Prevalence and severity of psychological distress among early-career researchers: a systematic review and multilevel meta-analysis

Aljoscha Dreisoerner[1, 4], Vanessa Goetz[2,4], David Frohnmayer[2], Ulrich S. Tran[1,2], Martin Voracek[1,2], Urs M. Nater[1,3]

1) Research Platform: The Stress of Life (SOLE) – Processes and Mechanisms underlying Everyday Life Stress, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

2) Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

3) Department of Clinical and Health Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

4) These authors contributed equally: Aljoscha Dreisoerner, Vanessa Goetz

***Corresponding Authors:**
Dr. Aljoscha Dreisoerner & Dr. Urs M. Nater
Research Platform: The Stress of Life (SOLE) – Processes and Mechanisms underlying Everyday Life Stress, University of Vienna
Liebiggasse 5, 1010 Vienna, Austria
Email address: aljoscha.dreisoerner@univie.ac.at; urs.nater@univie.ac.at
Tel: +43-1-4277-47238
ORCID: 0000-0003-0786-7267; 0000-0002-2430-5090

This paper has not yet undergone peer review.

Abstract

Mental health concerns among early-career researchers (ECRs) have gained attention, but prevalence and severity estimates remain fragmented. We conducted a pre-registered systematic review and multilevel meta-analysis to quantify the burden of psychological distress in this population (PROSPERO: CRD42024596813). A total of 139 studies and four primary databases were identified (k = 212, N = 128,684) via Scopus and supplementary searches through 31 December 2024. Based on established diagnostic thresholds, pooled prevalence estimates indicated that 28.9% (95% confidence interval 25.3–32.7%) of ECRs met the criteria for at least one mental health syndrome or symptom. Specifically, 28.9% (25.4–32.5%) met criteria for depression, 29.1% (25.4–32.9%) for generalized anxiety disorder, 24.9% (20.6–29.4%) for an eating disorder, 21.4% (17.2–26.0%) for alcohol abuse, 18.2% (14.2–22.6%) engaged in nonsuicidal self-injury, and 17.5% (14.2–21.0%) reported suicidal ideation. Complementing prevalence estimates, analyses of normalized severity scores indicated that ECRs experienced mild degrees of depression, anxiety, and moderate degrees of stress. Severity estimates for eating disorders and suicidal ideation did not exceed clinical cut-offs. Moderator analyses indicated that the type of syndrome and the type of measurement scale explained the largest part of the variance in both the prevalence and the severity models. Most individual-level factors, such as gender, ethnicity, academic discipline, country, career stage, social support, and supervisor relationship, did not significantly moderate prevalence or severity estimates. Age showed a small, but significant association with lower prevalence (β = -0.01, CI -0.03 – -0.0009). Higher study quality and random or full sampling were also associated with lower severity and prevalence levels of psychological distress. These findings underscore both the need for structural reforms and for the implementation of evidence-based support services and interventions for ECRs.

Keywords: mental health, early-career researchers, psychological distress, meta-analysis, systematic review, depression, anxiety

Mental health concerns among early-career researchers (ECRs), including predoctoral and early postdoctoral researchers, have been gaining attention and point towards a potential mental health crisis in academia[1–7]. While ECRs generally report high job satisfaction and intrinsic motivation[8], recent studies suggest that this population experiences a disproportionate degree of psychological distress, including clinically defined syndromes such as depression, general anxiety disorder (GAD), and burnout, or individual symptoms such as stress or suicidal ideation[1,9–11]. ECRs cite issues such as short-term employment contracts and related job insecurity, a high degree of competition, perceived isolation, lack of feedback, and power abuse or discrimination as factors contributing to poor mental health[12–17]. These structural and individual stressors are accompanied by several concerning trends. Doctoral attrition remains at alarmingly high rates, with estimates between 30 and 50% of doctoral candidates leaving their programs before graduating[18–21], pointing towards wasted research potential and associated mental health issues. Individual studies further suggest that a significant portion of doctoral students have contemplated suicide[22,23]. In response to these trends, awareness and interest in mental health issues among ECRs have increased in recent years, prompting some universities, governments, and professional organizations to call for reforms and increased support for ECRs[24–26]. However, despite this increase in awareness, the full extent of psychological distress among ECRs is not known, as several important gaps remain open.

Existing meta-analyses estimate that 21–24% of the predoctoral researchers meet the clinical cut-off for depression and 17–29% for GAD, with average symptom severity falling in the mild range[9–11]. However, there is also variability in the severity of symptoms. One meta-analysis of doctoral students reported that 36% experience mild, 12% moderate, and 9% severe depressive symptoms[10]. Another meta-analysis of doctoral students found that 22.9% experience mild, 16.5% moderate, and 13.5% severe anxiety symptoms[23]. Despite the growing literature on this topic[27–34], most reviews and meta-analyses have focused on depression, anxiety, or suicidal ideation, with prevalence and severity estimates limited to depression[10] and GAD[9]. A broader systematic evaluation of other relevant mental health issues remains lacking, including clinically defined syndromes such as social phobia, obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), attention-deficit/hyperactivity disorder (ADHD), personality disorders, eating disorders (ED), alcohol and drug abuse, and burnout, as well as common symptoms of psychological distress such as stress, sleep disturbance, nonsuicidal self-injury (NSSI), suicidal

ideation, and somatic symptoms. In addition, there was a surge in research on mental health in ECRs during and after the COVID-19 pandemic, and no meta-analysis has synthesized these newer studies. For example, the most recent estimates on depression include data up to November 2019[11]. It also remains unknown whether psychological distress among ECRs increased during the pandemic or how it evolved in the post-pandemic period.

The high prevalence of psychological distress among ECR underscores the urgent need for a systematic investigation of individual study-level factors that moderate their prevalence and severity. To date, only two meta-analyses have empirically examined such moderators—and only in relation to depression[10] and anxiety[9]. These studies identified several *study-level moderators*, showing that prevalence estimates were generally lower in studies with random sampling and higher study quality, and that both prevalence and severity estimates varied depending on which type of measurement scale was used. In contrast, *sample-level moderators* such as country income level, gender, and calendar year showed no consistent effects. Systematic and narrative reviews highlight additional sample-level moderators, including protective and risk factors[27–30,32,34–36]. These include financial challenges[37], self-doubt[38], social and academic loneliness[14,39] as risk factors, and suggest social support, a positive and future-oriented mindset, and self-care[14] as potential protective factors. Supervisor relationships have been described as either a risk or protective factor depending on their quality[14,40]. Despite these insights, many potentially important sample-level moderators, including demographic characteristics (e.g., age, ethnicity), contextual variables (country, academic discipline, career stage), and individual factors (pre-existing mental health diagnoses, life or job satisfaction), have yet to be systematically assessed across the broader spectrum of mental health outcomes beyond depression and anxiety. Evaluating these moderators is critical for understanding psychological distress in ECRs, and for developing evidence-based interventions.

This systematic review and multi-level meta-analysis[41] aimed to assess the extent of mental health issues among ECRs, including the severity and prevalence of a broad range of mental health syndromes and symptoms—subsumed under the umbrella term ‘psychological distress’. Our analysis includes both predoctoral and postdoctoral researchers and provides pooled estimates of prevalence and severity across a spectrum of conditions in addition to depression and anxiety (e.g., social phobia, OCD, PTSD, ADHD, personality disorders, ED, alcohol and drug abuse, and burnout) and individual symptoms (e.g., stress, sleep disturbance,

NSSI, suicidal ideation, and somatic symptoms). We also systematically evaluated the moderating role of both individual-level factors (gender, age, career stage, country, ethnicity, academic discipline, pre-existing mental health diagnosis, job satisfaction, life satisfaction, supervisor relationship quality, social support, and loneliness) and study-level characteristics (study quality, sampling method, type of measurement scale).

Results

Study identification and selection

The PRISMA flowchart shows the study identification and selection process for the meta-analysis (Fig. 1). Of the 3,701 records screened, 139 records and four databases qualified for inclusion, yielding a total of 212 samples. More information on the respective samples and their characteristics and references can be found in Supplementary Table 1.

Characteristics of the studies

The final, cumulative sample consisted of 128,684 individuals (125,642 doctoral and 2,690 post-doctoral ECRs). The majority of the data were collected between 2018 and 2024, with most data having been collected in 2020 during the COVID-19 pandemic (Supplementary Fig. 1). The overall sample was 63% female with a mean age of 30.4 years (SD = 6.1). Reported ethnicities included Caucasian (61%), Asian (22%), Latin (9%), and Black (8%). Doctoral fields were grouped according to broader categories of academic disciplines (Supplementary Table 2), with social sciences representing the largest share (33%), followed by natural sciences (27%), applied sciences (26%), the humanities (11%), and formal sciences (3%). Within the social sciences, psychology accounted for 32% of samples. Among doctoral studies in applied sciences, medicine represented the second biggest category (13%), followed by engineering (6%), communication science (6%), and lastly law (2%).

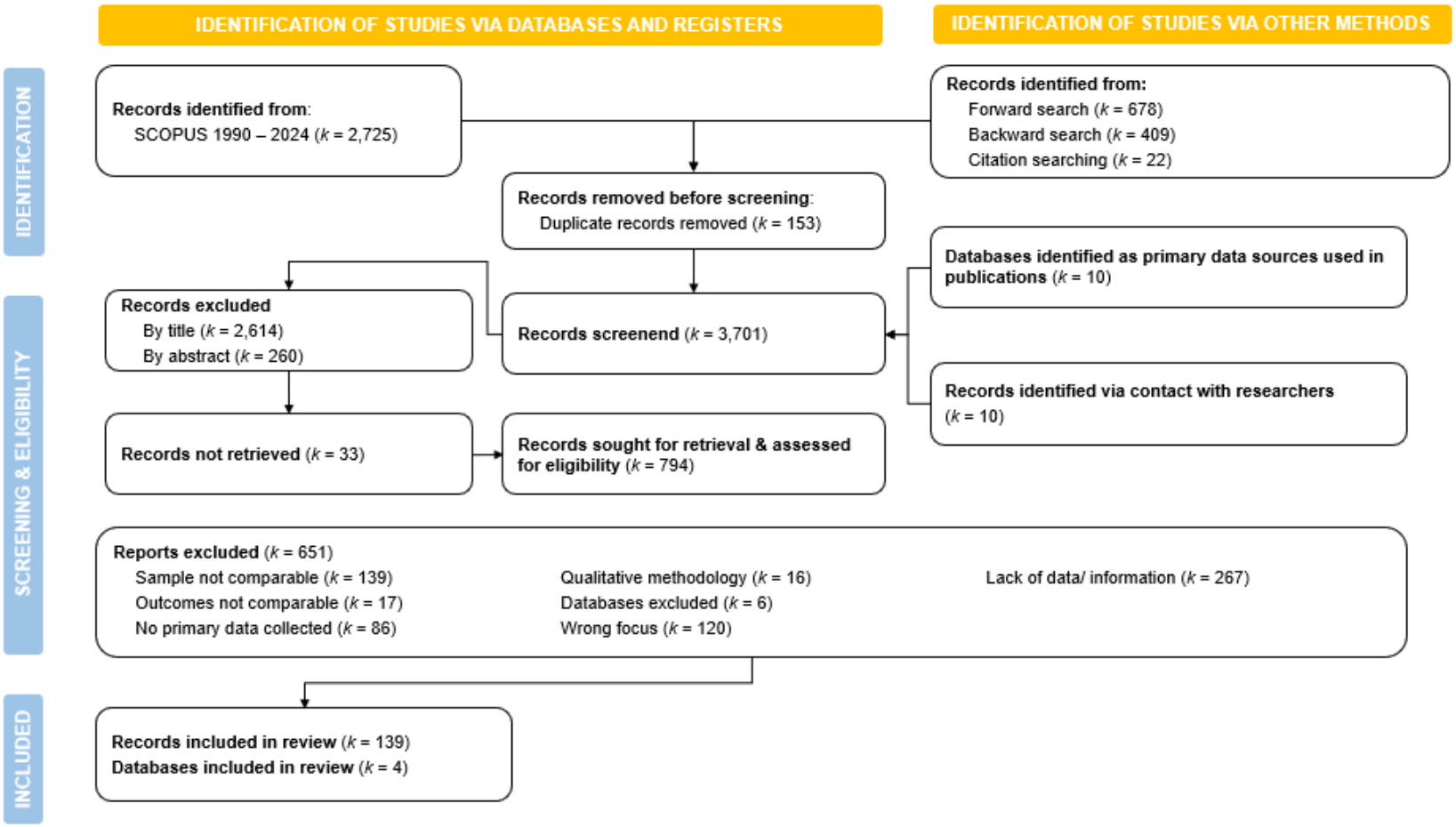
Countries were clustered based on similar working conditions for ECRs. These clusters were created using generative artificial intelligence tools (see Supplementary Table 3 for methodological details and for a full list of countries per cluster). Most samples (103 samples, 49%) originated from countries with flexible, grant-dependent research careers (Cluster 1), such as the United States (k = 88) and the United Kingdom (k = 13). The second largest cluster (Cluster 2, k = 41, 19%) included countries with high government support, structured contracts and career security (Cluster 2), such as France (k = 13), Germany (k = 10), and the Netherlands (k = 5). Cluster 3 (k = 29, 14%) included countries with self-funded or low-funding systems,

such as Italy (k = 8) and Spain (k = 7). Cluster 4 (k = 22, 10%) included countries with emerging research systems with variable funding sources and career structures, such as China (k = 7), or Brazil (k = 6). Cluster 5 (k = 14, 7%) served as a catch-all category for countries that did not fit any of the other categories. Last, 3 samples (1%) originated from Cluster 6, characterized by competitive fellowships with limited benefits, and short-term contracts, including Japan (k = 1) and Hong Kong (k = 2).

Most studies collected data on depression (k = 92), GAD (k = 78), stress (k = 55), psychological distress in general (k = 33), burnout (k = 21), suicidal ideation (k = 17), sleep quality (k = 7), alcohol abuse (k = 6), somatic symptoms (k = 4), eating disorders and drug abuse (k = 3 each), ADHD, bipolar disorder, NSSI (k = 2 each), PTSD and personality disorders (k = 1 each). The prevalence and severity models were based on overlapping, but not identical, sets of studies, and consequently, their study characteristics also differed. A detailed comparison of the study characteristics between the models can be found in Supplementary Table 4. In addition, reporting on sample-level demographics and other moderators was inconsistent among the studies included in the prevalence and severity models, with missing data variables such as gender, age, ethnicity, and academic discipline, ranging from 5% to 58% (see Supplementary Table 5).

Figure 1.

PRISMA 2020 Flow diagram of the literature search and the screening of records

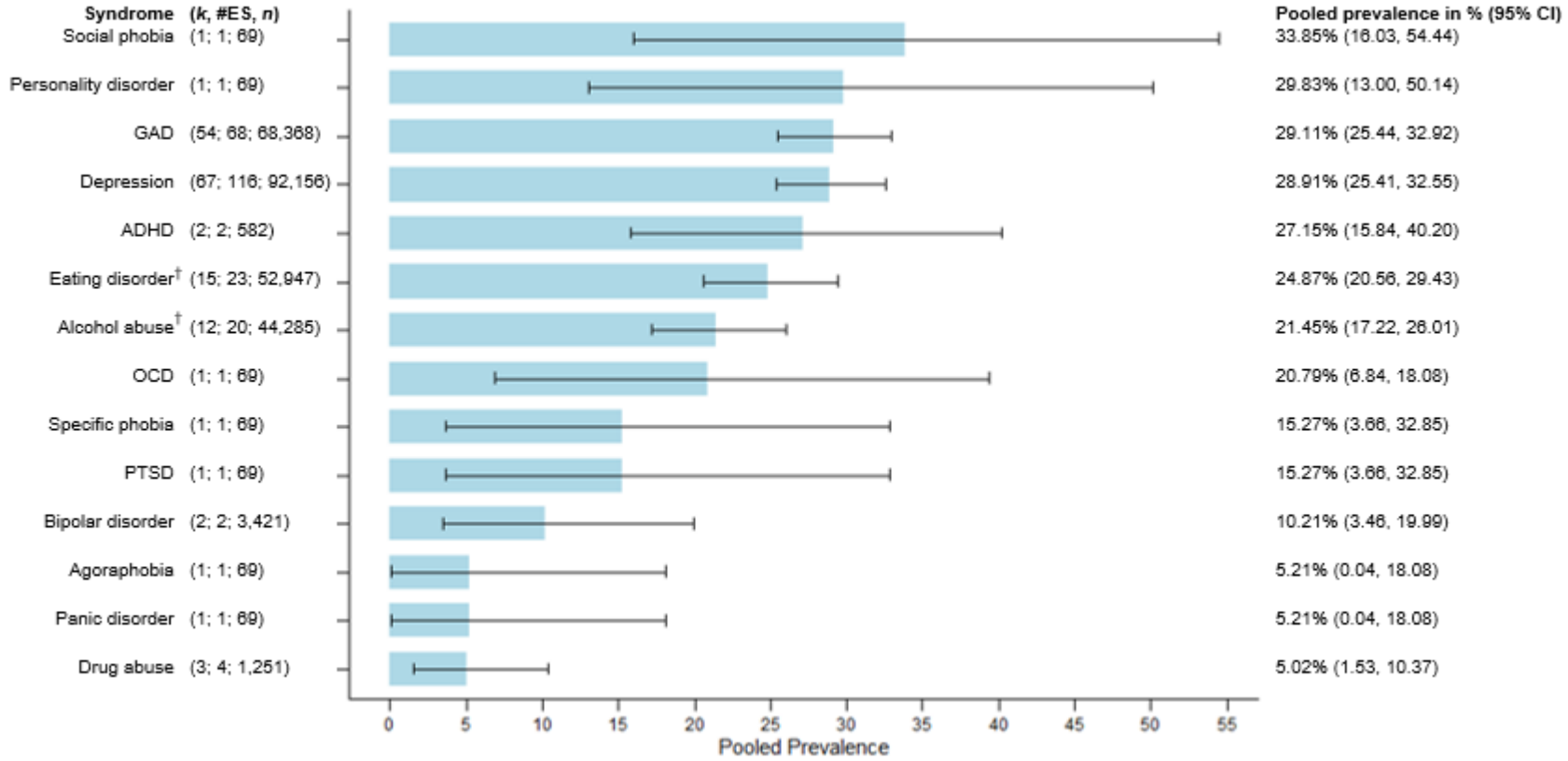


Analytic procedures

We fitted two multi-level models, one for prevalence and one for severity. The prevalence model was based on outcomes with continuous measures for which there were psychometrically validated or commonly used clinical thresholds (e.g., PHQ-9 scores above 10), or binary thresholds (yes/no) response formats (Supplementary Tables 1, 10–18). Severity was analyzed with the percent of maximum possible score (POMP) method, which transforms continuous measures into a normalized scale from 0 to 100 (see Methods for detail). For example, a POMP score of 50 represents the mean value on the original scale and POMP scores above 50 indicate higher and below 50 indicate lower symptom severity. Both models included data on alcohol abuse, drug abuse, ADHD, depression, GAD, ED, sleep quality, general mental health, somatic symptoms, and suicidal ideation. The severity model additionally covered stress and burnout, including the subscales of emotional exhaustion, depersonalization, and personal accomplishment as subfactors of burnout. The prevalence model also included bipolar disorder, personality disorder, PTSD, OCD, social and specific phobia, agoraphobia, panic disorder, and NSSI. To assess the effect of potential moderators, such as participant and study characteristics, syndrome or symptom type, type of measurement scale, study quality, and POMP scores of risk and protective factors, we conducted univariate moderator analyses. To account for multiple testing, the false discovery rate (FDR) was controlled at 5%. In the subsequent multivariate moderator analysis for the severity and prevalence model, only moderators remained that were significant after controlling for the FDR (see Methods for details). All confidence intervals reported in this paper are 95% confidence intervals (CI), and all tests were two-tailed. More information on the moderators and adjusted *p*-values can be found in Supplementary Tables 19 and 20.

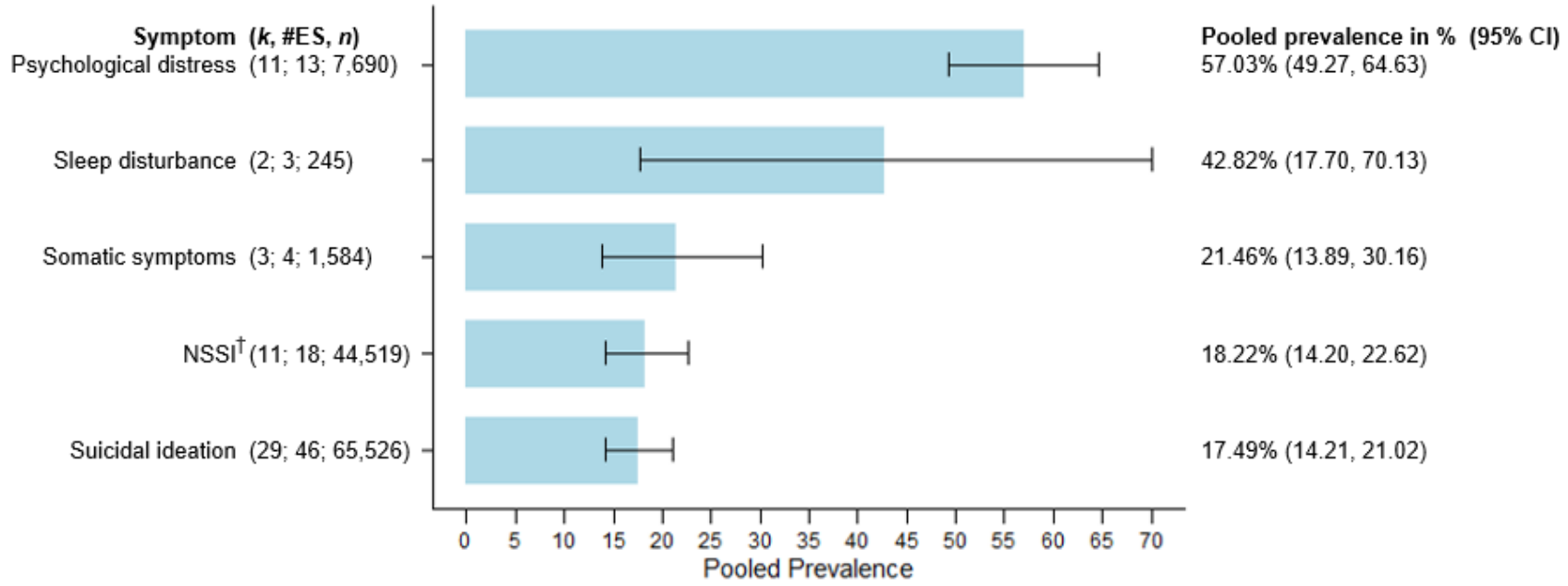
Figure 2.

(A) Prevalence rates by syndrome



(B)

(B) Prevalence rates by symptom



Note. All of the syndromes were significant on a $p < .001$ level, except panic disorder and agoraphobia which were significant on a $p < .05$ level; *k* = number of studies included in the analysis; #ES = number of effect sizes included in the analysis; *n* = cumulative sample included in the analysis per moderator; GAD = general anxiety disorder; ADHD = attention-deficit hyper-activity disorder; OCD = obsessive compulsive disorder; PTSD = post-traumatic stress disorder; NSSI = non-suicidal self-injury; CI = 95% confidence interval; †pooled prevalence estimates stem predominantly from the Healthy Minds Network[42]. Psychological distress here refers to information on overall mental health and psychological distress that was reported in some studies.

Main results

Pooled prevalence and severity estimates

The pooled prevalence of psychological distress across all outcomes was 28.9% (95% CI, 25.3–32.7%; $k = 82$). The pooled severity of psychological distress across all outcomes was 34 (95% CI, 31.7–36.3; $k = 141$) on a scale from 0–100. Prevalence and severity estimates displayed substantial heterogeneity, with 95% prediction intervals ranging from 2.8–67.6% for prevalence and 1.1–67 for severity, indicating wide ranges into which the results of future individual studies can be expected. Forest and caterpillar plots for both models can be found in Supplementary Figures 3–9.

Severity estimates from the PHQ-9 and GAD-7

The most common measurement scales to assess depression and GAD were the PHQ-9 and GAD-7. Based on the PHQ-9, severity estimates indicate that 33.2% (95% CI, 30.7–35.8) of ECRs reported mild symptoms, 17.3% (95% CI, 15.3–19.3) moderate, 8.2% (95% CI, 6.8–9.8) moderately severe, and 4.8% (95% CI, 3.7–6.1) severe depressive symptoms. Similarly, for GAD, 30.9% (95% CI, 28.2–33.6) reported mild, 16.1% (95% CI, 14.0–18.3) moderate, and 9.6% (95% CI, 7.9–11.4) severe anxiety.

Moderators

Pooled prevalence and severity estimates by syndromes and symptoms

Moderator analyses showed that the type of syndrome and symptom influenced both prevalence and severity estimates. In the prevalence model, syndrome or symptom type explained 27.37% of the variance ($F[18, 307] = 13.11$, adjusted $p = .006$). Prevalence estimates varied strongly between the types of syndromes or symptoms, but all deviated significantly from zero (Fig. 2). In the severity model, syndrome or symptom type explained 38.97% of true effect size heterogeneity ($F[11, 477] = 25.01$, adjusted $p = .004$).

In exploratory analyses, we examined the severity across the three different burnout subscales (without FDR correction). Depersonalization showed the lowest severity (2.2, 95% CI, -6.8–11.2), followed by emotional exhaustion (11.1, 95% CI, 3.4–18.8), while reduced personal accomplishment had the highest severity (23.0, 95% CI, 12.3–33.8). These results suggest that ECRs struggle most with the feeling of being ineffective or lacking personal accomplishment.

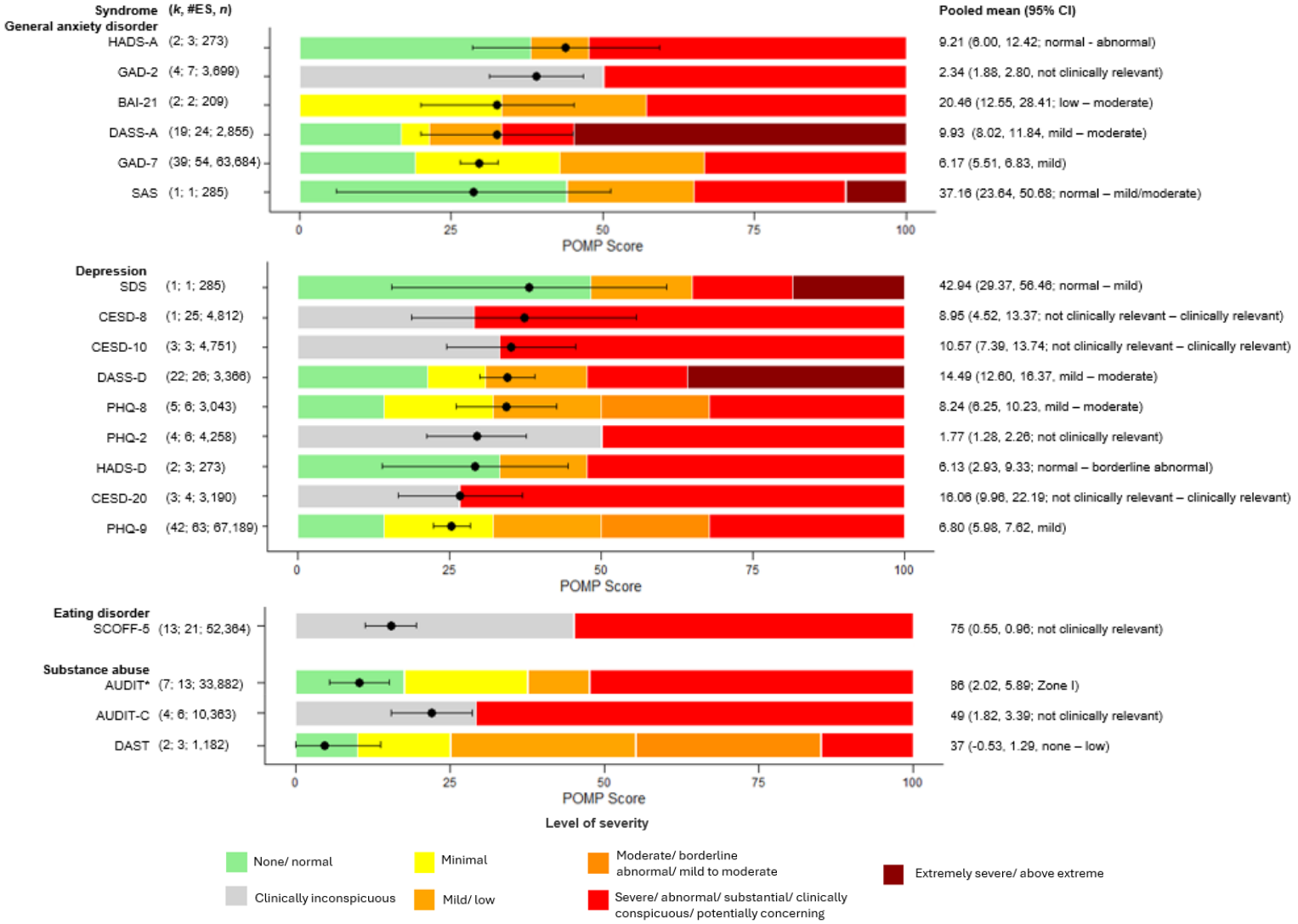
The moderating effects of the type of measurement scale used

The type of measurement scale to assess various mental health syndromes and symptoms significantly moderated estimates of both prevalence (F[49, 276] = 11.07, adjusted p = .006) and severity (F[90, 398] = 9.69, adjusted p = .004), explaining 44.9% and 50.9% of the true effect size heterogeneity for prevalence and severity, respectively. As shown in Fig. 3, severity levels varied widely depending on which measurement scale was used. Whereas severity estimates for depression ranged from mild to severe, and for GAD from mild and borderline abnormal, assessments of ED and substance abuse were consistently below clinical thresholds. At a symptom level, moderate levels of stress and psychological distress[1] were observed. However, scores on broad psychological distress screening instruments did not conclusively exceed clinical cut-offs. Likewise, suicidal ideation did not exceed the clinical cut-off. However, clinical thresholds for suicidal ideation have been increasingly criticized, as systematic reviews show that these cut-off scores have low predictive value of actual suicide. Importantly, subthreshold levels of suicidal ideation still include suicidal thoughts, which can be clinically relevant[43,44]. Sleep quality ranged between unproblematic and poor, though additional data is needed for more precise estimates. Full effect size estimates per scale are available in Supplementary Table 21 and 22, including comparison to the most recent meta-analyses on depression and GAD. Scale references and severity thresholds per scale are available in Supplementary Tables 10-18.

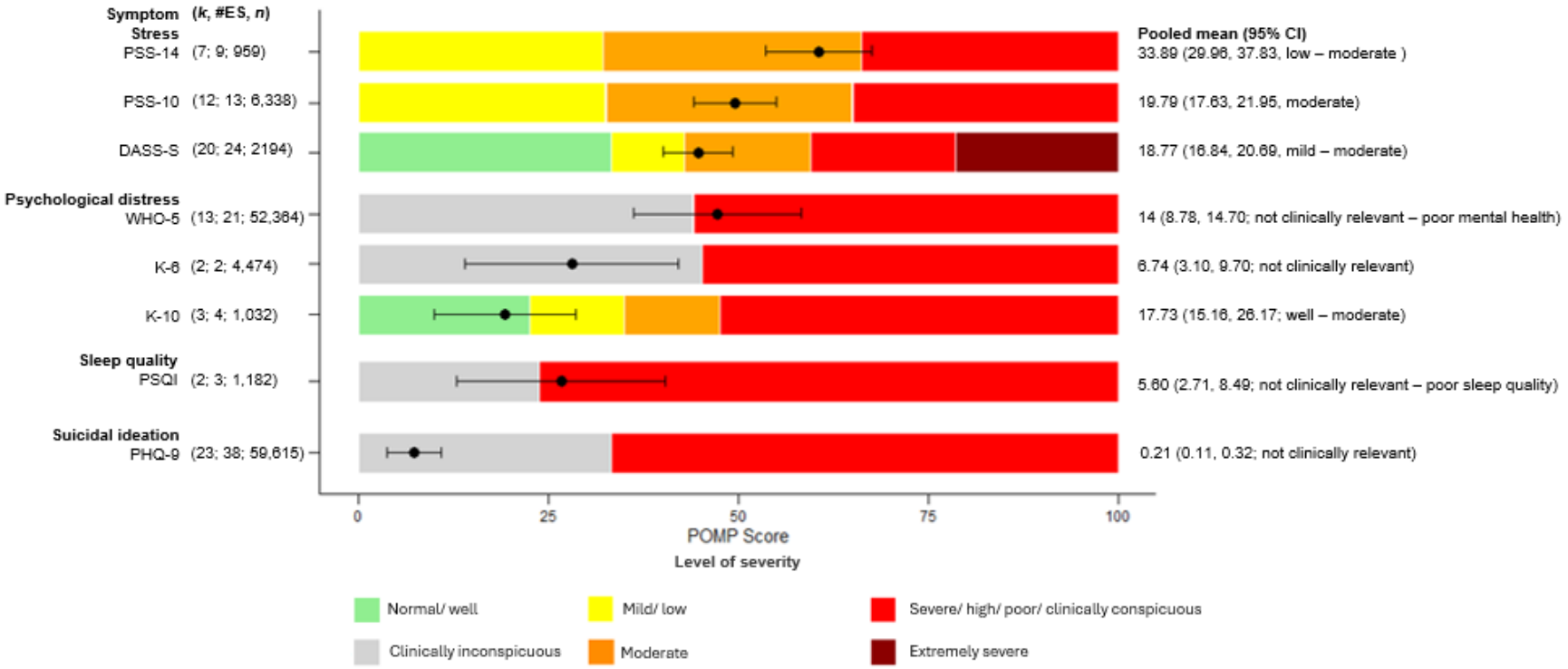
[1] Here, psychological distress is referred to the outcomes of broad screening tools from the WHO-5, K-6, and K-10 (rather than as an umbrella term for the rest of the meta-analysis).

Figure 3.

(A) Severity levels for the most frequently utilized scales per syndrome



(B) Severity levels for the most frequently utilized scales per symptom



Note. All of the scales used were significant at the $p < .001$ level, except for the DAST, which was not significant. Categorization of the pooled estimate and confidence interval was provided based on the respective scales, and category labelling was provided on the right-hand side. In case the confidence interval did not span more than one category, only one category label was provided. Estimates pertain to the most frequently used scales for each syndrome in cases where several alternative scales were encountered in primary studies. The WHO-5 (World Health Organization well-being index, 5-item version) has been reverse-scored to enable an interpretation in the same direction as K-6 and K-10, namely that an increase in points corresponds to an increase in the severity of psychological distress. k = number of studies included in the analysis; #ES = number of effect sizes included in the analysis; n = cumulative sample

included in the analysis per moderator; HADS-A = Hospital Anxiety and Depression scale, anxiety subscale; GAD-2 = General Anxiety Disorder, 2-item version; DASS-A = Depression, Anxiety and Stress scale, 21- and 42-item version, anxiety subscale; GAD-7 = Generalized Anxiety Disorder, 7-item version; SAS = Zung Self-rating Anxiety Scale; BAI-21 = Beck anxiety inventory; SDS = Zung Self-rating Depression Scale; CESD-8 = Center for Epidemiological Studies Depression scale, 8-item version; CESD-10 = Center for Epidemiological Studies Depression scale, 10-item version; DASS-D = Depression, Anxiety and Stress scale, 21- and 42-item version, depression subscale; PHQ-8 = Patient Health Questionnaire, 8-item version; PHQ-2 = Patient Health Questionnaire, 2-item version; HADS-D = Hospital Anxiety and Depression scale, depression subscale; CESD-20 = Center for Epidemiological Studies Depression scale, 20-item version; PHQ-9 = Patient Health Questionnaire, 9-item version; SCOFF-5 = Sick, Control, One, Fat, Food Questionnaire, 5-item version; AUDIT = Alcohol Use Disorder Identification Test, 10-item version; AUDIT-C = Alcohol Use Disorder Identification Test – Consumption; DAST = Drug Abuse Screening Test, 10-item version; PSS-14 = Perceived Stress Scale, 14-item version; PSS-10 = Perceived Stress Scale, 10-item version; DASS-S = Depression, Anxiety and Stress Scale, 21- and 42-item version, stress subscale; K-6 = Kessler Psychological Distress Scale, 6-item version; K-10 = Kessler Psychological Distress Scale, 10-item version; PSQI = Pittsburgh Sleep Quality Index.

The moderating effects of sample characteristics, including risk and protective factors

We tested a broad range of sample-level characteristics as potential moderators of psychological distress in both the prevalence and severity models. These included academic position (pre- vs. postdoctoral), gender, ethnicity, academic discipline, age, or country cluster, history of mental health diagnosis, social support, supervisor relationship quality, job satisfaction, and life satisfaction. In the prevalence model, only age was a significant moderator. With an increase in ten years of age, the prevalence of psychological distress decreased by 1% (see Table 1 for more information). Surprisingly, in the severity model, none of the sample characteristics remained statistically significant after correcting for a 5% FDR. These findings indicate that the prevalence and severity of psychological distress were largely similar across demographics and individual protective and risk factors. The full moderator model results, including all raw and adjusted *p*-values, are available in Supplementary Tables 19 and 20. Descriptive data on the protective and risk factors can be found in Supplementary Table 23.

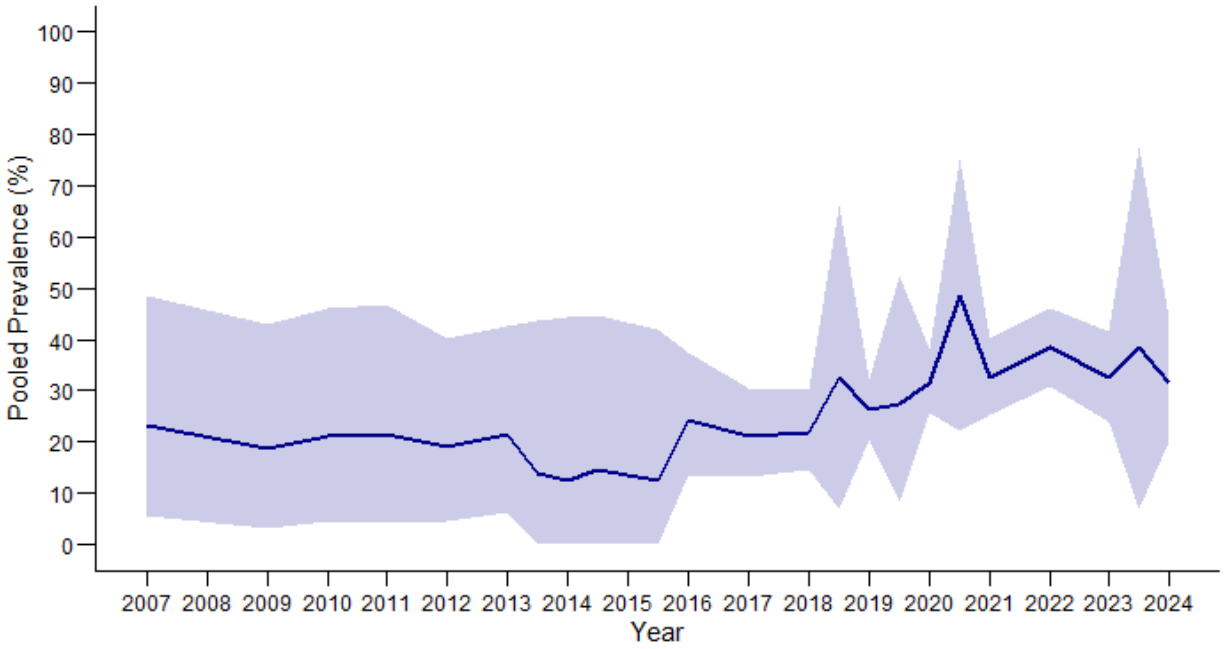
The moderating effect of COVID-19

The analysis on how psychological distress changed from before to during and after the COVID-19 pandemic yielded mixed results. In the prevalence model, the time of COVID-19 significantly moderates the pooled prevalence of psychological distress, and Figure 4 displays this change over time. Prevalence was significantly higher during the COVID-19 period (33.6% [29.1, 38.3]) compared to before (23.9% [19.8, 28.4]). Although prevalence remained elevated post-pandemic (30.5% [23.1, 38.5]), this difference was no longer statistically significant compared to the pre-pandemic period. Assessing the linear and quadratic association between prevalence and year of data collection indicated a similar effect of the time of the COVID-19 pandemic (see Supplementary Fig. 10).

The severity model indicated that psychological distress was higher during the COVID-19 period (2020–2022; POMP score = 34.5 [31.3, 37.8) compared to the period after (2023–2024; POMP score = 31.0 [24.4, 37.7]), but these estimates were not significantly different from the pre-pandemic period (≤2019; POMP score = 34.1 [30.9, 37.4]). Additionally, a polynomial moderator analysis with year as a linear and quadratic predictor revealed no significant time trend (see Supplementary Table 19).

Figure 4.

Pooled prevalence of psychological distress over the years 2007–2024



Note. This line graph illustrates the pooled prevalence of psychological distress across the years 2007 to 2024. The dark blue line represents the annual point estimates of prevalence, while the shaded region indicates the 95% confidence intervals. The graph shows a general decrease in prevalence from 2007 to approximately 2015, followed by a noticeable upward trend from 2016 onwards, with peaks observed in 2019, 2020, and 2023. The widening of the confidence intervals in later years reflects increased variability or fewer data points contributing to those estimates.

The moderating effects of study design and quality

Study quality (as measured by the overall JBI score) significantly moderates prevalence and severity estimates, with higher study quality indicating both lower prevalence and severity. Sampling strategy also significantly moderates psychological distress, with studies using full or random sampling reporting lower prevalence and severity estimates than those using convenience sampling. Assessment strategy (external assessment vs. self-report scales) was not a significant moderator, neither in the prevalence model ($F[1, 324] = 4.55$, adjusted $p = .06$; nor in the severity model ($F[1, 487] = 0.01$, adjusted $p = .95$) (see Table 1 and Supplementary Table 20 for details). Pooled prevalence estimates were lower when external assessment instead of self-report techniques were used (-12.4% [-40.6, -0.1]), while pooled severity estimates were slightly higher with external assessments (1.6 [-24.0, 27.2]), though neither effect reached statistical significance after controlling for the FDR.

Table 1.

Significant moderators of prevalence and severity

Model/ Moderator	*k*	#ES	β	95% CI		$F(df_1,df_2)$	adjusted *p*	R^2	*Q* test*(df)*
				LL	UL				
Prevalence									
Age (mean-centered)	57	270	-0.01**	-0.03	-0.00	7.58(1, 268)	.015	23.86%	274,19.32(268) ***
Covid-19						6.56(2, 323)	.007	8.95%	266,73.93(323) ***
Before covid	37	150	23.97 ***	19.81	28.39				
During covid	39	136	1.12 ***	0.23	2.67				
After covid	11	40	0.55	-0.04	2.80				
Year						5.66(2, 324)	.012	5.82%	264,56.67(324) ***
Year (set to zero, mean-centered)	82	326	-0.00	-0.15	-0.09				
Year²	82	326	0.00	-0.00	0.00				
Overall study quality	82	326	-0.16 **	-0.41	-0.02	10.01(1, 324)	.007	13.86%	248,89.42(324) ***
Sampling strategy						11.49(1, 324)	.006	9.58%	
Convenience	65	195	32.20***	28.15	36.39				244,87.89(324) ***
Random or full	13	131	-2.42***	-5.98	-0.43				
Assessment strategy						9.40(1, 324)	.066	3.94%	291,59.09(325) ***
Severity									
Overall study quality	1	487	-3.21***	-4.85	-1.56	14.72(1, 487)	.004	8.09%	592,915.03(487) ***
Sampling strategy						39.83(1, 487)	.004	16.02%	622,128.94(486) ***
Convenience	124	373	36.56***	34.36	38.76				
Random or full	18	116	-18.01***	-23.62	12.40				

Note. This table only presents such moderators that remained significant after controlling for the false discovery rate. For categorical moderators, baseline categories are provided in the first line, and estimates of the other categories are deviations from the respective baseline category. For continuous moderators, estimates are equal to the slope. Data on prevalence have been back-transformed into percent. The false discovery rate adjusted *p*-value is presented for the *F*-tests. *k* = number of studies; *#ES* = number of effect sizes; β =

pooled effect size estimate; CI = Confidence interval; LL = lower limit; UL = upper limit; R^2 = amount of explained true effect-size variance; * = p-value < .05; ** = p-value < .01; *** = p-value < .001

Multivariate moderator analyses

Lastly, to evaluate whether significant moderators (after controlling for the 5% FDR) without missingness collectively explained variance above the sampling error, we conducted a multivariate moderator analysis for both prevalence and severity of psychological distress. To control for small-study effects, the standard error was included as a moderator in both models as well. For the prevalence model, we included the same moderators as for the severity model, and added the assessment strategy, COVID-19 period, and age. The multivariate moderator model for prevalence also reached significance, explaining 30.7% of true effect size heterogeneity, $F(18, 307) = 5.10$, $p < .001$. For the severity model, we included the type of syndrome and symptom, the type of scale used, overall study quality, and sampling strategy. The moderator model for severity ($F[23, 450] = 11.41$, $p < .001$) explained 41.2% of true effect size heterogeneity.

Risk of bias

Risk of bias assessment using nine criteria from the JBI checklist: sample frame, sampling, sample size, study description, data analysis, validity of methods, measuring of condition, statistical analysis, and response rate (JBI, 2020; Supplementary Table 6). Overall study quality was low to moderate and varied substantially ($M = 4.6$, $SD = 1.2$, range = 2.5–8.2; see Supplementary Table 7 and Supplementary Fig. 2). This may be partly because in many of the studies, ECRs were not the focal group of interest, but graduate students or university staff in general. Only 60% of studies had sufficient sample sizes ($k = 127$). Only one study utilized external assessments to measure psychological distress[46] and random sampling was used rarely (for exceptions see[42,47–54]). Also, only a fraction ($k = 23$; 11%) of the studies reported an adequate response rate (≥ 45%), and prevalence estimates were almost never accompanied by confidence intervals (CI).

Risk of publication bias

A preliminary qualitative assessment of asymmetry in the funnel plot for the severity model was indicative of publication bias (Supplementary Fig. 11), which was confirmed by Egger's multilevel meta-analysis regression test ($F[1, 487] = 22.45$, $p < .001$; 2.8 [1.6, 4.0]). A sensitivity analysis revealed that this publication bias did not impact the significance of the already significant moderators (Supplementary Table 24). The visual inspection for publication bias in the prevalence model was inconclusive, due to the transformed effect sizes and variance

(Supplementary Fig. 12), Egger's test revealed no publication bias (F[1, 294] = 1.42, p = .23; 0.7 [-0.5, 1.9]), as had already been the case in previous meta-analyses on depression[10] and GAD[9].

Discussion

This multilevel meta-analysis assessed the prevalence and severity of psychological distress among ECRs across a broad range of mental health syndromes and symptoms. To understand the underlying factors affecting the prevalence and severity of psychological distress, we conducted moderation analyses on a wide set of sample characteristics as well as study-level aspects. From 212 effect sizes drawn from 139 studies and four databases (N = 128,684), our results indicate that 28.9% of ECRs met the criteria for at least one syndrome or symptom. In terms of severity, ECRs reported mild levels of depression and anxiety, moderate stress, and generally poor sleep quality, while levels of severity of suicidal ideation and eating disorders fell below clinical cut-offs. Moderation analyses indicated that prevalence and severity estimates were largely unaffected by demographic or individual factors (except for age affecting prevalence), but rather due to study-level variables such as sampling strategy, study quality, and type of measurement scale. The consistency of these findings across sample characteristics highlights the widespread extent of mental health issues among ECRs.

Our results indicate that the prevalence rates for depression (29%) and GAD (29%) among ECRs are five to six times higher compared to similarly aged individuals from the general population (5% and 6%, respectively[55]). These findings are broadly consistent with other meta-analyses, but fall on the higher end of previously reported ranges of 21–24% for depression and 17–30% for GAD[9–11]. While the prevalence of psychological distress moderately increased during the COVID-19 pandemic, our findings indicate that elevated rates of depression in particular and psychological distress in general predate the pandemic and have persisted over time[10]. Beyond depression and GAD, our meta-analysis also examined the prevalence of other important syndromes and symptoms. Ranging from highest to lowest, nearly 43% of ECRs reported poor sleep quality, 25% met thresholds for an ED, 21% for alcohol abuse, 18% for both suicidal ideation and NSSI. These estimates are largely based on self-reports, which tend to overestimate prevalence[56], and were influenced by the type of measurement scale used[10]. Nonetheless, the consistency and magnitude of elevated prevalence estimates across a wide range of mental health outcomes provide support to the growing concerns regarding a mental health crisis among ECRs.

Another central aim of this meta-analysis was to quantify the severity of psychological distress across a wide range of mental health syndromes and symptoms. According to the current data from 141 samples, a large proportion of ECRs reported elevated levels of psychological distress (overall POMP score = 34, on a scale from 0–100). However, POMP scores must be interpreted within the context of the scales that were employed for the evaluation of each syndrome or symptom individually.

Across studies, the classification of depressive symptomatology among ECRs differed based on which type of measurement scale was used to measure depression. In particular, depression was classified as mild with the PHQ-9, moderate with the DASS, and severe with the CES-D. The PHQ-9 and PHQ-8 combined provided the largest sample with approximately 70,000 individuals, and therefore the most precise estimate. The findings from the PHQ-9 and PHQ-8 are in line with those of ref.[10], who have found that 36% of PhD students suffer from mild depressive symptoms, 12% experience moderate symptoms, and 9% experience severe symptoms. The PHQ-9 has also demonstrated strong psychometric properties[57] while other scales, such as the CESD-8, were found to overestimate prevalence in comparison with a diagnostic interview[58]. The CESD-8 was used only by the COVID-19 International Student Well-Being Study in 2020, which found higher depression severity[59], but our meta-analysis found no systematic and significant change in overall psychological distress severity during the pandemic period. These results suggest that the type of measurement scale, rather than pandemic timing, is the primary driver of variation in severity estimates. When depression was assessed using large, representative samples and well-validated instruments such as the PHQ-9, severity levels for depression among ECRs can be described as mild.

Results on GAD were more consistent, with symptomology generally categorized as mild, corroborating findings from ref.[9], who reported that 23% of PhD students experienced mild, 16% moderate, and 13% severe GAD. The GAD-7 was used in the largest sample in the meta-analysis (n = 63,684), compared to the DASS-42 and DASS-21 (n = 2,855), thus making it the most precise estimate. However, the findings should be interpreted with similar caution, as the GAD-7 and the DASS-42 have been found to overestimate severity[9,60]. For other syndromes such as ED or alcohol abuse, the overall sample did not exceed clinical cut-offs. Notably, data on ED (assessed via the SCOFF-5) and alcohol abuse (AUDIT, AUDIT-C) originated exclusively from the Healthy Minds Network[42] and were thus limited to the United States. At the symptom

level, stress was consistently categorized as moderate across all scales used (PSS-10, PSS-14, DASS-42, DASS-21), underscoring the wide range of stressors experienced by ECRs[61]. Sleep quality (assessed via the PSQI), was rated as poor. However, the estimate lacked statistical power (n = 245), and the wide CI extended into the normal range. Fortunately, severity estimates for suicidal ideation indicate that most ECRs are below clinical severity cut-offs. Taken together, while severity estimates for syndromes and symptoms other than depression and GAD indicated mild symptoms, ECRs still face elevated levels of stress and poor sleep quality. More research is needed to generalize the findings beyond the United States and to obtain more precise estimates on lesser-studied conditions such as PTSD, alcohol use, and sleep problems.

In addition to estimating the prevalence and severity, this meta-analysis assessed sample-level moderators, including gender, age, career stage, country, ethnicity, history of mental health disorders, as well as protective and risk factors such as social support, supervisor relationship quality, job and life satisfaction, and loneliness. None of these variables significantly moderated severity estimates, and only age significantly moderated prevalence estimates, with a small effect of 1% lower prevalence per 10 years of age. These results are broadly consistent with other meta-analyses, which also reported no consistent effect of gender or country income levels. While one meta-analysis found that anxiety was more prevalent in Saudi Arabia and India and lower in China and the United States[9], our meta-analysis did not find effects of country across a broad range of mental health outcomes. The absence of significant sample-level moderators, including career stage, gender, academic discipline, individual risk and protective factors, suggests that psychological distress among ECRs is not limited to specific subgroups, but instead reflects a broader and systematic issue in academia. For example, the results indicate that psychological distress was just as prevalent and severe in the natural sciences as in the social sciences, between men and women, or between postdocs and PhD students. However, since there was limited data available for many sample-level moderators (except gender and age), this conclusion must be interpreted with caution, and more research is needed. Another related and important question is the direction of causality. Is psychological distress elevated among ECRs because of a career in academia, or do individuals with a tendency prone to psychological distress self-select into academia? Although the cross-sectional nature of the data of the included studies does not allow causal inference, individual studies point towards deteriorating mental health over time for doctoral students. For example, a large cohort study in Sweden with 20,000

participants found that the share of PhD students who received prescribed psychiatric medicine increased by 40% by their fifth year, with a parallel increase in hospitalizations due to mental health problems[62]. Future research needs to include measures of individual-level moderators, enabling more precise meta-analytic estimates of moderators. In addition, longitudinal designs are needed to establish temporal dynamics, mechanistic pathways, and causality of risk and protective factors of psychological distress among ECRs.

While sample-level moderators did not consistently influence the results, several study-level moderators did. In particular, the type of syndrome and symptom explained the largest proportion of the variance. In other words, levels of psychological distress vary depending on which syndrome or symptom is measured. Similarly, the type of measurement scale explained a substantial proportion of the variance, with important differences in prevalence and severity estimates depending on which scale was used. Finally, prevalence and severity were lower with higher study quality and with non-random assessment. These findings are in line with previous meta-analyses, which have also found that studies of low quality or those that used nonrandom sampling techniques reported higher prevalence rates for depression[10] and anxiety[9]. Taken together, these results underscore the methodological challenges in estimating psychological distress among ECRs. To obtain a more accurate estimate of the true prevalence and severity of psychological distress among ECRs, future studies should incorporate structured clinical interviews conducted by trained professionals, which represent the gold standard of mental health assessment[63,64]. For regular monitoring, large-scale assessments, or when structured clinical interviews are not feasible, future research should at least rely on well-validated measurement scales.

The present meta-analysis has several implications for research institutions and universities, funders and policymakers, supervisors, and ECRs themselves. For institutions, the findings underscore the need for systemic changes or adaptations to improve working conditions. Institutions should expand access to support services (e.g., coaching, ombudspersons, staff training) and dedicated mental health counseling (e.g., through collaborations with existing medical or psychology departments), provide supervisor training, and implement strategies to reduce power abuse and discrimination. Annual mental health surveys can help identify institution-specific problems and track progress. Academic institutions should increase transparency about mental health challenges (e.g., for prospective PhD students) and provide

career guidance, preparing ECRs for alternative careers outside academia. These measures are especially important given that the number of PhD graduates continues to outpace the availability of academic positions[65], leading to a culture of extreme competition, job insecurity, and pressure to game success metrics[4]. Policy makers and funders can address some of these structural issues by increasing the number of permanent positions, providing alternative career paths beyond tenure, promoting and developing fairness in promotion criteria, and establishing safeguards against power abuse and discrimination (e.g., safe whistleblower channels). Nationwide surveys on mental health and academic working conditions could be installed to monitor systemic progress over time (e.g., the Healthy Minds Study[42]). Supervisors can actively address mental health issues and contribute to a supportive working environment in which employees feel safe to voice their concerns[66]. Finally, for ECRs themselves, the findings can help normalize conversations about mental health issues and destigmatize seeking help. While systemic change is essential, the findings also highlight the need for ECRs to develop strategies for coping with the immediate stressors of working in academia[67,68].

We note four methodological limitations and suggest extensions for future research. First, the quality of the studies included varied greatly. For example, only one of the included studies conducted an external assessment via clinical-diagnostic interviews instead of relying on self-reported rating scales[69]. Additionally, only a few studies (predominantly databases) utilized random instead of convenience sampling techniques[42,48,49,70–73]. However, both moderators appeared to inflate the true effect sizes[10,74]. Furthermore, due to convenience sampling, generalizing the results to a greater sample frame remains difficult, as the pooled effect sizes might be biased by self-selection[75]. Second, postdoctoral ECRs made up less than 3% of the overall sample[42,73,76–86], yet they face slightly different challenges than doctoral students, and thus, more research should be dedicated to this group. In doing so, we recommend that future research clearly differentiate between the different student levels and academic positions, as this would have enabled the inclusion of more records. For example, many studies assessed graduate students, including predoctoral ECRs, but provided no information on the exact sample composition. Third, for some syndromes and symptoms, the available evidence was drawn exclusively from studies conducted in the US (e.g., ED, NSSI, alcohol abuse), and for syndromes or symptoms, it was limited overall (e.g., sleep quality). Poor sleep quality, for instance, can impair learning ability, and thus academic performance[87], and future research is needed on

lesser-studied outcomes, including data on non-US contexts. Finally, our review has underscored the need for further research into potential risk and protective factors, as existing studies often did not report these variables. Although this meta-analysis did not find any significant buffering or exacerbating effects of supervisor relationships, social support, or loneliness, life and work satisfaction, this does not mean that such effects do not exist. As almost all tests of these moderators lacked power, future research should direct more efforts to determine potential protective and risk factors in a congruent manner and thereby deduce effective and evidence-based interventions and implement systemic changes to reduce the psychological distress of ECRs. Such moderators could be the contract[59] or funding type[23,88], supervisor variables[14] (e.g., leadership style, perceived support from the supervisor), financial strain[48,90], socio-economic[91], or minority status, and additional predictors of psychological distress that may be uniquely intensified or moderated in the ECR population.

To conclude, this meta-analysis found that nearly one in three ECRs meets the criteria for at least one syndrome, with mild symptom severity for depression and GAD, moderate stress, and poor sleep quality. These patterns were consistent across academic disciplines, genders, and career stages, underscoring the widespread and systemic nature of psychological distress among ECRs. Targeted reforms at the institutional and policy level, ongoing monitoring, and strengthened individual coping resources could help improve mental health and working conditions in academia.

Methods

This review was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)[92] guidelines. The study protocol was registered on October 13, 2024, with the International Prospective Register for Systematic Review (PROSPERO ID: CRD42024596813). Deviations from the preregistration were recorded in the Supplementary Table 26.

Outcomes and definitions

For this systematic review and multilevel meta-analysis, two key terms necessitated a more precise operationalization, namely ECRs and mental health. ECRs, as defined by the European Council[93], are researchers enrolled in, or working for, higher education institutions who are actively pursuing a PhD or researchers with a PhD who work in the initial stages of their academic research career (corresponding to R1 and R2 levels, respectively). This includes pre-

and postdoctoral students, research associates, graduate students, doctoral candidates, PhD students, postdocs, postdoctoral researchers, postdoctoral fellows, (junior) lecturers, readers, and junior researchers. In the present analysis, a cut-off of no more than six years of post-doctoral working experience was applied to distinguish between early- and mid-career researchers, as this criterion was adopted from one of the studies included in the review[81]. In terms of mental health outcomes, the review was guided by definitions of mental health syndromes and diagnostic criteria outlined in the 11th revision of the International Classification of Diseases[94].

Search strategy

With the aim of identifying existing studies assessing mental health outcomes of ECRs, a detailed search strategy was developed. Although our preregistration listed several databases, the final search was conducted in Scopus, complemented by forward and backward citation search of relevant systematic and narrative reviews and meta-analyses (references of all included studies and databases are listed in Supplementary Table 1). We selected Scopus because it is the largest abstract and citation database of peer-reviewed literature and because it provides broad coverage of biomedical, psychological, and educational research. With the addition of references retrieved through backward and forward citation search, this systematic review and meta-analysis synthesized data from 212 samples and 128,684 individuals, representing the largest quantitative synthesis of psychological distress among ECRs to date. We are therefore confident that the final approach adequately captured the relevant literature (see Supplementary Table 26 for details on this deviation from the preregistration). For the Scopus search a combination of Boolean operators ('AND', 'OR', '*') was utilized, combining key terms defining ECRs, such as 'PhD' or 'post-doctoral researcher', mental health disorders, like 'GAD', 'depression', 'suicidal ideation' or 'eating disorder', further diagnostic criteria, for example 'stress' or 'burnout', as well as methodological determinants like 'survey' or 'quantitative' (Supplementary Table 27).

Inclusion and exclusion criteria

Due to the explorative nature of this analysis, any study that evaluated and reported the prevalence or summary scores on any mental-health disorder among ECRs was included. No exclusion criteria were formulated in regard to the record type; thus, journal articles, preprints, conference papers, reports, book chapters, and theses were all eligible for inclusion. Overall 27 non-peer-reviewed records were included[42,95–120]. Additionally, articles of any language were included and translated using DeepL, if necessary. Records were excluded if

(1) the sample did not correspond to our samples of interest (i.e., undergraduate students; graduate students pursuing a Master's degree or a professional doctorate; mid-stage to senior-stage researchers, like senior lecturers, senior scientists, senior research fellows, associate professors, assistant professors, full professors; research group leaders; principal investigators; adjunct faculty; adjunct professors; research directors; heads of departments) or compromised less or equal than two participants;

(2) data of samples of interest were not reported separately from the data of the overall, mixed samples assessed within the study;

(3) the record utilized a scale which was not utilized in at least one other record or did not provide information on the percentage of individuals scoring above a certain threshold; and lastly

(4) if the record did not provide information on any of the outcomes of interest, neither within its main text, supplementary material, any other related material, nor via direct contact with the authors.

Study selection and eligibility

Identification of studies via databases and the other described methods yielded a total of 3,846 articles published up to December 31, 2024. An independent screening based on title, abstract, and full text was conducted. Following this screening, 135 records met the predefined inclusion criteria. During the screening of the articles, it became apparent that several publications utilized the same datasets. We decided to include the primary datasets rather than derivative articles, resulting in a screening of the databases. Four databases[14,42,99,100] were included.

Data extraction

Data extraction followed two steps and was conducted by one scholar (VG). First, all screened articles were documented in Excel, and their subsequent exclusion based on 'title', 'abstract', and 'full text' was tracked. First, we utilized the 'duplicate values' function integrated in Excel to detect duplicates. For reports screened in full text, a reason was reported. In cases where the original report was not accessible or the report lacked information on the exact sample composition, the corresponding authors were contacted via a standardized mail template (Supplementary Table 28), and their responses were documented via Excel. In the subsequent second step, information from the initially included articles was extracted in an exhaustive and explorative manner, for which a coding scheme was developed and iteratively adjusted

(Supplementary Table 29). Extracted information encompassed data on the source (authors, publishing year, the type of report and whether or not it was peer-reviewed), study characteristics (year in which data was collected, country, sample size, scales used to assess the outcomes and moderators), sociodemographic information on the participants (academic position, age, percentage female and diverse, ethnicity, faculty affiliation), mental health information of the participants (history of psychiatric disorders, summary scores, proportion of those above a predetermined threshold) and potential additional diagnostic criteria, as well as risk and protective factors (*M* and *SD* for continuous scales). To ensure intercoder reliability and to mitigate inconsistencies in the data extraction[121], 30 articles were randomly selected and independently coded by a second scholar (DF). Any inconsistencies were discussed and agreed upon.

Effect-size measures

To retrieve information on the severity as well as the prevalence of certain syndromes and symptoms, two effect-size measures were used: POMP scores and prevalence rates. Two types of studies were included: studies providing information on the percentages of individuals scoring above a certain cut-off, and studies providing information on summary scores, in the form of means and standard deviations. Secondly, given that primary studies employed a wide range of measures of psychological distress as well as its syndromes and symptoms, the summary scores were not comparable across studies. Therefore, to increase the comparability of the summary scores, we followed the example of previous meta-analyses[122,123] and converted summary scores into POMP scores[124]. In case of several scales (i.e., WEMWBS, SWEMWBS, WHO-5 Index), the POMP scores were directly reverse-coded, so that higher scores no longer represented higher but lower well-being.

Calculation of POMP scores

POMP scores are calculated via these formulas:

$$POMP = \frac{observed - minimum}{maximum - minimum} * 100$$

$$SD(POMP) = \frac{SD}{minimum - maximum} * 100$$

In case information on summary scores was not provided, yet information on the percentage per rung or category given[23,68,70,71,82,98,115,116,125–136], we recalculated the expected

value of the mean and standard deviation in four steps. First, in case percentages were provided per category, but not per value, we determined the mean value per category. In a second step, we estimated the probability distribution based on a normal distribution, yet this was only applied in case the probability distribution had not been fully provided. Last, we recalculated the expected values of the mean and standard deviation based on the following formulas[137]:

$$M = \sum_{i=1}^{n} xi * pi$$

$$SD = \sqrt{\sum_{i=1}^{n} pi * (xi - \mu)^2}$$

For example, in the case of ref.[70], 3.1% scored above the cut-off of ≥ 1 on the PHQ-9, item 9. As the scoring on item 9 of the PHQ-9 can only be between 0 and 1, we concluded 96.9% must have scored 0, then we assumed the remaining 3.1% to be equally distributed between the scores 1, 2, and 3. By multiplying the respective scores by their probability and summing these scores, we calculated the *M* and *SD*, as in the above formula. More information on the individual steps taken in case *M*, and *SD* were not provided can be found in Supplementary Table 30. The terms POMP-score and severity model are interchangeable, yet we use the term severity model to enhance comprehensibility.

Transformation of prevalence

Similar to the POMP scores, the thresholds identifying a clinically relevant amount or severity of symptoms varied depending on the scale used to assess these. For a better overview of the varying thresholds across the different scales used, a comparative overview of these scales, their scoring method, range, and threshold was created (Supplementary Tables 10–18). Thresholds were either based on psychometric validation studies or on common practice (e.g., PHQ-9, item 9 to assess suicidal ideation).

However, the use of prevalence rates as effect sizes is not unproblematic, as the range of the proportion is doubly-bounded, in that it can only range from zero up to unity [138]. This causes the standard error to become artificially compressed, in case the percentage value is close to either 0% or 100%. Thereby, the precision of the prevalence is overestimated in the meta-analytic model, as the study receives a higher weight when estimating the pooled effect size. Hence, before fitting the multilevel meta-analysis, we stabilized the variance via a single (not

double) arcsine transformation, which has been proven to generate more adequate results than the double arcsine transformation when dealing with multiple and varying sample sizes [139,140].

$$transformed\ ES = ASIN(\sqrt{ES})$$

$$transformed\ var = \frac{1}{4 * n}$$

The transformed values were then back-transformed for interpretation of the results using the inverse formula:

$$backtransformed\ prevalence = \sin(ES)^2$$

Critical appraisal

The Joanna Briggs Institute (JBI) critical appraisal tool for epidemiological studies was utilized[141], which evaluates the methodological quality of a study based on nine categories: sample frame, sampling, sample size, description of study subjects and setting, sufficient coverage for data analysis of the identified sample, validity of methods used for identifying the condition/mental health syndromes, reliability and standardization of measures used, appropriateness of the statistical analysis, adequacy of the response rate. The JBI was used to assess the quality of the studies themselves, and not the unique effect sizes; for example, if data were provided on depression, GAD, severity, and prevalence alike, all the effect sizes received the same score. The JBI critical appraisal tool was further refined by VG and UT to assess the included studies and databases more rigorously. For instance, the extent to which the study subjects and setting were described in sufficient detail was taken into consideration. Rather than awarding zero or full points, studies were able to score quarter points based on whether they provided key information, such as the age of the participants, the country in which the study was conducted, and the gender composition, as well as additional information. In order to ascertain whether the response rate had been adequate, an external cut-off criterion of 45% was employed, based on the average online survey response rate[142]. Studies could score full points if the response rate was adequate (≥ 45%) or a quarter point if the response rate was below 45% yet provided, and no points if no information was given. In addition to the JBI coding scheme and the actual scoring per study provided in Supplementary Tables 6 and 7, the elaboration of these refinements allows the reader to understand the scoring of each study. To assess the overall study quality via moderator analysis, the scores on the individual categories were summarized.

Analysis

Analyses were conducted in R, Quarto (version 4.4.3) using the following packages: Metafor[143], MetaForest[144], tidyverse[145], and the integrated ggplot2 for data visualization[146]. For parameter estimation, the restricted maximum likelihood estimation was applied, and for standard errors, p-value, and 95% CIs estimation, Knapp and Hartung's adjustment was utilized (41,138). For further specifications about the analysis, the R script and the data are available in Supplementary Material 1–3 (https://osf.io/r9nkd/?view_only=5cd03fdac50c480caa4a422a8579c97b).

Model fit and heterogeneity

As studies reported more than one effect size (e.g., depression and suicidal ideation), we opted for a multilevel meta-analysis with three levels (e.g. ref.[41]) as it is a random-effect model, which additionally enables modeling the dependency between effect sizes derived from the same study. Multilevel meta-analysis partitions the variation in effect sizes into three levels: namely, participants (Level 1; sampling error), the outcomes (Level 2; here the syndromes and symptoms), and the studies themselves (Level 3). Model fit was determined via an ANOVA comparing the fit of a classic two-level model and the fit of a three-level random effects model. The final decision was based on the Akaike and Bayesian Information Criterion, and the likelihood ratio test. For parameter estimation, restricted maximum likelihood estimation was applied. Heterogeneity was assessed using the Q test, the I^2 statistic, and τ, as well as PI. Cochran's Q indicates whether there is excess variance in the data exceeding the sampling error variance, and the I^2 statistic quantifies the amount of variation beyond variance attributable to the sampling error[138]. An overall I^2 greater than 50% was considered to be indicative of substantial heterogeneity [147]. Additionally, we provide information on τ, and PI as these do not systematically increase with k and n. τ can be interpreted as the standard deviation of the true effect sizes; prediction intervals provide information on the range into which outcomes of future studies should fall[138].

Effect-size heterogeneity within- and between-studies

For both the prevalence and severity models, the data showed a significantly better fit to the three-level model than the reduced model (Supplementary Table 8). For the severity model, 51.4% of the overall variance could be attributed to Level 2 (within-study variance), and 48.5% to Level 3 (between-study variance; total $I^2 = 99.9\%$). The estimated between-effect size

variance (Level 2, $\tau = 12.0$) and between-study variance (Level 3, $\tau = 11.6$) indicated substantial heterogeneity across effect sizes and studies. For the prevalence model, 29.5% of the overall variance could be attributed to Level 2, and 70% to Level 3 (total $I^2 = 99.5\%$). The estimated between-effect size variance (Level 2, $\tau = 0.1$) and between-study variance (Level 3, $\tau = 0.2$) also indicated substantial heterogeneity (Supplementary Table 9).

Moderator analysis

As effects were heterogeneous between and within studies, univariate and bivariate moderator analyses were conducted per moderator and per severity and prevalence model, respectively. To account for multiple testing, the Benjamini-Hochberg[148] correction was applied, and FDR was controlled at 5%. Here, adjusted *p*-values were reported; a comparison between adjusted and non-adjusted *p*-values can be found in Supplementary Table 20.

Like the severity estimate, several moderators (loneliness, social support, job or life satisfaction) had been transformed into POMP-scores beforehand due to the diverse scales used. POMP scores were always calculated in such a way that the higher scores represented lower satisfaction or higher ratings of loneliness.

In addition, those moderators that had remained significant after having controlled the FDR were assessed independently for each model via a multivariate moderator analysis. Originally, a random forest analysis [149] had been conducted to retrieve information on the variable importance and explained variance, but was replaced by a multivariate moderator analysis due to the high missingness in many moderators. Information on the MetaForest analysis can be found in the Supplementary Material 4 and Figure 14–15.

Outlier and sensitivity analysis

To determine outliers, we assessed Cook's distance (Supplementary Fig. 13) separately per model, based on the 4/n cut-off. Six outliers were found in the prevalence model [88,126,150–152], and 13 in the POMP-score model[95,103,108,114,125,153–158]. There was more than one outlier per study at times, which is why the number of outliers does not match the number of listed publications. No striking differences were found between the outliers and non-outliers, thus a removal of these datapoints did not substantially alter the pooled effect, nor reduce heterogeneity (Supplementary Table 25).

Publication bias

Publication bias was visually assessed via multilevel meta-analysis adjusted funnel plots using the code provided by Fernández-Castilla and colleagues (2020; see Supplementary Fig. 11 & 12), and the three-level Egger's regression test[160,161].

References

1. Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T. & Vanderford, N. L. Evidence for a mental health crisis in graduate education. *Nature Biotechnology* **36**, 282–284 (2018).
2. Woolston, C. PhDs: The tortuous truth. *Nature* **575**, 403–406 (2019).
3. Busch, C. A., Wiesenthal, N. J., Gin, L. E. & Cooper, K. M. Behind the graduate mental health crisis in science. *Nat Biotechnol* **42**, 1749–1753 (2024).
4. Rahal, R.-M. *et al.* Quality research needs good working conditions. *Nature Human Behaviour* **7**, 164–167 (2023).
5. Trang, D. *et al.* A survey of the severity of mental health symptoms in the planetary science community. *Nat Astron* **8**, 691–696 (2024).
6. Woolston, C. Stress and uncertainty drag down graduate students' satisfaction. *Nature* **610**, 805–808 (2022).
7. Gewin, V. Postdoc depression and anxiety rates are rising, finds survey of 872 researchers. *Nature* https://www.nature.com/articles/d41586-025-02450-9 (2025).
8. Woolston, C. Satisfaction in science. *Nature* **562**, 611–615 (2018).
9. Chi, T., Cheng, L. & Zhang, Z. Global prevalence and trend of anxiety among graduate students: A systematic review and meta-analysis. *Brain and Behavior* **13**, e2909 (2023).
10. Guo, L. *et al.* Prevalence and changes in depressive symptoms among postgraduate students: A systematic review and meta-analysis from 1980 to 2020. *Stress and Health* **37**, 835–847 (2021).
11. Satinsky, E. N. *et al.* Systematic review and meta-analysis of depression, anxiety, and suicidal ideation among Ph.D. students. *Sci Rep* **11**, 14370 (2021).
12. Jaremka, L. M. *et al.* Common academic experiences no one talks about: Repeated rejection, impostor syndrome, and burnout. *Perspectives on Psychological Science* **15**, 519–543 (2020).
13. Levecque, K., Anseel, F., De Beuckelaer, A., Van Der Heyden, J. & Gisle, L. Work organization and mental health problems in PhD students. *Research Policy* **46**, 868–879 (2017).
14. Hazell, C. M. *et al.* Understanding the mental health of doctoral researchers: a mixed methods systematic review with meta-analysis and meta-synthesis. *Syst Rev* **9**, 197 (2020).
15. Keashly, L. & Neuman, J. H. Faculty experiences with bullying in higher education: causes, consequences, and management. *Administrative Theory & Praxis* **32**, 48–70 (2010).
16. Abbott, A. Germany's prestigious Max Planck Society conducts huge bullying survey. *Nature* **571**, 14–15 (2019).
17. Else, H. Bullying in science: largest-ever national survey reveals bleak reality. *Nature* **607**, 431–431 (2022).

18. Wollast, R. *et al.* Facing the dropout crisis among PhD candidates: the role of supervisor support in emotional well-being and intended doctoral persistence among men and women. *Studies in Higher Education* **48**, 813–828 (2023).
19. Hasgall, A., Saenen, B. & Borrell-Damian, L. *Doctoral Education in Europe Today: Approaches and Institutional Structures*. (European University Association, 2019).
20. Golde, C. M. The role of the department and discipline in doctoral student attrition: lessons from four departments. *The Journal of Higher Education* **76**, 669–700 (2005).
21. Glorieux, A., Spruyt, B., Minnen, J. & van Tienoven, T. P. Calling it quits: a longitudinal study of factors associated with dropout among doctoral students. *Studies in Continuing Education* **47**, 155–173 (2025).
22. Barreira, P., Basilicoz, M. & Bolotnyy, V. Graduate student mental health: Lessons from American economics departments. (2018).
23. Mills, L., Read, G. J. M., Bragg, J. E., Hutchinson, B. T. & Cox, J. A. A study into the mental health of PhD students in Australia: Investigating the determinants of depression, anxiety, and suicidality. *Sci Rep* **14**, 22636 (2024).
24. Woolston, C. 'The problem is greater than it's ever been': US universities urged to invest in mental-health resources. *Nature* **590**, 171–172 (2021).
25. Sperl, M. F. J. *et al.* Gute Betreuung, faire bewertungen und angemessene Arbeitsbedingungen ermöglichen exzellente promotionen in der Psychologie: Kommentar der Jungmitgliedervertretung der DGPs zum Positionspapier. 196–199 (2025).
26. OECD. Reducing the precarity of academic research careers. *OECD Science, Technology and Industry Policy Papers* **113**, (2021).
27. Inguaggiato, G. *et al.* The experience of women researchers during the Covid-19 pandemic: a scoping review. *Research Ethics* 17470161241231268 (2024) doi:10.1177/17470161241231268.
28. Jackman, P. C., Jacobs, L., Hawkins, R. M. & Sisson, K. Mental health and psychological wellbeing in the early stages of doctoral study: a systematic review. *European Journal of Higher Education* **12**, 293–313 (2022).
29. Lokhtina, I. A. *et al.* The impact of the COVID-19 pandemic on early career researcher activity, development, career, and well-being: the state of the art. *SGPE* **13**, 245–265 (2022).
30. Mackie, S. A. & Bates, G. W. Contribution of the doctoral education environment to PhD candidates' mental health problems: a scoping review. *Higher Education Research & Development* **38**, 565–578 (2019).
31. Martínez-García, I., De Witte, H., García-Martínez, J. & Cano-García, F. J. A systematic review and a comprehensive approach to PhD students' wellbeing. *Applied Psych Health & Well* aphw.12541 (2024) doi:10.1111/aphw.12541.

32. Melián, E., Reyes, J. I. & Meneses, J. The online PhD experience: A qualitative systematic review. *IRRODL* **24**, 137–158 (2023).
33. Nicholls, H., Nicholls, M., Tekin, S., Lamb, D. & Billings, J. The impact of working in academia on researchers' mental health and well-being: A systematic review and qualitative meta-synthesis. *PLoS ONE* **17**, e0268890 (2022).
34. Schmidt, M. & Hansson, E. Doctoral students' well-being: a literature review. *International Journal of Qualitative Studies on Health and Well-being* **13**, 1508171 (2018).
35. Martinez, S. M. *et al.* COVID-19-related stressors exacerbate food insecurity and depressive symptoms among graduate students receiving campus basic needs services: Cross-sectional findings from seven California public universities. *Stress and Health* **40**, e3345 (2024).
36. Nicholls, H., Nicholls, M., Tekin, S., Lamb, D. & Billings, J. The impact of working in academia on researchers' mental health and well-being: A systematic review and qualitative meta-synthesis. *PLoS ONE* **17**, e0268890 (2022).
37. Eisenberg, D., Golberstein, E. & Hunt, J. B. Mental health and academic success in college. **9**, (2009).
38. Richardson, C. M. E., Trusty, W. T. & George, K. A. Trainee wellness: self-critical perfectionism, self-compassion, depression, and burnout among doctoral trainees in psychology. *Counselling Psychology Quarterly* **33**, 187–198 (2020).
39. The Graduate Assembly, University of California, Berkeley. *Graduate Student Happiness & Well-Being Report*. (2014).
40. Barry, K. M., Woods, M., Martin, A., Stirling, C. & Warnecke, E. A randomized controlled trial of the effects of mindfulness practice on doctoral candidate psychological status. *Journal of American College Health* **67**, 299–307 (2019).
41. Assink, M. & Wibbelink, C. J. M. Fitting three-level meta-analytic models in R: A step-by-step tutorial. *TQMP* **12**, 154–174 (2016).
42. Healthy Minds Network, University of Michigan, University of California Los Angeles, Boston University & Wayne State University. Healthy minds study among colleges and universities 2007-2024. (2024).
43. Carter, G. *et al.* Predicting suicidal behaviours using clinical instruments: Systematic review and meta-analysis of positive predictive values for risk scales. *British Journal of Psychiatry* **210**, 387–395 (2017).
44. Fergusson, D. M., Horwood, L. J., Ridder, E. M. & Beautrais, A. L. Subthreshold depression in adolescence and mental health outcomes in adulthood. *Arch Gen Psychiatry* **62**, 66–72 (2005).
45. JBI. *Critical Appraisal Checklist for Prevalence Studies*. https://jbi.global/sites/default/files/2020-08/Checklist_for_Prevalence_Studies.pdf (2020).

46. Nagy, G. A. *et al.* Burnout and mental health problems in biomedical doctoral students. *LSE* **18**, ar27 (2019).
47. Ahalli, S., Fort, E., Bridai, Y., Baborier, N. & Charbotel, B. Mental health and working constraints of first-year PhD students in health and science in a French university: A cross-sectional study in the context of occupational health monitoring. *BMJ Open* **12**, e057679 (2022).
48. Charles, S. T., Karnaze, M. M. & Leslie, F. M. Positive factors related to graduate student mental health. *Journal of American College Health* **70**, 1858–1866 (2022).
49. Farrer, L. M., Gulliver, A., Bennett, K., Fassnacht, D. B. & Griffiths, K. M. Demographic and psychosocial predictors of major depression and generalised anxiety disorder in Australian university students. *BMC Psychiatry* **16**, 241 (2016).
50. Lt, L., Hense, S., Kodali, P. B. & Thankappan, K. R. Prevalence and underlying factors of depressive disorders among PhD students: A mixed-method study in the Indian context. *JARHE* **14**, 1704–1717 (2022).
51. Panger, G., Tyron, J. & Smith, A. *Graduate Student Happiness & Well-Being Report*. (2014).
52. Sverdlik, A., C. Hall, N. & McAlpine, L. PhD imposter syndrome: Exploring antecedents, consequences, and implications for doctoral well-being. *IJDS* **15**, 737–758 (2020).
53. Talkhi, N., Ghavami, V., Jamali, M. R., Jamali, H. & Jamali, J. Relationship between minor psychiatric disorders and health literacy among students: A latent class regression. *J Health Lit* **5**, (2020).
54. Thiagarajah, K., Sit, N. W. & Chee, H. P. Prevalence of sleep quality among academic staff of a private university in Malaysia. *MJMHS* **19**, 28–34 (2023).
55. Institute for Health Metrics and Evaluation (IHME). *Global Burden of Disease 2021: Findings from the GBD 2021 Study*. (IHME, 2024).
56. Levis, B. *et al.* Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. *BMC Med* **17**, 65 (2019).
57. Miller, P. *et al.* The performance and accuracy of depression screening tools capable of self-administration in primary care: A systematic review and meta-analysis. *The European Journal of Psychiatry* **35**, 1–18 (2021).
58. Dang, L., Dong, L. & Mezuk, B. Shades of Blue and Gray: A Comparison of the Center for Epidemiologic Studies Depression Scale and the Composite International Diagnostic Interview for Assessment of Depression Syndrome in Later Life. *The Gerontologist* **60**, e242–e253 (2020).
59. Carr, E. *et al.* Mental health among UK university staff and postgraduate students in the early stages of the COVID-19 pandemic. *Occup Environ Med* **79**, 259–267 (2022).

60. Collins, C. R., Schnell, P. & Cronin, R. Evaluating Self vs. Interviewer-Administered Screening for Anxiety in Sickle Cell Disease. *Journal of Sickle Cell Disease* **1**, yoae002.029 (2024).
61. Cornwall, J. *et al.* Stressors in early-stage doctoral students. *Studies in Continuing Education* **41**, 363–380 (2019).
62. Bergvall, S., Fernström, C., Ranehill, E. & Sandberg, A. The impact of PhD studies on mental health—A longitudinal population study. Preprint at https://ssrn.com/abstract=4920527 (2024).
63. Nordgaard, J., Revsbech, R., SÆBYE, D. & Parnas, J. Assessing the diagnostic validity of a structured psychiatric interview in a first-admission hospital sample. *World Psychiatry* **11**, 181–185 (2012).
64. First, M. B. *et al. User's Guide for the SCID-5-PD Structured Clinical Interview for DSM-5 Personality Disorders*. (American Psychiatric Association Publishing, Arlington, VA, 2016).
65. Sarrico, C. S. The expansion of doctoral education and the changing nature and purpose of the doctorate. *Higher Education* **84**, 1299–1315 (2022).
66. Newman, A., Donohue, R. & Eva, N. Psychological safety: a systematic review of the literature. *Human Resource Management Review* **27**, 521–535 (2017).
67. Dreisoerner, A., Klaic, A., van Dick, R. & Junker, N. M. Self-compassion as a means to improve job-related well-being in academia. *Journal of Happiness Studies* Advance Online Publication (2022) doi:10.1007/s10902-022-00602-6.
68. Marais, G. A. B., Shankland, R., Haag, P., Fiault, R. & Juniper, B. A survey and a positive psychology intervention on French PhD student well-being. *IJDS* **13**, 109–138 (2018).
69. Nagy, G. A. *et al.* Burnout and mental health problems in biomedical doctoral students. *LSE* **18**, ar27 (2019).
70. Ahalli, S., Fort, E., Bridai, Y., Baborier, N. & Charbotel, B. Mental health and working constraints of first-year PhD students in health and science in a French university: A cross-sectional study in the context of occupational health monitoring. *BMJ Open* **12**, e057679 (2022).
71. Lt, L., Hense, S., Kodali, P. B. & Thankappan, K. R. Prevalence and underlying factors of depressive disorders among PhD students: A mixed-method study in the Indian context. *JARHE* **14**, 1704–1717 (2022).
72. Talkhi, N., Ghavami, V., Jamali, M. R., Jamali, H. & Jamali, J. Relationship between minor psychiatric disorders and health literacy among students: A latent class regression. *J Health Lit* **5**, (2020).
73. Thiagarajah, K., Sit, N. W. & Chee, H. P. Prevalence of sleep quality among academic staff of a private university in Malaysia. *MJMHS* **19**, 28–34 (2023).

74. Levis, B. *et al.* Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. *BMC Med* **17**, 65 (2019).
75. Keiding, N. & Louis, T. A. Web-Based Enrollment and Other Types of Self-Selection in Surveys and Studies: Consequences for Generalizability. *Annu. Rev. Stat. Appl.* **5**, 25–47 (2018).
76. Bitter, A. N. & McCrea, S. M. "The ship is sinking, yet the band plays on": COVID-19 and academia. *Journal of American College Health* **72**, 1876–1885 (2024).
77. Castelló, M., McAlpine, L. & Pyhältö, K. Spanish and UK post-PhD researchers: Writing perceptions, well-being and productivity. *Higher Education Research & Development* **36**, 1108–1122 (2017).
78. Cilli, E., Ranieri, J., Guerra, F. & Di Giacomo, D. Early career researchers and mental health: Observational study of challenge and wellbeing. *Health Science Reports* **6**, e1649 (2023).
79. Dhont, J. *et al.* Conducting research in radiation oncology remotely during the COVID-19 pandemic: Coping with isolation. *Clinical and Translational Radiation Oncology* **24**, 53–59 (2020).
80. Gloria, C. T. & Steinhardt, M. A. Flourishing, languishing, and depressed postdoctoral fellows: differences in stress, anxiety, and depressive symptoms. *Journal of Postdoctoral Affairs* **3**, 1–9 (2013).
81. Hanitzsch, T., Markiewitz, A. & Bødker, H. Publish and perish: Mental health among communication and media scholars. *Journal of Communication* jqae012 (2024) doi:10.1093/joc/jqae012.
82. Hu, X. *et al.* Using Multilevel Mediation Model to Measure the Contribution of Beliefs to Judgments of Learning. *Front. Psychol.* **11**, 637 (2020).
83. Pitt, R. N., Taskin Alp, Y. & Shell, I. A. The mental health consequences of work-life and life-work conflicts for STEM postdoctoral trainees. *Front. Psychol.* **12**, 750490 (2021).
84. Trang, D. *et al.* A survey of the severity of mental health symptoms in the planetary science community. *Nat Astron* **8**, 691–696 (2024).
85. Van Der Weijden, I. & Teelken, C. Precarious careers: Postdoctoral researchers and wellbeing at work. *Studies in Higher Education* **48**, 1595–1607 (2023).
86. Ysseldyk, R. *et al.* A leak in the academic pipeline: Identity and health among postdoctoral women. *Front. Psychol.* **10**, 1297 (2019).
87. Seoane, H. A. *et al.* Sleep disruption in medicine students and its relationship with impaired academic performance: A systematic review and meta-analysis. *Sleep Medicine Reviews* **53**, 101333 (2020).

88. Aristeidou, M. & Aristidou, A. Research disruption during PhD studies and its impact on mental health: Implications for research and university policy. *PLoS ONE* **18**, e0291555 (2023).
89. Felden, E. & Sanders IV, L. Machtmissbrauch an Max-Planck-Instituten – DW – 13.03.2025. *dw.com* https://www.dw.com/de/deutschland-forschung-wissenschaft-machtmissbrauch-max-planck-institute-v2/a-71903094 (2025).
90. Borgogna, N. C., Smith, T., Berry, A. T. & McDermott, R. C. An evaluation of the mental health, financial stress, and anticipated debt-at-graduation across clinical and counseling PhD and PsyD students. *Teaching of Psychology* **48**, 328–338 (2021).
91. Cullinan, J., Walsh, S. & Flannery, D. Socioeconomic disparities in unmet need for student mental health services in higher education. *Appl Health Econ Health Policy* **18**, 223–235 (2020).
92. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71 (2021) doi:10.1136/bmj.n71.
93. European Commission. Commission Recommendation (EU) 2023/1640 of 31 July 2023 on fostering social dialogue at national and Union level. (2023).
94. *International Classification of Diseases, 11th Revision (ICD-11)*. (Geneva, 2024).
95. Andrade, D. M. B. The effect of burnout syndrome on graduate students during the COVID-19 pandemic. (University of Pécs, Pécs, 2023).
96. Barreira, P., Basilicoz, M. & Bolotnyy, V. Graduate student mental health: Lessons from american economics departments. (2018).
97. Bonetto, E. & Adam-Troian, J. The benefits of group identification for PhD candidates' health and work. *Psicologia e Saber Social* **11**, 485–512 (2022).
98. Bork-Young, S. J. Examining factors related to engineering graduate students' mental health experiences. (University of Michigan, Michigan, 2023).
99. Buffel, V. & Velde, S. V. de. Cross-country dataset C19 ISWS. Zenodo https://doi.org/10.5281/zenodo.5946964 (2022).
100. Byrom, N. & Metcalfe, J. Impact of COVID 19 on doctoral and early career researchers. 4461473 Bytes figshare https://doi.org/10.6084/M9.FIGSHARE.12361493 (2021).
101. Dumitrescu, G. A. Self-effifficacy, locus of control, perceived stress and student satisfaction as correlates of dissertation completion. (Andrews University, 2016).
102. Gakhal, S. Enhancing the mental wellbeing of PhD students. (2024).
103. Gérard, L. & Nagels, M. La gestion du stress chez les doctorants : La surconsommation de certains produits qui pourraient nuire à leur santé. in (Universités de Montpellier, Montpellier, 2013).
104. Giménez Dasí, M. *et al. La salud mental en nuestras manos: Una acción para los estudiantes de doctorado desde la Facultad de Psicología*.

https://docta.ucm.es/rest/api/core/bitstreams/b637aae8-91b9-4e8c-ba30-76f173242203/content (2024).

105. Hazell, C. M. *et al.* Nationwide assessment of the mental health of UK Doctoral Researchers. Palgrave https://doi.org/10.1057/s41599-021-00983-8 (2021).
106. Hill, L. M. Perceived stress, academic support, social support, and professional support factors as predictors of student s success in distributed-learning doctoral education. (Fielding Graduate University, 2011).
107. Jing, J. The association between stress, coping strategies, individual characteristics and well-being among mainland Chinese PhD students in Hong Kong – A mixed-methods study. (Chinese University of Hong Kong, 2021).
108. Kaufman, J. A. Personal perceptions of stress and self-perceived need for social support among doctoral psychology students in a distance education university sample. (Capella University, 2004).
109. Lowe, R. L. The relationship between personality, self-care, stress, and perceived wellness in psychology doctoral students. (Tennessee State University, 2015).
110. Nelson, K. Academic progress in doctoral students: Levels of hope, subjective well-being, and stress. (Walden University, 2014).
111. Nickel, S. C. Write, talk and feel good: Evaluation einer peergeführten Schreibgruppenintervention für Promovierende zur Verbesserung der psychischen Gesundheit. (Otto-Friedrich-Universität Bamberg, Bamberg, 2014).
112. Nottingham, K. A predictive analysis of the psychosocial outcomes of doctoral students. (Capella University, 2017).
113. Panger, G., Tyron, J. & Smith, A. *Graduate Student Happiness & Well-Being Report*. (2014).
114. Peters, B. M. The relationships among physiological and perceived stress, quality of life, self-care, and impairment in doctoral students. (State University of New York, Buffalo, 2007).
115. Ramadoss, D., McCord, M. & Horn, J. P. Climate surveys of biomedical PhD students and training faculty members in the time of Covid. Preprint at https://doi.org/10.1101/2022.01.06.475246 (2022).
116. Ramadoss, D. & Horn, J. P. Second climate survey of biomedical PhD students in the time of Covid. Preprint at https://doi.org/10.1101/2022.01.13.476194 (2022).
117. Rukmanee, N. *et al.* Prevalence of stress, anxiety, depression and associated risk factors among graduate students in the faculty of tropical medicine, Mahidol university. in *JITMM Proceedings* 68–83 (2024).
118. Scheidler, J. A. Effects of perceived stress and perceived social support on marital satisfaction in doctoral students. (Walden University, 2008).

119. Sundström, L. Väitöskirjatutkij oiden kuorm ittuminen ja sosiaali nen tuki Ohjaajalta ja tiedeyhteisöltä saadun tuen yhteys uupumukseen ja stressiin. (University of Helsinki, Helmikuu, 2022).
120. Williams, M. D. HBCU vs. PWI: Institutional integration at PWIs and black doctoral student depression, anxiety, and stress. (UNIVERSITY OF MINNESOTA, MINNESOTA, 2014).
121. Cooper, H. M. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. (Sage, Los Angeles London New Dehli Singapore Washington DC, 2017).
122. Buecker, S., Mund, M., Chwastek, S., Sostmann, M. & Luhmann, M. Is loneliness in emerging adults increasing over time? A preregistered cross-temporal meta-analysis and systematic review. *Psychological Bulletin* **147**, 787–805 (2021).
123. Bühler, J. L., Krauss, S. & Orth, U. Development of relationship satisfaction across the life span: A systematic review and meta-analysis. *Psychological Bulletin* **147**, 1012–1053 (2021).
124. Cohen, P., Cohen, J., Aiken, L. S. & West, S. G. The problem of units and the circumstance for POMP. *Multivariate Behavioral Research* **34**, 315–346 (1999).
125. Almasri, N., Read, B. & Vandeweerdt, C. Mental health and the PhD: Insights and implications for political science. *APSC* **55**, 347–353 (2022).
126. Bastos, A. F. *et al.* Academic-related stressors predict depressive symptoms in graduate students: A machine learning study. *Behavioural Brain Research* **478**, 115328 (2025).
127. Doom, J. R., Seok, D., Narayan, A. J. & Fox, K. R. Adverse and benevolent childhood experiences predict mental health during the COVID-19 pandemic. *ADV RES SCI* **2**, 193–204 (2021).
128. Estupiñá, F. J., Santalla, Á., Prieto-Vila, M., Sanz, A. & Larroy, C. Mental health in doctoral students: Individual, academic, and organizational predictors. *Psicothema* **36**, 123–132 (2024).
129. Fernández-Rodríguez, C., Romero, L. S., Moreno, S. S. & Cuesta, M. Academic and emotional challenges beyond COVID-19. analysis of years and branches of study in 35 Spanish universities. *Psicología Educativa* **30**, 75–84 (2024).
130. Gaudel, J. *et al.* Impact du confinement sur la santé mentale des doctorants, une étude de cohorte dans une université française. *L'Encéphale* **50**, 75–84 (2024).
131. Hulac, D. M., Ryan, A. M., Pratt, A., Nyberg, J. & Kriescher, S. Understanding depression and anxiety rates of school psychology graduate students. *School Psychology* (2024) doi:10.1037/spq0000655.
132. Li, G. & Su, Q. Exploring the role of personal quantification in alleviating generalized anxiety disorder among Chinese PhD students: A cross-sectional study. *Medicine* **103**, e38449 (2024).

133. Mavrogalou-Foti, A. P., Kambouri, M. A. & Çili, S. The supervisory relationship as a predictor of mental health outcomes in doctoral students in the United Kingdom. *Front. Psychol.* **15**, 1437819 (2024).
134. Panayidou, F. & Priest, B. Enhancing postgraduate researcher wellbeing through support groups. *SGPE* **12**, 42–57 (2021).
135. Schad, A., Layton, R. L., Ragland, D. & Cook, J. G. Mental health in medical and biomedical doctoral students during the 2020 COVID-19 pandemic and racial protests. *eLife* **11**, e69960 (2022).
136. Wang, X. *et al.* Investigating mental health of US college students during the COVID-19 pandemic: Cross-sectional survey study. *J Med Internet Res* **22**, e22817 (2020).
137. Perry, D. L. *Introduction to Statistics*.
138. Harrer, M., Cuijpers, P., Furukawa, T. A. & Ebert, D. D. *Doing Meta-Analysis in R*. (Chapman & Hall/CRC Press, Boca Raton, FL and London, 2021).
139. Röver, C. & Friede, T. Double arcsine transform not appropriate for meta-analysis. *Research Synthesis Methods* **13**, 645–648 (2022).
140. Schwarzer, G., Chemaitelly, H., Abu-Raddad, L. J. & Rücker, G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Research Synthesis Methods* **10**, 476–483 (2019).
141. Munn, Z., Moola, S., Lisy, K., Riitano, D. & Tufanaru, C. Chapter 5: Systematic reviews of prevalence and incidence. in *JBI Manual for Evidence Synthesis* (eds Aromataris, E. & Munn, Z.) (JBI, 2020).
142. Wu, M.-J., Zhao, K. & Fils-Aime, F. Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports* **7**, 100206 (2022).
143. Viechtbauer, W. metafor: Meta-Analysis Package for R. (2025).
144. Lissa, C. J. van. metaforest: Exploring Heterogeneity in Meta-Analysis using Random Forests. (2024).
145. Wickham, H. & RStudio. tidyverse: Easily Install and Load the 'Tidyverse'. (2023).
146. Wickham, H. *et al.* ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. (2024).
147. *Cochrane Handbook for Systematic Reviews of Interventions*. vol. Version 6.5 (2024).
148. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57**, 289–300 (1995).
149. van Lissa, C. J. Small sample meta-analyses exploring heterogeneity using MetaForest. in *Small sample size solutions: a guide for applied researchers and practitioners* (Routledge, an imprint of the Taylor & Francis Group, an informa business, Abingdon, Oxon New York, NY, 2020).

150. Jackman, P. C., Jacobs, L., Hawkins, R. M. & Sisson, K. Mental health and psychological wellbeing in the early stages of doctoral study: a systematic review. *European Journal of Higher Education* **12**, 293–313 (2022).
151. Martinez, S. M. *et al.* COVID-19-related stressors exacerbate food insecurity and depressive symptoms among graduate students receiving campus basic needs services: Cross-sectional findings from seven California public universities. *Stress and Health* **40**, e3345 (2024).
152. van Beuningen, L. Talent management of doctoral students: Focus on well-being, or how to deal with the mental health crisis in graduate Education. in *Talent management in higher education* (eds Thunnissen, M. & Boselie, P.) 77–98 (Emerald Publishing, Leeds, 2024). doi:10.1108/9781802626858.
153. Chen, J., King, R. B., Li, Y. & Xu, W. The role of the research environment and motivation in PhD students' well-being: A perspective from self-determination theory. *Higher Education Research & Development* **43**, 809–826 (2024).
154. Feizi, S., Knäuper, B. & Elgar, F. Perceived stress and well-being in doctoral students: Effects on program satisfaction and intention to quit. *Higher Education Research & Development* **43**, 1259–1276 (2024).
155. McGregor, B. A., Antoni, M. H., Ceballos, R. & Blomberg, B. B. Very low CD19+ B-lymphocyte percentage is associated with high levels of academic stress among healthy graduate students. *Stress and Health* **24**, 413–418 (2008).
156. Sverdlik, A., C. Hall, N. & McAlpine, L. PhD imposter syndrome: Exploring antecedents, consequences, and implications for doctoral well-being. *IJDS* **15**, 737–758 (2020).
157. Sverdlik, A. & Hall, N. C. Not just a phase: Exploring the role of program stage on well-being and motivation in doctoral students. *Journal of Adult and Continuing Education* **26**, 97–124 (2020).
158. Talarowska, M. *et al.* Mental health of students at Polish universities after two years of the outbreak of COVID-19. *IJERPH* **20**, 1921 (2023).
159. Fernández-Castilla, B. *et al.* Visual representations of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology* **16**, 299–315 (2020).
160. Nakagawa, S. *et al.* Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods Ecol Evol* **13**, 4–21 (2022).
161. Yang, Y., Lagisz, M., Williams, C., Noble, D. W. A. & Nakagawa, S. A two-step approach for robust point and variance estimation for meta-analyses with selective reporting and dependent effect sizes. https://yefeng0920.github.io/BiasRobustMA_tutorial/#evidence-of-publication-bias (2023).