



Universiteit
Leiden
The Netherlands

Comparison of patient health questionnaire-9, Edinburgh Postnatal Depression Scale and hospital anxiety and depression - depression subscale scores by administration mode: an individual participant data differential item functioning meta-analysis

Harel, D.; Wu, Y.; Levis, B.; Fan, S.Q.; Sun, Y.; Xu, M.Y.; ... ; Thombs, B.D.

Citation

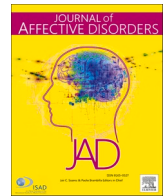
Harel, D., Wu, Y., Levis, B., Fan, S. Q., Sun, Y., Xu, M. Y., ... Thombs, B. D. (2024). Comparison of patient health questionnaire-9, Edinburgh Postnatal Depression Scale and hospital anxiety and depression - depression subscale scores by administration mode: an individual participant data differential item functioning meta-analysis. *Journal Of Affective Disorders*, 361, 674-683. doi:10.1016/j.jad.2024.06.033

Version: Publisher's Version
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/4301651>

Note: To cite this publication please use the final published version (if applicable).

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper



Comparison of Patient Health Questionnaire-9, Edinburgh Postnatal Depression Scale and Hospital Anxiety and Depression – Depression subscale scores by administration mode: An individual participant data differential item functioning meta-analysis

Daphna Harel^{a,*}, Yin Wu^b, Brooke Levis^b, Suiqiong Fan^b, Ying Sun^b, Mingyao Xu^a, Danielle B. Rice^{c,d}, Jill Boruff^e, Sarah Markham^f, John P.A. Ioannidis^g, Yemisi Takwoingi^h, Scott B. Patten^{i,j,k}, Roy C. Ziegelstein^l, Pim Cuijpers^m, Simon Gilbodyⁿ, Simone Vigod^o, Dickens Akena^p, Andrea Benedetti^{q,r,s}, Brett D. Thombs^{b,q,r,t,u,v}, DEPRESSion Screening Data (DEPRESSD) Collaboration

^a Department of Applied Statistics, Social Science, and Humanities, New York University, New York, NY, USA

^b Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada

^c Department of Psychology, St. Joseph's Healthcare, Hamilton, Ontario, Canada

^d Department of Psychiatry and Behavioural Neurosciences, McMaster University, Ontario, Canada

^e Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Québec, Canada

^f King's College London, UK

^g Department of Medicine, Department of Epidemiology and Population Health, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, CA, USA

^h Institute of Applied Health Research, University of Birmingham, Birmingham, UK

ⁱ Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

^j Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

^k O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada

^l Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

^m Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, the Netherlands

ⁿ Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK

^o Department of Psychiatry, Women's College Hospital and University of Toronto, Toronto, Ontario, Canada

^p Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda

^q Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada

^r Department of Medicine, McGill University, Montréal, Québec, Canada

^s Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada

^t Department of Psychiatry, McGill University, Montréal, Québec, Canada

^u Department of Psychology, McGill University, Montréal, Québec, Canada

^v Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada

ARTICLE INFO

Keywords:

Depression

EPDS

HADS

Mode of administration

PHQ-9

ABSTRACT

Administration mode of patient-reported outcome measures (PROMs) may influence responses. We assessed if Patient Health Questionnaire-9 (PHQ-9), Edinburgh Postnatal Depression Scale (EPDS) and Hospital Anxiety and Depression Scale – Depression subscale (HADS-D) item responses and scores were associated with administration mode. We compared (1) self-administration versus interview-administration; within self-administration (2) research or medical setting versus private; and (3) pen-and-paper versus electronic; and within interview-administration (4) in-person versus phone. We analysed individual participant data meta-analysis datasets with item-level data for the PHQ-9 ($N = 34,529$), EPDS ($N = 16,813$), and HADS-D ($N = 16,768$). We used multiple indicator multiple cause models to assess differential item functioning (DIF) by administration mode. We found statistically significant DIF for most items on all measures due to large samples, but influence on total scores was negligible. In 10 comparisons conducted across the PHQ-9, EPDS, and HADS-D, Pearson's correlations

* Corresponding author at: New York University, 246 Greene Street, New York, NY 10003, USA.

E-mail address: daphna.harel@nyu.edu (D. Harel).

<https://doi.org/10.1016/j.jad.2024.06.033>

Received 16 April 2024; Received in revised form 23 May 2024; Accepted 13 June 2024

Available online 21 June 2024

0165-0327/© 2024 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

and intraclass correlation coefficients between latent depression symptom scores from models that did or did not account for DIF were between 0.995 and 1.000. Total PHQ-9, EPDS, and HADS-D scores did not differ materially across administration modes. Researcher and clinicians who evaluate depression symptoms with these questionnaires can select administration methods based on patient preferences, feasibility, or cost.

1. Introduction

Patient-reported outcome measures (PROMs) are used in mental health research and, increasingly, in measurement-based mental health care, which involves the systematic use of patient-reported outcomes to inform treatment decisions (Lewis et al., 2019). Responses to PROMs that evaluate sensitive mental health information, including depressive symptoms, however, may be influenced by how they are administered. Administration modes may include (1) self-administration via pen-and-paper or electronic forms via computer, tablet, or handheld devices and (2) interview-administration, whereby items are read aloud in-person or via phone or videoconference. Historically, in-person methods were most commonly used, but use of alternative methods has increased in recent years, accelerated by the COVID-19 pandemic (Santomauro et al., 2021). It is unclear, however, if different administration modes generate equivalent and comparable responses or if PROM responses may differ across modes because of social desirability biases (Kreuter et al., 2008), stigma (Dillman and Christian, 2005), different respondent or researcher burden (Greenleaf et al., 2017), different visual layouts (Dillman and Christian, 2005), or other factors.

We searched PubMed on January 26, 2024 and identified 11 systematic reviews that have compared PROM results obtained via different administration modes. See appendices pp. 2–8 for search terms and characteristics of the 11 systematic reviews. Among the 11 reviews, a systematic review and meta-analysis of 6 studies did not find statistically significant differences in the disclosure of intimate partner violence between face-to-face interview-administration, paper-and-pencil self-administration, and computer-assisted self-administration, but each comparison included only 2–3 studies and small numbers of women who disclosed intimate partner violence (Hussain et al., 2015). Among the other 10 systematic reviews, most reported that scores were comparable across administration modes; when differences were reported, they were small, and patterns of differences did not repeat consistently across reviews. None of the systematic reviews, however, focused on mental health symptoms or analysed mental health symptom PROMs separately, and the percentage of mental health symptom PROMs per review was between 0 % and 30 %.

The limited evidence on whether administration mode may influence mental health PROM responses is important because theories underlying possible mode-related differences (e.g., social desirability bias) suggest that differences may be largest when measured constructs are sensitive, as with mental health symptoms (Kreuter et al., 2008; Nederhof, 1985). In practice, it is difficult to fund and conduct studies where large numbers of participants are administered the same PROM via multiple modes (within person) or where participants are randomized to receive the same PROM via different modes (between persons). An alternative is to use large, existing datasets to compare responses on PROMs obtained via different administration modes.

Differential item functioning (DIF) provides a framework to understand whether item responses and scores obtained from PROMs may be influenced by factors, such as administration mode, that are unrelated to the construct being measured (Thissen et al., 1993). When DIF is present, participants with the same level of an underlying construct, such as depression, but who are administered the PROM differently, have different expected item scores. If DIF due to administration mode is present, responses from the same PROM administered via different modes may not be directly comparable. DIF by administration mode might occur if more sensitive experiences or symptoms, such as items about self-harm, are differentially reported depending on factors such as

the level of privacy or relative anonymity associated with different administration modes.

The DEPRESSion Screening Data (DEPRESSD) Project is an international collaboration that conducts individual participant data meta-analyses (IPDMAs) (Riley et al., 2010) of depression screening questionnaire accuracy. DEPRESSD has synthesized large databases of commonly used depression symptom PROMs, including the Patient Health Questionnaire-9 (PHQ-9) (Levis et al., 2019; Negeri et al., 2021), Edinburgh Postnatal Depression Scale (EPDS) (Levis et al., 2020a), and Hospital Anxiety and Depression Scale – Depression subscale (HADS-D) (Wu et al., 2021). The objective of the present study was to compare item responses on the PHQ-9, EPDS, and HADS-D by administration mode and determine if total scores are comparable across administration modes. To do this, we first confirmed the unidimensionality of all measures. We then assessed DIF on each PROM across administration modes and evaluated whether total scores were comparable across modes. We assessed DIF by comparing (1) self-administration versus interview-administration. Then, among respondents for whom a PROM was self-administered, we compared (2) administration in research or medical settings to private administration outside of a research setting (e.g., home) and (3) administration via pen-and-paper versus electronic forms. Among interview-administered assessments, we compared (4) in-person versus phone administration.

2. Methods

This was a secondary analysis that used data from IPDMAs on depression screening accuracy of the PHQ-9, EPDS, and HADS-D (Levis et al., 2019; Levis et al., 2020a; Levis et al., 2020b; Negeri et al., 2021; Wu et al., 2021). The main IPDMAs were registered in PROSPERO (CRD42014010673, CRD42015024785, CRD42015016761), and protocols were published. We used databases compiled for those IPDMAs to conduct the present study. Prior to initiation, we posted a protocol online (<https://osf.io/kp4m6/>).

2.1. Dataset eligibility

For the main PHQ-9, EPDS, and HADS-D IPDMAs, data from articles in any language were eligible for inclusion if (1) they included diagnostic classification for current Major Depressive Disorder or Major Depressive Episode using Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 2013) or International Classification of Diseases (ICD) (World Health Organization, 1992) criteria based on a validated semi-structured or fully structured interview; (2) they include total scores for the PHQ-9, EPDS, or HADS-D PROMs; (3) the diagnostic interview and PROM were administered within two weeks of each other; (4) participants were ≥ 18 years of age and not recruited from youth or psychiatric settings; and (5) participants were not recruited because they were already identified as having symptoms of depression or were receiving treatment for depression since screening is done to identify previously unrecognized cases (Rice and Thombs, 2016; Thombs et al., 2011). Datasets where not all participants were eligible were included if primary data allowed selection of eligible participants. For the present study, we included only primary datasets with item-level scores because item scores are needed to conduct DIF analyses.

2.2. Search strategy and study selection

We used peer-reviewed search strategies (McGowan et al., 2016) developed by an experienced health sciences librarian to identify eligible studies for each DEPRESSD database. We searched MEDLINE All (Ovid) PsycINFO (Ovid), and Web of Science Core Collection from January 1, 2000 to May 9, 2018 for the PHQ (Levis et al., 2019; Levis et al., 2020b; Negeri et al., 2021) inception to October 3, 2018 for the EPDS (Levis et al., 2020a), and inception to February 21, 2020 for the HADS-D (Wu et al., 2021). See appendices pp. 9–17 for search strategies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA). After de-duplication, unique citations were uploaded into DistillerSR (Evidence Partners, Ottawa, Canada). We additionally reviewed reference lists of relevant reviews and queried contributing authors about non-published studies. Two investigators reviewed titles and abstracts independently for eligibility. If either deemed a study potentially eligible, full-text review was completed, with disagreements resolved by consensus, consulting a third investigator when necessary.

2.3. Data contribution, extraction, and synthesis

Authors of eligible datasets were invited to contribute de-identified primary data. We emailed corresponding authors of eligible primary studies at least three times, as necessary. If no response was received, we emailed co-authors then attempted to contact corresponding authors by phone. Individual participant data that were obtained were transferred to a standard format and merged into a single dataset with study-level data. Any discrepancies between published primary study results and raw datasets were resolved in consultation with primary study investigators.

Participant-level data included age, sex, major depression status, and item-level scores for the PHQ-9, EPDS, or HADS-D. Study-level data extracted from published reports by two investigators independently, with any disagreements resolved by consensus, included country, recruitment setting (non-medical care, primary care or specialty outpatient care, inpatient care), World Health Organization region, and United Nation Human Development Index (HDI) classification. The HDI is a statistical composite index that includes indicators of life expectancy, education, and income, for the country in the year of the study publication (Human Development Reports, n.d.). For four studies with multiple recruitment settings, recruitment setting was coded at the participant level.

When datasets included statistical weights to reflect sampling procedures, we used provided weights. When studies' sampling procedures merited weighting, but no weights were available, we constructed appropriate weights using inverse selection probability values based on the sampling procedure.

PROM administration modes were extracted from published study reports by two independent investigators with any disagreements resolved by consensus, consulting with a third investigator as necessary. We contacted study authors if administration mode was not specified or was unclear. If administration mode could not be determined and investigators did not respond, the study was not included in the analyses. Administration modes were classified as (a) self-administration in a research or medical setting with pen-and-paper forms; (b) self-administration in a private setting (e.g., at home) with pen-and-paper forms; (c) self-administration in a research or medical setting with electronic forms (e.g., handheld device, tablet, computer); (d) self-administration in a private setting with electronic forms; (e) interview-administration face-to-face in any setting; (f) interviewer administration by phone in any setting. No included studies administered PROMs by videoconference.

2.4. Statistical analyses

We used confirmatory factor analysis (CFA) to ensure that

unidimensional factor structures could be assumed, separately for each PROM, using the weighted least squares estimator (diagonal weight matrix), robust standard errors, and a mean- and variance-adjusted chi-square statistic with delta parameterization (Muthen and Muthen, 2017). Modification indices were calculated to identify item pairs for which measurement errors correlated highly. We allowed error terms of item pairs with high modification indices and similar item content or wording to co-vary until good model fit was attained. We assessed model fit with the Tucker-Lewis Index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA). Good fit of models was indicated by TLI and CFI ≥ 0.95 and RMSEA ≤ 0.08 (Hu and Bentler, 1999).

To determine whether PHQ-9, EPDS, and HADS-D items, separately, exhibited DIF between administration modes, we used multiple indicator multiple cause (MIMIC) models. The base MIMIC model consists of a CFA model with an added direct effect of administration mode on the depressive symptoms latent factor to control for differences between studies that used different administration mode on the latent factor. We regressed each item separately on the administration mode variable to assess potential DIF (Kwakkenbos et al., 2013). DIF was confirmed by a statistically significant ($P < 0.05$) association between administration mode with the item, controlling for differences in the overall latent factor level. Any item that displayed statistically significant DIF in bivariate analysis was included in a final MIMIC model, with an additional effect between that item and the underlying latent depression symptoms factor. All MIMIC models were fit with the variance of the latent factor constrained to be equal to one so that unstandardized factor loadings and latent factor standardized mean differences would be interpretable.

For the PHQ-9, EPDS, and HADS-D, separately, we conducted MIMIC analyses to assess possible DIF for (1) self-administration versus interview-administration. Among respondents for whom the questionnaire was self-administered, we compared (2) administration in research or medical settings to private administration outside of such settings (e.g., at home) and (3) administration via pen-and-paper versus electronic forms. Among interview-administered assessments, we compared (4) in-person versus phone administration. For the HADS-D, we could not compare self-administration pen-and-paper versus electronic and interview-administration in-person versus phone because no studies reported data in the required categories. Therefore, 10 total analyses were conducted.

MIMIC models controlled for participant characteristics potentially related to depressive symptoms, including age (<45 years, 45 to 65 years, >65 years; not for EPDS), dichotomized sex or gender (as defined by the primary study data definitions; not for EPDS), HDI (low to medium, high, very high), and health care setting (non-medical care, mixed inpatient and outpatient care, inpatient care, outpatient care).

Because of the large number of participants included in each analysis, we expected to detect statistically significant but potentially low-magnitude, inconsequential DIF on many items. Thus, we assessed the influence of item-level DIF on total scores for each PROM. First, we calculated the strength of the association between latent factor scores obtained from baseline and DIF-adjusted models via the Pearson's product-moment correlation coefficient and agreement via the intraclass correlation coefficient (ICC). Second, we evaluated the standardized mean difference on the latent factor between administration modes in the baseline model and after controlling for DIF.

2.5. Involvement of people with lived experience

Dr. Sarah Markham is a member of the DEPRESSD Steering Committee who has lived experience as a person with major depressive disorder. She was a co-applicant on the funding proposal that supported this work, provided input into the study protocol, and provided comments on this manuscript. She is a co-author.

3. Results

3.1. Sample characteristics

Analyses used data from 34,529 participants for the PHQ-9 (88 studies), 16,813 for the EPDS (41 studies), and 16,768 for the HADS-D (69 studies), including 4327 individuals with data for more than one PROM. Table 1 shows participant characteristics. See appendices pp. 18–20 for flowcharts of included studies and participants. See appendices pp. 21–46 for characteristics of included studies and mode classifications for each study for the PHQ-9, EPDS, and HADS-D.

3.2. Unidimensionality of PROMs

A unidimensional CFA model of PHQ-9 items, where covariance of item residuals was restricted to zero, resulted in less-than-ideal fit because of relatively high RMSEA (TLI = 0.962, CFI = 0.972, RMSEA = 0.083). Modification indices suggested allowing residual errors of item 3 (sleep disturbances) and item 4 (tired or little energy) to covary. The CFA model was refitted allowing these residuals to covary, and fit became sufficiently good (TLI = 0.973, CFI = 0.980, RMSEA = 0.071). Similarly, for the EPDS, the initial model resulted in less-than-ideal fit (TLI = 0.940, CFI = 0.953, RMSEA = 0.114). Modification indices suggested allowing error measurements of items 1 (able to laugh) and 2 (enjoyment of things) plus items 4 (anxious) and 5 (scared or panicky) to covary. The refitted model with covarying error terms indicated a good fit (TLI = 0.980, CFI = 0.985, RMSEA = 0.066). A unidimensional CFA model of the HADS-D, where covariance of item residuals was restricted to zero, resulted in a good fit (TLI = 0.992, CFI = 0.995, RMSEA = 0.041).

3.3. Differential item functioning

Factor loadings for baseline (unadjusted) and final (DIF-adjusted) MIMIC models are shown for all comparisons in appendices pp. 47–49 for the PHQ-9, appendices pp. 50–52 for the EPDS, and appendices pp. 53–54 for the HADS-D. As shown in the appendices (pp 47–54), after controlling for age (PHQ-9 and HADS-D only), sex or gender (PHQ-9 and HADS-D only), HDI, and health care setting, most items were identified as having statistically significant but small-magnitude DIF. Some items showed larger magnitude DIF, and the largest was for item 4 of the PHQ-9 (tired or little energy) between self-administered pen-and-paper versus electronic. Perhaps the most sensitive items for respondents on the PROMs we evaluated, item 9 of the PHQ-9, which assesses thoughts of death or self-harm, and item 10 of the EPDS, which assesses thoughts of self-harm, had minimal DIF.

As shown in Table 2, in all comparisons for all PROMs, latent factor scores obtained from unadjusted and DIF-corrected models were closely aligned. ICCs and correlation coefficients ranged from 0.995 to 1.000 in the 10 analyses we conducted, indicating near perfect concordance between latent depression symptom scores when DIF was or was not considered. As shown in Table 3, consistent with this, the standardized mean differences on the latent factor between mode of administration in the unadjusted baseline models and in model controlling for DIF were similar. Unadjusted and adjusted estimates were within 0.05 SMD of each other in 8 of 10 comparisons, 0.07 in one, and 0.13 in another, all minimal to small.

4. Discussion

We evaluated the influence of DIF based on the mode by which three depression symptom questionnaires, the PHQ-9 ($N = 34,529$), EPDS ($N = 16,813$), and HADS-D ($N = 16,768$), were administered. We compared (1) self-administration versus interview-administration. Within self-administration, we compared (2) administration in research or medical settings to private administration outside of a research setting (e.g.,

Table 1
Participant characteristics.

	N (%) or mean (SD)		
	PHQ-9	EPDS	HADS-D
Total participants	34,529	16,813	16,768
Age			
18–24	2871 (8.3 %)	3202 (19.0 %)	588 (3.5 %)
25–34	6908 (20.0 %)	9217 (54.8 %)	1177 (7.0 %)
35–44	6675 (19.3 %)	4316 (25.7 %)	2098 (12.5 %)
45–54	6573 (19.0 %)	63 (0.4 %)	3432 (20.5 %)
55–64	5748 (16.6 %)	–	3847 (22.9 %)
65–74	3478 (10.1 %)	–	3203 (19.1 %)
>74	2276 (6.6 %)	–	2423 (14.5 %)
Missing		15 (0.1 %)	–
Sex			
Female	20,372 (59.0 %)	16,813 (100 %)	8859 (52.8 %)
Male	14,157 (41.0 %)	–	7909 (47.2 %)
Region where the study was conducted ^a			
African Region	4302 (12.5 %)	923 (5.5 %)	0 (0 %)
Eastern Mediterranean Region	200 (0.6 %)	40 (0.2 %)	135 (0.8 %)
European Region	6018 (17.4 %)	4495 (26.7 %)	9676 (57.7 %)
Region of Americas	10,268 (29.7 %)	8910 (53.0 %)	1967 (11.7 %)
South-East Asia Region	834 (2.4 %)	1106 (6.6 %)	0 (0 %)
Western Pacific Region	12,907 (37.4 %)	1339 (8.0 %)	4990 (29.8 %)
Human Development Index ^b			
Very high	26,144 (75.7 %)	13,426 (79.9 %)	16,379 (97.7 %)
High	4750 (13.8 %)	2002 (11.9 %)	389 (2.3 %)
Low or medium	3635 (10.5 %)	1385 (8.2 %)	0 (0 %)
Setting			
Non-medical care	10,654 (30.9 %)	28 (0.2 %)	1813 (10.8 %)
Mixed inpatient and outpatient care	130 (0.4 %)	972 (5.8 %)	247 (1.5 %)
Inpatient care	2929 (8.5 %)	504 (3.0 %)	7498 (44.7 %)
Outpatient care	20,816 (60.3 %)	15,309 (91.1 %)	7210 (43.0 %)
Administration mode			
Self-administration			
Research or medical setting			
Pen-and-paper	15,176 (44.0 %)	2618 (15.6 %)	9721 (58.0 %)
Electronic	369 (1.1 %)	450 (2.7 %)	164 (1.0 %)
Private setting (e.g., home)			
Pen-and-paper	4571 (13.2 %)	2013 (12.0 %)	3697 (22.0 %)
Electronic	1142 (3.3 %)	0 (0 %)	0 (0 %)
Interview-administration			
In-person	7494 (21.7 %)	10,169 (60.4 %)	2653 (15.8 %)
Videoconference	0 (0 %)	0 (0 %)	0 (0 %)
Phone	5185 (15.0 %)	731 (4.3 %)	0 (0 %)
Mixed methods	592 (1.7 %)	832 (4.9 %)	533 (3.2 %)
Major depression based on diagnostic interview	3757 (10.9 %)	1630 (9.7 %)	1676 (9.7 %)
Total PHQ-9, EPDS, or HADS-D score	5.3 (5.4)	6.4 (5.6)	6.0 (4.0)

PHQ-9: Patient Health Questionnaire-9; EPDS: Edinburgh Postnatal Depression Scale; HADS-D: Hospital Anxiety and Depression Scale – Depression subscale.

^a Regions based on World Health Organization designations (www.who.int/countries).

^b United Nations Human Development Index.

at home); and (3) administration via pen-and-paper forms versus electronic forms. Among interview-administered assessments, we compared (4) in-person administration versus phone administration.

As expected, given the large sample size in each analysis, we found statistically significant, but mostly small magnitude DIF. The presence of DIF did not meaningfully influence latent depression symptom factor levels for any of the measures. Pearson’s correlations and ICCs between latent factor levels that did and did not account for DIF ranged from 0.995 and 1.000, indicating near perfect correlation and near perfect agreement. The standardized mean differences for the latent factor across modes of administration assessed in each model did not meaningfully change between models that did and did not account for DIF, similarly indicating a negligible impact of DIF on latent factor scores.

This is by far the largest study that has compared mental health symptom PROMs across administration modes. We found that the PHQ-9, EPDS, and HADS-D performed similarly across different modes and that total scores were comparable, which is generally consistent with previous meta-analyses that did not focus on mental health measures (see appendices pp. 3–8).

This study has several research and clinical implications. Researchers may choose to administer the PHQ-9, EPDS, or HADS-D using a range of administration modes without being concerned that the measures should be scored differently. Factors such as research participant preferences, feasibility, and cost can be considered in selecting administration mode. When conducting primary or secondary analyses, researchers

may synthesize evidence and results obtained from these PROMs across the different administration modes. Clinically, depression symptom PROMs are often collected as part of diagnostic assessments or to inform treatment decisions through progress monitoring and measurement-based care (Tasca et al., 2019). Based on our results, clinicians may administer the PHQ-9, EPDS, or HADS-D via multiple modes without undue concern that scores will be influenced. For instance, they can be administered in person or as part of a telehealth appointment, and scores can be interpreted in the same way.

There are limitations to consider. First, in our data synthesis, we were unable to obtain data from all investigators we contacted. For the PHQ-9, we could not include data from 31 of 126 eligible published studies; for the EPDS, 25 of 82; and for the HADS-D, 72 of 165 (see appendices pp. 18–20). However, pooled estimates of screening accuracy from previous meta-analyses in these datasets that incorporated aggregate and IPDMA data versus analyses that only used IPDMA were similar (Levis et al., 2020a; Negeri et al., 2021; Wu et al., 2021). Second, the datasets that we analysed, although providing the best evidence available on this topic to date, were not originally designed to compare administration mode. Large, randomized experiments that assess the effects of different administration modes for these scales would provide strong evidence. Third, the datasets that we compiled for our IPDMAs excluded participants known to be undergoing depression treatment, although approximately 10 % of participants across PROM datasets met criteria for major depression, which is likely similar to what would occur in many non-specialist settings where mental health care is provided. Fourth, for the HADS-D, we were unable to conduct comparison between self-administration pen-and-paper and electronic modes and between in-person and phone interviews due to a lack of data. Fifth, we assessed the influence of administration mode on continuous measure scores and

Table 2
Correlation and intraclass correlation coefficients for the PHQ-9, EPDS, and HADS-D.

Comparison	PHQ-9		EPDS		HADS-D	
	Pearson correlation (95 % confidence interval)	ICC (95 % confidence interval)	Pearson correlation (95 % confidence interval)	ICC (95 % confidence interval)	Pearson correlation (95 % confidence interval)	ICC (95 % confidence interval)
Self- versus interview-administration	0.999 (0.999, 0.999)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
Self-administration: research or medical versus private	0.999 (0.999, 0.999)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
Self-administration: pen-and-paper versus electronic	0.997 (0.997, 0.997)	0.999 (0.999, 0.999)	0.998 (0.998, 0.998)	0.995 (0.995, 0.995)	–	–
Interview-administration: in-person versus phone	0.999 (0.999, 0.999)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	0.999 (0.999, 0.999)	–	–

EPDS: Edinburgh Postnatal Depression Scale; HADS-D: Hospital Anxiety and Depression Scale - Depression subscale; ICC: Intraclass Correlation Coefficient; PHQ-9: Patient Health Questionnaire-9.

Table 3
Standardized mean difference (SMD) based on administration mode with and without DIF adjustment.

Comparison	PHQ-9		EPDS		HADS-D	
	Unadjusted SMD (95 % confidence interval)	DIF-adjusted SMD (95 % confidence interval)	Unadjusted SMD (95 % confidence interval)	DIF-adjusted SMD (95 % confidence interval)	Unadjusted SMD (95 % confidence interval)	DIF-adjusted SMD (95 % confidence interval)
Self- versus interview-administration	–0.057 (–0.086, –0.027)	–0.013 (–0.062, 0.038)	0.273 (0.233, 0.314)	0.276 (0.226, 0.326)	0.022 (–0.026, 0.071)	0.009 (–0.042, 0.061)
Self-administration: research or medical versus private	0.238 (0.188, 0.289)	0.229 (0.177, 0.281)	0.123 (0.028, 0.217)	0.157 (0.055, 0.259)	0.000 (–0.044, 0.045)	–0.008 (–0.063, 0.047)
Self-administration: pen-and-paper versus electronic	0.274 (0.208, 0.341)	0.144 (0.047, 0.240)	–0.206 (–0.326, –0.087)	–0.140 (–0.298, 0.017)	–	–
Interview-administration: in-person versus phone	0.597 (0.521, 0.674)	0.638 (0.526, 0.750)	–0.380 (–0.467, –0.292)	–0.367 (–0.497, –0.237)	–	–

EPDS: Edinburgh Postnatal Depression Scale; HADS-D: Hospital Anxiety and Depression Scale - Depression subscale; ICC: Intraclass Correlation Coefficient; PHQ-9: Patient Health Questionnaire-9; SMD: standardized mean difference.

not diagnostic accuracy compared to a reference standard. A future study could address diagnostic accuracy, but, presently, given the very high correlations between unadjusted and DIF-adjusted scores, it would be reasonable to assume that mode may not influence accuracy. Sixth, our focus in this manuscript was on DIF, and we investigated observed differences in item-level parameters and their impact on total scores. This is consistent with the overall unidimensionality of the measures we studied and the use of their total scores in practice. A multigroup confirmatory factor analysis could be used in future research to determine whether the underlying factor structure is consistent across modes of administration and if measurement invariance is achieved.

In conclusion, we investigated whether mode of administration used for depressive symptom PROMs, specifically the PHQ-9, EPDS, and HADS-D, resulted in substantively different score estimates if the PROMs were administered through (1) self-administration versus interview, (2) in research or medical settings versus in private settings for self-administration, (3) with pen-and-paper or electronic forms in self-administration, and (4) through in-person versus phone interviews. We found that, despite statistically significant DIF on items across administration modes, the overall influence of different administration modes on total PROM scores was negligible. Our findings suggest that that scores on sensitive mental health measures, specifically the PHQ-9, EPDS, and HADS-D, are not influenced meaningfully by administration mode.

Statement of ethics

The Research Ethics Committee of the Jewish General Hospital declared that research ethics approval was not required since the study involved IPDMA of de-identified previously collected data. However, for each included dataset, we confirmed that the original study received ethics approval and that participants provided informed consent.

Funding sources

Compilation of the datasets used in this study was funded by the Canadian Institutes of Health Research (KRS-134297, PCG-155468, PJT-162206, KRS-140994, KRS-144045, PJT-183746), and the analyses were funded by the New Frontiers in Research Fund (NFRFR-2021-00336). Dr. Patten was supported by the Cuthbertson & Fischer Chair in Pediatric Mental Health. Dr. Thoms was supported by a Tier 1. Canada Research Chair, outside of the present work.

The primary studies by Fiest et al., Patten et al., Amoozegar et al. and Prsnie et al. were supported by the Cumming School of Medicine, University of Calgary, and Alberta Health Services through the Calgary Health Trust, as well as the Hotchkiss Brain Institute. Dr. Jetté was supported by a Canada Research Chair in Neurological Health Services Research and an AIHS Population Health Investigator Award. She is Icahn School of Medicine at Mount Sinai Bludhorn Professor of International Medicine. The primary study by Alvarado et al. was supported by the Ministry of Health of Chile. The primary study by Amtmann et al. was supported by a grant from the Department of Education (NIDRR grant number H133B080025) and by the National Multiple Sclerosis Society (MB 0008). Collection of data for the study by Arroll et al. was supported by a project grant from the Health Research Council of New Zealand. Data collection for the study by Ayalon et al. was supported from a grant from Lundbeck International. The primary study by Khamseh et al. was supported by a grant (M-288) from Tehran University of Medical Sciences. The primary study by Barnes et al. was supported by a grant from the Health Foundation (1665/608). The primary study by Beck, CT et al. was supported by the Patrick and Catherine Weldon Donaghue Medical Research Foundation and the University of Connecticut Research Foundation. The primary studies by Marrie et al. and Bernstein et al. were supported by CIHR (THC-135234) and Crohn's and Colitis Canada. Dr. Marrie was supported by the Waugh Family Chair in Multiple Sclerosis and the Research Manitoba Chair during the

conduct of the study. Dr. Bernstein was supported in part by the Bingham Chair in Gastroenterology. The primary studies by Kohrt et al. and Nakku et al. were output of the PProgramme for Improving Mental health care (PRIME) and were supported by the UK Department for International Development (201446). The views expressed do not necessarily reflect the UK Government's official policies. The primary study by Helle et al. was supported by the Werner Otto Foundation, the Kroschke Foundation, and the Feindt Foundation. The primary study by Bombardier et al. was supported by the Department of Education, National Institute on Disability and Rehabilitation Research, Spinal Cord Injury Model Systems: University of Washington (grant no. H133N060033), Baylor College of Medicine (grant no. H133N060003), and University of Michigan (grant no. H133N060032). The primary study by Buji et al. was supported by grants from the UKMMC Fundamental Research Fund (FF-2015-051) and the Fundamental Research Grant Scheme by the Malaysian Ministry of Higher Education (FRGS/2/2014/SKK09/UKM/02/1). The primary study by Butnoriene et al. was supported by a grant from the Research Council of Lithuania (LIG-03/2011). Collection of data for the primary study by Kiely et al. was supported by National Health and Medical Research Council (grant number 1002160) and Safe Work Australia. Dr. Butterworth was supported by Australian Research Council Future Fellowship FT130101444. Collection of data for the primary study by Zhang et al. was supported by the European Foundation for the Study of Diabetes, the Chinese Diabetes Society, Lilly Foundation, Asia Diabetes Foundation and Liao Wun Yuk Diabetes Memorial Fund. The primary study by Chen et al. was supported by the National Science Council, Taiwan (NSC 96–2314-B-182A-090-MY2). The primary study by Chibanda et al. was supported by a grant from Grand Challenges Canada (0087-04). The primary study by Chorwe-Sungani et al. was supported by the University of Malawi through grant QZA-0484 NORHED 2013. The primary study by Cheung et al. was supported by the Waikato Clinical School, University of Auckland, the Waikato Medical Research Foundation and the Waikato Respiratory Research Fund. The primary study by Conway et al. was supported by the Institute of Health and Biomedical Innovation at Queensland University of Technology and the Sigma Theta Tau International Honour Society of Nursing (ID: 8580). Dr. Conwell received support from NIMH (R24MH071604) and the Centers for Disease Control and Prevention (R49 CE002093). The primary study by Figueira et al. was supported by the Brazilian Ministry of Health and by the National Counsel of Technological and Scientific Development (CNPq) (Grant no. 403433/2004-5). The primary study by Couto et al. was supported by the National Counsel of Technological and Scientific Development (CNPq) (Grant no. 444254/2014-5) and the Minas Gerais State Research Foundation (FAPEMIG) (Grant no. APQ-01954-14). The primary study by Cukor et al. was supported in part by a Promoting Psychological Research and Training on Health-Disparities Issues at Ethnic Minority Serving Institutions Grants (ProDIGs) awarded to Dr. Cukor from the American Psychological Association. The primary study by De la Torre et al. was supported by a Research Grant “Ramón Carrillo-Arturo Oñativa for Multicentric Studies” (2015) from the commission “Salud Investiga” of the Ministry of Health and Social Action of Argentina (Grant n° 1853). The primary study by de Figueiredo et al. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo.724: The primary study by Groot et al. was supported by a programme grant from The Netherlands Heart Foundation (2007B027). The primary study by Zuithoff et al. was supported by The European Commission (PREDICTQL4-CT2002-00683) and The Netherlands Organization for Scientific Research (ZonMw 016.046.360). The primary study by De Souza et al. was supported by Birmingham and Solihull Mental Health Foundation Trust. Collection of data for the primary study by Delgadillo et al. was supported by grant from St. Anne's Community Services, Leeds, United Kingdom. The primary study by Dorow et al. was supported by the German Federal Ministry of Education and Research (Grant/Award Number: 01GY1155A). Collection of data for the primary study by Fann et al. was supported by grant RO1 HD39415 from the US

National Center for Medical Rehabilitation Research. The primary study by Tissot et al. was supported by the Swiss National Science Foundation (grant 32003B 125493). The primary study by Fernandes et al. was supported by grants from the Child: Care Health and Development Trust and the Department of Psychiatry, University of Oxford, Oxford, UK, and by the Ashok Ranganathan Bursary from Exeter College, University of Oxford. Dr. Fernandes is supported by a University of Southampton National Institute for Health Research (NIHR) academic clinical fellowship in Paediatrics. The primary study by van Heyningen et al. was supported by the Medical Research Council of South Africa (fund no. 415865), Cordaid Netherlands (Project 103/10002 G Sub 7) and the Truworths Community Foundation Trust, South Africa. The primary study by Tendais et al. was supported under the project POCI/SAU-ESP/56397/2004 by the Operational Program Science and Innovation 2010 (POCI 2010) of the Community Support Board III and by the European Community Fund FEDER. The primary study by Fischer et al. was funded by the German Federal Ministry of Education and Research (01GY1150). The primary study by Fisher et al. was supported by grants from the National Health and Medical Research Council (APP1026550), the Australian Government Department of Social Services - Families, Housing, Community Services and Indigenous Affairs, and the Victorian Department of Education and Early Childhood Development. Dr. Fisher was supported by a Monash Professorial Fellowship and the Jean Hailes Professorial Fellowship, which is supported by a grant to the Jean Hailes Foundation from the H and L Hecht Trust managed by Perpetual Trustees. The primary study by Gagnon et al. was supported by the Drummond Foundation and the Department of Psychiatry, University Health Network. The primary study by Akechi et al. was supported in part by a Grant-in-Aid for Cancer Research (11-2) from the Japanese Ministry of Health, Labour and Welfare and a Grant-in-Aid for Young Scientists (B) from the Japanese Ministry of Education, Culture, Sports, Science and Technology. The primary study by Kugaya et al. was supported in part by a Grant-in-Aid for Cancer Research (9-31) and the Second-Term Comprehensive 10-year Strategy for Cancer Control from the Japanese Ministry of Health, Labour and Welfare. The primary study Ryan et al. was supported by the Irish Cancer Society (Grant CRP08GAL). The primary study by Baron et al. was supported by the National Income Dynamics Study (NIDS), at the University of Cape Town, South Africa. The NIDS is implemented by the Southern Africa Labour and Development Research Unit, and is funded by the Department of Planning, Monitoring and Evaluation. The funding body was involved in the design of the primary study. Data for the primary study by Gelaye et al. was supported by a grant from the NIH (T37 MD001449). The primary study by Gholizadeh et al. was supported by University of Technology Sydney under UTS Research Reestablishment Grants. The primary study by Grassi et al. was supported by the European Commission DG Health and Consumer Protection (Agreement with the University of Ferrara — SI2.307317 2000CVGG2-026), the University of Ferrara, and the Fondazione Cassa di Risparmio di Ferrara. The primary study by Green et al. (2018) was supported by a grant from the Duke Global Health Institute (453-0751). The primary study by Eack et al. was funded by the NIMH (R24 MH56858). Collection of data for the primary study by Hobfoll et al. was made possible in part from grants from NIMH (RO1 MH073687) and the Ohio Board of Regents. Dr. Hall received support from a grant awarded by the Research and Development Administration Office, University of Macau (MYRG2015-00109-FSS). The primary study by Garabiles et al. was supported by the Macao (SAR) Government, through the University of Macau RSKTO grants: MYRG-2014-111. The primary study by Hantsoo et al. was supported by K23 MH107831-02, Brain and Behavior Research Foundation NARSAD Young Investigator Award. The primary study by Haroz et al. was supported by the United States Agency for International Development Victims of Torture Fund: AID-DFD A-00-08-00308. Dr. Haroz was supported by a NIMH T32 predoctoral training grant (MH014592-38) and postdoctoral training grant (MH103210) during the conduct of primary study. Collection of data provided by Drs. Härter and Reuter was

supported by the Federal Ministry of Education and Research (grants No. 01 GD 9802/4 and 01 GD 0101) and by the Federation of German Pension Insurance Institute. The primary study by Henkel et al. was funded by the German Ministry of Research and Education. The primary study by Hides et al. was funded by the Perpetual Trustees, Flora and Frank Leith Charitable Trust, Jack Brockhoff Foundation, Grosvenor Settlement, Sunshine Foundation and Danks Trust. The primary study by Howard et al. was supported by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Numbers RP-PG-1210-12002 and RP-DG-1108-10012) and by the South London Clinical Research Network. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Collection of data for the primary study by Hyphantis et al. (2014) was supported by grant from the National Strategic Reference Framework, European Union, and the Greek Ministry of Education, Lifelong Learning and Religious Affairs (ARISTEIA-ABREVIATE, 1259). The primary study by Paika et al. was supported by the European Economic Area (EEA) Financial Mechanism 2009–2014 (EEA GR07/3767) and National funds as part of the program “Dissimilarity, Inequality and Social Integration” (132324/I4-25/8/2015). The primary study by Inagaki et al. was supported by the Ministry of Health, Labour and Welfare, Japan. The primary study by Twist et al. was funded by the UK National Institute for Health Research under its Programme Grants for Applied Research Programme (grant reference number RP-PG-0606-1142). The primary study by Kim et al. was supported by a grant from the Korean Mental Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HM14C2567), an Institute for Information & Communications Technology Promotion grant funded by the Korea government (MSIP) (B0132-15-1003: the development of skin adhesive patches for the monitoring and prediction of mental disorders), and by the Original Technology Research Program for Brain Science through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2016M3C7A1947307). The primary study by Kettunen et al. was supported with an Annual EVO Financing (Special government subsidies from the Ministry of Health and Welfare, Finland) by North Karelia Central Hospital and Päijät-Häme Central Hospital. Dr. Kiely was supported by funding from an Australian National Health and Medical Research Council fellowship (grant number 1088313). The primary study by Jang et al. was supported by a grant from the Korea Health 21 R&D, Ministry of Health and Welfare, Republic of Korea. The primary study by Kang et al. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0087344), and was supported by a Grant of the Korea Health 21 R&D, Ministry of Health and Welfare, Republic of Korea (A102065). The primary study by Phillips et al. was supported by a scholarship from the National Health and Medical Research Council (NHMRC). The primary study by Lamers et al. was funded by the Netherlands Organisation for Health Research and Development (grant number 945-03-047). The primary study by Lara et al. was supported by the Consejo Nacional de Ciencia y Tecnología/National Council for Science and Technology (CB-2009-133923-H). The primary study by Liu et al. (2011) was funded by a grant from the National Health Research Institute, Republic of China (NHRI-EX97-9706PI). The primary study by Lotrakul et al. was supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand (grant number 49086). The primary studies by Osório et al. (2012) were funded by Reitoria de Pesquisa da Universidade de São Paulo (grant number 09.1.01689.17.7) and Banco Santander (grant number 10.1.01232.17.9). The primary study by Love et al. (2004) was supported by the Kathleen Cuninghame Foundation (National Breast Cancer Foundation), the Cancer Council of Victoria and the National Health and Medical Research Council. The primary study by Love et al. (2002) was supported by a grant from the Bethlehem Griffiths Research Foundation. Dr. Bernd Löwe received research grants from Pfizer, Germany, and from

the medical faculty of the University of Heidelberg, Germany (project 121/2000) for the study by Gräfe et al. The primary study by Roomruangwong et al. was supported by the Ratchadaphiseksomphot Endowment Fund 2013 of Chulalongkorn University (CU-56-457-HR). Collection of data for the primary study by Williams et al. was supported by a NIMH grant to Dr. Marsh (RO1-MH069666). The primary study by Martínez et al. was supported by Iniciativa Científica Milenio, Chile, process # IS130005 and by Fondo Nacional de Desarrollo Científico y Tecnológico, Chile, process # 1130230. Dr. Marx was supported by the Department of Defense (W81XWH-08-2-0100/W81XWH-08-2-0102 and W81XWH-12-2-0117/W81XWH-12-2-0121). The primary study by Matsuoka et al. was supported by the Japanese Ministry of Health, Labor, and Welfare through Research on Psychiatric and Neurological Disease and Mental Health (16190501, 19230701 and 20300701). The primary study by Hartung et al. was supported by the German Cancer Aid within the psychosocial oncology funding priority program (grant number 107465). The primary study by Mohd Sidik et al. was funded under the Research University Grant Scheme from Universiti Putra Malaysia, Malaysia and the Postgraduate Research Student Support Accounts of the University of Auckland, New Zealand. The primary study by Santos et al. was funded by the National Program for Centers of Excellence (PRONEX/FAPERGS/CNPq, Brazil). The primary study by Muramatsu et al. (2007) was supported by an educational grant from Pfizer US Pharmaceutical Inc. The primary study by Muramatsu et al. (2018) was supported by grants from Niigata Seiryō University. The primary study by Nakić Radoš et al. was supported by the Croatian Ministry of Science, Education, and Sports (134-0000000-2421). The primary study by Usuda et al. was supported by Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (primary investigator: Daisuke Nishi, MD, PhD), and by an Intramural Research Grant for Neurological and Psychiatric Disorders from the National Center of Neurology and Psychiatry, Japan. The primary study by McFarlane et al. was supported by an Australian Government National Health and Medical Research Council program grant. Dr. O'Donnell was supported by grants from NHMRC Program (1073041) during the conduct of the study. The primary study by O'Rourke et al. was supported by the Scottish Home and Health Department, Stroke Association, and Medical Research Council. Dr. Osório was supported by Productivity Grants (PQ-CNPq-2-number 301321/2016-7). The primary study by Sia et al. (Pls: Pasco and Williams) was supported by the Victorian Health Promotion Foundation (ID 91-0095) and the National Health and Medical Research Council (ID 299831; 251638; 454356, 509103; 628582; 1026265). Williams is supported by a NHMRC Emerging Leader Fellowship (1174060). The primary study by Pawlby et al. was supported by a Medical Research Council UK Project Grant (number G89292999N). Collection of primary data for the study by Pence et al. was provided by NIMH (R34MH084673). The primary study by Persoons et al. was partly funded by a grant from the Belgian Ministry of Public Health and Social Affairs and supported by a limited grant from Pfizer Belgium. The primary study by Picardi et al. was supported by funds for current research from the Italian Ministry of Health. The primary study by Gould et al. was supported by the Transport Accident Commission Grant. The primary study by Wagner et al. was supported by grants U10CA21661, U10CA180868, U10CA180822, and U10CA37422 from the National Cancer Institute. The study was also funded in part by a grant from the Pennsylvania Department of Health. The department specifically disclaims responsibility for any analyses, interpretations, or conclusions of the primary study. The primary study by Bayon-Perez et al. was supported by a grant from the Instituto de Investigación Hospital 12 de Octubre (i + 12). Dr. Pulido was an investigator from the Intensification of Research Activity Program of the Instituto de Investigación Hospital 12 de Octubre (i + 12) during the conduct of the study. The primary study by Quispel et al. was supported by Stichting Achmea Gezondheid (grant number z-282). The primary study by Reme et al. was supported by the Research Council of Norway. The primary study by Rochat et al. was supported by grants from the

University of Oxford (HQ5035), the Tuixen Foundation (9940), the Wellcome Trust (082384/Z/07/Z and 071571), and the American Psychological Association. The primary study by Rooney et al. was funded by the United Kingdom National Health Service Lothian Neuro-Oncology Endowment Fund. The primary study by Schwarzbald et al. was supported by PRONEX Program (NENASC Project) and PPSUS Program of Fundação de Amparo a pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and the National Science and Technology Institute for Translational Medicine (INCT-TM). Dr. Shaaban was supported by funding from Universiti Sains Malaysia. The primary study by Schellekens et al. was supported by the Dutch Cancer Society (KUN 2011-5077) and Alpe d'HuZes Foundation. The primary study by Shinn et al. was supported by grant NCI K07 CA 093512 and the Lance Armstrong Foundation. The primary study by Sidebottom et al. was funded by a grant from the United States Department of Health and Human Services, Health Resources and Services Administration (grant number R40MC07840). The primary study by Simard et al. was supported by IDEA grants from the Canadian Prostate Cancer Research Initiative and the Canadian Breast Cancer Research Alliance, as well as a studentship from the Canadian Institutes of Health Research. The primary study by Singer et al. (2009) was supported by a grant from the German Federal Ministry for Education and Research (no. 01ZZ0106). The primary study by Singer et al. (2008) was supported by grants from the German Federal Ministry for Education and Research (# 7DZAIQTX) and of the University of Leipzig (# formel. 1-57). The primary study by Meyer et al. was supported by the Federal Ministry of Education and Research (BMBF). The primary study by Comasco et al. was supported by funds from the Swedish Research Council (VR: 521-2013-2339, VR:523-2014-2342), the Swedish Council for Working Life and Social Research (FAS: 2011-0627), the Marta Lundqvist Foundation (2013, 2014), and the Swedish Society of Medicine (SLS-331991). The primary study by Smith-Nielsen et al. was supported by a grant from the charitable foundation Tryg Foundation (Grant ID no 107616). The primary study by Spangenberg et al. was supported by a junior research grant from the medical faculty, University of Leipzig. The primary study by Stafford et al. (2014) was supported in part by seed funding from the Western and Central Melbourne Integrated Cancer Service. The primary study by Stafford et al. (2007) was supported by the University of Melbourne. Dr. Stafford received PhD scholarship funding from the University of Melbourne. The primary study by Prenoveau et al. was supported by The Wellcome Trust (grant number 071571). The primary study by Stewart et al. was supported by Professor Francis Creed's Journal of Psychosomatic Research Editorship fund (BA00457) administered through University of Manchester. Dr. Stewart was supported by MRC GCRF grant MR/S035818/1. The primary study by Phan et al. was supported by The Government of Western Australia, Department of Health (Grant number G1000794). The primary study by Su et al. was supported by grants from the Department of Health (DOH94F044 and DOH95F022) and the China Medical University and Hospital (CMU94-105, DMR-92-92 and DMR94-46). The primary study by Tandon et al. was funded by the Thomas Wilson Sanitarium. The primary study by de Oliveira et al. was supported by CNPq and Fapemig, Brazil. The primary study by Pedroso et al. (2018) was supported by FAPEMIG (APq-03539-13). The primary study by Pedroso et al. (2016) was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) (APq-03539-13). The primary study by Tiringier et al. was supported by the Hungarian Research Council (ETT 395). The primary study by Tran et al. was supported by the Myer Foundation who funded the study under its Beyond Australia scheme. Dr. Tran was supported by an early career fellowship from the Australian National Health and Medical Research Council. The primary study by Tschorn et al. is part of the study "CDCare - Supply of patients with coronary artery disease and comorbid Depression: A Patient Oriented Needs Analysis". CDCare was funded by the Federal Ministry of Education and Research (BMBF; 01GY1154). The primary study by Turner et al. was supported by a bequest from Jennie Thomas through Hunter Medical Research Institute. The study by van

Steenbergen-Weijenburg et al. was funded by Innovatiefonds Zorgverzekeraars. The primary study by Volker et al. was supported by The Netherlands organization for Health Research and Development (ZonMw) and from Achmea SZ, a Dutch insurance company. Dr. van Heyningen was supported by the National Research Foundation of South Africa and the Harry Crossley Foundation. VHYTHE001/1232209. The study by Wittkamp et al. was funded by The Netherlands Organization for Health Research and Development (ZonMw) Mental Health Program (nos. 100.003.005 and 100.002.021) and the Academic Medical Center/University of Amsterdam. The primary study by Vega-Dienstmaier et al. was supported by Tejada Family Foundation, Inc., and Peruvian-American Endowment, Inc. The primary study by Lee et al. (2017) was supported by a grant from the Kaohsiung Chang Gung Memorial Hospital, Taiwan (CMRPG8A0581). The primary study by Lee et al. (2016) was supported by a grant from Kaohsiung Chang Gung Memorial Hospital, Taiwan (CMRPG891321). The primary study by Liu et al. (2015) was supported by CIHR (MOP-114970). The primary study by Thombs et al. was done with data from the Heart and Soul Study. The Heart and Soul Study was funded by the Department of Veterans Epidemiology Merit Review Program, the Department of Veterans Affairs Health Services Research and Development service, the National Heart Lung and Blood Institute (R01 HL079235), the American Federation for Aging Research, the Robert Wood Johnson Foundation, and the Ischemia Research and Education Foundation. The primary study by Yonkers et al. was supported by a National Institute of Child Health and Human Development grant (5 R01HD045735). The primary study by Liu et al. (2016) was supported by Shanghai Municipal Health and Family Planning Commission Bureau-level Project - Preliminary Exploration of Depression Risk Prediction Model for Outpatients in General Hospitals (Project No. 2010105).

Members of the DEPRESSD Collaboration Group contributed

To data extraction, coding, and synthesis: MA, PMB, MJC, CH, MIhman, AK, ZN, DNeupane, KER, XWY. Via the design and conduct of database searches: LAK. As knowledge user consultants: MHenry, ZI, CGL, NDM. By contributing included datasets: SBP, SAA, RA, DAmtmann, BA, LA, HRB, JBarnes, KRB, CTB, CNB, CB, CHB, BB, NBD, RIB, ABunevicius, PB, CC, GC, MHC, JCNC, LFC, CKC, DChibanda, GCS, KC, RMC, AC, YC, HC, TCeC, DCukor, FMD, JMdMvG, JDS, MGD, VE, JRF, NF, EF, GF, PPF, MFernandes, SField, BF, FHF, JRWF, AJF, MFujimori, DF, PG, MG, BG, LGholizadeh, LJG, FGS, LGrassi, EPG, CGG, BJH, LHantsoo, EEH, MHärter, UH, NH, AH, LHides, SEH, SH, LMH, TH, MIG, Minagaki, JJ, HJJ, NJ, MJ, PAK, MEK, KMK, SWK, MKjærgaard, JK, BAK, HHK, ZK, YK, FL, MAL, AAL, HFLA, SIL, MLöbner, WLL, MLo-trakul, SRL, AWLove, BLöwe, NPL, CL, MM, UFM, RAM, LM, PM, BPM, YM, AMcGuire, AMehnert, IM, SMS, JMN, KM, SNR, LN, CJN, CGN, DNishi, MLO, SJO, FLO, APabst, JAP, SJP, JP, BWP, PP, IP, APicardi, JLP, SLP, FP, TJQ, CQ, SDR, SER, KR, SGRH, AGR, ISS, RMS, MPJS, MLS, VSC, JS, DJS, LSharpe, EHS, ASidebottom, SSimard, SSinger, ASkalkidou, JSN, LSpangenberg, LStafford, AStein, RCS, NAS, KPS, SSultan, ISP, SCS, KS, MTadinac, PLLT, SDT, MTR, ALT, ITendais, ITiringer, ATörekli, TDT, KT, MTschorn, ATurner, MSV, CMvdFC, TvH, JMVD, MW, LIW, LJWang, JLW, DW, SBW, JW, MAW, BW, LJWilliams, KWinkley, KWynter, MY, KAY, QZZ, YZ.

All authors, including group authors, provided a critical review and approved the final manuscript. DH had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analyses. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Role of funding

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.

CRedit authorship contribution statement

Daphna Harel: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Yin Wu:** Writing – review & editing, Project administration, Investigation, Data curation. **Brooke Levis:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Suiqiong Fan:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Data curation. **Ying Sun:** Writing – review & editing, Project administration, Investigation, Data curation. **Mingyao Xu:** Writing – review & editing, Methodology, Formal analysis. **Danielle B. Rice:** Writing – review & editing, Data curation, Conceptualization. **Jill Boruff:** Writing – review & editing, Data curation, Conceptualization. **Sarah Markham:** Writing – review & editing, Funding acquisition, Conceptualization. **John P.A. Ioannidis:** Writing – review & editing, Conceptualization. **Yemisi Takwoingi:** Writing – review & editing, Conceptualization. **Scott B. Patten:** Writing – review & editing, Conceptualization. **Roy C. Ziegelstein:** Writing – review & editing, Conceptualization. **Pim Cuijpers:** Writing – review & editing, Conceptualization. **Simon Gilbody:** Writing – review & editing, Conceptualization. **Simone Vigod:** Writing – review & editing, Conceptualization. **Dickens Akena:** Writing – review & editing, Conceptualization. **Andrea Benedetti:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Brett D. Thombs:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

All authors have completed the ICJME uniform disclosure form. Dr. Vigod declares that she receives royalties from UpToDate, outside the submitted work. Dr. Beck CT declares that she receives royalties for her Postpartum Depression Screening Scale published by Western Psychological Services. Dr. Bernstein declares that he has consulted to Abbvie Canada, Amgen Canada, Bristol Myers Squibb Canada, Eli Lilly, Ferring Canada, JAMP Pharmaceuticals, Janssen Canada, Pendopharm Canada, Pfizer Canada, Sandoz Canada, Takeda Canada. He has also received unrestricted educational grants from Abbvie Canada, Bristol Myers Squibb Canada, Eli Lilly, Ferring Canada, Organon Canada, Janssen Canada, Pfizer Canada, and Takeda Canada; as well as been on speaker's bureau of Abbvie Canada, Janssen Canada, Takeda Canada and Pfizer Canada, all outside the submitted work. Dr. Chan J CN is a steering committee member and/or consultant of Astra Zeneca, Bayer, Lilly, MSD and Pfizer. She has received sponsorships and honorarium for giving lectures and providing consultancy and her affiliated institution has received research grants from these companies. Dr. Chan LF declares personal fees and non-financial support from Otsuka, Lundbeck, and Johnson and Johnson; and non-financial support from Ortho-McNeil-Janssen, and Menarini, outside the submitted work. Dr. Hegerl declares that within the last three years, he was an advisory board member for Janssen and received a research grant from Medice, all outside the submitted work. Dr. Howard declares that she has received personal fees from NICE Scientific Advice, outside the submitted work. Dr. Inagaki declares that he has received personal fees from Meiji, Mochida, Takeda, Novartis, Yoshitomi, Pfizer, Eisai, Otsuka, MSD, Sumitomo Dainippon, Janssen, and Eli Lilly, all outside of the submitted work. Dr. Marrie declares that she is a co-investigator on a study funded in part by Biogen Idec and Roche (no funds to her), outside the submitted work. Dr. Pugh declares that she received salary support from Pfizer-Astellera and Millennium, outside the submitted work. Dr. Sultan declares funding from Sanofi-Aventis Corporation, during conduct of the primary study. Dr.

Sundström-Poromaa declares that she has served on advisory boards and acted as invited speaker at scientific meetings for MSD, Novo Nordisk, Bayer Health Care, and Lundbeck A/S. Dr. Lynne Wagner declares that she receives personal fees from Celgene, outside the submitted work. Dr. Yonkers declares that she receives royalties from UpToDate, outside the submitted work. All other authors declare no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Data availability

Requests to access data should be made to the corresponding author.

Acknowledgements

The authors would like to thank all participants of the primary studies for contributing their data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jad.2024.06.033>.

References

- American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. <https://doi.org/10.1176/appi.books.9780890425596> Arlington, VA.
- Dillman, D.A., Christian, L.M., 2005. Survey mode as a source of instability in responses across surveys. *Field Methods* 17, 30–52. <https://doi.org/10.1177/1525822X04269550>.
- Greenleaf, A.R., Gibson, D.G., Khattar, C., Labrique, A.B., Pariyo, G.W., 2017. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy across survey modalities. *J. Med. Internet Res.* 19, e140 <https://doi.org/10.2196/jmir.7331>.
- Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. <https://doi.org/10.1080/10705519909540118>.
- Human Development Reports, n.d. Human Development Index (HDI). <http://hdr.undp.org/en/content/human-development-index-hdi> (accessed 28 February 2024).
- Hussain, N., Sprague, S., Madden, K., Hussain, F.N., Pindiprolu, B., Bhandari, M., 2015. A comparison of the types of screening tool administration methods used for the detection of intimate partner violence: a systematic review and meta-analysis. *Trauma Violence Abuse* 16, 60–69. <https://doi.org/10.1177/1524838013515759>.
- Kreuter, F., Presser, S., Tourangeau, R., 2008. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opin. Q.* 72, 847–865. <https://doi.org/10.1093/poq/nfn063>.
- Kwakkenbos, L., Arthurs, E., Hoogen, F.H.J. van den, Hudson, M., van Lankveld, W.G.J. M., Baron, M., et al., 2013. Cross-language measurement equivalence of the Center for Epidemiologic Studies Depression (CES-D) Scale in systemic sclerosis: a comparison of Canadian and Dutch patients. *PLOS ONE* 8, e53923. <https://doi.org/10.1371/journal.pone.0053923>.
- Levis, B., Benedetti, A., Thombs, B.D., DEPRESSion Screening Data (DEPRESSD) Collaboration, 2019. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 365, 11476. <https://doi.org/10.1136/bmj.11476>.
- Levis, B., Negeri, Z., Sun, Y., Benedetti, A., Thombs, B.D., DEPRESSion Screening Data (DEPRESSD) EPDS Group, 2020a. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *BMJ* 371, m4022. <https://doi.org/10.1136/bmj.m4022>.
- Levis, B., Sun, Y., He, C., Wu, Y., Krishnan, A., Bhandari, P.M., et al., 2020b. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. *JAMA* 323, 2290–2300. <https://doi.org/10.1001/jama.2020.6504>.
- Lewis, C.C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., et al., 2019. Implementing measurement-based care in behavioral health: a review. *JAMA Psychiatry* 76, 324–335. <https://doi.org/10.1001/jamapsychiatry.2018.3329>.
- McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V., Lefebvre, C., 2016. PRESS peer review of electronic search strategies: 2015 guideline statement. *J. Clin. Epidemiol.* 75, 40–46. <https://doi.org/10.1016/j.jclinepi.2016.01.021>.
- Muthen, L., Muthen, B., 2017. Mplus statistical analysis with latent variables. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf. (Accessed 28 February 2024).
- Nederhof, A.J., 1985. Methods of coping with social desirability bias: a review. *Eur. J. Soc. Psychol.* 15, 263–280. <https://doi.org/10.1002/ejsp.2420150303>.
- Negeri, Z.F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., et al., 2021. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *BMJ* 375, n2183. <https://doi.org/10.1136/bmj.n2183>.
- Rice, D.B., Thombs, B.D., 2016. Risk of bias from inclusion of currently diagnosed or treated patients in studies of depression screening tool accuracy: a cross-sectional analysis of recently published primary studies and meta-analyses. *PLoS One* 11, e0150067. <https://doi.org/10.1371/journal.pone.0150067>.
- Riley, R.D., Lambert, P.C., Abo-Zaid, G., 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 340, c221. <https://doi.org/10.1136/bmj.c221>.
- Santomauro, D.F., Herrera, A.M.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., et al., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 398, 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7).
- Tasca, G.A., Angus, L., Bonli, R., Drapeau, M., Fitzpatrick, M., Hunsley, J., et al., 2019. Outcome and progress monitoring in psychotherapy: report of a Canadian Psychological Association Task Force. *Can. Psychol.* 60, 165–177. <https://doi.org/10.1037/cap0000181>.
- Thissen, D., Steinberg, L., Wainer, H., 1993. Detection of differential item functioning using the parameters of item response models. In: Holland, P.W., Wainer, H. (Eds.), *Differential Item Functioning*. Lawrence Erlbaum Associates Inc., New York, pp. 67–113.
- Thombs, B.D., Arthurs, E., El-Baalbaki, G., Meijer, A., Ziegelstein, R.C., Steele, R.J., 2011. 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* 343, d4825. <https://doi.org/10.1136/bmj.d4825>.
- World Health Organization, 1992. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. <https://apps.who.int/iris/handle/10665/37958> (accessed 28 February 2024).
- Wu, Y., Levis, B., Sun, Y., He, C., Krishnan, A., Neupane, D., et al., 2021. Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: systematic review and individual participant data meta-analysis. *BMJ* 373, n972. <https://doi.org/10.1136/bmj.n972>.