Original Investigation

Prevalence of Depression and Depressive Symptoms Among Resident Physicians
A Systematic Review and Meta-analysis

Douglas A. Mata, MD, MPH; Marco A. Ramos, MPhil, MSEd; Narinder Bansal, PhD; Rida Khan, BS; Constance Guille, MD, MS; Emanuele Di Angelantonio, MD, PhD; Srijan Sen, MD, PhD

IMPORTANCE Physicians in training are at high risk for depression. However, the estimated prevalence of this disorder varies substantially between studies.

OBJECTIVE To provide a summary estimate of depression or depressive symptom prevalence among resident physicians.

DATA SOURCES AND STUDY SELECTION Systematic search of EMBASE, ERIC, MEDLINE, and PsycINFO for studies with information on the prevalence of depression or depressive symptoms among resident physicians published between January 1963 and September 2015. Studies were eligible for inclusion if they were published in the peer-reviewed literature and used a validated method to assess for depression or depressive symptoms.

DATA EXTRACTION AND SYNTHESIS Information on study characteristics and depression or depressive symptom prevalence was extracted independently by 2 trained investigators. Estimates were pooled using random-effects meta-analysis. Differences by study-level characteristics were estimated using meta-regression.

MAIN OUTCOMES AND MEASURES Point or period prevalence of depression or depressive symptoms as assessed by structured interview or validated questionnaire.

RESULTS Data were extracted from 31 cross-sectional studies (9447 individuals) and 23 longitudinal studies (8113 individuals). Three studies used clinical interviews and 51 used self-report instruments. The overall pooled prevalence of depression or depressive symptoms was 28.8% (4969/17 560 individuals, 95% CI, 25.3%-32.5%), with high between-study heterogeneity (Q = 1247, τ^2 = 0.39, I^2 = 95.8%, P < .001). Prevalence estimates ranged from 20.9% for the 9-item Patient Health Questionnaire with a cutoff of 10 or more (741/3577 individuals, 95% CI, 17.5%-24.7%, Q = 14.4, τ^2 = 0.04, I^2 = 79.2%) to 43.2% for the 2-item PRIME-MD (1349/2891 individuals, 95% CI, 37.6%-49.0%, Q = 45.6, τ^2 = 0.09, I^2 = 84.6%). There was an increased prevalence with increasing calendar year (slope = 0.5% increase per year, adjusted for assessment modality; 95% CI, 0.03%-0.9%, P = .04). In a secondary analysis of 7 longitudinal studies, the median absolute increase in depressive symptoms with the onset of residency training was 15.8% (range, 0.3%-26.3%; relative risk, 4.5). No statistically significant differences were observed between cross-sectional vs longitudinal studies, studies of only interns vs only upper-level residents, or studies of nonsurgical vs both nonsurgical and surgical residents.

CONCLUSIONS AND RELEVANCE In this systematic review, the summary estimate of the prevalence of depression or depressive symptoms among resident physicians was 28.8%, ranging from 20.9% to 43.2% depending on the instrument used, and increased with calendar year. Further research is needed to identify effective strategies for preventing and treating depression among physicians in training.

Editorial page 2357

Supplemental content at jama.com

CME Quiz at jamanetworkcme.com and CME Questions page 2407

Author Affiliations: Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Mata); Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, Connecticut (Ramos); Department of Public Health and Primary Care, University of Cambridge, Cambridge, England (Bansal, Di Angelantonio); Department of Medicine, Baylor College of Medicine, Texas Medical Center, Houston (Khan); Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston (Guille); Molecular and Behavioral Neuroscience Institute, Department of Psychiatry, University of Michigan, Ann Arbor (Sen).

Corresponding Author: Douglas A. Mata, MD, MPH, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115 (dmata@bwh.harvard.edu).

JAMA. 2015;314(22):2373-2383. doi:10.1001/jama.2015.15845

Studies have suggested that resident physicians experience higher rates of depression than the general public.[1-5] Beyond the effects of depression on individuals, resident depression has been linked to poor-quality patient care and increased medical errors.[6-8] However, estimates of the prevalence of depression or depressive symptoms vary across studies, from 3% to 60%.[9,10] Studies also report conflicting findings about resident depression depending on specialty, postgraduate year, sex, and other characteristics.[4,11-13] A reliable estimate of depression prevalence during medical training is important for informing efforts to prevent, treat, and identify causes of depression among residents.[14] We conducted a systematic review and meta-analysis of published studies of depression or depressive symptoms in graduate medical trainees.

Methods

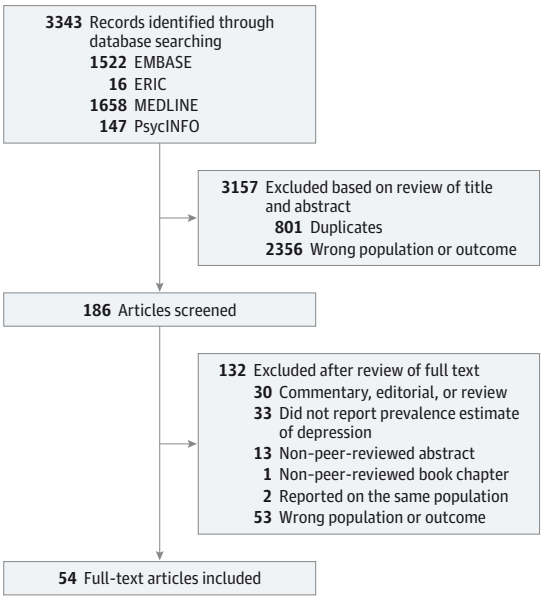
Search Strategy and Study Eligibility

Cross-sectional and longitudinal studies published between January 1963 and September 2015 that reported on the prevalence of depression or depressive symptoms in interns, resident physicians, or both were identified using EMBASE, ERIC, MEDLINE, and PsycINFO (independently performed by D.A.M. and M.A.R.); by screening the reference lists of articles identified; and by correspondence with study investigators using the approach recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (**Figure 1**).[15] The computer-based searches combined terms related to interns, resident physicians, and study design with those related to depression, without language restriction (full details of the search strategy are provided in eMethods 1 in the Supplement). Studies were included if they reported data on resident physicians, were published in peer-reviewed journals, and used a validated method to assess for depression or depressive symptoms.[16]

Data Extraction and Quality Assessment

The following information was independently extracted from each article by 2 trained investigators (D.A.M. and M.A.R.) using a standardized form: study design, geographic location, years of survey, specialty, postgraduate level, sample size, average age of participants, number and percentage of male participants, diagnostic or screening method used, outcome definition (ie, specific diagnostic criteria or screening instrument cutoff), and reported prevalence of depression or depressive symptoms. The most comprehensive publication was used when there were several involving the same population of residents. A modified version of the Newcastle-Ottawa Scale was used to assess the quality of nonrandomized studies included in systematic reviews and meta-analyses.[17] This scale assesses quality in several domains: sample representativeness and size, comparability between respondents and nonrespondents, ascertainment of depressive symptoms, and statistical quality (full details regarding scoring are provided in eMethods 2 in the Supplement). Studies were judged to be at low risk of bias (≥3 points) or high risk of bias (<3 points). All discrepancies were resolved by discussion and adjudication of a third reviewer (S.S.).

Figure 1. Flow Diagram for Identifying Studies on the Prevalence of Depression or Depressive Symptoms Among Resident Physicians



All studies identified by hand searching reference lists were found in the database search. For simplicity, this number is not duplicated in the diagram.

Data Synthesis and Analysis

Prevalence estimates of depression or depressive symptoms were calculated by pooling the study-specific estimates using random-effects meta-analysis that accounted for between-study heterogeneity.[18] Binomial proportion confidence intervals for individual studies were calculated using the Clopper-Pearson method, which allows for asymmetry. When longitudinal studies reported prevalence estimates made at different time periods within the year, the overall period prevalence for the time period was used. Between-study heterogeneity was assessed by standard χ^2 tests and the I^2 statistic (ie, the percentage of variability in prevalence estimates due to heterogeneity rather than sampling error, or chance, with values ≥75% indicating considerable heterogeneity)[19,20] and by comparing results from studies grouped according to prespecified study-level characteristics (study design, country, year of baseline survey, specialty, postgraduate level, Newcastle-Ottawa Scale components, age, sex, and diagnostic method) using stratified meta-analysis and meta-regression.[21,22] The influence of individual studies on the overall prevalence estimate was explored by serially excluding each study in a sensitivity analysis. A secondary analysis restricted to longitudinal studies reporting both preresidency and intraresidency depressive symptom prevalence estimates was performed to better isolate associations with the residency experience from associations with

assessment tools. Bias secondary to small study effects was investigated by funnel plot and Egger test.[23,24] All analyses were performed using R version 3.2.2 (R Foundation for Statistical Computing).[25] Statistical tests were 2-sided and used a significance threshold of $P < .05$.

Results

Study Characteristics

Thirty-one cross-sectional[10-13,26-52] and 23 longitudinal[4,6-8,53-71] studies involving a total of 17 560 individuals were included in the study (Figure 1, **Table 1**, and **Table 2**). Thirty-five took place in North America, 9 in Asia, 5 in Europe, 4 in South America, and 1 in Africa. Twenty-eight studies recruited residents from multiple specialties, while 26 recruited exclusively from single specialties. Thirteen studies included interns only, 36 included both interns and residents, and 5 included upper-level residents only. The median number of participants per study was 141 (range, 27-2323). Eleven studies assessed for depressive symptoms using the Beck Depression Inventory (BDI),[72] 11 used the Center for Epidemiologic Studies Depression Scale (CES-D),[73] 8 used the 2-item Primary Care Evaluation of Mental Disorders questionnaire (PRIME-MD),[74] 7 used the 9-item Patient Health Questionnaire (PHQ-9),[75] 4 used the Zung Self-rating Depression Scale (SDS),[76] 3 used the Harvard Department of Psychiatry/National Depression Screening Day Scale (HANDS),[77] and 7 used other methods.[78-82] Three assessed for depression using structured interviews.[83] The diagnostic criteria and scoring cutoffs used by the studies are summarized in Table 1. When evaluated by Newcastle-Ottawa quality assessment criteria, out of 5 possible points, 3 studies received 5 points, 13 received 4 points, 23 received 3 points, 10 received 2 points, 4 received 1 point, and 1 received 0 points (scores for individual studies are presented in eTable 1 in the Supplement).

Prevalence of Depression or Depressive Symptoms Among Resident Physicians

Meta-analytic pooling of the prevalence estimates of depression or depressive symptoms reported by the 54 studies yielded a summary prevalence of 28.8% (4969/17 560 individuals, 95% CI, 25.3%-32.5%), with significant evidence of between-study heterogeneity ($Q = 1247$, $P < .001$, $\tau^2 = 0.39$, $I^2 = 95.8\%$) (**Figure 2**). Sensitivity analysis, in which the meta-analysis was serially repeated after exclusion of each study, demonstrated that no individual study affected the overall prevalence estimate by more than 1% (eTable 2 in the Supplement).

To provide a range of the depression or depressive symptom prevalence estimates identified by these methodologically diverse studies, estimates were stratified by screening instrument and cutoff score (**Figure 3**). Summary prevalence estimates ranged from 20.9% for the PHQ-9 with cutoff of 10 or more (741/3577 individuals, 95% CI, 17.5%-24.7%, $Q = 14.4$, $\tau^2 = 0.04$, $I^2 = 79.2\%$) to 43.2% for the 2-item PRIME-MD (1349/2891 individuals, 95% CI, 37.6%-49.0%, $Q = 45.6$, $\tau^2 = 0.09$, $I^2 = 84.6\%$). The 8 studies using the 2-item PRIME-MD yielded significantly higher estimates than did the others ($Q = 69.0$, $P < .001$). In contrast, there were no significant differences between estimates made using the CES-D, PHQ-9, HANDS, BDI, or Zung SDS ($Q = 8.65$, $P = .12$), suggesting that variation between instruments did not explain the heterogeneity in the observed depression or depressive symptom prevalence estimates. A model including only those studies[4,7,34,47,48,50,60,66] using inventories with specificities greater than 88% yielded a prevalence estimate of 20.2% (1119/5425, 95% CI, 18.0%-22.6%, $Q = 22.0$, $P < .01$, $\tau^2 = 0.02$, $I^2 = 68.2\%$).

Prevalence of Depression or Depressive Symptoms by Study-Level Characteristics

Among all 54 studies, the prevalence of depression or depressive symptoms significantly increased with baseline survey year (slope = 0.5% per calendar-year increase; 95% CI, 0.03%-0.9%; test of moderator, $Q = 4.4$, $P = .04$). This association persisted when studies using the 2-item PRIME-MD were excluded and the analysis was restricted to the 23 studies using the CES-D, PHQ-9, HANDS, BDI, or Zung SDS presented in Figure 3 (slope = 0.6% per calendar-year increase; 95% CI, 0.1%-1.2%, $P = .02$).

Among the full set of studies, no statistically significant differences in prevalence estimates were noted between cross-sectional vs longitudinal studies (2851/9447, 29.1% [95% CI, 23.9% to 34.9%] vs 2111/8113, 28.4% [95% CI, 24.2% to 33.0%]; test for subgroup differences, $Q = 0.04$, $P = .85$), studies in the United States vs elsewhere (3026/10 883, 26.6% [95% CI, 21.9% to 31.9%] vs 1936/6677, 31.1% [95% CI, 26.0% to 36.7%]; $Q = 1.4$, $P = .23$), studies of nonsurgical vs both nonsurgical and surgical residents (1570/5841, 28.9% [95% CI, 24.7% to 33.4%] vs 3392/11 719, 28.8% [95% CI, 23.6% to 34.7%]; $Q = 0$, $P = .98$), or studies of only interns vs those of only upper-level residents (1411/5127, 31.9% [95% CI, 25.4% to 39.1%] vs 211/1061, 26.6% [95% CI, 14.9% to 42.8%]; $Q = 0.9$, $P = .62$) (**Figure 4**). There were no significant associations between prevalence and mean or median age (slope = −1.0% per year [95% CI, −2.8% to 0.8%]; $Q = 1.2$, $P = .28$) or percentage of males (slope = 3.4% per percentage increase in males [95% CI, −28.9% to 22.1%]; $Q = 0.1$, $P = .79$).

When evaluated by Newcastle-Ottawa criteria, studies with lower total overall quality scores yielded higher depression estimates (660/1658, 36.7% [95% CI, 30.2%-43.7%] vs 4302/15 902, 26.1% [95% CI, 22.4%-30.2%]; $Q = 7.3$, $P = .007$) (**Figure 5**). In terms of individual quality assessment criteria, higher prevalence estimates were found among studies with less representative participant populations (569/1472, 37.7% [95% CI, 32.4%-43.2%] vs 4393/16 088, 26.8% [95% CI, 23.1%-30.9%]; $Q = 10.4$, $P = .001$) and less valid assessment methods (1835/4425, 36.2% [95% CI, 29.9%-43.0%] vs 3127/13 135, 25.7% [95% CI, 22.6%-29.0%]; $Q = 8.6$, $P = .003$). No statistically significant differences in prevalence estimates were noted when studies were stratified by respondent/nonrespondent comparability criteria ($Q = 0.11$, $P = .75$) or by quality of descriptive statistic reporting ($Q = 0.23$, $P = .63$).

Table 1. Selected Characteristics of the 31 Cross-sectional Studies Included in This Systematic Review and Meta-analysis

Source	Country	Survey Years	Specialty	PGY	No. of Participants	Age, y	Men, No. (%)	Diagnostic Method	Outcome Definition	NOS
de Oliveira et al,[47] 2013	United States	2011	Anesthesia	1-4	1384	No. (%) ≤30 y: 779 (54.0)	850 (57.0)	HANDS	>9	5
Waldman et al,[43] 2009	Argentina	2007	Cardiology	3-4	106	Mean (SD), 29.1 (2.4)	70 (66.0)	21-Item BDI	≥10	3
Hasanović and Herenda,[39] 2008	Bosnia and Herzegovina	2004	Family medicine	≥1	78	Median (range), NR (30-45)	12 (15.4)	HSCL-25	≥1.75	3
Godenick et al,[29] 1995	United States	1992	Family medicine	1-4	164	Mean (SD), 30.3 (4.6)	133 (74.7)	21-Item BDI	≥10	3
Oriel et al,[33] 2004	United States	NR	Family medicine	1-4	185	Mean (range), 33 (26-57)	87 (47.0)	9-Item survey	*DSM-IV* criteria	1
Earle and Kelly,[34] 2005	Canada	2002	Family medicine	≥1	254	Mean (SD), 29 (NR)	90 (35.4)	PHQ-9	≥10	4
Hainer and Palesch,[30] 1998	United States	1993-1996	Family medicine	1-3	268	Mean (SD), 30.4 (5.2)	239 (68.3)	21-Item BDI	≥10	4
Lam et al,[44] 2010	Hong Kong	2005	General internship	1	95	Mean (range), 24.4 (23-28)	48 (49.5)	DASS-21	≥10	3
Sakata et al,[40] 2008	Japan	2005	General internship	1-2	196	Mean (SD), 27.3 (2.9)	149 (76)	CES-D	≥19	3
Hsieh et al,[13] 2011	Taiwan	2004-2005	General internship	1	302	NR	216 (71.5)	Zung SDS	≥41	2
Costa et al,[45] 2012	Brazil	2008	Internal medicine	1	84	Mean (SD), 24.6 (3.8)	45 (53.6)	21-Item BDI	≥10	3
Shanafelt et al,[32] 2002	United States	2001	Internal medicine	1-3	115	NR	54 (47.0)	PRIME-MD	Yes to either item	0
Yi et al,[37] 2006	United States	2003	Medical and pediatric	≥1	227	Mean (SD), 28.7 (3.8)	95 (42)	CES-D	≥10	3
Raviola et al,[31] 2002	Kenya	1997-1999	Medical and surgical	3-4	50	Mean (SD), 33 (NR)	NR	Structured interview	*DSM-IV* criteria	2
Valko and Clayton,[27] 1975	United States	1972	Medical and surgical	1	53	NR	NR	Structured interview	*DSM-II* criteria	2
Kirsling et al,[12] 1989	United States	1987-1988	Medical and surgical	1	58	NR	38 (62.3)	21-Item BDI	≥10	3
Cruz EP,[36] 2006	Mexico	NR	Medical and surgical	1-6	80	Mean (SD), 27.5 (1.8)	53 (66.3)	Zung SDS	≥41	1
Demir et al,[38] 2007	Turkey	2004	Medical and surgical	≥1	86	Mean (SD), 28.2 (3.2)	38 (44.2)	21-Item BDI	≥11	3
Sánchez et al,[41] 2008	Mexico	2007-2008	Medical and surgical	1-3	90	Mean (SD), 28.6 (0.5)	49 (54.4)	HAM-D	≥8	4
Al Ghafri et al,[48] 2014	Oman	2011	Medical and surgical	1-4	132	73%<30 y	42 (31.8)	PHQ-9	≥12	3
Al-Maddah et al,[51] 2015	Saudi Arabia	2012	Medical and surgical	1-5	171	Median (range), NR (25-35)	72 (42)	21-Item BDI	≥10	3
Yousuf et al,[10] 2011	Pakistan	2008	Medical and surgical	≥1	172	No. (%) <30 y: 104 (70.3)	111 (64.5)	Zung SDS	≥45	2
Steinert et al,[28] 1991	Canada	1984	Medical and surgical	1-6	255	Mean (range), 27.7 (21-52)	182 (71.4)	Zung SDS	≥50	4
Stoesser and Cobb,[50] 2014	United States	2009	Medical and surgical	≥1	260	Mean (range), 30.8 (25-55)	126 (50.2)	PHQ-9	≥10	4
Pereira-Lima and Loureiro,[52] 2015	Brazil	2012	Medical and surgical	1-5	305	Mean (SD), 28 (2.5)	159 (52.1)	PHQ-4	≥3	4
Goebert et al,[42] 2009	United States	2003-2004	Medical and surgical	1-4	532	NR	254 (48)	CES-D	≥16	3
Dyrbye et al,[49] 2014	United States	2011-2012	Medical and surgical	1-7	1701	Median (range), 31 (NR)	824 (48.6)	PRIME-MD	Yes to either item	3
Hsu and Marshall,[11] 1987	Canada	1984-1985	Medical and surgical	≥1	1785	Mean (SD), 29 (4.2)	1184 (66.3)	CES-D	≥16	4
Govardhan et al,[46] 2012	United States	2009	Ob/gyn	1-4	56	Mean (SD), 30.1 (3.0)	5 (8.8)	CES-D	>16	3
Becker et al,[35] 2006	United States	2004	Ob/gyn	1-4	120	Mean (SD), 29.3 (3.0)	26 (20.8)	CES-D	≥16	3
Waring EM,[26] 1974	United Kingdom	NR	Psychiatry	≥1	83	NR	NR	GHQ	≥12	2

Abbreviations: BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; DASS-21, 21-item Depression, Anxiety, and Stress Scale; *DSM*, *Diagnostic and Statistical Manual of Mental Disorders*; GHQ, General Health Questionnaire; HADS-D, Hospital Anxiety and Depression Scale; HAM-D, Hamilton Depression Rating Scale; HANDS, Harvard Department of Psychiatry/National Depression Screening Day Scale; HSCL-25, 25-item Hopkins Symptom Checklist; NOS, Newcastle-Ottawa score; NR, not reported; PGY, postgraduate year; PHQ-9, 9-item Patient Health Questionnaire; PRIME-MD, 2-item Primary Care Evaluation of Mental Disorders questionnaire; SSTDS, Spielberger State-Trait Depression Scale; Zung SDS, Zung Self-rating Depression Scale.

Table 2. Selected Characteristics of the 23 Longitudinal Studies Included in This Systematic Review and Meta-analysis

Source	Country	Survey Years	Specialty	PGY	No. of Participants	Age, y	Men, No. (%)	Diagnostic Method	Outcome Definition	NOS
Katz et al,[57] 2006	United States	2003-2004	Emergency medicine	1-4	31	Median (range), 29 (24-49)	33 (66.0)	CES-D	>14	3
Revicki et al,[55] 1993	United States	1989-1992	Emergency medicine	1-3	1117	Mean (SD), 30 (3.6)	827 (74.0)	CES-D	>16	4
Kleim et al,[68] 2014	Switzerland	NR	General rotating internship	1	47	Mean (SD), 24 (2)	20 (42.5)	PHQ-9	≥5	2
Ito et al,[70] 2015	Japan	2011	General rotating internship	1	1209	Mean (SD), 26 (3)	668 (65.5)[a]	CES-D	≥16	4
Rosen et al,[58] 2006	United States	2002-2003	Internal medicine	1	47	NR	28 (48.3)	13-Item BDI	≥8	2
Reuben DB,[54] 1985	United States	1981-1982	Internal medicine	1-3	68	NR	NR	CES-D	≥16	1
Campbell et al,[62] 2010	United States	2003-2008	Internal medicine	1-3	86	Mean (SD), NR (26-40)	44 (51.1)	PRIME-MD	Yes to either item	1
Wada et al,[59] 2007	Japan	2005-2006	Internal medicine	1	99	Median (range), NR (24-39)	71 (71.7)	CES-D	≥19	4
Gopal et al,[56] 2005	United States	2003-2004	Internal medicine	1-3	121	Median (range), NR (26-40)	53 (43.8)	PRIME-MD	Yes to either item	2
West et al,[6] 2006	United States	2003-2006	Internal medicine	1-3	149	No. (%) ≤30 y: 129 (70.1)	94 (51.1)	PRIME-MD	Yes to either item	2
Beckman et al,[63] 2012	United States	2009-2010	Internal medicine	1-3	202	≥24	116 (57.4)	PRIME-MD	Yes to either item	3
West et al,[8] 2009	United States	2003-2009	Internal medicine	1-3	239	No. (%) ≤30 y: 240 (63.2)	236 (62.1)	PRIME-MD	Yes to either item	3
West et al,[65] 2012	United States	2007-2011	Internal medicine	1-3	278	No. (%) ≤30 y: 209 (84.3)	208 (61.2)	PRIME-MD	Yes to either item	3
Ford and Wentz,[53] 1984	United States	NR	Medical and surgical	1	27	Median (range), 26 (NR)	22 (81.4)	Structured interview	*DSM-III* criteria	3
Jiménez-López et al,[71] 2015	Mexico	NR	Medical and surgical	2	100	Mean (SD), 26.4 (1.8)	70 (64.8)	13-Item BDI	≥5	2
Buddeberg-Fischer et al,[61] 2009	Switzerland	2001-2007	Medical and surgical	2, 4, 6	390	Mean (SD), 33 (2.2)	176 (45.1)	HADS-D	≥8	3
Weigl et al,[64] 2012	Germany	NR	Medical and surgical	2-3	415	Mean (SD), 30.5 (2.7)	218 (52.5)	10-Item SSTDS	>24.21	4
Sen et al,[4] 2010	United States	2007-2009	Medical and surgical	1	740	Mean (SD), 27.9 (2.8)	337 (45.6)	PHQ-9	≥10	5
Sen et al,[66] 2013	United States	2009-2011	Medical and surgical	1	2323	Mean (SD), 27.6 (2.9)	1140 (49.1)	PHQ-9	≥10	5
Cubero et al,[69] 2015	Brazil	2010-2011	Medical oncology	≥1	50	Median (IQR), 28.4 (27.4-29.7)	29 (53.7)	21-Item BDI	≥16[b]	3
Velásquez-Pérez et al,[67] 2013	Mexico	2010-2011	Neurology, neurosurgery, psychiatry	1	43	Mean (range), 25 (24-41)	26 (60.5)	21-Item BDI	≥10	3
Fahrenkopf et al,[7] 2008	United States	2003	Pediatrics	1-3	123	No. (%) <30 y: 76 (62.0)	37 (30.1)	HANDS	≥9	4
Landrigan et al,[60] 2008	United States	2003-2004	Pediatrics	1-3	209	Mean (SD), 29.7 (NR)	64 (30.4)	HANDS	>9	4

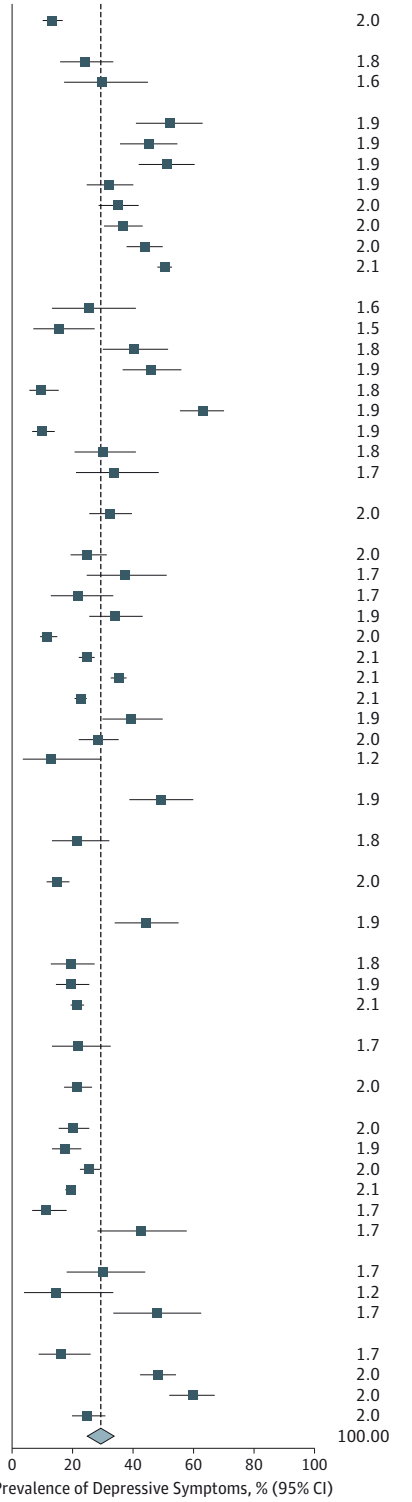
Abbreviations: BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; *DSM-III*, *Diagnostic and Statistical Manual of Mental Disorders* (Third Edition); HADS-D, Hospital Anxiety and Depression Scale; HANDS, Harvard Department of Psychiatry/National Depression Screening Day Scale; NOS, Newcastle-Ottawa score; NR, not reported; PGY, postgraduate year; PHQ-9, 9-item Patient Health Questionnaire; PRIME-MD, 2-item Primary Care Evaluation of Mental Disorders questionnaire; SSTDS, Spielberger State-Trait Depression Scale.

[a] Based on a subset of participants.

[b] The authors do not explicitly report a cutoff, but the study they cite suggests that it is 16.

Figure 2. Meta-analysis of the Prevalence of Depression or Depressive Symptoms Among Resident Physicians

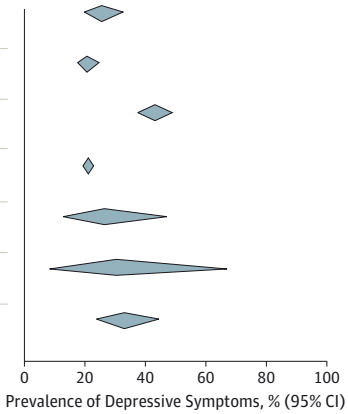
Source	Diagnostic Criteria or Instrument Cutoff	No. of Participants With Depressive Symptoms	Total No. of Participants	Prevalence of Depressive Symptoms, % (95% CI)	Weight, %
10-Item SSTDS					
Weigl et al,[64] 2012	>24.21	55	415	13.3 (10.1-16.9)	2.0
13-Item BDI					
Jiménez-López et al,[71] 2015	≥5	24	100	24.0 (16.0-33.6)	1.8
Rosen et al,[58] 2006	≥8	14	47	29.8 (17.3-44.9)	1.6
2-Item PRIME-MD					
Campbell et al,[62] 2010		45	86	52.3 (41.3-63.2)	1.9
Shanafelt et al,[32] 2002		52	115	45.2 (35.9-54.8)	1.9
Gopal et al,[56] 2005		62	121	51.2 (42.0-60.4)	1.9
West et al,[6] 2006		48	149	32.2 (24.8-40.4)	1.9
Beckman et al,[63] 2012		71	202	35.1 (28.6-42.2)	2.0
West et al,[8] 2009		88	239	36.8 (30.7-43.3)	2.0
West et al,[65] 2012		122	278	43.9 (38.0-49.9)	2.0
Dyrbye et al,[49] 2014		861	1701	50.6 (48.2-53.0)	2.1
21-Item BDI	≥10				
Velásquez-Pérez et al,[67] 2013		11	43	25.6 (13.5-41.2)	1.6
Kirsling et al,[12] 1989		9	58	15.5 (7.3-27.4)	1.5
Costa et al,[45] 2012		34	84	40.5 (29.9-51.7)	1.8
Waldman et al,[43] 2009		49	106	46.2 (36.5-56.2)	1.9
Godenick et al,[29] 1995		16	164	9.8 (5.7-15.4)	1.8
Al-Maddah et al,[51] 2015		108	171	63.2 (55.5-70.4)	1.9
Hainer and Palesch,[30] 1998		27	268	10.1 (6.7-14.3)	1.9
Demir et al,[38] 2007	≥11	26	86	30.2 (20.8-41.1)	1.8
Cubero et al,[69] 2015	≥16	17	50	34.0 (21.2-48.8)	1.7
9-Item survey					
Oriel et al,[33] 2004	*DSM-IV*	60	185	32.4 (25.7-39.7)	2.0
CES-D					
Yi et al,[37] 2006	≥10	57	227	25.1 (19.6-31.3)	2.0
Govardhan et al,[46] 2012	≥16	21	56	37.5 (24.9-51.5)	1.7
Reuben DB,[54] 1985		15	68	22.1 (12.9-33.8)	1.7
Becker et al,[35] 2006		41	120	34.2 (25.8-43.4)	1.9
Goebert et al,[42] 2009		63	532	11.8 (9.2-14.9)	2.0
Revicki et al,[55] 1993		277	1117	24.8 (22.3-27.4)	2.1
Ito et al,[70] 2015		427	1209	35.3 (32.6-38.1)	2.1
Hsu and Marshall,[11] 1987		407	1785	22.8 (20.9-24.8)	2.1
Wada et al,[59] 2007	≥19	39	99	39.4 (29.7-49.7)	1.9
Sakata et al,[40] 2008		56	196	28.6 (22.4-35.4)	2.0
Katz et al,[57] 2006	>14	4	31	12.9 (3.6-29.8)	1.2
DASS-21					
Lam et al,[44] 2010	≥10	47	95	49.5 (39.1-59.9)	1.9
GHQ					
Waring EM,[26] 1974	≥12	18	83	21.7 (13.4-32.1)	1.8
HADS-D					
Buddeberg-Fischer et al,[61] 2009	≥8	59	390	15.1 (11.7-19.1)	2.0
HAM-D					
Sánchez et al,[41] 2008	≥8	40	90	44.4 (34.0-55.3)	1.9
HANDS					
Fahrenkopf et al,[7] 2008	≥9	24	123	19.5 (12.9-27.6)	1.8
Landrigan et al,[60] 2008		41	209	19.6 (14.5-25.7)	1.9
de Oliveira et al,[47] 2013		298	1384	21.5 (19.4-23.8)	2.1
HSCL-25					
Hasanović and Herenda,[39] 2008	≥75	17	78	21.8 (13.2-32.6)	1.7
PHQ-4					
Pereira-Lima and Loureiro,[52] 2015	≥3	66	305	21.6 (17.1-26.7)	2.0
PHQ-9					
Earle and Kelly,[34] 2005	≥10	51	254	20.1 (15.3-25.5)	2.0
Stoesser and Cobb,[50] 2014		46	260	17.7 (13.3-22.9)	1.9
Sen et al,[4] 2010		190	740	25.7 (22.6-29.0)	2.0
Sen et al,[66] 2013		454	2323	19.5 (17.9-21.2)	2.1
Al Ghafri et al,[48] 2014	≥12	15	132	11.4 (6.5-18.0)	1.7
Kleim et al,[68] 2014	≥5	20	47	42.6 (28.3-57.8)	1.7
Structured interview, *DSM* criteria					
Valko and Clayton,[27] 1975	*DSM-II*	16	53	30.2 (18.3-44.3)	1.7
Ford and Wentz,[53] 1984	*DSM-III*	4	27	14.8 (4.2-33.7)	1.2
Raviola et al,[31] 2002	*DSM-IV*	24	50	48.0 (33.7-62.6)	1.7
Zung SDS					
Cruz EP,[36] 2006	≥41	13	80	16.2 (8.9-26.2)	1.7
Hsieh et al,[13] 2011		146	302	48.3 (42.6-54.1)	2.0
Yousuf et al,[10] 2011	≥45	103	172	59.9 (52.1-67.3)	2.0
Steinert et al,[28] 1991	≥50	64	255	25.1 (19.9-30.9)	2.0
Pooled summary estimate: I^2 = 95.8%, τ^2 = 0.39, P <.001		4969	17 560	28.8 (25.3-32.5)	100.00



Contributing studies are stratified by screening modality and ordered by increasing sample size. The area of each square is proportional to the inverse variance of the estimate. The dotted line marks the overall summary estimate for all studies, 28.8% (4969/17 560 individuals, 95% CI, 25.3%-32.5%, Q = 1247.11, τ^2 = 0.39, I^2 = 95.8% [95% CI, 95.0%-96.4%], P < .001). (Refer to footnotes of Table 1 and Table 2 for expanded names of diagnostic instruments.)

Figure 3. Meta-analyses of the Prevalence of Depressive Symptoms Among Resident Physicians in Subsets of Studies Stratified by Screening Modality and Cutoff Score

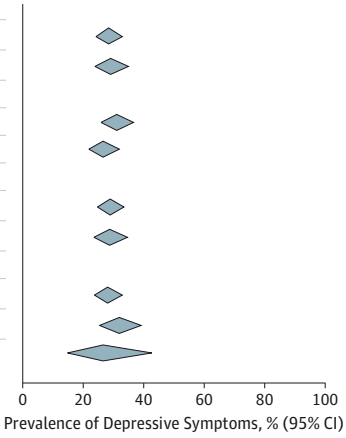
Instrument	Diagnostic Cutoff	No. of Studies	No. of Participants With Depressive Symptoms	Total No. of Participants	Prevalence of Depressive Symptoms, % (95% CI)
CES-D I^2 = 95.1%, τ^2 = 0.18, P <.001	≥16	7	1251	4887	25.6 (19.7-32.5)
PHQ-9 I^2 = 79.2%, τ^2 = 0.04, P = .002	≥10	4	741	3577	20.9 (17.5-24.7)
2-Item PRIME-MD I^2 = 84.6%, τ^2 = 0.09, P <.001	Yes to either item	8	1349	2891	43.2 (37.6-49.0)
HANDS I^2 = 0%, τ^2 = 0, P = .74	≥9	3	363	1716	21.2 (19.3-23.2)
21-Item BDI I^2 = 96.4%, τ^2 = 1.40, P <.001	≥10	7	254	894	26.6 (12.9-47.1)
Zung SDS I^2 = 95.8%, τ^2 = 1.19, P <.001	≥41	2	159	382	30.4 (8.6-67.1)
CES-D I^2 = 71.4%, τ^2 = 0.08, P = .06	≥19	2	95	295	33.4 (23.8-44.6)



The area of each diamond is proportional to the inverse variance of the estimate. BDI indicates Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; HANDS, Harvard Department of Psychiatry/National Depression Screening Day Scale; PHQ-9, 9-item Patient Health Questionnaire; PRIME-MD, 2-item Primary Care Evaluation of Mental Disorders questionnaire; Zung SDS, Zung Self-rating Depression Scale.

Figure 4. Meta-analyses of the Prevalence of Depression or Depressive Symptoms Among Resident Physicians Stratified by Study-Level Characteristics

	No. of Studies	No. of Participants With Depressive Symptoms	Total No. of Participants	Prevalence of Depressive Symptoms, % (95% CI)	*P* Value
Study design					
Cohort	23	2111	8113	28.4 (24.2-33.0)	.85
Cross-sectional	31	2851	9447	29.1 (23.9-34.9)	
Country					
Not United States	26	1936	6677	31.1 (26.0-36.7)	.23
United States	28	3026	10 883	26.6 (21.9-31.9)	
Specialty					
Nonsurgical only	27	1570	5841	28.9 (24.7-33.4)	.98
Nonsurgical and surgical	27	3392	11 719	28.8 (23.6-34.7)	
Postgraduate level					
Interns and upper levels	36	3340	11 372	28.1 (23.7-32.9)	.62
Interns only	13	1411	5127	31.9 (25.4-39.1)	
Upper levels only	5	211	1061	26.6 (14.9-42.8)	



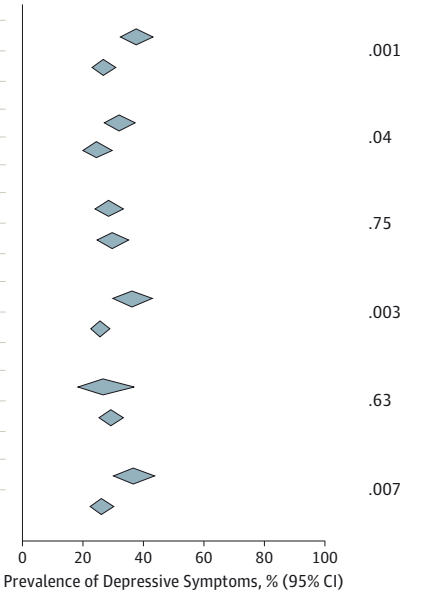
The area of each diamond is proportional to the inverse variance of the estimate.

Heterogeneity Within Screening Instruments

To identify potential sources of heterogeneity independent of assessment modality, heterogeneity was examined within the studies using common instruments when at least 5 studies were available and at least 2 studies were in each comparator subgroup. Among the 7 studies using the CES-D and a cutoff of 16 or greater, heterogeneity was not accounted for by study design (Q = 0.3, P = .61), baseline survey year (Q = 1.3, P = .25), specialty (Q = 0.2, P = .70), sample size (Q = 2.1, P = .15), age (Q = 0.7, P = .41), or sex (Q = 0.7, P = .41) (full results are provided in eTable3 in the Supplement). Among the 8 studies using the 2-item PRIME-MD, heterogeneity was partially explained by study design (cross-sectional studies yielded higher estimates, 49.8% vs 41.3%; Q = 5.2, P = .02) and respondent/nonrespondent comparability (studies that established comparability yielded lower estimates, 39.6% vs 50.4%; Q = 10.3, P = .001) but was not significantly explained by sample size (Q = 0.2, P = .64), sex (Q = 2.7, P = .10), baseline survey year (Q = 0.1, P = .80), or Newcastle-Ottawa score (Q = 0.2, P = .64). Among 7 studies using the 21-item BDI with cutoff of 10 or greater, heterogeneity was in part explained by country (United States vs other, 10.7% vs 44.6%; Q = 30.7, P < .001), baseline survey year (Q = 13.4, P < .001), and sex (Q = 10.7, P = .001), but not by specialty (Q = 0.3, P = .58), postgraduate year (Q = 0,

Figure 5. Meta-analyses of the Prevalence of Depression or Depressive Symptoms Among Resident Physicians Stratified by Newcastle-Ottawa Scale Components and by Total Score

Newcastle-Ottawa Component	No. of Studies	No. of Participants With Depressive Symptoms	Total No. of Participants	Prevalence of Depressive Symptoms, % (95% CI)	P Value
Sample representativeness					
Less	11	569	1472	37.7 (32.4-43.2)	.001
More	43	4393	16088	26.8 (23.1-30.9)	
Sample size					
<200 Participants	33	1092	3165	32.0 (27.1-37.4)	.04
≥200 Participants	21	3870	14359	24.5 (20.0-29.7)	
Respondent and nonrespondent comparability					
Less comparable	37	3443	11482	28.5 (24.1-33.4)	.75
More comparable	17	1519	6078	29.7 (24.8-35.1)	
Ascertainment of depression					
Less valid	17	1835	4425	36.2 (29.9-43.0)	.003
More valid	37	3127	13135	25.7 (22.6-29.0)	
Descriptive statistics					
Less detail	12	434	1600	26.7 (18.5-37.0)	.63
More detail	42	4528	15960	29.3 (25.4-33.4)	
Total score					
<3 Points	15	660	1658	36.7 (30.2-43.7)	.007
≥3 Points	39	4302	15902	26.1 (22.4-30.2)	



The area of each diamond is proportional to the inverse variance of the estimate.

Table 3. Secondary Analysis of 7 Longitudinal Studies Reporting Prevalence Estimates Both Prior to and During Internship

Source	Instrument	Cutoff	Follow-up	Baseline			Follow-up			Comparison	
				No. Depressed	Total No.	Prevalence, % (95% CI)	No. Depressed	Total No.	Prevalence, % (95% CI)	Absolute Increase, % (95% CI)	Relative Increase Ratio, (95% CI)
Velásquez-Pérez et al,[67] 2013	21-Item BDI	≥10	1 y	1	43	2.3 (0.1-12.3)	5	32	15.6 (5.3-32.8)	13.3 (13.2-13.4)	6.7 (6.6-7.0)
Rosen et al,[58] 2006	13-Item BDI	≥8	1 y	2	58	3.4 (0.4-11.9)	14	47	29.8 (17.3-44.9)	26.3 (26.3-26.5)	8.6 (8.6-8.9)
Kleim et al,[68] 2014	PHQ-9	≥5	3 mo	12	47	25.5 (13.9-40.4)	20	47	42.6 (28.3-57.8)	17.0 (17.0-17.3)	1.7 (1.7-1.7)
Wada et al,[59] 2007	CES-D	≥19	1 y	16	62	25.8 (15.5-38.5)	12	46	26.1 (14.3-41.1)	0.3 (0.1-0.5)	1.0 (1.0-1.0)
Sen et al,[4] 2010	PHQ-9	≥10	1 y	29	740	3.9 (2.6-5.6)	190	740	25.7 (22.6-29.0	21.8 (21.8-21.8)	6.6 (6.6-6.6)
Ito et al,[70] 2015	CES-D	≥16	3 mo	189	1209	15.6 (13.6-17.8)	238	1020	23.3 (20.8-26.1)	7.7 (7.7-7.7)	1.5 (1.5-1.5)
Sen et al,[66] 2013	PHQ-9	≥10	1 y	86	2323	3.7 (3.0-4.6)	454	2323	19.5 (18.0-21.2)	15.8 (15.8-15.8)	5.3 (5.3-5.3)

Abbreviations: BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; PHQ-9, 9-item Patient Health Questionnaire.

P = .99), age (Q = 1.3, P = .26), or respondent/nonrespondent comparability (Q = 0, P = .99).

Secondary Analysis of Longitudinal Studies

In a secondary analysis of 7 longitudinal studies,[4,58,59,66-68,70] the temporal relationship between exposure to residency training and increased depressive symptoms was assessed (**Table 3**). Because studies used different assessment instruments, the relative change in depressive symptoms was calculated for each study individually (ie, follow-up divided by baseline prevalence), and then the relative changes derived from individual studies were meta-analyzed. Overall, the median absolute increase in depressive symptoms with the onset of residency training was 15.8% (range, 0.3%-26.3%; relative risk, 4.5).

Assessment of Publication Bias

Although visual inspection of the funnel plot revealed relatively minimal asymmetry (eFigure in the Supplement), there was evidence of small studies effect (Egger test P = .02), with

smaller studies (<200 participants) reporting more extreme depression prevalence estimates than larger studies (32.0% [95% CI, 27.1%-37.4%] vs 24.5% [95% CI, 20.0%-29.7%]; $Q = 4.2$, $P = .04$) (Figure 5).

Discussion

This systematic review and meta-analysis of 54 studies involving 17 560 physicians in training demonstrated that between 20.9% and 43.2% of trainees screened positive for depression or depressive symptoms during residency. Because the development of depression has been linked to a higher risk of future depressive episodes and greater long-term morbidity, these findings may affect the long-term health of resident doctors.[84,85] Depression among residents may also affect patients, given established associations between physician depression and lower-quality care.[6-8] These findings highlight an important issue in graduate medical education.

In interpreting the results of this meta-analysis, it is important to note that the vast majority of participants were assessed through self-report inventories that measured depressive symptoms, rather than gold-standard diagnostic clinical interviews for major depressive disorder. The sensitivity and specificity of these instruments for diagnosing major depressive disorder vary substantially (eTable 4 in the Supplement).[86] Instruments such as the 2-item PRIME-MD have low specificity (66%, 95% CI, 48%-84%) and should be viewed as screening tools. In contrast, other commonly used instruments, such as the PHQ-9, have high sensitivity (88%, 95% CI, 74%-96%) and specificity (88%, 95% CI, 85%-90%) for diagnosing major depressive disorder and have been shown to be comparable with clinician-administered assessments. Furthermore, although self-report measures of depressive symptoms have limitations, there is evidence that among medical trainees the absence of anonymity in formal diagnostic assessments may compromise accurate assessment of sensitive personal information such as depressive symptoms.[87] To reflect the heterogeneity of the measures included in this meta-analysis, a range of prevalence estimates (ie, 20.9%-43.2%) was reported in addition to a single measure (ie, 28.8%).

This study found an increase in depressive symptoms among residents over time that in part explained the heterogeneity between studies. This increase, while modest, is notable given efforts by the Accreditation Council for Graduate Medical Education,[88] European Working Time Directive,[89] and others[90] to limit trainee duty hours and improve work conditions. The identified trend may reflect the medical community's increased awareness of depression or developments external to medical education.[91] Future studies should explore specific factors that may explain this trend.

A secondary analysis restricted to longitudinal studies found a significant increase in depressive symptoms among trainees after the start of residency. The median absolute increase in depressive symptoms among trainees was 15.8% (range, 0.3%-26.3%) within a year of beginning training. This finding, in combination with evidence that the prevalence of depressive symptoms is similar across specialties and countries, suggests that the underlying causes of depressive symptoms are common to the residency experience. Identifying the factors that negatively affect trainee mental health may help inform the development of effective interventions for the reduction of depression that would be generalizable to different countries and specialties.

Variation in study sample size contributed importantly to the observed heterogeneity in the data. Studies with fewer participants generally yielded more extreme prevalence estimates, suggesting the presence of publication bias. Furthermore, some studies used screening instruments in nonstandard ways (eg, with cutoff scores that have not been validated). These variations were captured in part by Newcastle-Ottawa score, which assessed the risk of bias in each study. Studies with higher risk of bias yielded higher prevalence estimates of depressive symptoms. Study design (ie, cross-sectional vs longitudinal), country, survey years, specialty, postgraduate level, age, and sex also contributed to the heterogeneity between studies.

Limitations should be considered when interpreting the findings of this study. First, a substantial amount of the heterogeneity among the studies remained unexplained by the variables examined. Unexamined factors, such as the institutional cultures of specific residency programs, may contribute to the risk for depressive symptoms among trainees. A better understanding of program culture and working environments may help elucidate some of the root causes of depressive symptoms. Second, the data were derived from studies that used different designs and involved different groups of trainees (eg, from different countries, specialties, and years of training). For example, all but 3 studies used screening tools to measure depressive symptoms, and the 3 that employed structured interviews used convenience samples not representative of the resident population at large. Because the studies were heterogeneous with respect to screening inventories and resident populations, the prevalence of major depressive disorder could not be precisely determined. However, a secondary meta-analysis of studies using validated, high-specificity (>88%) inventories involving 5425 participants yielded a prevalence of 20.2%, which may better reflect the true prevalence of major depression. Third, the analysis relied on aggregated published data. A multicenter prospective study using a single validated measure of depression and structured diagnostic interviews in a random subset of participants would provide a more accurate estimate of the prevalence of depression among physicians in training.

Conclusions

In this systematic review, the summary estimate of the prevalence of depression or depressive symptoms among resident physicians was 28.8%, ranging from 20.9% to 43.2% depending on the instrument used, and increased with time. Further research is needed to identify effective strategies for preventing and treating depression among physicians in training.

ARTICLE INFORMATION

Author Contributions: Dr Mata had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Mata.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Mata, Ramos.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Mata, Bansal, Di Angelantonio.
Obtained funding: Guille, Sen.
Administrative, technical, or material support: Guille, Sen.
Study supervision: Guille, Sen.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This work was supported in part by a US Department of State Fulbright Scholarship (D.A.M.), National Institutes of Health (NIH) funding (R01MH101459 to S.S.), and NIH Medical Scientist Training Program funding (TG 2T32GM07205 to M.A.R.).

Role of the Funder/Sponsor: The study funders had no 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.

Disclaimer: The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources.

REFERENCES

1. Schneider SE, Phillips WM. Depression and anxiety in medical, surgical, and pediatric interns. *Psychol Rep*. 1993;72(3 pt 2):1145-1146.

2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602.

3. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Acad Med*. 2006;81(4):354-373.

4. Sen S, Kranzler HR, Krystal JH, et al. A prospective cohort study investigating factors associated with depression during medical internship. *Arch Gen Psychiatry*. 2010;67(6):557-565.

5. Joules N, Williams DM, Thompson AW. Depression in resident physicians: a systematic review. *Open J Depress*. 2014;03(03):89-100.

6. West CP, Huschka MM, Novotny PJ, et al. Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *JAMA*. 2006;296(9):1071-1078.

7. Fahrenkopf AM, Sectish TC, Barger LK, et al. Rates of medication errors among depressed and burnt out residents: prospective cohort study. *BMJ*. 2008;336(7642):488-491.

8. West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD. Association of resident fatigue and distress with perceived medical errors. *JAMA*. 2009;302(12):1294-1300.

9. Buddeberg-Fischer B, Klaghofer R, Buddeberg C. Stress at work and well-being in junior residents [in German]. *Z Psychosom Med Psychother*. 2005;51(2):163-178.

10. Yousuf A, Ishaque S, Qidwai W. Depression and its associated risk factors in medical and surgical post graduate trainees at a teaching hospital: a cross sectional survey from a developing country. *J Pak Med Assoc*. 2011;61(10):968-973.

11. Hsu K, Marshall V. Prevalence of depression and distress in a large sample of Canadian residents, interns, and fellows. *Am J Psychiatry*. 1987;144(12):1561-1566.

12. Kirsling RA, Kochar MS, Chan CH. An evaluation of mood states among first-year residents. *Psychol Rep*. 1989;65(2):355-366.

13. Hsieh Y-H, Hsu C-Y, Liu C-Y, Huang T-L. The levels of stress and depression among interns and clerks in three medical centers in Taiwan: a cross-sectional study. *Chang Gung Med J*. 2011;34(3):278-285.

14. Shanafelt TD, Sloan JA, Habermann TM. The well-being of physicians. *Am J Med*. 2003;114(6):513-519.

15. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med*. 2009;3(3):e123-e130.

16. Kerr LK, Kerr LD Jr. Screening tools for depression in primary care: the effects of culture, gender, and somatic symptoms on the detection of depression. *West J Med*. 2001;175(5):349-352.

17. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*. 2010;25(9):603-605.

18. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111.

19. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.

20. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.

21. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med*. 2002;21(11):1513-1524.

22. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002;21(4):589-624.

23. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.

24. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001;54(10):1046-1055.

25. The R project for statistical computing. http://www.R-project.org/. Accessed November 6, 2015.

26. Waring EM. Emotional illness in psychiatric trainees. *Br J Psychiatry*. 1974;125(0):10-11.

27. Valko RJ, Clayton PJ. Depression in the internship. *Dis Nerv Syst*. 1975;36(1):26-29.

28. Steinert Y, Magonet G, Rubin G, Carson K. Emotional well-being of house staff: comparison of residency training programs. *Can Fam Physician*. 1991;37:2130-2138.

29. Godenick MT, Musham C, Palesch Y, Hainer BL, Michels PJ. Physical and psychological health of family practice residents. *Fam Med*. 1995;27(10):646-651.

30. Hainer BL, Palesch Y. Symptoms of depression in residents: a South Carolina Family Practice Research Consortium study. *Acad Med*. 1998;73(12):1305-1310.

31. Raviola G, Machoki M, Mwaikambo E, Good MJD. HIV, disease plague, demoralization and "burnout": resident experience of the medical profession in Nairobi, Kenya. *Cult Med Psychiatry*. 2002;26(1):55-86.

32. Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. *Ann Intern Med*. 2002;136(5):358-367.

33. Oriel K, Plane MB, Mundt M. Family medicine residents and the impostor phenomenon. *Fam Med*. 2004;36(4):248-252.

34. Earle L, Kelly L. Coping strategies, depression, and anxiety among Ontario family medicine residents. *Can Fam Physician*. 2005;51:242-243.

35. Becker JL, Milad MP, Klock SC. Burnout, depression, and career satisfaction: cross-sectional study of obstetrics and gynecology residents. *Am J Obstet Gynecol*. 2006;195(5):1444-1449.

36. Cruz EP. Burnout syndrome as a risk factor of depression in medical residents. *Med Interna Mex*. 2006;22(4):282-286.

37. Yi MS, Luckhaupt SE, Mrus JM, et al. Religion, spirituality, and depressive symptoms in primary care house officers. *Ambul Pediatr*. 2006;6(2):84-90.

38. Demir F, Ay P, Erbaş M, Ozdil M, Yaşar E. The prevalence of depression and its associated factors among resident doctors working in a training hospital in Istanbul [in Turkish]. *Turk Psikiyatri Derg*. 2007;18(1):31-37.

39. Hasanović M, Herenda S. Post traumatic stress disorder, depression and anxiety among family medicine residents after 1992-95 war in Bosnia and Herzegovina. *Psychiatr Danub*. 2008;20(3):277-285.

40. Sakata Y, Wada K, Tsutsumi A, et al. Effort-reward imbalance and depression in Japanese medical residents. *J Occup Health*. 2008;50(6):498-504.

41. Sánchez MJB, Islas CLR, Escobar IV, Rico LES. Symptoms of anxiety and depression in resident physicians at high risk of stress. *Psiquiatr Biol*. 2008;15(5):147-152.

42. Goebert D, Thompson D, Takeshita J, et al. Depressive symptoms in medical students and residents: a multischool study. *Acad Med*. 2009;84(2):236-241.

43. Waldman SV, Diez JCL, Arazi HC, Linetzky B, Guinjoan S, Grancelli H. Burnout, perceived stress,

and depression among cardiology residents in Argentina. *Acad Psychiatry*. 2009;33(4):296-301.

44. Lam TP, Wong JGWS, Ip MSM, Lam KF, Pang SL. Psychological well-being of interns in Hong Kong: what causes them stress and what helps them. *Med Teach*. 2010;32(3):e120-e126.

45. Costa EF, Santana YS, Santos AT, Martins LA, Melo EV, Andrade TM. Depressive symptoms among medical intern students in a Brazilian public university [in Portuguese]. *Rev Assoc Médica Bras*. 2012;58(1):53-59.

46. Govardhan LM, Pinelli V, Schnatz PF. Burnout, depression and job satisfaction in obstetrics and gynecology residents. *Conn Med*. 2012;76(7):389-395.

47. de Oliveira GS Jr, Chang R, Fitzgerald PC, et al. The prevalence of burnout and depression and their association with adherence to safety and practice standards: a survey of United States anesthesiology trainees. *Anesth Analg*. 2013;117(1):182-193.

48. Al-Ghafri G, Al-Sinawi H, Al-Muniri A, et al. Prevalence of depressive symptoms as elicited by Patient Health Questionnaire (PHQ-9) among medical trainees in Oman. *Asian J Psychiatr*. 2014;8:59-62.

49. Dyrbye LN, West CP, Satele D, et al. Burnout among US medical students, residents, and early career physicians relative to the general US population. *Acad Med*. 2014;89(3):443-451.

50. Stoesser K, Cobb NM. Self-treatment and informal treatment for depression among resident physicians. *Fam Med*. 2014;46(10):797-801.

51. Al-Maddah EM, Al-Dabal BK, Khalil MS. Prevalence of sleep deprivation and relation with depressive symptoms among medical residents in King Fahd University Hospital, Saudi Arabia. *Sultan Qaboos Univ Med J*. 2015;15(1):e78-e84.

52. Pereira-Lima K, Loureiro SR. Burnout, anxiety, depression, and social skills in medical residents. *Psychol Health Med*. 2015;20(3):353-362.

53. Ford CV, Wentz DK. The internship year: a study of sleep, mood states, and psychophysiologic parameters. *South Med J*. 1984;77(11):1435-1442.

54. Reuben DB. Depressive symptoms in medical house officers: effects of level of training and work rotation. *Arch Intern Med*. 1985;145(2):286-288.

55. Revicki DA, Gallery ME, Whitley TW, Allison EJ. Impact of work environment characteristics on work-related stress and depression in emergency medicine residents: a longitudinal study. *J Community Appl Soc Psychol*. 1993;3(4):273-284.

56. Gopal R, Glasheen JJ, Miyoshi TJ, Prochazka AV. Burnout and internal medicine resident work-hour restrictions. *Arch Intern Med*. 2005;165(22):2595-2600.

57. Katz ED, Sharp L, Ferguson E. Depression among emergency medicine residents over an academic year. *Acad Emerg Med*. 2006;13(3):284-287.

58. Rosen IM, Gimotty PA, Shea JA, Bellini LM. Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns. *Acad Med*. 2006;81(1):82-85.

59. Wada K, Sakata Y, Fujino Y, et al. The association of needlestick injury with depressive symptoms among first-year medical residents in Japan. *Ind Health*. 2007;45(6):750-755.

60. Landrigan CP, Fahrenkopf AM, Lewin D, et al. Effects of the Accreditation Council for Graduate Medical Education duty hour limits on sleep, work hours, and safety. *Pediatrics*. 2008;122(2):250-258.

61. Buddeberg-Fischer B, Stamm M, Buddeberg C, Klaghofer R. Anxiety and depression in residents: results of a Swiss longitudinal study [in German]. *Z Psychosom Med Psychother*. 2009;55(1):37-50.

62. Campbell J, Prochazka AV, Yamashita T, Gopal R. Predictors of persistent burnout in internal medicine residents: a prospective cohort study. *Acad Med*. 2010;85(10):1630-1634.

63. Beckman TJ, Reed DA, Shanafelt TD, West CP. Resident physician well-being and assessments of their knowledge and clinical performance. *J Gen Intern Med*. 2012;27(3):325-330.

64. Weigl M, Hornung S, Petru R, Glaser J, Angerer P. Depressive symptoms in junior doctors: a follow-up study on work-related determinants. *Int Arch Occup Environ Health*. 2012;85(5):559-570.

65. West CP, Tan AD, Shanafelt TD. Association of resident fatigue and distress with occupational blood and body fluid exposures and motor vehicle incidents. *Mayo Clin Proc*. 2012;87(12):1138-1144.

66. Sen S, Kranzler HR, Didwania AK, et al. Effects of the 2011 duty hour reforms on interns and their patients: a prospective longitudinal cohort study. *JAMA Intern Med*. 2013;173(8):657-662.

67. Velásquez-Pérez L, Colin-Piana R, González-González M. Coping with medical residency: depression burnout [in Spanish]. *Gac Med Mex*. 2013;149(2):183-195.

68. Kleim B, Thörn HA, Ehlert U. Positive interpretation bias predicts well-being in medical interns. *Front Psychol*. 2014;5:640.

69. Cubero DIG, Fumis RRL, de Sá TH, et al. Burnout in medical oncology fellows: a prospective multicenter cohort study in Brazilian institutions [published online May 9, 2015]. *J Cancer Educ*. doi:10.1007/s13187-015-0850-z.

70. Ito M, Seo E, Ogawa R, Sanuki M, Maeno T, Maeno T. Can we predict future depression in residents before the start of clinical training? *Med Educ*. 2015;49(2):215-223.

71. Jiménez-López JL, Arenas-Osuna J, Angeles-Garay U. Depression, anxiety and suicide risk symptoms among medical residents over an academic year [in Spanish]. *Rev Med Inst Mex Seguro Soc*. 2015;53(1):20-28.

72. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561-571.

73. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385-401.

74. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. 1999;282(18):1737-1744.

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

76. Biggs JT, Wylie LT, Ziegler VE. Validity of the Zung Self-rating Depression Scale. *Br J Psychiatry*. 1978;132:381-385.

77. Baer L, Jacobs DG, Meszler-Reizes J, et al. Development of a brief screening instrument: the HANDS. *Psychother Psychosom*. 2000;69(1):35-41.

78. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56-62.

79. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci*. 1974;19(1):1-15.

80. Spielberger CD. *Manual for the State-Trait Personality Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press; 1983.

81. Wilkinson MJB, Barczak P. Psychiatric screening in general practice: comparison of the general health questionnaire and the hospital anxiety depression scale. *J R Coll Gen Pract*. 1988;38(312):311-313.

82. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behav Res Ther*. 1997;35(1):79-89.

83. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.

84. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust*. 2009;190(7)(suppl):S54-S60.

85. Glassman AH, Shapiro PA. Depression and the course of coronary artery disease. *Am J Psychiatry*. 1998;155(1):4-11.

86. Williams JW Jr, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry*. 2002;24(4):225-237.

87. Myers M. On the importance of anonymity in surveying medical student depression. *Acad Psychiatry*. 2003;27(1):19-20.

88. Accreditation Council for Graduate Medical Education. Report of the ACGME Work Group on Resident Duty Hours. June 11, 2002.

89. Organisation of working time: Council Directive 93/104/EC. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31993L0104. Accessed November 6, 2015.

90. Final report on doctor work reform. Steering Committee on Doctor Work Hour, Hospital Authority. https://www.ha.org.hk/haho/ho/hesd/FinalReportonDWR_2009_10_Full_Version.pdf. Accessed November 6, 2015.

91. Daskivich TJ, Jardine DA, Tseng J, et al. Promotion of wellness and mental health awareness among physicians in training: perspective of a national, multispecialty panel of residents and fellows. *J Grad Med Educ*. 2015;7(1):143-147.