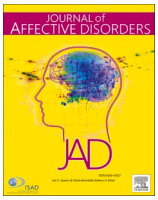
Journal of Affective Disorders 399 (2026) 121106

Contents lists available at ScienceDirect

Journal of Affective Disorders

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



Research paper

Does baseline severity interact with the effects of psychotherapy for depression? A meta-analytic review

Pim Cuijpers [a,b,*], Mathias Harrer [a], Clara Miguel [a], Eirini Karyotaki [a,c], Davide Papola [d]

[a] Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, the Netherlands
[b] Babeș-Bolyai University, International Institute for Psychotherapy, Cluj-Napoca, Romania
[c] Programme group Clinical Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, the Netherlands
[d] WHO Collaborating Centre for Research and Training in Mental Health and Service Evaluation, Department of Neuroscience, Biomedicine and Movement Science, Section of Psychiatry, University of Verona, Verona, Italy

ARTICLE INFO

Keywords:
Depression
Psychotherapy
Meta-analysis
Meta-regression analysis
Baseline severity

ABSTRACT

Introduction: It is not yet clear whether baseline severity is associated with the effects of psychotherapies. We examined baseline severity at the study level in a large sample of randomized controlled trials comparing psychotherapies against a control condition for the treatment of depression.
Methods: We used an existing large database of randomized trials comparing psychotherapies for depression with control groups (www.metapsy.org). We converted baseline severity scores across different depression measures into a common metric. We ran bivariable and multivariable meta-regression analyses to examine the association of effect sizes with baseline severity. We also examined response rates in treatment and control conditions.
Results: We included 387 randomized trials (463 comparisons; 47,315 patients). The pooled effect size of the psychotherapies was g = 0.77 (95 % CI, 0.70; 0.84). In the main analyses, we found a highly significant association between the effect size and baseline severity (bivariable coefficient: 0.024 (SE = 0.006; $p < 0.0001$), multivariable coefficient: 0.022 (SE = 0.007; $p = 0.002$)). This was confirmed in some but not all sensitivity analyses. Absolute response rates in the control conditions remained stable across different levels of baseline severity (bivariable metaregression analyses: $p = 0.545$), or showed a negative association (multivariable analyses: $p = 0.002$). In the therapy conditions the response rates were significantly larger with increasing levels of baseline severity (bivariable: $p \leq 0.0001$; multivariable: $p = 0.006$).
Conclusion: The effects of psychotherapies are probably associated with baseline severity. Response rates in control conditions remained relatively stable across different levels of baseline severity, while in the treatment conditions the response rates increased with increasing levels of baseline severity.

1. Introduction

It is well-established that several types of psychotherapy are effective in the treatment of depression. A large network meta-analysis of eight major types of therapy found that all therapies are more effective than usual care and waitlist control groups (Cuijpers et al., 2021a), without significant differences between the therapies. It is less clear, however, if the effects of therapies are associated with the severity of depression at baseline. Early research suggested that cognitive behavior therapy (CBT) was not more effective than pill placebo in more severely depressed patients (Elkin et al., 1989, 1995). Because antidepressants did result in significant outcomes compared to placebo in severely depressed patients, it was suggested that antidepressants are the treatment of choice for this group of patients.

Later meta-analytic research did not confirm this finding. A meta-regression analysis of 132 randomized controlled trials of psychotherapies for depression did not find that baseline severity was associated with the effect size (Driessen et al., 2010). This meta-analysis could only examine baseline severity at the study level, and did not account for the differences in severity within each of the included studies. This means that baseline severity could still be very well associated with the effect sizes, even though that was not found in this study.

The question whether baseline severity is associated with the effects of therapy can be better examined in 'individual patient data' (IPD)

* Corresponding author at: Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-9, 1081 BT Amsterdam, the Netherlands.
E-mail address: p.cuijpers@vu.nl (P. Cuijpers).

https://doi.org/10.1016/j.jad.2025.121106
Received 21 May 2025; Received in revised form 23 December 2025; Accepted 30 December 2025
Available online 3 January 2026

Table 1
Meta-analyses and metaregression analyses of the association between baseline severity and outcome of psychotherapies for adult depression.

	k	Main analyses						Bivariable metaregression			Multivariable metaregression		
		g	*95 % CI*	I^2	95 % CI	*95 % PI*	*NNT*	*Coef*	*SE*	*p*	*Coef*	*SE*	*p*
Main analyses													
All studies	463	0.77	0.70; 0.84	84	83; 85	−0.44; 1.98	3.98	0.024	0.006	<0.0001	0.022	0.007	0.002
Sensitivity analyses													
One ES/study (lowest)[a]	387	0.73	0.65; 0.80	85	83; 86	−0.49; 1.94	4.27	0.031	0.006	<0.0001	0.029	0.008	0.0002
One ES/study (highest)[a]	387	0.78	0.71; 0.86	85	84; 86	−0.47; 2.03	3.91	0.030	0.006	<0.0001	0.027	0.008	<0.001
All outliers removed	315	0.69	0.66; 0.72	25	14; 35	0.46; 0.92	4.53	0.003	0.003	0.343	−0.004	0.004	0.420
Influence Analysis	450	0.68	0.63; 0.73	78	76; 79	−0.18; 1.54	4.62	0.015	0.005	0.003	0.015	0.006	0.012
Only low risk of bias	170	0.58	0.49; 0.67	83	81; 85	−0.42; 1.58	5.53	−0.003	0.009	0.718	0.006	0.014	0.653
Common metric 60 to 80[b]	449	0.76	0.70; 0.82	83	82; 84	−0.40; 1.92	4.05	0.022	0.007	<0.001	0.027	0.008	<0.001
ES > 2 excluded	429	0.62	0.57; 0.66	74	71; 76	−0.09; 1.32	5.18	0.011	0.004	0.009	0.010	0.005	0.052
Separate analyses per instrument													
BDI-2	129	0.85	0.70; 0.99	87	85; 89	−0.64; 2.33	3.56	0.026	0.014	0.062	0.008	0.016	0.640
BDI	123	0.86	0.73; 1.00	76	71; 80	−0.35; 2.08	3.48	0.004	0.015	0.784	−0.002	0.015	0.880
HAM-D-17	110	0.83	0.69; 0.98	82	79; 85	−0.49; 2.15	3.63	0.024	0.011	0.033	0.040	0.013	0.002
PHQ-9	88	0.64	0.53; 0.75	84	80; 86	−0.27; 1.55	4.95	0.026	0.016	0.116	0.050	0.019	0.011
CES-D	70	0.76	0.58; 0.94	91	89; 92	−0.65; 2.17	4.06	0.013	0.020	0.501	0.025	0.23	0.279
EPDS	40	0.66	0.49; 0.83	86	82; 89	−0.28; 1.60	4.81	0.034	0.021	0.117	0.084	0.034	0.020
Publication bias correction													
- Trim-and-fill method	589	0.45	0.37; 0.53	90	89; 90	−1.36; 2.27	7.51						
- Limit meta-analysis	463	0.38	0.28; 0.47	84	–	−0.84; 1.59	9.25						
- Selection model	463	0.65	0.55; 0.75	93	91; 94	−0.82; 2.11	4.90						

Abbreviations: BDI: Beck Depression Inventory; CES—D: Center for Epidemiological Studies – depression; CI: confidence interval; Coef: coefficient; EPDS: Edinburgh Postnatal Depression scale; ES: effect size; HAM—D: Hamilton Depression Rating Scale; NNT: numbers-needed-to-treat; PHQ-9: Patient Health Questionnaire – 9; PI: prediction interval; SE: standard error. Underlined values are significant ($p < 0.05$).

[a] In these analyses we used only one effect size from studies in which multiple psychological interventions were compared to one control group (in one analysis we used only the only smallest and in another only the largest effect size).

[b] In these analyses we only included only studies in which the baseline 'common metric' was between 60 and 80 (between 1 and 3 standard deviations above the population mean). This was aimed at examining if exclusion of very low (<1 SD) or very high (>3 SD) levels of baseline severity had an impact on the outcomes.

meta-analyses, because these analyses do allow to examine the association between baseline severity and outcome at the individual level. In a systematic review of 10 IPD meta-analyses of psychological treatments of depression (Cuijpers et al., 2022), we did find that several of these meta-analyses showed that higher baseline depression severity was associated with better outcomes compared to control conditions (Bower et al., 2013; Karyotaki et al., 2018; Kuyken et al., 2016; Reins et al., 2021), and this was confirmed in more recent meta-analyses (Driessen et al., 2023; Buntrock et al., 2024). However, this was not confirmed in all included IPD meta-analyses, although statistical power to find this association was considerable.

Although IPD meta-analyses are superior to study-level meta-analyses for examining baseline severity, it is still important to conduct such a study for at least three important reasons. First, the evidence from IPD meta-analyses still does not consistently support the association between baseline severity and outcome. There are now more than 500 randomized controlled trials comparing psychotherapies for depression with control conditions (Cuijpers et al., 2025). It is unlikely that all of these trials will be examined in a large IPD meta-analysis in the short term. This means that IPD meta-analyses will only be able to examine subsets of this large body of research, which is at odds with the second predicament of evidence-based medicine, postulating that "the pursuit of truth is best accomplished by evaluating the totality of the evidence" (Djulbegovic and Guyatt, 2017). It should also be noted that IPD-MA moderator analyses, if properly conducted, explicitly "factor out" study-level mean differences in the baseline symptom severity, and focus on within-study variability instead (Fisher et al., 2017). If depression scores in studies are restricted in their variability at baseline, this can make it harder to detect moderator effects on a patient level. Using study-level information may be helpful in this case, since it provides a broader assessment if the average symptom severity in a sample predicts higher or lower effects of psychotherapy.

Second, the last meta-analysis examining this question was conducted quite a few years ago (Driessen et al., 2010) and included only a relatively small set of studies (132), compared to the several hundreds of trials that are currently available. This previous meta-analysis also had to examine the association between baseline severity and outcome separately for each depression instrument (Driessen et al., 2010) and separate analyses were conducted for the BDI, the BDI-II and the HDRS. Since then, "common metrics" have been developed for depression measures (Wahl et al., 2014), allowing all major depression measures to be transformed into a common metric thereby increasing the statistical power to detect differences between groups. This common metric indicates the mean of the general population with a score of 50, and each standard deviation increases the score by 10 (so a score of 60 indicates one standard deviation above the population mean). Several other similar studies have been conducted (Furukawa et al., 2020; Leucht et al., 2018; Choi et al., 2014) and it is now possible to convert most measures of depression into one common metric and examine the hundreds of trials reporting baseline severity in one big meta-analysis.

Finally, baseline severity is typically not included in meta-analyses and meta-regression analyses of psychotherapies for depression as a predictor of outcome (Cuijpers et al., 2023). The reason is that such an operation cannot be carried out when different depression measures are used. The common metrics approach, however, solves many of these issues. Most meta-analyses of therapies for depression have found very high levels of heterogeneity, and this heterogeneity has not yet been explained by characteristics of the participants, interventions or studies (Cuijpers et al., 2023). If baseline severity is a significant predictor, this could potentially help to explain the high levels of heterogeneity.

In the current study, we use the Metapsy database of randomized controlled trials comparing psychotherapies for depression with control condition (www.metapsy.org), convert measures of baseline depression into a common metric and examine in metaregression analyses if

Table 2
Multivariable metaregression of baseline severity and other characteristics of randomized controlled trials of psychotherapies for depression.

		Coef	SE	P[a]
Baseline severity (continuous)		0.022	0.007	0.002
Recruitment	Only clinical	Ref.		
	Community	0.113	0.100	0.240
	Other	0.116	0.110	0.292
Diagnosis	Mood disorder	Ref.		
	Above cut-off	−0.055	0.074	0.460
	Subthreshold	0.309	0.136	0.023
Target group	Adults	Ref.		
	General medical patients	0.003	0.117	0.981
	Older adults	−0.107	0.182	0.558
	Perinatal depression	0.172	0.151	0.909
	College students	0.193	0.166	0.246
	Other specific group	0.083	0.107	0.439
Mean age (continuous)		0.004	0.004	0.281
Proportion women (continuous)		0.450	0.167	0.007
Therapy	CBT	Ref.		
	Third wave therapy	−0.98	0.113	0.385
	Behavioural activation	0.046	0.128	0.720
	Psychodynamic	−0.188	0.238	0.430
	Interpersonal	−0.369	0.143	0.011
	Life review therapy	−0.097	0.233	0.676
	Problem-solving therapy	0.029	0.142	0.841
	Supportive therapy	−0.025	0.190	0.895
	Other therapy	−0.262	0.101	0.010
Format	Individual	Ref.		
	Group	0.080	0.080	0.321
	Guided self-help	−0.042	0.092	0.652
	Other/mixed	−0.023	0.123	0.850
Number of sessions (continuous)		0.001	0.007	0.853
Control group	Care-as-usual	Ref.		
	Waiting list	0.324	0.078	<0.0001
	Other control	0.010	0.100	0.922
Risk of bias (low versus other)		−0.044	0.017	0.010
Country	North America	Ref.		
	Australia	0.043	0.138	0.754
	East Asia	0.453	0.133	<0.001
	Europe	−0.068	0.085	0.242
	Other country	0.546	0.116	<0.0001
Instrument	BDI	Ref.		
	BDI-2	0.132	0.093	0.154
	CESD	0.131	0.124	0.292
	EPDS	−0.370	0.170	0.030
	HADS-D	0.025	0.234	0.914
	HAMD-17	−0.038	0.119	0.749
	IDS	0.532	0.423	0.209
	MADRS	0.007	0.295	0.981
	PHQ-9	−0.014	0.118	0.908
	QIDS-SR	0.277	0.318	0.383

Abbreviations: BDI: Beck Depression Inventory; CBT: cognitive behavior therapy; CES—D: Center for Epidemiological Studies – depression; Coef: coefficient; EPDS: Edinburgh Postnatal Depression scale; HADS-D: Hospital Anxiety and Depression Scale – depression; HAM—D: Hamilton Depression Rating Scale; IDS: Inventory of Depressive Symptomatology; MADRS: Montgomery-Asberg Depression Rating Scale; PHQ-9: Patient Health Questionnaire – 9; QIDS-SR: Quick Inventory of Depressive Symptomatology – self-report; SE: standard error. Underlined values are significant ($p < 0.05$).

[a] The *p*-values in this column indicate the significance of the association between the characteristic that is examined and the effect size, based on metaregression analyses with Knapp-Hartung adjustments.

baseline severity is associated with the effect size of the intervention.

2. Methods

2.1. Identification and selection of studies

The current study is part of a larger meta-analytic project on psychological treatments of depression that was registered at the Open Science Framework (Cuijpers & Karyotaki, 2021; doi:10.17605/OSF.IO/825C6; supplemental materials are available at www.metapsy.org). This database has been used in a series of earlier meta-analyses (Cuijpers et al., 2023). The protocol for the current meta-analysis was published at the Open Science Framework (Cuijpers, 2025; doi:10.17605/OSF.IO/37E8F).

The studies included in the current study were identified through the larger, existing database of randomized trials on the psychological treatment of depression. We searched four major bibliographical databases (PubMed, PsycINFO, Embase and the Cochrane Library; from 1966 to September 1st, 2024) by combining index and free terms indicative of depression and psychological treatments, with filters for randomized controlled trials. Full search strings can be found in Supplement A. The database is updated every four months. All records were screened by two independent researchers and all papers that could possibly meet inclusion criteria according to one of the researchers were retrieved as full-text. The decision to include or exclude a study in the database was done by the two independent researchers. Disagreements were resolved through discussion.

For the current meta-analysis, we selected all randomized trials in which a psychological intervention for people with depression was compared with a control condition (waitlist, care-as-usual, other inactive control), and which reported baseline severity according to one of the following depression measures: BDI, BDI-II, CESD, EPDS, HADS-D, PHQ-9, HAMD-17, MADRS, IDS or QIDS-sr. We selected these instruments because these measures can be converted to a common metric for depression (Wahl et al., 2014; Blackwell et al., 2021; Carmody et al., 2006; Furukawa et al., 2020).

Depression could be defined as meeting criteria for a depressive disorder according to a diagnostic interview or as a score above the cut-off on a validated self-report depression measure. We only included individual, group, telephone, and guided self-help interventions. Interventions without any human interaction (unguided self-help) were not included, because these have been found to be significantly less effective than other treatment formats (Cuijpers et al., 2019; Karyotaki et al., 2021; Tong et al., 2024; Papola et al., 2023). We excluded studies in children and adolescents because we previously found that psychological treatment is significantly less effective in these age groups. We also excluded studies in inpatient settings (Cuijpers et al., 2021b).

2.2. Quality assessment and data extraction

We assessed the validity of included studies using four criteria of the 'Risk of bias' (RoB) assessment tool, version 1, developed by the Cochrane Collaboration (Higgins et al., 2011). We used version 1 of this tool because this meta-analysis is included in the broader meta-analytic project of psychological treatments of depression (Sterne et al., 2019). The RoB tool assesses possible sources of bias in randomized trials, including the adequate generation of allocation sequence; the concealment of allocation to conditions; the prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat analyses were conducted, meaning that all randomized patients were included in the analyses). We considered trials as having low risk of bias when they scored positive on all four domains. Assessment of the validity of the included studies was conducted by two independent researchers, and disagreements were solved through discussion.

We also coded participant characteristics (diagnostic method for participant inclusion; recruitment method; target group; mean age; the proportion of women); characteristics of the psychological treatments (type of therapy; treatment format; the number of sessions), as well as general characteristics of the studies (type of control group; publication year; the country where the study was conducted). The details of these characteristics can be found at the website of the project (www.metapsy.org).

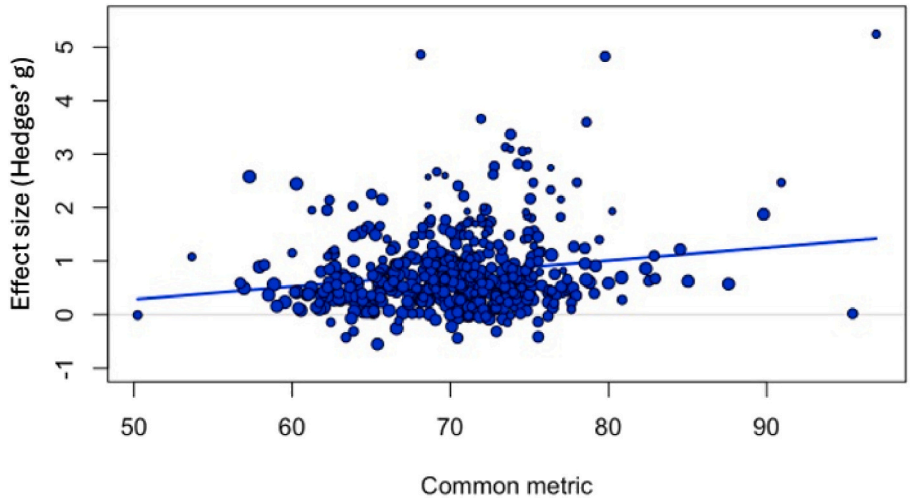


Fig. 1. The association between baseline severity and the effect size for psychotherapies for depression: Bubble plot [a) b) c). a)] The horizontal axis gives the common metric at baseline, where 50 indicates the population mean, 60 indicates one standard deviation above the population mean, 70 indicates two standard deviations above the population mean, etc. [b)] the regression line is based on a bivariable metaregression analysis with the effect size as dependent variable and baseline severity as predictor. [c)] The size of the bubbles is based on the inverse-variance weight of the studies.

2.3. Outcome measures

For each comparison between a psychological treatment and a control condition, the effect size indicating the difference between the two groups at post-test was calculated (Hedges' *g*). Effect sizes were calculated by subtracting (at post-test) the average score of the treatment group from the average score of the control group and dividing the result by the pooled standard deviation. Because some studies were expected to have relatively small sample sizes, we corrected the effect size for small sample bias. When the means and standard deviations were not reported in a study, we used change scores. If these were not reported either, we converted binary outcomes to Hedges' *g* or used other statistics (e.g., *p* value, *t* value) to calculate the effect size.

If a study used more than one depression measure, we selected one of them for the main analyses, so that each study had only one baseline depression score. The selection was done based on the frequency with which the measure was used in the total set of studies. This resulted in the following selection order: BDI-II; BDI; HAMD-17; PHQ-9; CESD; HADS-d; QIDS-sr; MADRS; IDS. We also conducted analyses separately for each depression measure. In these analyses all available studies were included in which that measure was used.

To examine baseline severity across studies, we converted the depression measures (BDI, BDI-II, CESD, EPDS, HADS-D, PHQ-9, HAMD-17, MADRS, IDS or QIDS-sr) to a common metric using the crosswalk tables provided in Wahl et al. (2014). For some measures, such as the HAMD-17, MADRS and EPDS, values are first transformed to BDI-I equivalents using other established crosswalks (Furukawa et al., 2020; Blackwell et al., 2021; Carmody et al., 2006). All conversions were performed using the 'ipdconverters' package, which implements these crosswalks (Harrer et al., 2025).

Common metric scores are normed so that values of 50 indicate the general population mean, with a population standard deviation (SD) of 10 (so a common metric score of 60 indicates one standard deviation above the general population mean). We also categorized the studies according to their common metric score at baseline (<1 SD above population mean; 1 to 1.5 SD, ≥1.5 to 2 SD, ≥2 to 2.5 SD, ≥ 2.5 to 3 SD, ≥3 SD).

To further explore the association between baseline severity and outcome, we also calculated the response rates (a 50 % reduction of depressive symptoms compared to baseline) with a validated method using the baseline means, the post-test means, the post-test *SD* and *N* (Furukawa et al., 2005). In these analyses we conservatively assumed that all dropouts were non-responders.

We also extracted study drop-out for any reason as an outcome, that could potentially be associated with baseline severity.

2.4. Meta-analyses

The meta-analyses were conducted using the 'metapsyTools' package in R (version 4.1.1; Harrer et al., 2022) and Rstudio (version 1.1.463 for Mac). This package was specifically developed for our meta-analytic project and it imports functionality of the 'meta' (Balduzzi et al., 2019), 'metafor' (Viechtbauer, 2010), and 'dmetar' (Harrer et al., 2019) packages.

We pooled the effect sizes using a random-effects model. Between-study heterogeneity variance (components) was estimated using restricted maximum likelihood. We applied the Knapp-Hartung method to obtain robust confidence intervals and significance tests of the overall effect (IntHout et al., 2014). As a test of homogeneity of effect sizes, we calculated the I^2-statistic and its *95 % CI*, which is an indicator of heterogeneity in percentages (Higgins et al., 2011). We also added the prediction interval (*PI*) which indicates the range in which the true effect size of 95 % of all populations will fall (Borenstein et al., 2017; Borenstein et al., 2009). We estimated the number-needed-to-treat (*NNT*) for depression using the formulae provided by Furukawa (1999) (assuming the control group's event rate at a conservative 16 %; Cuijpers et al., 2021c).

We conducted several sensitivity analyses. First, we pooled effects while excluding outliers, using the "non-overlapping confidence intervals" approach, in which a study is defined as an outlier when the *95 % confidence interval (CI)* of the effect size does not overlap with the *95 % CI* of the pooled effect size (Harrer et al., 2019). Second, we pooled effects while excluding influential cases as defined by the diagnostics in Viechtbauer and Cheung (2010). Third, we calculated the effect when only the smallest or largest effect in each study was considered. Fourth, we estimated the pooled effect using only studies with low risk of bias. We also used three different methods to assess and adjust for potential publication bias (Harrer et al., 2021; Maier et al., 2022): Duval and Tweedie's trim and fill procedure (Duval and Tweedie, 2000), Rücker's 'Limit meta-analysis method' (Rücker et al., 2011) and a step function selection model (McShane et al., 2016; Carter et al., 2019). We also conducted separate analyses for the studies using each of the included depression measures.

We examined the association between the effect size and baseline

Table 3
Subgroup analysis of the effect sizes of psychotherapies across categories of baseline severity.

Standard deviation above population mean	*k*	*g*	*95 % CI*	I^2	*95 % CI*	*p*[a]
		Effect size of therapy versus control				
< 1	11	0.72	0.24; 1.2	92	88; 95	0.003
≥ 1 to 1.5	68	0.57	0.43; 0.71	85	82; 88	
≥ 1.5 to 2	131	0.74	0.63; 0.84	80	76; 83	
≥ 2 to 2.5	182	0.78	0.68; 0.89	84	81; 86	
≥ 2.5 to 3	56	0.97	0.72; 1.21	85	81; 88	
≥ 3	15	1.21	0.52; 1.9	88	83; 92	
		Response in therapy[b]				
	k	*Prop*	*95 % CI*	I^2	*95 % CI*	*p*
All studies	457	0.39	0.37; 0.41	84	82; 85	
< 1	11	0.27	0.21; 0.34	68	39; 83	<0.0001
≥ 1 to 1.5	67	0.32	0.29; 0.36	74	66; 79	
≥ 1.5 to 2	129	0.39	0.35; 0.42	86	84; 88	
≥ 2 to 2.5	180	0.41	0.38; 0.44	83	81; 85	
≥ 2.5 to 3	55	0.44	0.37; 0.51	79	74; 84	
≥ 3	15	0.65	0.43; 0.82	84	76; 90	
		Response in control[b]				
All studies	381	0.19	0.17; 0.20	75	72; 77	
< 1	11	0.21	0.17; 0.26	54	9; 77	0.338
≥ 1 to 1.5	57	0.20	0.18; 0.22	54	38; 66	
≥ 1.5 to 2	113	0.18	0.16; 0.21	80	77; 83	
≥ 2 to 2.5	145	0.18	0.16; 0.20	73	69; 77	
≥ 2.5 to 3	41	0.17	0.13; 0.21	67	55; 76	
≥ 3	14	0.18	0.16; 0.21	83	73; 90	
		Relative risk of response				
	k	*RR*	*95 % CI*	I^2	*95 % CI*	*p*
All studies	457	2.03	1.92; 2.15	48	42; 53	
< 1	11	1.30	0.60; 2.84	61	26; 80	<0.0001
≥ 1 to 1.5	67	1.70	1.49; 1.93	34	10; 51	
≥ 1.5 to 2	129	2.08	1.88; 2.29	46	34; 56	
≥ 2 to 2.5	180	2.08	1.89; 2.28	51	42; 59	
≥ 2.5 to 3	55	2.33	1.92; 2.82	31	4; 51	
≥ 3	15	2.51	1.45; 4.35	66	42; 80	

[a] The p-values indicate whether the effect sizes differed significantly between the categories of baseline severity.
[b] Response rates were defined as a 50 % reduction of depressive symptoms compared to baseline, and were calculated using the method described by Furukawa et al. (2005); in these analyses it was conservatively assumed that all dropouts were non-responders.

severity using metaregression analyses. We also calculated this association for each of the sensitivity analyses that was conducted. We first conducted a bivariable metaregression analysis, with baseline severity as the only predictor. Then we conducted a multivariable metaregression analysis with baseline severity, as well as all major characteristics of the populations, interventions and studies, as predictor.

We also examined the association between the effect size and baseline severity with subgroup analyses, in which the different categories of baseline severity were used.

For dichotomous outcomes (the response and dropout rates) we calculated the pooled RRs (and NNTs) and the absolute rates within each condition using the "meta" package in R. In the analyses in which we pool the outcome rates within each condition, we synthesized the log-transformed risk ratios using a "normal-normal" random-effects model (Bakbergenuly et al., 2019). We then used the antilog to reconvert the pooled results back to the relative risk scale. The logit-transformed response rates were also pooled using the same approach, and final results were reconverted using the inverse of the logit function (expit). We then again examined the association between the outcomes and baseline severity with metaregression and subgroup analyses.

3. Results

3.1. Selection and inclusion of studies

After examining 37,440 records (26,497 after removal of duplicates), we retrieved 4652 full-text papers. We excluded 3537 of these. The PRISMA flowchart, including the reasons for exclusion, is presented in Supplement B. A total of 387 randomized trials (463 comparisons between a treatment and a control group) met inclusion criteria. The references of the included studies are given Supplement C.

3.2. Characteristics of included studies

An overview of aggregated characteristics is presented in Supplement D and key characteristics of the 387 included studies (and the 463 comparisons) are given in Supplement E. In the trials, 47,315 patients participated (25,078 in the interventions; 22,237 in the control conditions).

In 178 trials (46 %), participants met criteria for a depressive disorder according to a diagnostic interview, in 181 trials (47 %) included participants scored above a cut-off on a self-report scale, and in 28 trials (7.2 %) participants had subthreshold depression. In 166 studies (43 %), participants were recruited through the community, 85 (22 %) through clinical referrals and 136 (35 %) used other recruitment methods. A total of 141 studies (36 %) were aimed at adults in general, while the others were aimed at more specific target groups. In 181 studies (47 %), usual care was used as the control group, 152 studies (39 %) used waiting list control groups and the 54 remaining studies (14 %) used another control group. Most trials (149 (39 %)) were conducted in Europe, 120 (31 %) were conducted in North America, and the rest in other countries.

Cognitive behavior therapy was examined in 256 comparisons (55 %), while all other therapies were examined in less than 10 trials. 172 interventions (37 %) used an individual format, 144 (31 %) had a group

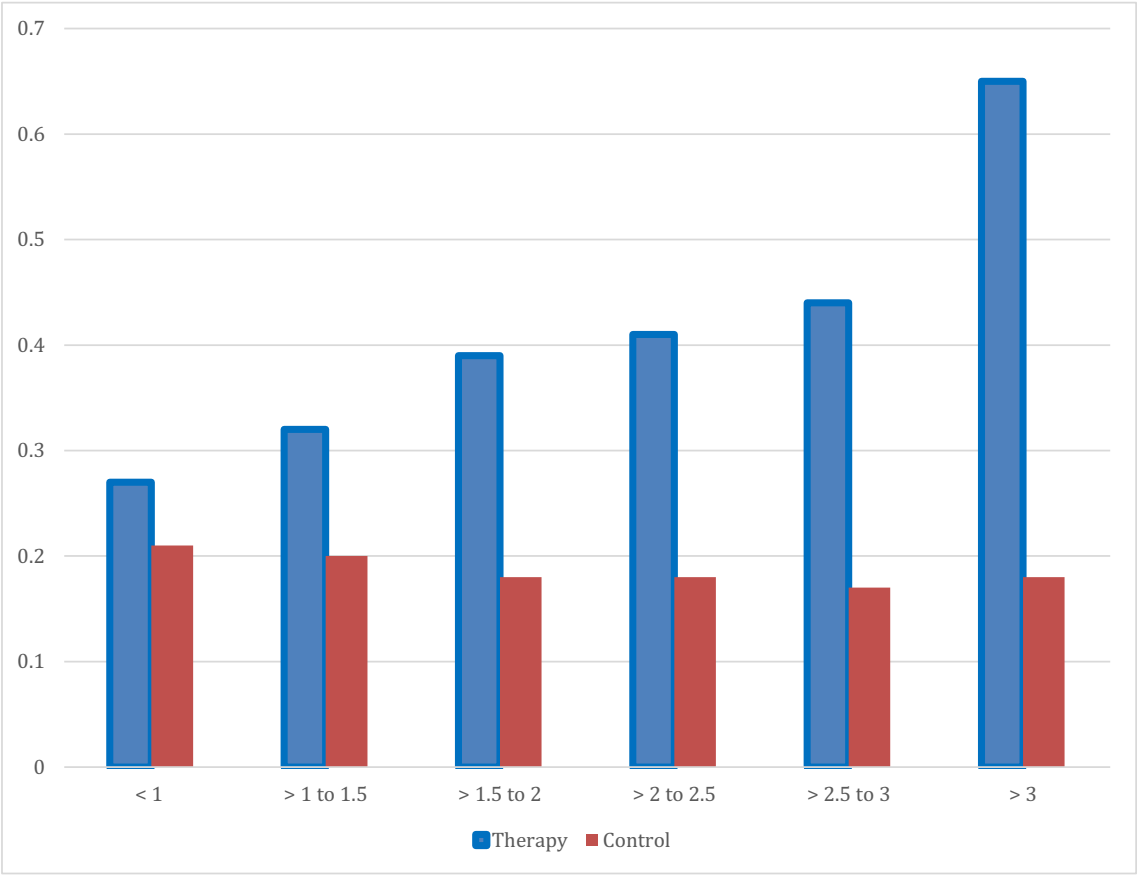


Fig. 2. Response rates in therapy and control groups in different categories of baseline severity [a) b). a)] Response rates were defined as a 50 % reduction of depressive symptoms compared to baseline, and were calculated using the method described by Furukawa et al. (2005); in these analyses it was conservatively assumed that all dropouts were non-responders. [b)] Sample sizes and confidence intervals for the different categories are reported in Table 3.

format, 111 (24 %) a guided self-help format and the remaining 36 comparisons (8 %) had a mixed or other format. The mean number of sessions was 9.

A total of 262 of the studies (68 %) reported an adequate sequence generation; 214 reported allocation to conditions by an independent party (55 %); 69 reported using blinded outcome assessors (18 %), while 310 (used only self-report outcomes (67 %). In 262 studies, intent-to-treat analyses were conducted (68 %). 151 (39 %) of the trials met all criteria for low risk of bias, 174 (45 %) met 2 or 3 criteria, and 62 met only one or no criterion (16 %).

3.3. The association between baseline severity and outcome: main analyses

The pooled effect size of the 463 comparisons was g = 0.77 (95 % CI: 0.70; 0.84), corresponding with an NNT of 3.98. Heterogeneity was very high (I^2 = 84; 95 % CI: 83; 85; prediction interval: −0.44; 1.98). The outcomes of the main analyses and sensitivity analyses are provided in Table 1. The bivariable and multivariable metaregression analyses indicated a significant association between baseline severity and the effect size (bivariable coefficient: 0.024 (SE = 0.006; $p < 0.0001$), multivariable coefficient: 0.022 (SE = 0.007; $p = 0.002$). The full results of the multivariable metaregression analyses are presented in Table 2. The bubble plot indicating the association between baseline severity and the effect size is in Fig. 1. The three analyses aimed at examining the results after adjustment for publication bias indicated that the adjusted effect sizes were smaller than the main analyses, but still significant (Table 1).

The results of the subgroup analyses of the effect sizes of psychotherapies across categories of baseline severity are presented in Table 3. The effect sizes ranged from g = 0.72 in the lowest category (baseline severity between 0 and 1 SD above the population mean) to g = 1.21 in the highest category (3 SD above the population mean). The difference between the subgroups was significant ($p = 0.30$).

The sensitivity analyses supported the main outcomes in some analyses (one effect size per study and the influence analyses) but not in others (the analyses in which the outliers were removed and the analyses of studies with low risk of bias). Because of these inconsistent results, we conducted two additional sensitivity analyses post-hoc. In one additional analysis, we excluded studies with a very large effect size (g > 2). We did that because the studies in which the outliers were removed resulted in a large group of removed studies (148 trials were removed as outliers, which is 32 % of all trials). In the set of studies in which the ones with a large effect size were removed (429 trials, 93 % of the total set of studies), we did find a significant association between the effect size and baseline severity in both the bivariable and the multivariable metaregression analyses. In the second post-hoc sensitivity analysis we excluded the studies with a very low baseline severity (less than one SD above the population mean), as well as the studies with a very high baseline severity (more than 3 SD above the population mean). We did that because the number of studies was small in both groups of studies (11 and 15 studies respectively), and because the baseline severity was at the extreme ends of the spectrum, they could have a disproportionate impact on the metaregression analyses. As can be seen in Table 1, the association between baseline severity and the pooled effect was significant in the bivariable metaregression analysis, but not in the multivariable analyses.

Most bivariable and multivariable metaregression analyses that were

done separately for each depression measure (Table 1) did not find a significant association between baseline severity and outcome, except for the bivariable analyses of the HAMD, as well as the multivariable analyses of the HAMD, the PHQ-9 and the EPDS. It must be noted that the number of studies in these analyses was considerable smaller than in the main analyses. We did not conduct separate metaregression analyses with the HADS-d (17 trials), the QIDS-sr (12 trials) and the IDS (6 trials), because we considered the number of trials too small for these analyses.

3.4. Association of response in psychotherapy and control conditions with baseline severity

The overall response rate in the psychotherapy conditions was 0.39 (95 % CI: 0.37; 0.41; $I^2 = 84$; 95 % CI: 82; 85) and in the control conditions 0.19 (95 % CI: 0.17; 0.20; $I^2 = 75$; 95 % CI: 72; 77).

In the psychotherapy conditions both the bivariable and multivariable metaregression analyses resulted in a significant association between the absolute response rate and baseline severity (bivariable: coefficient = 0.047 (SE = 0.007), $p \leq 0.0001$; multivariable: coefficient = 0.028 (SE = 0.010), $p = 0.006$). In the control conditions the association was not significant in the bivariable analyses (coefficient = −0.004 (SE = 0.007), $p = 0.545$), but there was a negative association in the multivariable analyses (coefficient = −0.031 (SE = 0.010), $p = 0.002$).

We have given a visual representation of the absolute response rates of the psychotherapy and control condition in the different severity categories in Fig. 2. This illustrates that the response rates increase with severity in the treatment groups but remain more or stable in the control conditions.

The overall relative risk of response was 2.03 (95 % CI: 1.92; 2.15; I2 = 48; 95 % CI: 42; 53). Both the bivariable and the multivariable metaregression resulted in a significant association between baseline severity and the RR of response (bivariable: coefficient = 0.02 (SE = 0.006), $p < 0.001$; multivariable: coefficient = 0.03 (SE = 0.007), $p < 0.0001$).

3.5. Relative risk and absolute rates of dropout in psychotherapy and control conditions

We examined if the dropout rates of the therapy and control groups were associated with baseline severity, but we found few indications that this was the case (Supplement F). The pooled dropout rate was 0.18 (95 % CI: 0.16; 0.19; $I^2 = 85$; 95 % CI: 83; 86) in the therapy conditions and 0.16 (95 % CI: 0.14; 0.17; I2 = 83;95 % CI: 82; 85) in the control conditions. We did not find any significant association between baseline severity and dropout in the therapy groups (bivariable: coefficient = −0.001 (SE = 0.009), $p = 0.88$; multivariable: coefficient = 0.010 (SE = 0.013), $p = 0.418$) and the control groups (bivariable: coefficient = 0.007 (SE = 0.10), $p = 0.470$; multivariable: coefficient = −0.012 (SE = 0.014), $p = 0.378$).

The RR of dropout was significant in the randomized treatment groups compared with the control groups(RR = 1.11; 95 % CI: 1.03; 1.18; I2 = 43; 95 % CI: 36; 50), but was not significantly associated with baseline severity (bivariable: coefficient = −0.010 (SE = 0.007), $p = 0.122$; multivariable: coefficient 0.018 (SE = 0.010), $p = 0.065$).

4. Discussion

We examined in a large sample of randomized controlled trials of psychotherapies for depression if baseline severity at the study level is associated with the outcomes. In order to do that, we converted the baseline scores of several depression measures to a common metric for depression using a consolidated and consistent methodology. In our main analysis we found indeed that baseline severity is significantly and strongly associated with the effect size of the studies. The findings of the multivariable metaregression analyses suggested that an increase of the baseline severity of one standard deviation was associated with an increase of the effect size of g = 0.22, which is substantial.

Although several sensitivity analyses supported the robustness of this finding, not all analyses were in agreement with the main findings. Some of these findings could be explained by low power to find a significant association, and all analyses with more than 400 studies were significant. However, several analyses for each depression measure separately and also the sample of studies with low risk of bias did not confirm the significant association. Although these were all on relatively small samples sizes, the overall results should be considered with caution.

This study also showed the importance of the common metrics approach to depression measures. It appeared to be feasible to convert most outcome measures to a common metric and to categorize them across the difference from the population mean. Although not perfect, this approach offers an important solution to the problem that trials in depression use many different outcome measures.

We also examined the absolute response rates in the psychotherapy and the control conditions, and found that the response rates in the control conditions remained relatively stable across the different levels of baseline severity. However, the response rates in the treatment groups increased significantly across the different levels of baseline severity. This suggests that treatments become more effective as baseline severity increases,. This is an important finding from a clinical perspective.

Heterogeneity was very high in the main analyses, meaning that other factors than baseline severity factors also contribute to outcome variability. Our metaregression analyses found that several other characteristics were significantly associated with the effect size, including subthreshold depression as inclusion criterion, the proportion of women, interpersonal psychotherapy as treatment, the use of waiting list control groups, risk of bias, the use of the EPDS as outcome measure and the country where the study was conducted. It should be noted, however, that these associations are always correlational and do not need to be causal. Interpretation of such findings should, therefore, always be done cautiously.

One important point to keep in mind is that we only examined baseline severity at the study level. IPD meta-analyses are more suited to examine baseline severity at the individual patient-level. However, as indicated in the Introduction, IPD meta-analyses are only conducted in relatively small subsets of studies examining the effects of therapies for depression. IPD meta-analyses also factor out study-level mean differences in the baseline symptom severity, and focus on within-study variability instead. Furthermore, their results are not straightforward with some studies reporting better outcomes with higher baseline severity, while others do not find this. Our meta-analysis contributes, therefore, substantially to the existing literature.This meta-analysis has several limitations that should be taken into account. One is that we could only examine baseline severity at the study level using aggregate-level data. This is an essential limitation, which means that the results cannot automatically be generalized to the individual patient level. Second, most studies had some kind of risk of bias and the analyses restricted only to low-risk of bias studies are not consistent with the main findings. Third, heterogeneity was very high in almost all analyses. Fourth, because we used only one outcome instrument per study, we were not able to conduct multilevel analyses as sensitivity analyses. The impact of this on our outcomes is probably limited, because we found hardly any difference between conventional analyses with multiple outcomes per study and multilevel analyses in earlier meta-analytic research (Cuijpers et al., 2025). A fifth limitation is that the transformation of depression scales into a common metric may have introduced measurement error and potential non-invariance in the dataset. Other problems that may have influenced the outcomes are potential publication bias and variation in follow-up periods. In meta-analyses like this, it is also impossible to examine therapist factors, such as competence, experience, and fidelity. Such factors may also have had an impact on the outcomes, the external validity and reproducability. A final limitation that we want to mention is that some multi-arm trials contribute more than one comparison sharing the same control group,

which introduces statistical dependence among effect sizes. We did conduct sensitivity analyses in which we included only one comparison from each study, and found that the overall effect sizes were very comparable to the main effect sizes. However, this should still be acknowledged as a limitation of the study. Because of these limitations, the results of this study should be considered with caution.

However, despite these limitations we found strong indications that the effects of psychotherapies are associated with baseline severity. Response rates in control conditions remained relatively stable across different levels of baseline severity, while in the treatment conditions the response rates increased with increasing levels of baseline severity.

CRediT authorship contribution statement

Pim Cuijpers: Methodology, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Mathias Harrer:** Methodology, Conceptualization, Writing – review & editing. **Clara Miguel:** Methodology, Data curation, Writing – review & editing. **Eirini Karyotaki:** Methodology, Data curation, Writing – review & editing. **Davide Papola:** Methodology, Data curation, Conceptualization, Writing – review & editing.

Role of funding

DP was funded by the European Union's Horizon-MSCA-2021-PF01 research program under grant agreement N 101061648. The funder had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

The other authors received no funding for this work.

Declaration of competing interest

None.

Acknowledgements

None

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2025.121106.

References

Bakbergenuly, I., Hoaglin, D.C., Kulinskaya, E., 2019. Pitfalls of using the risk ratio in meta-analysis. Res. Synth. Methods 10 (3), 398–419.
Balduzzi, S., Rücker, G., Schwarzer, G., 2019. How to perform a meta-analysis with R: A practical tutorial. Evid. Based Ment. Health 22, 153–160.
Blackwell, C.K., Tang, X., Elliott, A.J., Thomes, T., Louwagie, H., Gershon, R., Cella, D., 2021. Developing a common metric for depression across adulthood: linking PROMIS depression with the Edinburgh postnatal depression scale. Psychol. Assess. 33 (7), 610–621.
Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to meta-Analysis. Wiley.
Borenstein, M., Higgins, J.P.T., Hedges, L.V., Rothstein, H.R., 2017. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. Res. Synth. Methods 8, 5–18.
Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D.A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lovell, K., Liu, E. T.-H., 2013. Who should get low-intensity treatments for depression? An individual patient data meta-analysis. BMJ 346, Article f540.
Buntrock, C., Harrer, M., Sprenger, A.A., Ebert, D.D., Eiring, Ø., Fusar-Poli, P., et al., 2024. Psychological interventions to prevent the onset of major depression in adults: A systematic review and individual participant data meta-analysis. Lancet Psychiatry 11 (12), 990–1001.
Carmody, T.J., Rush, A.J., Bernstein, I., Warden, D., Brannan, S., Burnham, D., Woo, A., Trivedi, M.H., 2006. The Montgomery Asberg and the Hamilton ratings of depression: A comparison of measures. Eur. Neuropsychopharmacol. 16 (8), 601–611.
Carter, E.C., Schönbrodt, F.D., Gervais, W.M., Hilgard, J., 2019. Correcting for bias in psychology: A comparison of meta-analytic methods. Adv. Methods Pract. Psychol. Sci. 2 (2), 115–144.
Choi, S.W., Schalet, B., Cook, K.F., Cella, D., 2014. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol. Assess. 26 (2), 513–527.
Cuijpers, P., 2025. Does Baseline Severity Interact with the Effects of Psychotherapy for Depression? Open Science Foundation, A meta-analytic review. https://doi.org/10.17605/OSF.IO/37E8F.
Cuijpers, P., Karyotaki, E., 2021. A Meta-Analytic Database Of Randomized Trials On Psychotherapies For Depression. Open Science Foundation. https://doi.org/10.17605/OSF.IO/825C6.
Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., Furukawa, T., 2019. Individual, group, telephone, self-help and internet-based cognitive behavior therapy for adult depression: A network meta-analysis of delivery methods. JAMA Psychiatry 76 (7), 700–707.
Cuijpers, P., Ciharova, M., Miguel, C., Harrer, M., Ebert, D.D., Brakemeier, E.L., Karyotaki, E., 2021a. Psychological treatment of depression in institutional settings: A meta-analytic review. J. Affect. Disord. 286, 340–350.
Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., Furukawa, T.A., 2021b. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. Acta Psychiatr. Scand. 144 (3), 288–299.
Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I., Furukawa, T.A., 2021c. Psychotherapies for depression: A network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 20 (2), 283–293.
Cuijpers, P., Ciharova, M., Quero, S., Miguel, C., Driessen, E., Harrer, M., Purgato, M., Ebert, D., Karyotaki, E., 2022. The contribution of "individual participant data" meta-analyses of psychotherapies for depression to the development of personalized treatments: A systematic review. J. Pers. Med. 12(1), Article 93.
Cuijpers, P., Miguel, C., Harrer, M., Plessen, C.Y., Ciharova, M., Papola, D., Ebert, D., Karyotaki, E., 2023. Psychological treatment of depression: A systematic overview of a 'meta-analytic research domain'. J. Affect. Disord. 335, 141–151.
Cuijpers, P., Harrer, M., Miguel, C., Ciharova, M., Karyotaki, E., 2025. Five decades of research on psychological treatments of depression: A historical and meta-analytic overview. Am. Psychol. 80 (3), 297–310.
Djulbegovic, B., Guyatt, G.H., 2017. Progress in evidence-based medicine: A quarter century on. Lancet 390 (10092), 415–423.
Driessen, E., Cuijpers, P., Hollon, S.D., Dekker, J.J.M., 2010. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J. Consult. Clin. Psychol. 78 (5), 668–680.
Driessen, E., Fokkema, M., Dekker, J.J., Peen, J., Twisk, J.W., Van, H.L., Schoevers, R.A., Cuijpers, P., Kool, S., 2023. Which patients benefit from adding short-term psychodynamic psychotherapy to antidepressants in the treatment of depression? A systematic review and meta-analysis of individual participant data. Psychol. Med. 53 (13), 6090–6101.
Duval, S., Tweedie, R., 2000. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometr 56 (2), 455–463.
Elkin, I., Shea, M.T., Watkins, J.T., Imber, S.D., Sotsky, S.M., Collins, J.F., Glass, D.R., Pilkonis, P.A., Leber, W.R., Docherty, J.P., Fiester, S.J., Parloff, M.B., 1989. National Institute of Mental Health treatment of depression collaborative research program: general effectiveness of treatments. Arch. Gen. Psychiatry 46 (11), 971–982.
Elkin, I., Gibbons, R.D., Shea, M.T., Sotsky, S.M., Watkins, J.T., Pilkonis, P.A., Hedeker, D., 1995. Initial severity and differential treatment outcome in the National Institute of Mental Health treatment of depression collaborative research program. J. Consult. Clin. Psychol. 63 (5), 841–847.
Fisher, D.J., Carpenter, J.R., Morris, T.P., Freeman, S.C., Tierney, J.F., 2017. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ 356, Article j573.
Furukawa, T.A., 1999. From effect size into number needed to treat. Lancet 353 (9165), 1680. https://doi.org/10.1016/S0140-6736(99)01163-0.
Furukawa, T.A., Cipriani, A., Barbui, C., Brambilla, P., Watanabe, N., 2005. Imputing response rates from means and standard deviations in meta-analyses. Int. Clin. Psychopharmacol. 20 (1), 49–52.
Furukawa, T.A., Reijnders, M., Kishimoto, S., Sakata, M., DeRubeis, R.J., Dimidjian, S., Dozois, D.J.A., Hegerl, U., Hollon, S.D., Jarrett, R.B., Lespérance, F., Segal, Z.V., Mohr, D.C., Simons, A.D., Hollon, S.D., Vittengl, J.R., Dunlop, B.W., Dunner, D.L., Hirschfeld, R.M.A., Cuijpers, P., 2020. Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking. Epidemiol. Psychiatr. Sci. 29, Article e24.
Harrer, M., Cuijpers, P., Furukawa, T., 2019. Dmetar: companion R package for the guide 'doing meta-analysis in R' (R package version 0.0.9000). http://dmetar.protectlab.org.
Harrer, M., Kuper, P., Cuijpers, P., 2022. metapsyTools: several R helper functions for the "metapsy" database (R package version 0.3.2). https://tools.metapsy.org.
Harrer, M., Sprenger, A., Illing, S., Ciharova, M., Cuijpers, P., Buntrock, C., 2025. Ipdconverters: IPD common metric converters for Depression & Anxiety Scores [BETA] (R package version 0.0.9000). https://github.com/mathiasharrer/ipdconverters.
Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L., Sterne, J.A.C., 2011. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 343, Article d5928. https://doi.org/10.1136/bmj.d5928.
IntHout, J., Ioannidis, J.P., Borm, G.F., 2014. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med. Res. Methodol. 14, 25.

Karyotaki, E., Ebert, D.D., Donkin, L., Riper, H., Twisk, J., Burger, S., Rozental, A., Lange, A., Williams, A.D., Zarski, A.C., Geraedts, A., van Straten, A., Kleiboer, A., Meyer, B., Ünlü Ince, B.B., Buntrock, C., Lehr, D., Snoek, F.J., Andrews, G., Cuijpers, P., 2018. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin. Psychol. Rev. 63, 80–92.

Karyotaki, E., Efthimiou, O., Miguel, C., Genannt Bonsmann, M.G., Anttila, M., Araya, R., Buntrock, C., Cuijpers, P., David, D., Dowrick, C., Farrer, L., Gilbody, S., Høifødt, R. S., Kemmeren, L., Lam, R.W., Meyer, B., Moritz, S., Mira, A., Nobis, S., Cuijpers, P., 2021. Internet-based cognitive behavioral therapy for depression: an individual patient data network meta-analysis. JAMA Psychiatry 78 (4), 361–371.

Kuyken, W., Warren, F.C., Taylor, R.S., Whalley, B., Crane, C., Bondolfi, G., Hayes, R., Huijbers, M., Ma, H., Schweizer, S., Segal, Z., Speckens, A., Teasdale, J.D., Van Heeringen, K., Williams, M., Byford, S., Byng, R., Dalgleish, T., 2016. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 73 (6), 565–574.

Leucht, S., Fennema, H., Engel, R.R., Kaspers-Janssen, M., Szegedi, A., 2018. Translating the HAM-D into the MADRS and vice versa with equipercentile linking. J. Affect. Disord. 226, 326–331.

Maier, M., VanderWeele, T.J., Mathur, M.B., 2022. Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. Campbell Syst. Rev. 18(2), Article e1256.

McShane, B.B., Böckenholt, U., Hansen, K.T., 2016. Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. Perspect. Psychol. Sci. 11 (5), 730–749.

Papola, D., Ostuzzi, G., Tedeschi, F., Gastaldon, C., Purgato, M., Del Giovane, C., Pompoli, A., Pauley, D., Karyotaki, E., Sijbrandij, M., Furukawa, T.A., Cuijpers, P., Barbui, C., 2023. CBT treatment delivery formats for panic disorder: A systematic review and network meta-analysis of randomised controlled trials. Psychol. Med. 53 (3), 614–624.

Reins, J.A., Buntrock, C., Zimmermann, J., Grund, S., Harrer, M., Lehr, D., Baumeister, H., Weisel, K., Domhardt, M., Imamura, K., Kawakami, N., Spek, V., Nobis, S., Snoek, F., Cuijpers, P., Moritz, S., Berking, M., Riper, H., Bolinski, F., Ebert, D.D., 2021. Efficacy and moderators of internet-based interventions in adults with subthreshold depression: an individual participant data meta-analysis of randomized controlled trials. Psychother. Psychosom. 90 (2), 94–106.

Rücker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M., 2011. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics 12 (1), 122–142.

Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C. J., Cheng, H.Y., Corbett, M.S., Eldridge, S.M., Emberson, J.R., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., Higgins, J.P.T., 2019. RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ 366, l4898.

Tong, L., Panagiotopoulou, O.M., Cuijpers, P., Karyotaki, E., 2024. The effectiveness of self-guided interventions in adults with depressive symptoms: A systematic review and meta-analysis. EBioMed 105, 105208.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48.

Viechtbauer, W., Cheung, M.W.L., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 1 (2), 112–125.

Wahl, I., Löwe, B., Bjorner, J.B., Fischer, F., Langs, G., Voderholzer, U., Aita, S.A., Bergemann, N., Brähler, E., Rose, M., 2014. Standardization of depression measurement: A common metric was developed for 11 self-report depression measures. J. Clin. Epidemiol. 67 (1), 73–86.