Depression prevalence using the HADS-D compared to SCID major depression classification: an individual participant data meta-analysis

Brehaut, Eliana, Neupane, Dipika, Levis, Brooke, Wu, Yin, Sun, Ying, Krishnan, Ankur, He, Chen, Bhandari, Parash Mani, Negeri, Zelalem, Riehm, Kira E, Rice, Danielle B, Azar, Marleine, Yan, Xin Wei, Imran, Mahrukh, Chiovitti, Matthew J, Saadat, Nazanin, Cuijpers, Pim, Ioannidis, John PA, Markham, Sarah, Patten, Scott B, Ziegelstein, Roy C, Henry, Melissa, Ismail, Zahinoor, Loiselle, Carmen G, Mitchell, Nicholas D, Tonelli, Marcello, Boruff, Jill T, Kloda, Lorie A, Beraldi, Anna, Braeken, Anna PBM, Carter, Gregory, Clover, Kerrie, Conroy, Ronán M, Cukor, Daniel, da Rocha e Silva, Carlos E, De Souza, Jennifer, Downing, Marina G, Feinstein, Anthony, Ferentinos, Panagiotis P, Fischer, Felix H, Flint, Alastair J, Fujimori, Maiko, Gallagher, Pamela, Goebel, Simone, Jetté, Nathalie, Julião, Miguel, Keller, Monika, Kjærgaard, Marie, Love, Anthony, Löwe, Bernd, Martin-Santos, Rocio, Michopoulos, Ioannis, Navines, Ricard, O'Rourke, Suzanne J, Öztürk, Ahmet, Pintor, Luis, Ponsford, Jennie L, Rooney, Alasdair G, Sánchez-González, Roberto, Schwarzbold, Marcelo L, Sharpe, Michael, Simard, Sébastien, Singer, Susanne, Stone, Jon, Tung, Ka-Yee, Turner, Alyna, Walker, Jane, Walterfang, Mark, White, Jennifer, Benedetti, Andrea and Thombs, Brett D
(2020) Depression prevalence using the HADS-D compared to SCID major depression classification: an individual participant data meta-analysis. Journal of Psychosomatic Research, 139.

Depression Prevalence using the HADS-D Compared to SCID Major Depression Classification: an Individual Participant Data Meta-Analysis

Running head: Estimating Depression Prevalence using the HADS-D and SCID

Authors:

Eliana Brehaut[1]; Dipika Neupane, BPH[1,2]; Brooke Levis, PhD[1-3]; Yin Wu, PhD[1,2,4]; Ying Sun, MPH[1]; Ankur Krishnan, MSc[1]; Chen He, MScPH[1]; Parash Mani Bhandari, BPH[1,2]; Zelalem Negeri, PhD[1,2]; Kira E. Riehm, MSc[1,5]; Danielle B. Rice, MSc[1,6]; Marleine Azar, MSc[1,2]; Xin Wei Yan, MSc[1]; Mahrukh Imran, MScPH[1]; Matthew J. Chiovitti, MISt[1]; Nazanin Saadat, MSc[1]; Pim Cuijpers, PhD[7]; John P. A. Ioannidis, MD[8]; Sarah Markham, PhD[9]; Scott B. Patten, MD[10-12]; Roy C. Ziegelstein, MD[13]; Melissa Henry, PhD[1]; Zahinoor Ismail, MD[14-16]; Carmen G. Loiselle, PhD[1,17-19]; Nicholas D. Mitchell, MD[20,21]; Marcello Tonelli, MD[16]; Jill T. Boruff, MLIS[22]; Lorie A. Kloda, PhD[23]; Anna Beraldi, PhD[24]; Anna P. B. M. Braeken, PhD[25-27]; Gregory Carter, PhD[28,29]; Kerrie Clover, PhD[30]; Ronán M. Conroy, DSc[31]; Daniel Cukor, PhD[32]; Carlos E. da Rocha e Silva, MD[33]; Jennifer De Souza, PhD[34,35]; Marina G. Downing, PhD[36,37]; Anthony Feinstein, PhD[38,39]; Panagiotis P. Ferentinos, MD[40,41]; Felix H. Fischer, PhD[4,42]; Alastair J. Flint, M.B.[38,43]; Maiko Fujimori, PhD[44]; Pamela Gallagher, PhD[45]; Simone Goebel, PhD[46]; Nathalie Jetté, MD[10,14,47]; Miguel Julião, MD, MSc, PhD[48]; Monika Keller, MD, PhD[49]; Marie Kjærgaard, PhD[50,51]; Anthony W. Love, PhD[52]; Bernd Löwe, MD[53]; Rocio Martin-Santos, PhD[54,55]; Ioannis Michopoulos, MD[40]; Ricard Navines, PhD[54,55]; Suzanne J. O'Rourke, PhD[56]; Ahmet Öztürk, MD[57]; Luis Pintor, MD[58,59]; Jennie L. Ponsford, PhD[36,37]; Alasdair G. Rooney, PhD[60,61]; Roberto Sánchez-González, PhD[62-64]; Marcelo L. Schwarzbold, PhD[65]; Michael Sharpe, MD,

FRCPsych[66]; Sébastien Simard, PhD[67-69]; Susanne Singer, PhD[70]; Jon Stone, PhD[71]; Ka-Yee Tung, MBBS[72]; Alyna Turner, PhD[73,74]; Jane Walker, PhD[75]; Mark Walterfang, PhD, MD[76-78]; Jennifer White, PhD[79]; Andrea Benedetti, PhD[2,80,81]; Brett D. Thombs, PhD[1,2,4,6,81-83].

Affiliations:

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada;

[2]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada;

[3]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK;

[4]Department of Psychiatry, McGill University, Montréal, Québec, Canada;

[5]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA;

[6]Department of Psychology, McGill University, Montréal, Québec, Canada;

[7]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, The Netherlands;

[8]Department of Medicine, Department of Epidemiology and Population Health, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA;

[9]Department of Biostatistics and Health Informatics, King's College London, London, UK;

[10]Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada;

[11]Mathison Centre for Mental Health Research & Education, University of Calgary, Calgary, Canada;

[12]Cuthbertson & Fischer Chair in Pediatric Mental Health, University of Calgary, Calgary, Canada;

[13]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA;

[14]Hotchkiss Brain Institute and O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada;

[15]Department of Psychiatry, Clinical Neuroscience and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada;

[16]Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada;

[17]Ingram School of Nursing, McGill University, Montréal, Québec, Canada;

[18]Centre for Nursing Research, Jewish General Hospital, Montréal, Québec, Canada;

[19]Department of Oncology, Faculty of Medicine, McGill University, Montréal, Québec, Canada;

[20]Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada;

[21]Alberta Health Services, Edmonton, Alberta, Canada;

[22]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Quebec, Canada;

[23]Library, Concordia University, Montréal, Québec, Canada;

[24]kbo Lech-Mangfall-Klinik für Psychatrie, Psychotherapie und Psychsomatik, Garmisch-Partenkirchen, Bayern, German;

[25]Department of Radiation Oncology (MAASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands;

[26]Faculty of Psychology, Open University of the Netherlands, Heerlen, The Netherlands;

[27]Department of Health Services Research, CAPHRI School for Public Health and Primary, Maastricht University, Maastricht, The Netherlands;

[28]School of Medicine and Public Health, University of Newcastle, Callaghan NSW, Australia;

[29]Calvary Mater Newcastle, Australia;

[30]Centre for Brain and Mental Health Research, University of Newcastle, Callaghan NSW, Australia;

[31]Royal College of Surgeons in Ireland Division of Population Health Sciences, Dublin, Ireland;

[32]Rogosin Institute, New York, New York, USA;

[33]Clementino Fraga Filho University Hospital, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil;

[34]Birmingham and Solihull Mental Health Foundation Trust, Birmingham, UK;

[35]University of Birmingham, Birmingham, UK;

[36]School of Psychological Sciences, Monash University, Melbourne VIC, Australia;

[37]Monash Epworth Rehabilitation Research Centre, Epworth HealthCare, Melbourne VIC, Australia;

[38]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada;

[39]Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada;

[40]2nd Department of Psychiatry, Attikon General Hospital, National and Kapodistrian University of Athens, Athens, Greece;

[41]Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK;

[42]Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany;

[43]University Health Network, Toronto, Ontario, Canada;

[44]Section of Psychological Science, Division of Health Care Research, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan;

[45]School of Psychology, Dublin City University, Dublin, Ireland;

[46]Department of Clinical Psychology and Psychotherapy, Institute of Psychology, Christian-Albrechts University, Kiel, Germany;

[47]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA;

[48]Equipa Comunitária de Suporte em Cuidados Paliativos de Sintra, Portugal;

[49]Division of Psychooncology, Department of General Internal Medicine and Psychosomatics, University Hospital Heidelberg, Germany;

[50]Endocrinology Research Group, Medical Clinic, University Hospital of North Norway, Norway;

[51]Department of Internal Medicine, Kolding Hospital, Hospital Lillebaelt, Denmark;

[52]Department of Psychology, Victoria University, Victoria, Australia;

[53]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;

[54]Department of Psychiatry and Psychology, Hospital Clinic, IDIBAPS, CIBERSAM, Barcelona, Spain;

[55]Department of Medicine, Institute of Neuroscience, University of Barcelona, Barcelona, Spain;

[56]School of Health in Social Sciences, University of Edinburgh, Edinburgh, Scotland;

[57]Bezmialem Vakif University, Istanbul, Turkey;

[58]Instituto de Investigaciones Biomédicas Augusto Pi i Sunyer (IDIBAPS), Barcelona, Spain;

[59]Consultation Liaison Psychiatry Unit, Hospital Clínico de Barcelona, Barcelona, Spain;

[60]Division of Psychiatry, University of Edinburgh, Edinburgh, UK;

[61]Robert Fergusson Unit, Royal Edinburgh Hospital, NHS Lothian, Edinburgh, UK;

[62]Department of Psychiatry, Institut de Neuropsiquiatria i Addiccions, Centre Emili Mira, Parc de Salut Mar, Barcelona, Spain;

[63]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain;

[64]Centro de Investigación Biomédica En Red de Salud Mental (CIBERSAM), Barcelona, Spain;

[65]Department of Internal Medicine, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil;

[66]Department of Psychological Medicine, University of Oxford, Oxford, UK;

[67]Département des sciences de la santé, Université du Québec à Chicoutimi (UQAC), Québec, Canada;

[68]Centre intersectoriel en santé durable (CISD), Québec, Canada;

[69]Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ), Québec, Canada;

[70]University Medical Centre Mainz, Institute of Medical Biostatistics, Epidemiology and Informatics, Mainz, Germany;

[71]Department of Neurology, University of Edinburgh, Edinburgh, UK;

[72]Kwai Chung Hospital, Hong Kong SAR, China;

[73]Faculty of Health and Medicine, School of Medicine and Public Health, University of Newcastle, Callaghan NSW, Australia;

[74]Deakin University, IMPACT Strategic Research Centre and School of Medicine, Barwon Health, Geelong VIC, Australia;

[75]Department of Psychiatry, University of Oxford, Oxford, UK;

[76]Neuropsychiatry Unit, Royal Melbourne Hospital, Melbourne, Australia;

[77]Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia;

[78]Florey Institute of Neuroscience and Mental Health, Melbourne, Australia;

[79]Department of Physiotherapy, School of Primary and Allied Health Care, Monash University, Melbourne, Australia;

[80]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada;

[81]Department of Medicine, McGill University, Montréal, Québec, Canada;

[82]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada;

[83]Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada.

Word Count: 3, 649

Addresses for Correspondence:

Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Ste. Catherine Road; Montreal, Quebec, Canada H3T 1E4; Tel (514) 340-8222 ext. 25112; E-mail: brett.thombs@mcgill.ca

Author Emails

Full Name	Author Email
Eliana Brehaut	eliana.brehaut@mail.mcgill.ca
Dipika Neupane	dipika.neupane@mail.mcgill.ca
Brooke Levis	brooke.levis@mail.mcgill.ca
Yin Wu	yin.wu@mail.mcgill.ca
Ying Sun	ying.sun2@mail.mcgill.ca
Ankur Krishnan	ankur.krishnan@mail.mcgill.ca
Chen He	chen.he3@mail.mcgill.ca
Parash Mani Bhandari	parash.bhandari@mail.mcgill.ca
Zelalem Negeri	zelalem.negeri@mail.mcgill.ca
Kira E. Riehm	kirariehm@gmail.com
Danielle B. Rice	danielle.rice@mail.mcgill.ca
Marleine Azar	marleine.azar@mail.mcgill.ca
Xin Wei Yan	xin.yan@mail.mcgill.ca
Mahrukh Imran	mahrukh.imran@mail.mcgill.ca
Matthew J. Chiovitti	matthew.chiovitti@mail.mcgill.ca
Nazanin Saadat	nazanin.saadat@mail.mcgill.ca
Pim Cuijpers	p.cuijpers@vu.nl
John P. A. Ioannidis	jioannid@stanford.edu
Sarah Markham	sarah.markham@kcl.ac.uk
Scott B. Patten	patten@ucalgary.ca
Roy C. Ziegelstein	rziegel2@jhmi.edu

Melissa Henry	melissa.henry@mcgill.ca
Zahinoor Ismail	ismailz@ucalgary.ca
Carmen G. Loiselle	carmen.g.loiselle@mcgill.ca
Nicholas D. Mitchell	ndm@ualberta.ca
Marcello Tonelli	cello@ucalgary.ca
Jill T. Boruff	jill.boruff@mcgill.ca
Lorie A. Kloda	lorie.kloda@concordia.ca
Anna Beraldi	Anna.Beraldi@psychiatrie-gap.de
Anna P. B. M. Braeken	v.braeken@maastrichtuniversity.nl
Gregory Carter	Gregory.Carter@newcastle.edu.au
Kerrie Clover	Kerrie.Clover@calvarymater.org.au
Ronán M. Conroy	rconroy@rcsi.com
Daniel Cukor	dac9227@nyp.org
Carlos E. da Rocha e Silva	ceduardodarochaesilva@gmail.com
Jennifer De Souza	jennifer.desouza@nhs.net
Marina G. Downing	marina.downing@monash.edu
Anthony Feinstein	ant.feinstein@utoronto.ca
Panagiotis P. Ferentinos	pferentinos@med.uoa.gr
Felix H. Fischer	felix.fischer@charite.de
Alastair J. Flint	Alastair.Flint@uhn.ca
Maiko Fujimori	mfujimor@ncc.go.jp
Pamela Gallagher	pamela.gallagher@dcu.ie
Simone Goebel	simone.goebel@web.de
Nathalie Jetté	nathalie.jette@mssm.edu
Miguel Julião	migueljuliao@gmail.com
Monika Keller	monika.keller@uni-heidelberg.de
Marie Kjærgaard	marianguaq@gmail.com
Anthony W. Love	anthony.love@vu.edu.au
Bernd Löwe	b.loewe@uke.de
Rocio Martin-Santos	rmsantos@clinic.cat
Ioannis Michopoulos	yanmih@yahoo.com
Ricard Navines	rnavines@clinic.cat
Suzanne J. O'Rourke	Suzanne.O'Rourke@ed.ac.uk
Ahmet Öztürk	doktorahmet23@hotmail.com
Luis Pintor	LPINTOR@clinic.cat
Jennie L. Ponsford	jennie.ponsford@monash.edu
Alasdair G. Rooney	ally.rooney@ed.ac.uk
Roberto Sánchez-González	rsanchezgonzalez@psmar.cat
Marcelo L. Schwarzbold	schwlib@gmail.com

Michael Sharpe	michael.sharpe@psych.ox.ac.uk
Sébastien Simard	Sebastien1_Simard@uqac.ca
Susanne Singer	singers@uni-mainz.de
Jon Stone	jon.stone@ed.ac.uk
Ka-Yee Tung	tky028a@ha.org.hk
Alyna Turner	a.turner@deakin.edu.au
Jane Walker	jane.walker@psych.ox.ac.uk
Mark Walterfang	mark.walterfang@mh.org.au
Jennifer White	jenesiswhite@gmail.com
Andrea Benedetti	andrea.benedetti@mcgill.ca
Brett D. Thombs	brett.thombs@mcgill.ca

159

160

ABSTRACT

Objectives: Validated diagnostic interviews are required to classify depression status and estimate prevalence of disorder, but screening tools are often used instead. We used individual participant data meta-analysis to compare prevalence based on standard Hospital Anxiety and Depression Scale – depression subscale (HADS-D) cutoffs of ≥ 8 and ≥ 11 versus Structured Clinical Interview for DSM (SCID) major depression and determined if an alternative HADS-D cutoff could more accurately estimate prevalence.

Methods: We searched Medline, Medline In-Process & Other Non-Indexed Citations via Ovid, PsycINFO, and Web of Science (inception-July 11, 2016) for studies comparing HADS-D scores to SCID major depression status. Pooled prevalence and pooled differences in prevalence for HADS-D cutoffs versus SCID major depression were estimated.

Results: 6,005 participants (689 SCID major depression cases) from 41 primary studies were included. Pooled prevalence was 24.5% (95% Confidence Interval (CI): 20.5%, 29.0%) for HADS-D ≥ 8, 10.7% (95% CI: 8.3%, 13.8%) for HADS-D ≥ 11, and 11.6% (95% CI: 9.2%, 14.6%) for SCID major depression. HADS-D ≥ 11 was closest to SCID major depression prevalence, but the 95% prediction interval for the difference that could be expected for HADS-D ≥ 11 versus SCID in a new study was -21.1% to 19.5%.

Conclusions: HADS-D ≥ 8 substantially overestimates depression prevalence. Of all possible cutoff thresholds, HADS-D ≥ 11 was closest to the SCID, but there was substantial heterogeneity in the difference between HADS-D ≥ 11 and SCID-based estimates. HADS-D should not be used as a substitute for a validated diagnostic interview.

Key Words: depression, Hospital Anxiety and Depression Scale, individual participant data, meta-analysis, screening tools

INTRODUCTION

Accurately measuring depression prevalence in different populations is important to understand disease burden, interpret research on etiology, and utilize healthcare resources as efficiently as possible (Rogan & Gladen, 1978). In mental health research, diagnostic interviews are required for diagnosis of major depression (First, Spitzer, Gibbon, & Williams, 1995; Wittchen, 1994). These interviews, however, are costly to administer, especially in large groups, due to the time and trained personnel required to conduct them properly. Therefore, self-report screening questionnaires are sometimes used as an inexpensive alternative to evaluate depression prevalence, with the percentage of patients scoring above a cutoff threshold being described as the prevalence of depression (Levis et al., 2019; Thombs, Kwakkenbos, Levis, & Benedetti, 2018). Screening tool cutoffs, however, are typically set to cast a wide net and identify many more individuals for further assessment than will meet diagnostic criteria. Thus, commonly used screening tools tend to overestimate depression prevalence, sometimes substantially (Thombs et al., 2018).

A previous study used an individual participant data meta-analysis (IPDMA) approach to compare prevalence based on a depression screening tool with prevalence based on a validated diagnostic interview. That meta-analysis examined prevalence based on the Patient Health Questionnaire-9 (PHQ-9) using the standard cutoff of ≥ 10 compared to prevalence based on the Structured Clinical Interview for the DSM (SCID) among 9,242 participants from 44 primary studies (Levis et al., 2020). Compared to the SCID, PHQ-9 ≥ 10 overestimated prevalence by 11.9%; across included studies, the mean and median ratio of PHQ-9 prevalence to SCID-based prevalence were 2.5 and 1.9. In that study, the authors attempted to identify a PHQ-9 cutoff that

206 would match SCID-based prevalence, but heterogeneity was too high to generate consistently
207 accurate estimates in individual studies for any PHQ-9 cutoff.

208 The Hospital Anxiety and Depression Scale (HADS) is a self-report screening
209 questionnaire designed to be administered to non-psychiatric medical patients. It includes 14
210 items, with 7 assessing symptoms of depression (HADS-D) and 7 assessing symptoms of anxiety
211 (HADS-A) over the past week. To avoid overlap with physical illness, the HADS-D does not
212 include symptoms common to both physical and mental disorders, such as insomnia, loss of
213 appetite, or fatigue. Cutoff thresholds of ≥ 8 and ≥ 11 on the HADS-D are traditionally used as
214 standard cutoffs for identifying people who may have depression (Zigmond & Snaith, 1983).
215 Although not designed for this purpose, the HADS-D is also frequently used to report depression
216 prevalence in primary research studies. A review of recent studies listed in PubMed (2018-2019)
217 identified 32 studies that reported "prevalence" of depression based on a HADS-D cutoff, with ≥
218 8 and ≥ 11 used in 66% and 16% of the studies, respectively (see supplementary material
219 eMethods 1 and eTable 1).

220 Although other screening tools and commonly used cutoffs have been shown to
221 overestimate depression prevalence, it is not clear whether this would be the case with the
222 HADS-D. A previous study that investigated prevalence of major depression among survivors of
223 acute myocardial infarction found a prevalence of 20% (10,785 participants, 8 studies) using
224 structured interviews, compared to 16% using a HADS-D cutoff of ≥ 8 (863 participants, 4
225 studies), and 7% using ≥ 11 (830 participants, 4 studies) (Thombs et al., 2006). This was a
226 between-study comparison, however, and no included studies administered both the HADS-D
227 and a validated diagnostic interview.

The objectives of the present study were to use an IPDMA approach to (1) compare pooled prevalence based on HADS-D cutoffs of ≥ 8 and ≥ 11 with major depression prevalence based on the SCID; and (2) use a prevalence-matching approach to determine if any cutoff threshold on the HADS-D matches prevalence based on the SCID with sufficiently low heterogeneity that it could be used to accurately measure depression prevalence in future studies.

METHODS

This study used a subset of data collected for an IPDMA of the diagnostic accuracy of the HADS-D for screening to detect major depression. Detailed methods of the IPDMA were registered in PROSPERO (CRD42015016761), and a protocol was published (Thombs et al., 2016). The present analysis was not included in the original IPDMA protocol, which focused only on diagnostic accuracy. A protocol for the present study was published on the Open Science Framework prior to initiating the study (https://osf.io/n5a3e/).

Study Selection

In the main IPDMA, datasets from studies in any language were eligible for inclusion if (1) they included HADS-D scores; (2) they included diagnostic classifications for current Major Depressive Episode (MDE) or Major Depressive Disorder (MDD) based on the Diagnostic and Statistical Manual (DSM) or International Classification of Diseases criteria, using a validated semi-structured or fully structured interview; (3) the HADS-D 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; (4) participants were ≥ 18 years and not recruited from youth or school-based settings, since the main IPDMA was designed for adult screening, and although there are some adults in schools, the pathways for identification and management are likely very different from other adult settings; and (5) 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 selection of eligible participants.

For the present study, we included only primary studies that based diagnoses on the SCID (First et al., 1995). 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. The reason for including only studies that used the SCID is that in recent analyses using three large IPDMA databases (Levis et al., 2018, Levis et al., 2019, Wu et al., 2020) we found that, compared to semi-structured interviews, fully structured interviews, which are designed for administration by lay interviewers, may identify more patients with low-level symptoms as depressed but fewer patients with high-level symptoms. These results are consistent with the idea that semi-structured interviews most closely replicate clinical interviews done by trained professionals, whereas fully structured interviews are less rigorous reference standards; they are less resource-intensive options that can be administered by research staff without diagnostic skills but may misclassify major depression in substantial numbers of patients. An important feature of the SCID is that it allows the interviewer to probe to determine whether a symptom is merely a manifestation of a physical illness. In the HADS IPDMA database, the SCID was the most commonly used semi-structured interview; out of 83 studies, 45 used semi-structured interviews, and 41 of the 45 used the SCID. In sensitivity analyses, we also included the 4 studies from the IPDMA database that used semi-structured interviews other than the SCID.

Data Sources and Searches

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations via Ovid, PsycINFO, and Web of Science from inception to July 11, 2016, using a peer-reviewed search strategy (McGowan et al., 2016) (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 tracking search results.

Two investigators independently reviewed studies by title and abstract for eligibility. If either deemed a study 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 HADS-D scores and major depression classification status. 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 HADS-D ≥ 8 and ≥ 11 Prevalence with SCID Major Depression Prevalence

For each primary study, we estimated 7 values: (1) the percentage of participants who scored ≥ 8 on the HADS-D, (2) the percentage of participants who scored ≥ 11 on the HADS-D, (3) the percentage of participants classified as having major depression based on the SCID, (4) the difference between HADS-D ≥ 8 percentage and SCID percentage, (5) the ratio for HADS-D ≥ 8 percentage versus SCID percentage, and the corresponding (6) difference and (7) ratio for HADS-D ≥ 11 versus the SCID. Then, across all studies, we pooled prevalence for HADS-D ≥ 8, HADS-D ≥ 11, and SCID, and we pooled the HADS-D versus SCID differences in prevalence from each study.

Prevalence Matching

To identify which HADS-D cutoff best matches SCID-based prevalence, we estimated the pooled difference in prevalence for each possible HADS-D cutoff compared to the SCID. The HADS-D cutoff with the smallest pooled difference was chosen to be the "prevalence-matched cutoff." Then, for each included study, we estimated the difference and ratio in prevalence based on the prevalence-matched cutoff versus SCID major depression. We determined the mean and median absolute difference and the range of differences across all studies. To illustrate the range of difference values that would be expected if a new study were to compare prevalence based on the prevalence-matched cutoff to prevalence based on the SCID, we estimated a 95% prediction interval for the difference.

All meta-analyses were conducted in R (R version R 3.4.1 and R Studio version 1.0.143) using the lme4 package. To estimate pooled prevalence values, generalized linear mixed-effects models with a logit link function were fit using the glmer function. To estimate pooled difference values, linear mixed-effects models were fit using the lmer function. To account for correlation between subjects within the same primary study, random intercepts were fit for each primary

study. To quantify heterogeneity, for each analysis, we calculated τ^2, which is the estimate of between-study variance, and I^2, which quantifies the proportion of total variability due to the between-study heterogeneity.

We conducted two sets of post hoc analyses. First, some studies had high depression prevalence. Thus, to test whether differences in prevalence between the HADS-D and SCID might be influenced by heterogeneity in depression levels, we repeated the main analysis of prevalence excluding studies with SCID-based prevalence ≥ 20.0%. Second, we assessed whether differences in prevalence for the prevalence-matched cutoff and SCID were associated with study or patient characteristics. To do this, we fit an additional linear mixed-effects model for pooled prevalence difference, including age, sex, country human development index category ("very high" [reference group] or "high", based on the United Nation's Human Development Index for the year of publication), recruitment setting category (nonmedical care, inpatient care [reference group], outpatient care, or mixed inpatient and outpatient care), and sample size as fixed-effect covariates. For this analysis, we excluded 520 participants (8.7%) who were missing age or sex data. We repeated all analyses including 4 studies that used semi-structured interviews other than the SCID.

RESULTS

The initial search for the main IPDMA found 10,015 unique titles and abstracts for potential eligibility. Of these, we excluded 9,584 studies after reviewing titles and abstracts and 238 studies after full-text review. There were 193 eligible studies using data from 133 unique samples from which 75 (56.4%) contributed individual participant data. Authors also contributed data from 8 unpublished studies, resulting in a total of 83 datasets. For our main analyses, we excluded 42 studies that used diagnostic interviews other than the SCID. In total, the main

analyses included 41 primary studies involving 6,005 participants (689 SCID major depression cases; 11.5%; Figure 1). Of 58 eligible primary studies with unique samples that did not contribute individual participant data, 26 used the SCID (3,096 participants). Thus, the main analyses in the present study included 61.2% of eligible studies that used the SCID (41 of 67) and 66.0% of eligible participants (6,005 of 9,101). See Table 1 for characteristics of each included study.

There were 4 additional studies that used semi-structured diagnostic interviews other than the SCID (635 participants; 65 major depression cases; 10.2%), which we included in sensitivity analyses. Two of these studies used the Monash Interview for Liaison Psychiatry, one used the Schedule for Affective Disorders and Schizophrenia, and one used the Schedules for Clinical Assessment in Neuropsychiatry. Thus, these analyses included 45 primary studies (6,640 participants; 754 major depression cases; 11.4%; Table 1).

Objective 1: Comparison of HADS-D ≥ 8, HADS-D ≥ 11 and SCID Major Depression Prevalence

Pooled Prevalence

The results for individual studies are presented in Table 1. For the 41 studies included in our main analyses, the percentage of participants who scored ≥ 8 on the HADS-D ranged from 4.2% to 82.7%, with a pooled prevalence of 24.5% (95% CI: 20.5% to 29.0%, τ^2:0.49, I^2: 97.2%). The percentage of participants who scored ≥ 11 on the HADS-D ranged from 0.3% to 74.7%, with a pooled prevalence of 10.7% (95% CI: 8.3% to 13.8%, τ^2: 0.71, I^2: 97.1%). The percentage of participants classified as having SCID major depression ranged from 0% to 50.0%, with a pooled prevalence of 11.6% (95% CI: 9.2% to 14.6%, τ^2: 0.6, I^2:97.1%).

Excluding 8 studies (552 participants; 185 major depression cases; 33.5%) with SCID-based prevalence of 20.0% or over, prevalence based on the HADS-D ≥ 8 was 21.8% (95% CI: 18.4% to 25.6%, τ^2: 0.31, I^2= 96.4). Prevalence based on the HADS-D ≥ 11 was 9.2% (95% CI: 7.3% to 11.6%, τ^2: 0.41, I^2= 96.0). Prevalence based on the SCID was 8.9% (95% CI: 7.6% to 10.4%, τ^2: 0.14, I^2= 94.7).

Results were similar when the 4 studies using interviews other than the SCID were included.

Pooled Difference and Ratio

The difference between HADS-D ≥ 8 and SCID-based prevalence in the main analyses ranged from -9.5% to 41.3%, and the pooled difference was 12.4% (95% CI: 8.8% to 16%, τ^2: 0.01, I^2: 97.2%). The difference between HADS-D ≥ 11 and SCID-based prevalence ranged from -31.0% to 33.3%, and the pooled difference was -0.8% (95% CI: -4.1% to 2.5%, τ^2: 0.01, I^2: 97.2%).

Results were similar in the sensitivity analyses. Pooled difference for HADS-D ≥ 8 was 11.9% (95% CI: 8.6% to 15.2%, τ^2: 0.01, I^2: 97.4%), and pooled difference for HADS-D ≥ 11 was -1.0% (95% CI: -4.0% to 2.0%, τ^2: 0.01, I^2: 97.5%)). The ratio of HADS-D ≥ 8 prevalence to SCID major depression prevalence ranged from 0.4 to 7.7 times (mean: 2.6 times; median: 2 times). The ratio of HADS-D ≥ 11 prevalence to SCID major depression prevalence ranged from 0 to 3.8 times (mean: 1.2 times; median: 0.8 times).

Mean Ratio and Difference in Individual Studies

In the main analyses, the mean ratio of HADS-D to SCID-based prevalence was 0.73 times for the 3 studies with HADS-D ≥ 8-based prevalence < 10.0% (mean difference: -2.7%), 1.8 times for the 7 studies with HADS-D ≥ 8-based prevalence between 11.0% and 19.0% (mean

difference: 6.1%), and 2.9 times for the 31 studies with HADS-D ≥ 8-based prevalence of 20.0% or greater (mean difference: 15.2%). The mean ratio was 0.7 times for the 19 studies with HADS-D ≥ 11-based prevalence < 10.0% (mean difference: -4.4%), 1.5 times for the 15 studies with HADS-D ≥ 11-based prevalence between 11.0% and 19.0% (mean difference: -1.3), and 2 times for the 7 studies with HADS-D ≥ 11-based prevalence of 20.0% or greater (mean difference: 9.8%). Results were similar when the 4 additional studies were included.

Objective 2: Prevalence Matching

Of all possible HADS-D cutoffs, ≥ 11 produced the pooled prevalence estimate that most closely matched SCID major depression prevalence (HADS-D ≥ 11: 10.7%, SCID: 11.6%) (Figure 2). This cutoff underestimated depression prevalence compared to the SCID, but only slightly (pooled difference: -0.8%). HADS-D ≥ 10 produced a pooled prevalence of 14.7% (pooled difference: 3.1%), and HADS-D ≥ 12 a pooled prevalence of 7.9% (pooled difference: -3.7%). The mean absolute difference between HADS-D ≥ 11 and SCID was 8.2%, and the median absolute difference was 6.7%. The 95% prediction interval for the difference between HADS-D ≥ 11 and SCID-based prevalence was -21.1% to 19.5%. Results were similar in sensitivity analyses. In the post-hoc analysis, no participant or study characteristics were significantly associated with differences in prevalence for the HADS-D prevalence-match cutoff compared to the SCID.

DISCUSSION

Previous research has demonstrated that there may be substantial differences between screening tools and diagnostic tools in estimating depression prevalence (Levis et al., 2020, Thombs et al., 2018, Levis et al., 2019). In the present study, we found that the most commonly used HADS-D cutoff threshold for reporting depression prevalence of ≥ 8 overestimated

depression prevalence (24.5%) substantially compared to SCID major depression prevalence (11.6%). A HADS-D cutoff of ≥ 11 underestimated prevalence only slightly in aggregate compared to the SCID (10.7%), but heterogeneity in the difference between HADS-D ≥ 11 and SCID-based estimates in individual studies was high. The 95% prediction interval for difference between HADS-D ≥ 11 and SCID-based prevalence ranged from approximately -20% to 20%, which suggests that any single new study using HADS-D ≥ 11 may over or underestimate depression prevalence by up to 20%.

Results from the present study are partially consistent with what might be expected theoretically when comparing screening tools and diagnostic tools (Thombs et al., 2018). Since screening tools are designed to cast a wide net and identify individuals who might be depressed, they generally tend to overestimate depression prevalence when compared to diagnostic interviews, which are designed to determine who meets diagnostic criteria. This was indeed the case in our study for results from the HADS-D ≥ 8, which were in line with those from a previous study that found that the PHQ-9 similarly overestimated prevalence (Levis et al., 2020). A finding that was unique to the present study was that estimates based on another commonly used cutoff threshold, HADS ≥ 11, were in aggregate consistent with major depression prevalence based on the SCID. The findings from the present study differed from those in a previous synthesis of evidence from post-acute myocardial infarction patients in which depression prevalence estimates based on HADS-D ≥ 8 and ≥ 11 were both lower than estimates based on structured interviews (Thombs et al., 2006). This discrepancy may be due to the specific clinical population eligible for the review or because none of the studies included in that review administered both the HADS-D and a structured interview to the same group of individuals.

Identifying a HADS-D cutoff that consistently matches the SCID would allow researchers to use screening questionnaires rather than diagnostic interviews for prevalence estimation, thus conserving time and resources. However, when we used a prevalence-matching approach and identified the closest HADS-D cutoff (≥ 11) to the SCID, although the aggregate estimates were similar, heterogeneity between studies was too high to suggest that HADS-D ≥ 11 would accurately estimate prevalence in any particular future study. In fact, it may substantially under or overestimate prevalence in individual studies.

Researchers often describe the proportion of individuals scoring at or above a cutoff threshold as prevalence of "depressive symptoms" or "clinically significant depressive symptoms" rather than prevalence of "depression". However, this does not resolve the problem. There is no evidence that impairment becomes meaningful at or above these thresholds, which have been set for the purpose of screening, and not for impairment delineation. While individuals scoring above these thresholds have greater impairment on average than those scoring below the threshold, this would be the case for any threshold that is set. Reporting the proportion of individuals scoring above a threshold may be useful for comparisons between samples. However, it should not be characterized as "prevalence" or as the percentage of individuals who have "symptoms of depression" versus no symptoms.

Ideally, semi-structured interviews should be used for prevalence estimation, since they provide patient-specific details that help interviewers determine whether the diagnostic criteria for depression are met. They also most closely replicate full assessments done by trained professionals (Wu et al., 2020). However, these interviews are not always feasible as they are time-intensive compared to screening questionnaires. Diagnostic interviews also require trained research staff or mental health professionals to conduct them properly. Hiring clinicians or

training research staff to do this can be costly and time-consuming, especially when assessing large numbers of study participants. When determining which diagnostic interview to use, researchers should consider the advantages and disadvantages of each, including performance, cost, and required training (Wu et. al., 2020). When publishing studies, researchers should discuss their reasons for selecting a particular interview, as well as the implications of their selection.

To our knowledge, this is the first study to synthesize evidence and directly compare depression prevalence based on HADS-D scores versus the SCID. Strengths of this study are that we examined data from 41 primary research studies including 6,005 participants, and that we directly compared status based on HADS-D scores to status based on a validated diagnostic interview. A limitation is that we did not incorporate data from 39% of eligible studies that used the SCID (26 of 61) and 34% of eligible participants (3,096 of 9,101), since they did not provide individual participant data. Furthermore, since not all studies described the qualifications of the individuals administering the SCID, it is possible that interviewer skill-level contributed to heterogeneity. Since the objective of our study was to determine how accurate the HADS-D is for estimating depression prevalence, we did not evaluate whether the correct individuals were identified; that is beyond the scope of this study. Since diagnostic criteria for major depression are for symptoms experienced in the last two weeks, we ensured that all studies administered the HADS-D and SCID within two-weeks of each other. However, studies may not have administered the HADS-D and SCID on the same day. This may have contributed to variability in responses to the SCID and the HADS-D, but it would not be expected to contribute to bias. We included studies where diagnoses were based on DSM or ICD criteria, but only one study used ICD (De Souza et. al., 2009). This study did not use the SCID and was included only in

sensitivity analyses. Finally, this study considered only the HADS-D, which is one screening tool out of many that are commonly used in clinical practice. As shown in this study, the degree to which the use of screening tools may accurately estimate prevalence depends on the specific screening tool and cutoff threshold used.

In conclusion, we found that the standard HADS-D cutoff of ≥ 8, which is most commonly used by researchers to estimate depression prevalence, resulted in overestimation when compared to the SCID. The other standard screening cutoff of ≥ 11 most closely matched SCID prevalence, but heterogeneity in the difference between HADS-D and SCID-based estimates in individual studies was high and not associated with study or participant characteristics. Findings are consistent with evidence demonstrating that depression screening tools should not be used for diagnostic purposes. Studies should only report prevalence of depression if they used a validated diagnostic interview designed for case classification. Clinicians and researchers should be aware that the prevalence of depression reported in studies using depression screening tools may not be accurate.

Contributors:

- BLevis, PC, JPAI, SM, SBP, RCZ (DEPRESSD Steering Committee Members), MH, ZI, CGL, NDM, MT (DEPRESSD Knowledge Users), ABenedetti, and BDT (DEPRESSD Directors) were responsible for the conception, design and oversight of the main IPDMA project of which the present study is a part.
- EB, DN, BLevis, JPAI, ABenedetti, and BDT were responsible for the conception and design of the present study
- JTB and LAK designed and conducted database searches to identify eligible studies.
- ABeraldi, APBMB, GC, KC, RMC, DC, CEdReS, JDS, MGD, AF, PPF, FHF, AJF, MF, PG, SG, NJ, MJ, MKeller, MKjærgaard, AWL, BLöwe, RMS, IM, RN, SJO, AÖ, LP, JLP, AGR, RSG, MLS, MS, SSimard, SSinger, JS, KYT, AT, JWalker, MW, and JWhite contributed primary datasets that were included in this study.
- EB, DN, BLevis, YW, YS, AK, CH, PMB, ZN, KER, DBR, MA, XWY, MI, MJC, NS and BDT contributed to data extraction and coding for the meta-analysis.
- EB, DN, BLevis, ABenedetti, and BDT contributed to the data analysis and interpretation.
- EB, DN, BLevis, YW, and BDT contributed to drafting the manuscript.
- All authors provided a critical review and approved the final manuscript. ABenedetti and BDT are the guarantors; they had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analyses.

Declaration of Competing Interest:

All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf and declare that: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years with the following exceptions: (1) Dr. Ismail declares that he has received personal fees from Avanir, Janssen, Lundbeck, Otsuka, Sunovion, outside the submitted work. (2) Dr. Tonelli declares that he has received a grant from Merck Canada, outside the submitted work. (3) Dr. Feinstein reports that he received speaker's honorariums from Biogen, Sanofi-Genzyme, Merck-Serono, Novartis, Roche, and is on the advisory board for Akili Interactive, outside the submitted work; He has also received royalties from the Cambridge University Press for the Clinical Neuropsychiatry of Multiple Sclerosis, 2nd Edition. (4) Dr. Löwe declares that the primary study by Löwe et al. was supported by unrestricted educational grants from Pfizer, Germany. (5) Dr. Stone declares that he has received personal fees from UptoDate, outside the submitted work. No other relationships or activities that could appear to have influenced the submitted work.

Roles of the Funding Source:

This study was funded by the Canadian Institutes of Health Research (CIHR, KRS-144045 & PCG 155468). Ms. Neupane was supported by a G.R. Caverhill Fellowship from the Faculty of Medicine, McGill University. Drs. Levis and Wu were supported by Fonds de recherche du Québec - Santé (FRQS) Postdoctoral Training Fellowships. Mr. Bhandari was supported by a studentship from the Research Institute of the McGill University Health Centre. Ms. Rice was supported by a Vanier Canada Graduate Scholarship. Dr. Patten was supported by a Senior Health Scholar award from Alberta Innovates, Health Solutions. The primary study by Scott et al. was supported by the Cumming School of Medicine and Alberta Health Services through the Calgary Health Trust, and funding from the Hotchkiss Brain Institute. The primary study by Amoozegar et al. was supported by the Alberta Health Services, the University of Calgary Faculty of Medicine, and the Hotchkiss Brain Institute. 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 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 Souza et al. was supported by Birmingham and Solihull Mental Health Foundation Trust. The primary study by Honarmand et al. was supported by a grant from the Multiple Sclerosis Society of Canada. The primary study by Fischer et al. was supported as part of the RECODEHF study by the German Federal Ministry of Education and Research (01GY1150). 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 Keller et al. was supported by the Medical Faculty of the University of Heidelberg (grant no. 175/2000). The primary study by Love et al. (2004) was supported by the Kathleen Cuningham 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. The primary study by Löwe et al. was supported by the medical faculty of the University of Heidelberg, Germany (Project 121/2000). The primary study by Navines et al. was supported in part by the Spanish grants from the Fondo de Investigación en Salud, Instituto de Salud Carlos III (EO PI08/90869 and PSIGEN-VHC Study: FIS-E08/00268) and the support of FEDER (one way to make Europe). The primary study by O'Rourke et al. was supported by the Scottish Home and Health Department, Stroke Association, and Medical Research Council. The primary study by Sanchez-Gistau et al. was supported by a grant from the Ministry of Health of Spain (PI040418) and in part by Catalonia Government, DURSI 2009SGR1119. The primary study by Gould et al. was supported by the Transport Accident Commission Grant. The primary study by Rooney et al. was supported by the NHS Lothian Neuro-Oncology Endowment Fund. The primary study by Schwarzbold et al. was supported by PRONEX Program (NENASC Project) and PPSUS Program of Fundaçao de Amparo a esquisa e Inovacao do Estado de Santa Catarina (FAPESC)

and the National Science and Technology Institute for Translational Medicine (INCT-TM). 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 Stone et al. was supported by the Medical Research Council, UK and Chest Heart and Stroke, Scotland. The primary study by Turner et al. was supported by a bequest from Jennie Thomas through Hunter Medical Research Institute. The primary study by Walterfang et al. was supported by Melbourne Health. Drs. Benedetti and Thombs were supported by FRQS researcher salary awards. 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.

Acknowledgements

We thank Dr. Linda Kwakkenbos for helping with translation and coding throughout the project.

REFERENCES

Rogan, W. J., & Gladen, B. (1978). Estimating prevalence from the results of a screening test. *American Journal of Epidemiology, 107*(1), 71-76. DOI: https://doi.org/10.1093/oxfordjournals.aje.a112510.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1995). Structured clinical interview for DSM-IV axis I disorders. New York: New York State Psychiatric Institute, 9(2), 83-91. DOI: https://doi.org/10.1521/pedi.1995.9.2.83

Wittchen, H. (1994). Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): A critical review. *Journal of Psychiatric Research,* 28(1), 57-84. DOI: 10.1016/0022-3956(94)90036-1.

Levis, B., Yan, X. W., He, C., Sun, Y., Benedetti, A., & Thombs, B. D. (2019). Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: A meta-research review. *BMC Medicine,* 17(1), 65-019. DOI: 10.1186/s12916-019-1297-6.

Thombs, B. D., Kwakkenbos, L., Levis, A. W., & Benedetti, A. (2018). Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. *CMAJ : Canadian Medical Association Journal = Journal De L'Association Medicale Canadienne,* 190(2), E44-E49. DOI: doi: 10.1503/cmaj.170691.

Levis, B., et. al. (2020). Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: Individual participant data meta-analysis. *J. Clin. Epidemiol*. DOI:

10.1016/j.jclinepi.2020.02.002.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica,* 67(6), 361-370. DOI: 10.1111/j.1600-0447.1983.tb09716.x.

Thombs, B. D., Bass, E. B., Ford, D. E., Stewart, K. J., Tsilidis, K. K., Patel, U., et al. (2006). Prevalence of depression in survivors of acute myocardial infarction. *Journal of General Internal Medicine,* 21(1), 30-38. DOI: 10.1111/j.1525-1497.2005.00269.x.

Thombs, B. D., Benedetti, A., Kloda, L. A., Levis, B., Azar, M., Riehm, K. E., et al. (2016). Diagnostic accuracy of the depression subscale of the hospital anxiety and depression scale (HADS-D) for detecting major depression: Protocol for a systematic review and individual patient data meta-analyses. *BMJ Open,* 6(4), e011913-2016. DOI: 10.1136/bmjopen-2016-011913.

Levis, B, Benedetti, A, Riehm, KE, Saadat, N, Levis, AW, Azar, M, et al (2018). Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *Br J Psychiatry.* 212(6), 377-385. DOI: 10.1192/bjp.2018.54.

Levis, B., McMillan, D., Sun, Y, He, C, Rice, DB, Krishnan, A, et al (2019). 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(4), e1803. DOI: 10.1002/mpr.1803.

Wu, Y., Levis, B., Sun, Y., Krishnan, A., He, C., Riehm, K. E., et al. (2020). Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic

636 interviews controlling for hospital anxiety and depression scale - depression subscale scores:
637 An individual participant data meta-analysis of 73 primary studies. *J Psychosom Res*, 129,
638 109892. DOI: 10.1016/j.jpsychores.2019.109892.

639 McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016).
640 PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin*
641 *Epidemiol*, 75, 40-46. Retrieved from NLM; PRESS Peer Review of Electronic Search
642 Strategies: 2015 Guideline Statement database. DOI: 10.1016/j.jclinepi.2016.01.021.

Table 1. Characteristics of included studies.

Author, year	Country	Population	N total	N (%) Major Depression	Mean Age	% Female	N (%) HADS-D ≥ 8	% Difference: HADS-D ≥ 8 - Major Depression	Ratio: HADS-D ≥ 8 / Major Depression	N (%) HADS-D ≥ 11	% Difference: HADS-D ≥ 11 - Major Depression	Ratio: HADS-D ≥11 / Major Depression
Studies from IPDMA that used the SCID and were included in main analyses												
Akechi, 2006	Japan	Outpatients with cancer in palliative care	223	17 (8.0%)	61.1	65.0%	97 (43.0%)	35.9%	5.7	43 (19.0%)	11.7%	2.5
Amoozegar, 2017	Canada	Patients with migraines	102	51 (50.0%)	42.5	81.4%	53 (52.0%)	2.0%	1.0	32 (31.0%)	-18.6%	0.6
Beraldi, 2014	Germany	Patients of haemato-oncology	120	10 (8.0%)	52.1	32.5%	32 (27.0%)	18.3%	3.2	16 (13.0%)	5.0%	1.6
Braeken, 2010	Netherlands	Dutch cancer patients in radiotherapy	13	1 (8.0%)	69.4	NR	4 (31.0%)	23.1%	4.0	2 (15.0%)	7.7%	2.0
Cukor, 2008	USA	Patients with end-stage renal disease	70	14 (20.0%)	53.3	52.9%	18 (26.0%)	5.7%	1.3	7 (10.0%)	-10.0%	0.5
da Rocha e Silva, 2013	Brazil	Patients with stroke	47	14 (30.0%)	59.8	51.1%	16 (34.0%)	4.3%	1.1	7 (15.0%)	-14.9%	0.5
Ferentinos, 2011	Greece	Patients with amyotrophic	36	8 (22.0%)	62.0	41.7%	11 (31.0%)	8.3%	1.4	6 (17.0%)	-5.6%	0.7

		lateral sclerosis										
Fiest, 2014	Canada	Patients with epilepsy	180	30 (17.0%)	41.1	51.4%	31 (17.0%)	0.6%	1.0	18 (10.0%)	-6.7%	0.6
Fischer, 2014	Germany	Patients with heart failure	194	11 (6.0%)	65.9	20.6%	49 (25.0%)	19.6%	4.5	25 (13.0%)	7.2%	2.3
Gagnon, 2005	Canada	Patients admitted to hospital due to fall	108	14 (13.0%)	78.1	87.0%	22 (20.0%)	7.4%	1.6	7 (6.0%)	-6.5%	0.5
Goebel, 2011	Germany	Patients with brain tumors	26	0 (0.0%)	58.3	50.0%	5 (19.0%)	19.2%	_	1 (4.0%)	3.8%	_
Golden, 2006	Ireland	Outpatients with Hepatitis C	86	7 (8.0%)	37.7	25.6%	24 (28.0%)	19.8%	3.4	11 (13.0%)	4.7%	1.6
Gould, 2011	Australia	Patients with traumatic brain injury	189	15 (8.0%)	35.7	21.7%	35 (19.0%)	10.6%	2.3	12 (6.0%)	-1.6%	0.8
Honarmand, 2009	Canada	Patients with multiple sclerosis	140	9 (6.0%)	43.9	74.3%	26 (19.0%)	12.1%	2.9	10 (7.0%)	0.7%	1.1
Juliao, 2013	Portugal	Patients with advanced disease	75	31 (41.0%)	NR	NR	62 (83.0%)	41.3%	2.0	56 (75.0%)	33.3%	1.8
Keller, 2004	Germany	Inpatients with cancer at the department of surgery	76	4 (5.0%)	56.7	38.2%	22 (29.0%)	23.7%	5.5	15 (20.0%)	14.5%	3.8

Kjaergaard, 2014	Norway	Healthy population	357	20 (6.0%)	52.5	100.0%	15 (4.0%)	-1.4%	0.8	1 (0.3%)	-5.3%	0
Kugaya, 2000	Japan	Inpatients with Cancer	81	3 (4.0%)	61.2	25.9%	23 (28.0%)	24.7%	7.7	9 (11.0%)	7.4%	3.0
Lambert, 2015	Australia	Patients with cancer	164	25 (15.0%)	58.5	65.9%	33 (20.0%)	4.9%	1.3	16 (10.0%)	-5.5%	0.6
Löwe, 2002	Germany	Medical outpatients	497	64 (13.0%)	41.8	66.4%	193 (39.0%)	26.0%	3.0	100 (20.0%)	7.2%	1.6
Meyer, 2008	Germany	Patients undergoing laryngectomy	102	4 (4.0%)	60.4	93.1%	25 (25.0%)	20.6%	6.2	13 (13.0%)	8.8%	3.2
Michopoulos, 2010	Greece	Elderly inpatients	194	27 (14.0%)	74.0	47.9%	83 (43.0%)	28.9%	3.1	47 (24.0%)	10.3%	1.7
Navines, 2012	Spain	Patients with chronic hepatitis C	500	32 (6.0%)	43.4	30.6%	74 (15.0%)	8.4%	2.3	31 (6.0%)	-0.2%	1.0
Öztürk, 2013	Turkey	Patients with acne	45	7 (16.0%)	20.9	80.0%	14 (31.0%)	15.6%	2.0	5 (11.0%)	-4.4%	0.7
Patten, 2015	Canada	Patients with multiple sclerosis	42	20 (48.0%)	NR	28.6%	16 (38.0%)	-9.5%	0.8	7 (17.0%)	-31%	0.4
Pintor, 2006	Spain	Patients on waiting list for heart transplantation	73	13 (18.0%)	55.2	16.4%	15 (21.0%)	2.7%	1.2	8 (11.0%)	-6.8%	0.6
Rooney, 2013	UK	Adults with cerebral glioma	133	15 (11.0%)	53.7	42.9%	20 (15.0%)	3.8%	1.3	9 (7.0%)	-4.5%	0.6

Ryan, 2012	Ireland	Patients with advanced cancer	203	8 (4.0%)	61.6	49.3%	46 (23.0%)	18.7%	5.8	16 (8.0%)	3.9%	2.0
Sanchez-Gistau, 2012	Spain	Patients with epilepsy	296	35 (12.0%)	36.1	55.7%	74 (25.0%)	13.2%	2.1	40 (14.0%)	1.7%	1.1
Sanchez, 2012	Spain	Patients undergoing heart transplantation	22	3 (14.0%)	54.2	9.1%	6 (27.0%)	13.6%	2.0	2 (9.0%)	-4.5%	0.7
Sanchez, 2014	Spain	Candidates for heart transplantation	120	8 (7.0%)	55.6	22.5%	26 (22.0%)	15%	3.2	7 (6.0%)	-0.8%	0.9
Schwarzbold, 2014	Brazil	Patients with severe traumatic brain injury	44	14 (32.0%)	32.8	18.2%	12 (27.0%)	-4.5%	0.9	8 (18.0%)	-13.6%	0.6
Simard 2015	Canada	Survivors of cancer	60	7 (12.0%)	60.3	43.3%	3 (5.0%)	-6.7%	0.4	1 (2.0%)	-10%	0.1
Singer, 2008	Germany	Patients with laryngeal cancer	141	8 (6.0%)	63.7	8.5%	38 (27.0%)	21.3%	4.8	16 (11.0%)	5.7%	2.0
Singer, 2009	UK	Patients with cancer in acute care	580	55 (9.0%)	59.4	38.4%	200 (34.0%)	25%	3.6	101 (17.0%)	7.9%	1.8
Stone, 2004	UK	Outpatients after stroke	35	4 (11.0%)	71.2	31.4%	5 (14.0%)	2.9%	1.2	3 (9.0%)	-2.9%	0.8
Tung, 2015	China	Patients with diabetes	136	33 (24.0%)	39.8	56.6%	32 (24.0%)	-0.7%	1.0	12 (9.0%)	-15.4%	0.4

Turner, 2012	Australia	Patients after stroke	72	13 (18.0%)	66.7	47.2%	18 (25.0%)	6.9%	1.4	5 (7.0%)	-11.1%	0.4
Turner, Unpublished	Australia	Patients undergoing cardiac rehabilitation	52	4 (8.0%)	60.3	86.5%	4 (8.0%)	0%	1.0	3 (6.0%)	-1.9%	0.8
Walker, 2007	UK	Patients with cancer	361	30 (8.0%)	NR	23.5%	45 (12.0%)	4.2%	1.5	14 (4.0%)	-4.4%	0.5
Walterfang, 2007	Australia	Sample of Australian Patients with Adrenomyeloneuropathy	10	1 (10.0%)	43.8	10.0%	3 (30.0%)	20%	3.0	2 (20.0%)	10%	2.0
Studies that used other semi-structured interviews and were included in sensitivity analyses												
Love, 2002[1]	Australia	Outpatients with breast cancer	302	28 (9.0%)	46.3	100.0%	35 (12.0%)	2.3%	1.2	8 (3.0%)	-6.60%	0.3
Love, 2004[2]	Australia	Outpatients with breast cancer	227	16 (7.0%)	51.7	100.0%	43 (19.0%)	11.9%	2.7	16 (7.0%)	0%	1.0
O'Rourke, 1998[3]	UK	Patients with stroke	56	9 (16.0%)	67.1	33.9%	13 (23.0%)	7.1%	1.4	7 (13.0%)	-3.60%	0.8
De Souza, 2009[4]	UK	Outpatients with Huntington's disease	50	12 (24.0%)	50.8	48.0%	16 (32.0%)	8.0%	1.3	11 (22.0%)	-2.0%	0.9

Author, year	Country	Population	N total	N (%) Major Depression	Mean Age	Percent Female	N (%) HADS-D ≥ 8	% Difference: HADS-D ≥ 8 - Major Depression	Ratio: HADS-D ≥ 8 / Major Depression	N (%) HADS-D ≥ 11	% Difference: HADS-D ≥ 11 - Major Depression	Ratio: HADS-D ≥11 / Major Depression
Studies from IPDMA that used the SCID and were included in main analyses												
Akechi, 2006	Japan	Outpatients with cancer in palliative care	223	17 (8.0%)	61.1	65.0	97 (43.0%)	35.9%	5.7	43 (19.0%)	11.7%	2.5
Amoozegar, 2017	Canada	Patients with migraines	102	51 (50.0%)	42.5	81.4	53 (52.0%)	2.0%	1.0	32 (31.0%)	-18.6%	0.6
Beraldi, 2014	Germany	Patients of haemato-oncology	120	10 (8.0%)	52.1	32.5	32 (27.0%)	18.3%	3.2	16 (13.0%)	5.0%	1.6
Braeken, 2010	Netherlands	Dutch cancer patients in radiotherapy	13	1 (8.0%)	69.4	NR	4 (31.0%)	23.1%	4.0	2 (15.0%)	7.7%	2.0
Cukor, 2008	USA	Patients with end-stage renal disease	70	14 (20.0%)	53.3	52.9	18 (26.0%)	5.7%	1.3	7 (10.0%)	-10.0%	0.5
da Rocha e Silva, 2013	Brazil	Patients with stroke	47	14 (30.0%)	59.8	51.1	16 (34.0%)	4.3%	1.1	7 (15.0%)	-14.9%	0.5
Ferentinos, 2011	Greece	Patients with amyotrophic lateral sclerosis	36	8 (22.0%)	62.0	41.7	11 (31.0%)	8.3%	1.4	6 (17.0%)	-5.6%	0.7

Fiest, 2014	Canada	Patients with epilepsy	180	30 (17.0%)	41.1	51.4	31 (17.0%)	0.6%	1.0	18 (10.0%)	-6.7%	0.6
Fischer, 2014	Germany	Patients with heart failure	194	11 (6.0%)	65.9	20.6	49 (25.0%)	19.6%	4.5	25 (13.0%)	7.2%	2.3
Gagnon, 2005	Canada	Patients admitted to hospital due to fall	108	14 (13.0%)	78.1	87.0	22 (20.0%)	7.4%	1.6	7 (6.0%)	-6.5%	0.5
Goebel, 2011	Germany	Patients with brain tumors	26	0 (0.0%)	58.3	50.0	5 (19.0%)	19.2%	_	1 (4.0%)	3.8%	_
Golden, 2006	Ireland	Outpatients with Hepatitis C	86	7 (8.0%)	37.7	25.6	24 (28.0%)	19.8%	3.4	11 (13.0%)	4.7%	1.6
Gould, 2011	Australia	Patients with traumatic brain injury	189	15 (8.0%)	35.7	21.7	35 (19.0%)	10.6%	2.3	12 (6.0%)	-1.6%	0.8
Honarmand, 2009	Canada	Patients with multiple sclerosis	140	9 (6.0%)	43.9	74.3	26 (19.0%)	12.1%	2.9	10 (7.0%)	0.7%	1.1
Juliao, 2013	Portugal	Patients with advanced disease	75	31 (41.0%)	NR	NR	62 (83.0%)	41.3%	2.0	56 (75.0%)	33.3%	1.8
Keller, 2004	Germany	Inpatients with cancer at the department of surgery	76	4 (5.0%)	56.7	38.2	22 (29.0%)	23.7%	5.5	15 (20.0%)	14.5%	3.8
Kjaergaard, 2014	Norway	Healthy population	357	20 (6.0%)	52.5	100.0	15 (4.0%)	-1.4%	0.8	1 (0.3%)	-5.3%	0

Kugaya, 2000	Japan	Inpatients with Cancer	81	3 (4.0%)	61.2	25.9	23 (28.0%)	24.7%	7.7	9 (11.0%)	7.4%	3.0
Lambert, 2015	Australia	Patients with cancer	164	25 (15.0%)	58.5	65.9	33 (20.0%)	4.9%	1.3	16 (10.0%)	-5.5%	0.6
Löwe, 2002	Germany	Medical outpatients	497	64 (13.0%)	41.8	66.4	193 (39.0%)	26.0%	3.0	100 (20.0%)	7.2%	1.6
Meyer, 2008	Germany	Patients undergoing laryngectomy	102	4 (4.0%)	60.4	93.1	25 (25.0%)	20.6%	6.2	13 (13.0%)	8.8%	3.2
Michopoulos, 2010	Greece	Elderly inpatients	194	27 (14.0%)	74.0	47.9	83 (43.0%)	28.9%	3.1	47 (24.0%)	10.3%	1.7
Navines, 2012	Spain	Patients with chronic hepatitis C	500	32 (6.0%)	43.4	30.6	74 (15.0%)	8.4%	2.3	31 (6.0%)	-0.2%	1.0
Öztürk, 2013	Turkey	Patients with acne	45	7 (16.0%)	20.9	80.0	14 (31.0%)	15.6%	2.0	5 (11.0%)	-4.4%	0.7
Patten, 2015	Canada	Patients with multiple sclerosis	42	20 (48.0%)	NR	28.6	16 (38.0%)	-9.5%	0.8	7 (17.0%)	-31%	0.4
Pintor, 2006	Spain	Patients on waiting list for heart transplantation	73	13 (18.0%)	55.2	16.4	15 (21.0%)	2.7%	1.2	8 (11.0%)	-6.8%	0.6
Rooney, 2013	UK	Adults with cerebral glioma	133	15 (11.0%)	53.7	42.9	20 (15.0%)	3.8%	1.3	9 (7.0%)	-4.5%	0.6

Ryan, 2012	Ireland	Patients with advanced cancer	203	8 (4.0%)	61.6	49.3	46 (23.0%)	18.7%	5.8	16 (8.0%)	3.9%	2.0
Sanchez-Gistau, 2012	Spain	Patients with epilepsy	296	35 (12.0%)	36.1	55.7	74 (25.0%)	13.2%	2.1	40 (14.0%)	1.7%	1.1
Sanchez, 2012	Spain	Patients undergoing heart transplantation	22	3 (14.0%)	54.2	9.1	6 (27.0%)	13.6%	2.0	2 (9.0%)	-4.5%	0.7
Sanchez, 2014	Spain	Candidates for heart transplantation	120	8 (7.0%)	55.6	22.5	26 (22.0%)	15%	3.2	7 (6.0%)	-0.8%	0.9
Schwarzbold, 2014	Brazil	Patients with severe traumatic brain injury	44	14 (32.0%)	32.8	18.2	12 (27.0%)	-4.5%	0.9	8 (18.0%)	-13.6%	0.6
Simard 2015	Canada	Survivors of cancer	60	7 (12.0%)	60.3	43.3	3 (5.0%)	-6.7%	0.4	1 (2.0%)	-10%	0.1
Singer, 2008	Germany	Patients with laryngeal cancer	141	8 (6.0%)	63.7	8.5	38 (27.0%)	21.3%	4.8	16 (11.0%)	5.7%	2.0
Singer, 2009	UK	Patients with cancer in acute care	580	55 (9.0%)	59.4	38.4	200 (34.0%)	25%	3.6	101 (17.0%)	7.9%	1.8
Stone, 2004	UK	Outpatients after stroke	35	4 (11.0%)	71.2	31.4	5 (14.0%)	2.9%	1.2	3 (9.0%)	-2.9%	0.8
Tung, 2015	China	Patients with diabetes	136	33 (24.0%)	39.8	56.6	32 (24.0%)	-0.7%	1.0	12 (9.0%)	-15.4%	0.4

Turner, 2012	Australia	Patients after stroke	72	13 (18.0%)	66.7	47.2	18 (25.0%)	6.9%	1.4	5 (7.0%)	-11.1%	0.4
Turner, Unpublished	Australia	Patients undergoing cardiac rehabilitation	52	4 (8.0%)	60.3	86.5	4 (8.0%)	0%	1.0	3 (6.0%)	-1.9%	0.8
Walker, 2007	UK	Patients with cancer	361	30 (8.0%)	NR	23.5	45 (12.0%)	4.2%	1.5	14 (4.0%)	-4.4%	0.5
Walterfang, 2007	Australia	Sample of Australian Patients with Adrenomyeloneuropathy	10	1 (10.0%)	43.8	10.0	3 (30.0%)	20%	3.0	2 (20.0%)	10%	2.0
Stud71.2ies that used other semi-structured interviews and were included in sensitivity analyses 39.8												
Love, 2002[1]	Australia	Outpatients with breast cancer	302	28 (9.0%)	46.3	100.0	35 (12.0%)	2.3%	1.2	8 (3.0%)	-6.60%	0.3
Love, 2004[2]	Australia	Outpatients with breast cancer	227	16 (7.0%)	51.7	100.0	43 (19.0%)	11.9%	2.7	16 (7.0%)	0%	1.0
O'Rourke, 1998[3]	UK	Patients with stroke	56	9 (16.0%)	67.1	33.9	13 (23.0%)	7.1%	1.4	7 (13.0%)	-3.60%	0.8
De Souza, 2009[4]	UK	Outpatients with Huntington's disease	50	12 (24.0%)	50.8	48.0	16 (32.0%)	8.0%	1.3	11 (22.0%)	-2.0%	0.9

[1, 2] Diagnostic interview = Monash Interview for Liaison Psychiatry

[3] Diagnostic interview = Schedule for Affective Disorders and Schizophrenia

[4] Diagnostic interview = Schedules for Clinical Assessment in Neuropsychiatry

NR= Not reported

Figure 1. Study selection process.

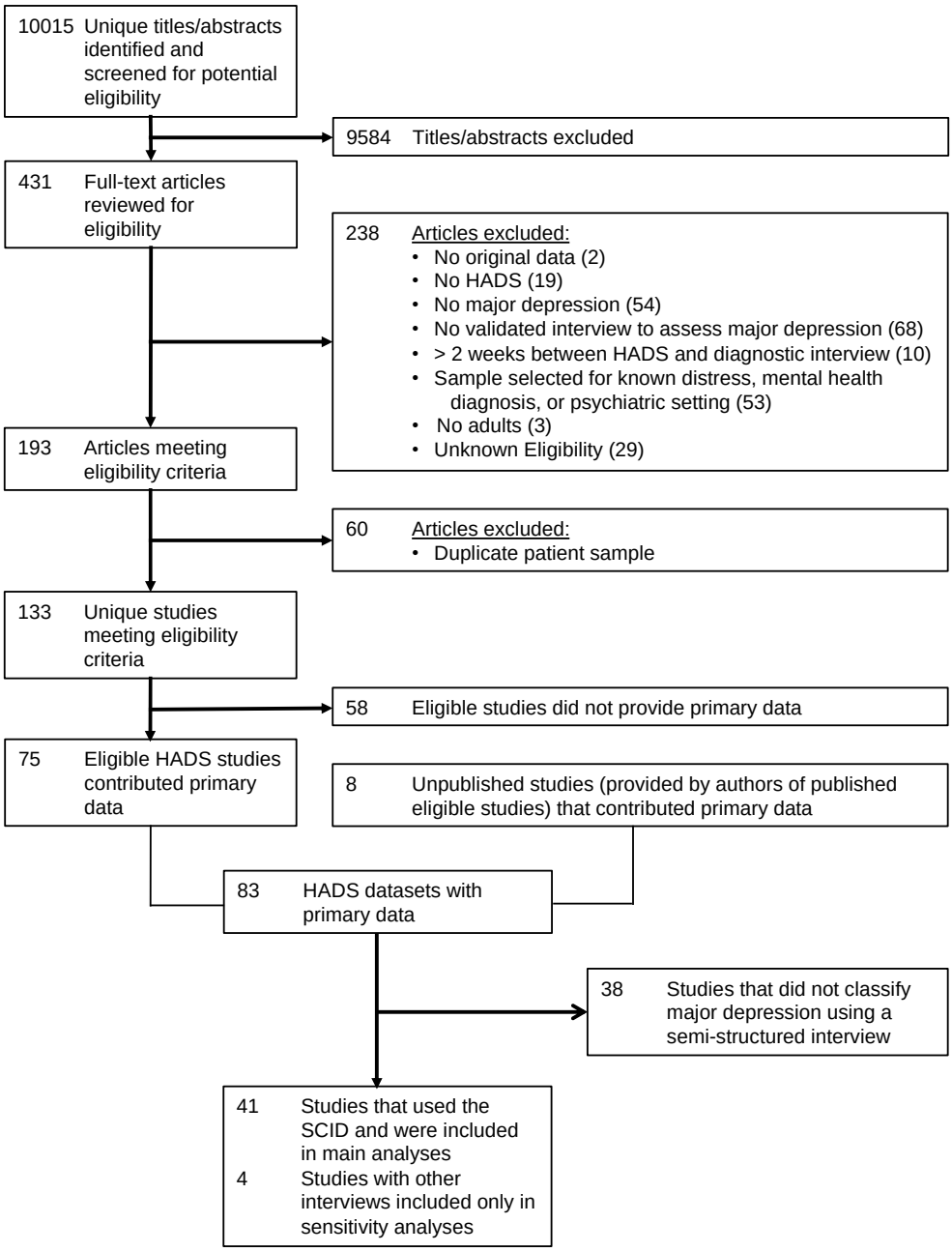

Figure 2. Proportion of participants (%) who scored at or above each possible HADS-D cutoff.