS-15 ≥ 5	% difference: GDS-15 ≥ 5 – SCID major depression	Ratio: GDS-15 ≥ 5 / SCID major depression	N (%) GDS-15 ≥ 8	% difference: GDS-15 ≥ 8 – SCID major depression	Ratio: GDS-15 ≥ 8 / SCID major depression
Allgaier, 2013	Germany	Nursing home residents	92	13 (14.1%)	84.5	73.9%	43 (46.7%)	32.6%	3.3	19 (20.7%)	6.5%	1.5
Chagas, 2013	Brazil	Parkinson disease outpatients	84	19 (22.6%)	59.4	46.4%	46 (54.8%)	32.1%	2.4	23 (27.4%)	4.8%	1.2
Dokuzlar, 2018	Turkey	Geriatric department outpatients	435	106 (24.4%)	72.9	63.9%	108 (24.8%)	0.5%	1.0	58 (13.3%)	−11.0%	0.5
Dorow, 2018	Germany	Outpatients in primary care	1183	51 (4.3%)	80.5	62.8%	264 (22.3%)	18.0%	5.2	90 (7.6%)	3.3%	1.8
Michopoulos, 2010	Greece	Elderly medical inpatients	190	27 (14.2%)	73.7	46.8%	80 (42.1%)	27.9%	3.0	44 (23.2%)	8.9%	1.6
Mougias, 2017	Greece	Patients with dementia	94	20 (21.3%)	77.3	72.3%	40 (42.6%)	21.3%	2.0	26 (27.7%)	6.4%	1.3
Prisnie, 2016	Canada	Outpatients in stroke clinic	110	11 (10.0%)	60.0	54.5%	32 (29.1%)	19.1%	2.9	15 (13.6%)	3.6%	1.4
Quinn, Unpublished	UK	Stroke patients	41	4 (9.8%)	70.2	33.3%	13 (31.7%)	22.0%	3.2	7 (17.1%)	7.3%	1.8
Saracino, 2017	USA	Outpatients with cancer	163	5 (3.1%)	76.9	51.9%	23 (14.1%)	11.0%	4.6	3 (1.8%)	−1.2%	0.6
Soysal, 2016	Turkey	Geriatric outpatients	186	53 (28.5%)	74.1	60.8%	44 (23.7%)	−4.8%	0.8	25 (13.4%)	−15.1%	0.5
Taycan, 2009	Turkey	Outpatients and nursing home patients with dementia	102	44 (43.1%)	78.3	63.7%	65 (63.7%)	20.6%	1.5	33 (32.4%)	−10.8%	0.7
Volz, 2016	Germany	Patients in stroke rehabilitation	88	14 (15.9%)	66.4	45.5%	27 (30.7%)	14.8%	1.9	15 (17.0%)	1.1%	1.1
Watson, 2009	USA	Residential care/assisted living residents	60	4 (6.7%)	83.2	73.3%	21 (35.0%)	28.3%	5.2	6 (10.0%)	3.3%	1.5
Williams, 2012	USA	Patients with Parkinsons	234	63 (26.9%)	66.3	33.3%	83 (35.5%)	8.5%	1.3	47 (20.1%)	−6.8%	0.7

Table 1. Characteristics of included studies. Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; GDS, Geriatric Depression Scale; SCID, Structured Clinical Interview for DSM.

Among the 54 studies that did not contribute individual participant data, 26 used the SCID, but only 5 studies published major depression prevalence based on the SCID and data on participants above one or more GDS-15 cutoffs.

Comparison of GDS-15 ≥ 5 prevalence and SCID major depression prevalence

In each of the 14 included studies, the percentage of participants with GDS-15 ≥ 5 ranged from 14.1 to 63.7%, with a pooled prevalence of 34.2% (95% confidence interval [CI]: 27.5–41.6%; τ^2 : 0.321; I^2 : 92.5). The percentage of participants with SCID major depression ranged from 3.1 to 43.1%, with a pooled prevalence of 14.8% (95% CI: 10.0–21.5%; τ^2 : 0.663; I^2 : 92.6).

Differences in prevalence (GDS-15 ≥ 5 minus SCID) ranged from −4.8 to 32.6%. The pooled difference was 17.6% (95% CI: 11.6–23.6%; τ^2 : 0.012; I^2 : 96.2). The ratio of GDS-15 ≥ 5 prevalence to SCID-based prevalence ranged from 0.8 to 5.2 times (mean: 2.7; median: 2.7). The mean ratio was 4.6 times for the 4 studies with SCID-based prevalence < 10% (mean difference: 19.8%), 2.8 times for the 4 studies with SCID-based prevalence between 10 and 20% (mean difference: 23.6%), and 1.5 times for the 6 studies with SCID-based prevalence of > 20% (mean difference: 13.0%). See eTable 2 for estimates of the proportion of participants at or above each GDS-15 cutoff and differences between these proportions and SCID major depression with 95% prediction intervals for both.

Prevalence matching

Of all possible GDS-15 cutoffs, GDS-15 ≥ 8 produced the pooled prevalence estimate that most closely matched SCID major depression prevalence (difference = −0.3%, 95% CI: −4.5 to 3.9%; τ^2 : 0.005; I^2 : 92.3; Fig. 2). GDS-15 ≥ 7 produced a pooled prevalence of 19.6% (pooled difference [95% CI]: 3.5% [−0.1%, 7.9%]), and GDS-15 ≥ 9

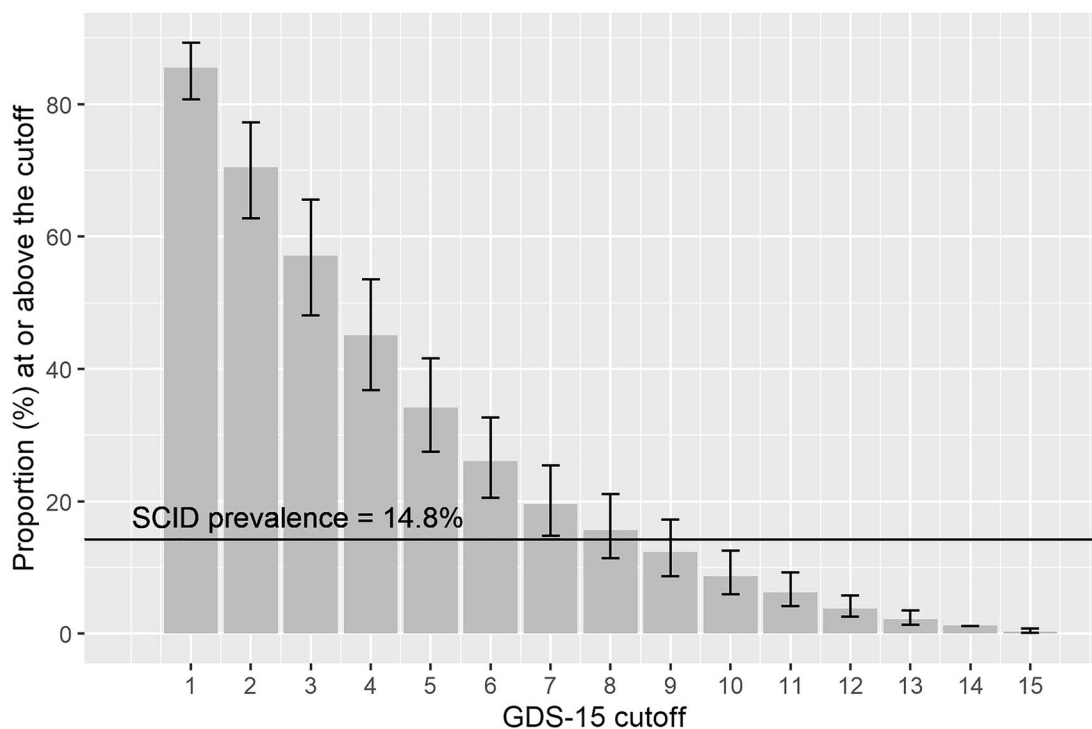


Figure 2. Prevalence estimates with 95% CIs based on a GDS-15 score at or above cutoff value of 1 to 15. CI, confidence interval; GDS, Geriatric Depression Scale; SCID, Structured Clinical Interview for DSM.

produced a pooled prevalence of 12.3% (pooled difference [95% CI]: -3.5% [-7.8 to 0.8%]). However, across the 14 individual studies, differences between GDS-15 ≥ 8 prevalence and SCID-based prevalence ranged from -15.1 to 8.9% (pooled difference [95% CI]: -0.3% [-4.5 to 3.9]). See Fig. 3. Specifically, 12 of the 14 studies (85.7%) had GDS-15 ≥ 8 prevalence estimates that were ≤ 0.75 times or ≥ 1.25 times the actual SCID-based prevalence (Table 1). The 95% prediction interval for the difference between GDS-15 ≥ 8 and SCID-based prevalence was -17.0 to 16.5% .

Sensitivity analyses

In sensitivity analyses ($N = 3,057$), no study or participant characteristics were significantly associated with differences in prevalence based on GDS-15 ≥ 8 versus SCID. See eTable 3.

Of the 5 eligible studies that published MDD prevalence based on the SCID and one or more GDS-15 cutoffs but which did not provide their data, 4 published results for GDS-15 ≥ 5 . For GDS-15 ≥ 5 , there was a weighted prevalence of 3.7% (95% CI: 3.1 – 4.2% ; τ^2 : 0.03 ; I^2 : 93.6) for SCID-diagnosed MDD and 9.9% (95% CI: 9.0 – 10.7% ; τ^2 : 0.11 ; I^2 : 99.4) for GDS-15 ≥ 5 . The weighted difference was 6.0% (95% CI: 5.1 – 7.0% ; τ^2 : 0.11 ; I^2 : 98.8). The mean ratio between the weighted prevalences was 2.7 . See eMethods 3 and eTable 4.

Discussion

We found that the most commonly used GDS-15 cutoff for reporting depression prevalence (GDS-15 ≥ 5) overestimated depression prevalence by 17.6% , or 2.7 times, compared to the SCID. A GDS-15 cutoff of ≥ 8 produced pooled prevalence estimates most similar to SCID-based prevalence (pooled difference: -0.3%). However, study-level differences showed considerable heterogeneity. Differences in estimated and real prevalence in individual studies ranged from -15.1 to 8.9% , and the prediction interval was -17.0 to 16.5% , suggesting that a new study using GDS-15 ≥ 5 may underestimate or overestimate prevalence by up to 17% . Many studies included people with health problems (e.g., Parkinson's, stroke, cancer), but the largest ratio of GDS-15 ≥ 5 to SCID prevalence was in a study of primary care patients. No study-level factors such as sample size or recruitment setting were associated with differences.

Our findings are consistent with what would be expected theoretically when screening tools are used to estimate prevalence³. Thombs et al. (2018) showed that questionnaires and cutoffs that are typically used for screening would be expected to exaggerate prevalence estimates, and that overestimation is inversely proportional to underlying prevalence. We found that GDS-15 ≥ 5 overestimated depression prevalence substantially and that the ratio of estimated to true prevalence was highest for studies with the lowest SCID-based prevalence. The present findings are also consistent with previous IPDMAs of other depression screening tools. IPDMAs that estimated prevalence based on commonly used cutoffs of the PHQ-9, EPDS, and HADS-D found that these tools overestimated depression prevalence by an average of 12 – 13% ^{5–7}. Furthermore, as in the present study, significant heterogeneity when using the prevalence-match cutoffs in the other IPDMAs precluded the possibility of identifying an alternative screening tool cutoff for prevalence estimation.

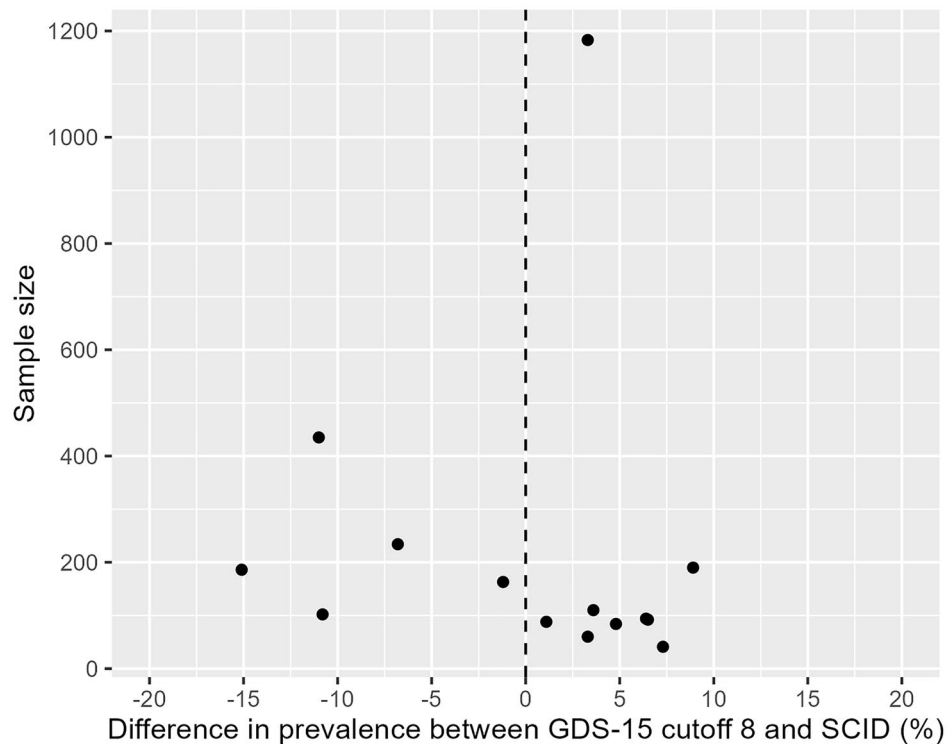


Figure 3. Plot of differences between prevalence estimates based on GDS-15 ≥ 8 and SCID major depression and sample size by included study. GDS, Geriatric Depression Scale; SCID, Structured Clinical Interview for DSM.

Many studies that report prevalence estimates based on screening cutoffs, do so with the caveat that they are prevalence estimates of “symptoms” of depression or “clinically significant” depressive symptoms rather than a depressive disorder³. However, this does not resolve the issue, as screening cutoff scores do not necessarily reflect a clinically meaningful difference between mental health impairment and non-impairment³. Indeed, some GDS items (e.g., “Do you prefer to stay at home, rather than going out and doing new things?”) may reflect personal preferences unrelated to depression. While it is true that patients who score higher than a given cutoff tend to have higher risk of depression than those scoring lower, this holds for any threshold that might be set. Reporting the proportion of individuals scoring above a certain threshold can be useful for comparisons between samples or across time points, but it should not be characterized as the percentage with depression or with symptoms of depression.

Researchers should use validated diagnostic interviews to estimate depression prevalence. However, such interviews are more time-intensive and costly to conduct compared to administering screening questionnaires. They require the hiring of specialised healthcare professionals or training of research staff which may not be feasible, especially in low-resource settings. Another option is to adopt a two-stage approach to prevalence estimation. This involves first administering a screening questionnaire to all participants^{3,35,36}. Afterwards, all patients screening positive but only a random sample of those screening negative, receive a diagnostic interview. Prevalence can be estimated by adding the number of true positives from the positive screen group to those from the negative group after weighting them by their overall proportion in the study sample. This reduces the total number of diagnostic interviews needed, with minimal effects on precision³.

This is the first study to synthesize differences in depression prevalence estimates based on GDS-15 cutoffs compared to results from a validated diagnostic interview. A strength of this study is that we were able to collect participant-level data from 14 primary studies, including 3,062 participants (434 cases of SCID major depression), and that we were able to directly compare classifications based on GDS-15 scores versus classifications from a validated diagnostic interview. A limitation is that we were not able to include data from 26 eligible datasets identified in our literature search which used the SCID to classify major depression because the authors did not provide individual participant data. In addition, we did not obtain sufficient data on the GDS-30 to include it in our analysis, although the GDS-30 is less commonly used to estimate prevalence.

In conclusion, we found that the most commonly used GDS cutoff of ≥ 5 overestimated prevalence substantially compared to a semi-structured diagnostic interview. Estimates based on a cutoff of ≥ 8 were most similar to SCID-based estimates, however, there was significant between-study heterogeneity, and differences were not associated with study or participant characteristics. Depression symptom questionnaires, including the GDS-15, may be used for multiple purposes, such as facilitating discussions between health care providers and patients,

supporting assessments if depression is suspected, or tracking progress for people receiving treatment³⁷. They should not be used to identify cases or estimate prevalence. Prevalence estimates should be based on assessments with validated diagnostic interviews. Estimating prevalence with screening tools like the GDS-15 misinforms evidence users, including healthcare decision-makers. Clinicians should be aware that studies that estimate prevalence based on standard cutoffs tend to generate much higher estimates than what they would expect to see in clinical practice.

Data availability

Data contribution agreements with primary study authors do not include permission to make their data publicly available, although the dataset used in this study will be archived through a McGill University repository (Borealis, <https://borealisdata.ca/dataverse.xhtml?alias=depressproject/>). Request to access for verification purposes can be sent to Drs. Brett Thombs and Andrea Benedetti via the “Access Dataset” function on the repository website.

Received: 21 January 2024; Accepted: 24 July 2024

Published online: 29 July 2024

References

- Rogan, W. J. & Gladen, B. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* **107**, 71–76 (1978).
- Wittchen, H. U. Reliability and validity studies of the WHO - Composite International Diagnostic Interview (CIDI): A critical review. *J. Psychiatr. Res.* **28**, 57–84 (1994).
- Thombs, B. D., Kwakkenbos, L., Levis, A. W. & Benedetti, A. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. *CMAJ Can. Med. Assoc. J. J. Assoc. Medicale Can.* **190**, E44–E49 (2018).
- Levis, B. *et al.* Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: A meta-research review. *BMC Med.* **17**, 65 (2019).
- Levis, B. *et al.* Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: Individual participant data meta-analysis. *J. Clin. Epidemiol.* **122**, 115–128.e1 (2020).
- Lyubanova, A. *et al.* Depression prevalence based on the Edinburgh Postnatal Depression Scale compared to Structured Clinical Interview for DSM Disorders classification: Systematic review and individual participant data meta-analysis. *Int. J. Methods Psychiatr. Res.* **30**, e1860 (2021).
- Brehaut, E. *et al.* Depression prevalence using the HADS-D compared to SCID major depression classification: An individual participant data meta-analysis. *J. Psychosom. Res.* **139**, 110256 (2020).
- Brink, T. L. *et al.* Screening tests for geriatric depression. *Clin. Gerontol.* **1**, 37–43 (1982).
- Mitchell, A. J., Bird, V., Rizzo, M. & Meader, N. Which version of the Geriatric Depression Scale is most useful in medical settings and nursing homes? Diagnostic validity meta-analysis. *Am. J. Geriatr. Psychiatry* **18**, 1066–1077 (2010).
- Mitchell, A. J., Bird, V., Rizzo, M. & Meader, N. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: A meta-analysis of GDS30 and GDS15. *J. Affect. Disord.* **125**, 10–17 (2010).
- Montorio, I. & Izal, M. The geriatric depression scale: A review of its development and utility. *Int. Psychogeriatr.* **8**, 103–112 (1996).
- Alden, D., Austin, C. & Sturgeon, R. A correlation between the Geriatric Depression Scale long and short forms. *J. Gerontol.* **44**, P124–P125 (1989).
- Stone, L. E., Granier, K. L. & Segal, D. L. Geriatric Depression Scale. in *Encyclopedia of Gerontology and Population Aging* (eds. Gu, D. & Dupre, M. E.) 1–8 (Springer International Publishing, 2019). https://doi.org/10.1007/978-3-319-69892-2_736-1.
- First, M. B. Structured Clinical Interview for the DSM (SCID). in *The Encyclopedia of Clinical Psychology* 1–6 (Wiley, 2015). <https://doi.org/10.1002/9781118625392.wbecp351>.
- Kelly, M. J., Dunstan, F. D., Lloyd, K. & Fone, D. L. Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: A comparison of five different methods. *BMC Psychiatry* **8**, 10 (2008).
- Benedetti, A. *et al.* Diagnostic accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5 and Geriatric Depression Scale-4 for detecting major depression: Protocol for a systematic review and individual participant data meta-analysis. *BMJ Open* **8**, e026598 (2018).
- Brugha, T. S., Jenkins, R., Taub, N., Meltzer, H. & Bebbington, P. E. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol. Med.* **31**, 1001–1013 (2001).
- Brugha, T. S., Bebbington, P. E. & Jenkins, R. A difference that matters: Comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychol. Med.* **29**, 1013–1020 (1999).
- McKay, D. *Handbook of Research Methods in Abnormal and Clinical Psychology*. (SAGE, 2008).
- Kurdyak, P. A. & Gnam, W. H. Small signal, big noise: Performance of the CIDI depression module. *Can. J. Psychiatry Rev. Can. Psychiatr.* **50**, 851–856 (2005).
- Levis, B. *et al.* Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *Br. J. Psychiatry J. Ment. Sci.* **212**, 377–385 (2018).
- Levis, B. *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.* **28**, e1803 (2019).
- Wu, 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.* **129**, 109892 (2020).
- Wu, Y. *et al.* Probability of major depression classification based on the SCID, CIDI, and MINI diagnostic interviews: A synthesis of three individual participant data meta-analyses. *Psychother. Psychosom.* **90**, 28–40 (2021).
- Lobbetael, J., Leurgans, M. & Arntz, A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother.* **18**, 75–79 (2011).
- Skre, I., Onstad, S., Torgersen, S. & Kringlen, E. High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatr Scand.* **84**, 167–173 (1991).
- National Collaborating Center for Mental Health. *The NICE Guideline on the Management and Treatment of Depression in Adults (updated edition)* (NICE, 2010).
- National Collaborating Centre for Mental Health. *Depression in Adults with a Chronic Physical Health Problem: The NICE Guideline on Treatment and Management* (The British Psychological Society & The Royal College of Psychiatrists, 2010).
- US Preventive Services Task Force recommendation Statement. Siu, A.L. and the US Preventive Services Task Force (USPSTF). Screening for depression in adults. *JAMA.* **315**, 380–387 (2016).

30. Joffres, M. *et al.* Recommendations on screening for depression in adults. *CMAJ*. **185**, 775–782 (2013).
31. McGowan, J. *et al.* PRESS peer review of electronic search strategies: 2015 guideline statement. *J. Clin. Epidemiol.* **75**, 40–46 (2016).
32. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2021).
33. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
34. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
35. Lehtinen, V. *et al.* The estimated incidence of depressive disorder and its determinants in the Finnish ODIN sample. *Soc. Psychiatry Psychiatr. Epidemiol.* **40**, 778–784 (2005).
36. Dunn, G., Pickles, A., Tansella, M. & Vázquez-Barquero, J. L. Two-phase epidemiological surveys in psychiatric research. *Br. J. Psychiatry* **174**, 95–100 (1999).
37. Thombs, B. D., Rice, D. B., Markham, S. & Ziegelstein, R. C. Depression screening in pregnancy and postpartum: just do something?. *Gen. Hosp. Psychiatry*. **82**, 14–18 (2023).

Author contributions

B.L., D.H., S.N.V., R.C.Z., Y.W., P.C., S.G., S.B.P., J.P.A.I. (DEPRESSD Steering Committee Members), S.M. (DEPRESSD Knowledge User), A.B., and B.D.T. (DEPRESSD Directors) were responsible for the conception, design and oversight of the main IPDMA project of which the present study is a part. M.P., L.Q., B.L., D.H., S.M., S.N.V., R.C.Z., Y.W., P.C., S.G., S.B.P., A.B., and B.D.T. were responsible for the conception and design of the present study. J.T.B. designed and conducted database searches to identify eligible studies. S.B.P., A.K.A., M.H.N.C., A.T.I., N.J., H.H.K., M.L., L.M., I.M., A.A.M., C.J.N., A.P., T.J.Q., S.G.R.H., R.S., M.S., M.T.R., M.V., M.W., K.W., and S.B.W. contributed primary datasets that were included in this study. M.P., L.Q., B.L., S.F., Y.S., L.S.N.A., Y.W., A.K., C.H., T.D.S., D.N., N.G.D., E.B., P.M.B., X.Q., L.L., M.L., E.L.N. and B.D.T. contributed to data extraction and coding for the meta-analysis. M.P., L.Q., B.L., A.B., and B.D.T. contributed to the data analysis and interpretation. M.P., L.Q., B.L., and B.D.T. contributed to drafting the manuscript. All authors provided a critical review and approved the final manuscript. A.B. and B.D.T. 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.

Funding

The GDS IPDMA research received funding from Canadian Institutes of Health Research grants PJT-156365, PJT-183746. Drs. Levis and Wu were supported by Fonds de recherche du Québec—Santé (FRQS) Postdoctoral Training Fellowships, Dr. Patten by the Cuthbertson & Fischer Chair in Pediatric Mental Health, and Dr. Thombs by a Tier 1 Canada Research Chair. The primary study by Prisni *et al.* was supported by the Cumming School of Medicine and Hotchkiss Brain Institute, University of Calgary and Alberta Health Services through the Calgary Health Trust. The primary study by Dorow *et al.* was funded by the German Federal Ministry of Education and Research (BMBF, grant number: 01GY1155A). Dr. Jetté holds a Canada Research Chair in Neurological Health Services Research and an AIHS Population Health Investigator Award. Dr. Marsh was supported by grants from National Institutes of Health, during the conduct of the study. The primary study by Lees *et al.* was supported by a 'start-up' research grant from the British Geriatric Society, Scotland. Dr. Quinn reports grants from Stroke Association (UK), grants from Chief Scientist Office, during the conduct of the study. No other authors reported funding for primary studies or for their work on this study. 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.

Competing interests

All authors have completed the ICJME uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work other than that described above; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years with the following exceptions: Dr. Vigod declares that she receives royalties from UpToDate, outside the submitted work. Dr. Quinn declares grants from BMS / Pfizer, outside the submitted work. All authors declare no other relationships or activities that could appear to have influenced the submitted work. No funder had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68496-3>.

Correspondence and requests for materials should be addressed to B.D.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

the DEPRESSion Screening Data (DEPRESSD) GDS Group

Ankur Krishnan³, Chen He³, Tiffany Dal Santo³, Dipika Neupane³,
Nadia González Domínguez³, Eliana Brehaut³, Parash M. Bhandari³, Xia Qiu¹⁸, Letong Li³,
Mahrukh Imran³, Elsa-Lynn Nassar³, John P. A. Ioannidis¹⁹, Antje-Kathrin Allgaier²⁰,
Marcos H. N. Chagas²¹, Ahmet Turan Isik²², Nathalie Jetté²³, Hans-Helmut König²⁴,
Margrit Löbner²⁵, Laura Marsh²⁶, Ioannis Michopoulos²⁷, Antonis A. Mougias²⁸,
Christian J. Nelson²⁹, Alexander Pabst²⁵, Terence J. Quinn³⁰, Steffi G. Riedel-Heller²⁵,
Rebecca Saracino²⁹, Martin Scherer³¹, Martin Taylor-Rowan³², Matthias Volz^{33,34},
Katja Werheid³⁵ & Siegfried B. Weyerer³⁶

¹⁸Department of Pediatrics, West China Second University Hospital, Sichuan University, Chengdu, China.

¹⁹Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, CA, USA. ²⁰Institute of Psychology, University of the Bundeswehr Munich, Neubiberg, Germany. ²¹Mental Health Graduate Program, University of Sao Paulo, Sao Paulo, Brazil. ²²Unit for Aging Brain and Dementia, Department of Geriatric Medicine, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkey. ²³Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada.

²⁴Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ²⁵Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, Germany. ²⁶Department of Psychiatry & Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA. ²⁷2nd Department of Psychiatry, Attikon General Hospital, National and Kapodistrian University of Athens, Athens, Greece. ²⁸Nestor Psychogeriatric Association, Athens, Greece. ²⁹Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ³⁰Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. ³¹Institute of Primary Medical Care, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ³²Institute of Health and Wellbeing, University of Glasgow, Scotland, UK, Glasgow, Scotland, UK. ³³Department of Psychology, University of Kassel, Kassel, Germany. ³⁴Michael Wagner, Department of Geriatric Psychiatry and Neurodegenerative Diseases, University Hospital Bonn, Bonn, Germany. ³⁵Department of Psychology, Bielefeld University, Northrhine-Westphalia, Germany. ³⁶Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany.