American Psychologist

ISSN: 0003-066X

2025, Vol. 80, No. 3, 297–310
https://doi.org/10.1037/amp0001250

Five Decades of Research on Psychological Treatments of Depression: A Historical and Meta-Analytic Overview

Pim Cuijpers[1, 2], Mathias Harrer[3], Clara Miguel[1], Marketa Ciharova[1], and Eirini Karyotaki[1]

[1] Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam

[2] International Institute for Psychotherapy, Babeș-Bolyai University

[3] Department of Psychology and Digital Mental Health Care, Technical University of Munich

Since the 1970s, hundreds of randomized trials have examined the effects of psychotherapies for depression, and this number is increasing every year. In this study, we report outcomes from a living systematic review of these studies. We use Poisson regression analyses to examine if the proportions of studies have changed over time across the characteristics of the participants, therapies, and studies. We also present a meta-analysis of the effects across the major types, formats, targets, and age groups. We included 562 randomized controlled trials (669 comparisons; 66,361 patients). Most trials are conducted in adults and the relative proportion of trials in children and adolescents, as well as in older patients is significantly decreasing. The effects in children and adolescents are also significantly smaller than in adults ($p = .007$). Cognitive behavior therapy (CBT) is by far the best examined type of therapy (52%), but not necessarily more effective than other therapies. Over time, the proportion of studies examining several other types of therapy is significantly decreased compared to CBT. The quality of trials has increased over time, but still, a majority do not meet basic quality criteria, not even in recent years. The effects found in studies with low risk of bias are significantly smaller than in other studies ($b = -0.21$; $SE = 0.05$; $p < .001$). Most trials are conducted in the United States, but the proportion of studies in other parts of the world is rapidly increasing. The evidence that psychotherapies are effective is strong and growing every year.

Public Significance Statement

Hundreds of randomized trials have examined the effects of psychological treatments for depression over the past 5 decades. They have shown that these therapies are effective, but research in children and adolescents is lagging behind, as well as research on other therapies than cognitive behavior therapy. The quality of trials is increasing over time, although there is still room for improvement.

Keywords: psychotherapy, depression, historical overview, meta-analysis, cognitive behavior therapy

Supplemental materials: https://doi.org/10.1037/amp0001250.supp

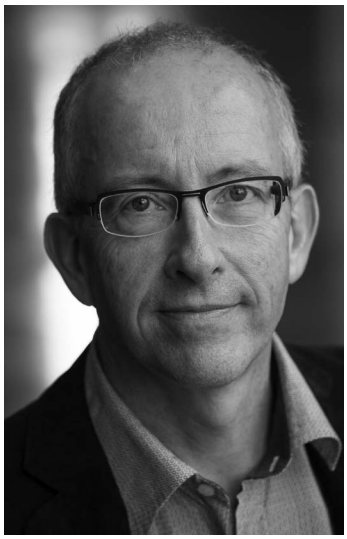
This article was published Online First November 16, 2023.

Pim Cuijpers https://orcid.org/0000-0001-5497-2743

This meta-analysis was registered on the Open Science Framework at https://doi.org/10.17605/OSF.IO/825C6.

Pim Cuijpers played a lead role in conceptualization, formal analysis, methodology, and writing–original draft and an equal role in data curation. Mathias Harrer played a lead role in methodology and an equal role in data curation, formal analysis, and writing–review and editing. Clara Miguel played a lead role in data curation, a supporting role in methodology, and an equal role in validation and writing–review and editing. Marketa Ciharova played an equal role in data curation, methodology, validation, and writing–review and editing. Eirini Karyotaki played an equal role in conceptualization, data curation, methodology, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Pim Cuijpers, 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. Email: p.cuijpers@vu.nl



Pim Cuijpers

Since the publication of the first randomized controlled trials examining the effects of psychotherapies in the treatment of depression in the 1970s (Covi et al., 1974; Klerman et al., 1974), hundreds of randomized controlled trials have been conducted to examine the effects of these therapies. This large body of evidence has shown that several types of psychotherapy are effective in the treatment of depression, including cognitive behavior therapy, interpersonal psychotherapy, third wave therapies, problem-solving therapy, psychodynamic therapy, and nondirective counseling (Cuijpers et al., 2023a). Comparative and controlled trials have also shown that the effects of these therapies do not significantly differ from each other (Cuijpers et al., 2023a), that they have comparable effects as antidepressant medication at the short term (Cuijpers et al., 2023a), but larger effects at the longer term (Cuijpers, Noma, et al., 2020; Furukawa et al., 2021). The effects of therapies are comparable across different specific target groups, such as women with perinatal depression, patients with comorbid general medical disorders, and older adults, but have significantly smaller effects in children and adolescents (Cuijpers et al., 2023a) and in people with comorbid substance use problems (Cuijpers, Miguel, Ciharova, et al., 2023).

The number of trials examining the effects of psychotherapies for depression has increased considerably over time, and the number of published trials is increasing every year. In our living systematic review of randomized trials examining the effects of psychological treatments of depression (Cuijpers et al., 2023a), we currently have included several hundreds of such trials. Over the past 15 years, this database has been used in a broad range of meta-analytic studies, examining many relevant research questions, such as the effects of different types of therapies, the differential effects between therapies, the comparative effects compared to pharmacotherapies and combined treatments, the different treatment formats, the effects across different age- and specific target groups, and a range of secondary outcomes (see the systematic review of all previous meta-analyses in Cuijpers et al., 2023a). However, to the best of our knowledge, no previous study has examined the development of randomized trials on psychotherapies for depression over time. Although many of the characteristics of the trials have been tested in meta-analyses, no study has examined whether the characteristics of trials have changed over time.

Such a historical overview is, however, important for several reasons. First, it can be assumed that the quality and, for example, sample sizes (and power) have improved over time, but this has not been examined empirically. Second, it is also not known which types of therapies and treatment formats have been examined over time, and whether the number of sessions of therapies has increased or decreased. The same is true for research in specific target and age groups. An overview of these developments over time is important because it can indicate gaps in research or redundant research in other areas. In the present study, we will provide such a historical overview of the research on psychotherapies for depression.

In addition to such an overview, we will also provide an updated meta-analysis of the effect sizes found in studies comparing psychotherapies with control groups. As indicated, the number of trials examining psychotherapies for depression has increased exponentially over time and includes now more than 550 trials, while previous meta-analyses have included less than 400 trials (Cuijpers, Karyotaki, et al., 2020). It is important, therefore, to give an updated overview of the effect sizes of these therapies across different types of therapies, treatment formats, target groups, and age groups. This will also allow to examine if the effect sizes of therapies have changed over time.

Method

Transparency and Openness

The data used in this study are part of the Metapsy initiative (https://www.metapsy.org), and the data are open and regularly updated. Researchers who want specific data sets can contact Pim Cuijpers.

Identification and Selection of Studies

The present study is part of a larger meta-analytic project on psychological treatments of depression that was registered at the Open Science Framework (Cuijpers et al., 2022; https://doi.org/10.17605/OSF.IO/825C6). The general methods of the project have been described in a separate article (Cuijpers et al., 2023b), and Supplemental Materials are available at the website of the project (https://www.metapsy.org). This database has been used in a series of earlier published meta-analyses (Cuijpers et al., 2023a). The protocol for the current review and meta-analysis has been published at the Open Science Framework (Cuijpers, 2023; https://osf.io/psda2).

Mathias Harrer

The studies included in the present study were identified through the larger, already existing database of randomized trials on the psychological treatment of depression. For this database, we searched four major bibliographical databases (Pubmed, APA PsycInfo, Embase, and the Cochrane Library) by combining index and free terms indicative of depression and psychotherapies with filters for randomized controlled trials. The full search strings can be found at the website of the project (https://www.metapsy.org and https://docs.metapsy.org/databases). Furthermore, we checked the references of earlier meta-analyses on psychological treatments of depression. The database is updated every 4 months and was developed through a comprehensive literature search (from 1966 to September 1, 2022). All records were screened by two independent researchers, and all articles 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 also done by the two independent researchers, and disagreements were resolved through discussion.

For the current systematic review and meta-analysis, we selected randomized controlled trials in which psychological treatments of depression were compared with a control group (waiting list, care-as-usual, pill placebo, other inactive control). We also included trials in which all participants received antidepressants (comparisons between combined treatment and pharmacotherapy only) because we found previously that concurrent use of antidepressants does not interact with the effects of psychotherapies (Cuijpers, Miguel, Harrer, et al., 2023). We defined psychological treatments as "the informed and intentional application of clinical methods and interpersonal stances derived from established psychological principles for the purpose of assisting people to modify their behaviors, cognitions, emotions, and/or other personal characteristics in directions that the participants deem desirable" (Campbell et al., 2013). Care-as-usual could be delivered in primary care, specialized mental health care, perinatal care, and general medical care (in patients with comorbid general medical disorders). We also included studies with no treatment control groups when they were not recruited from one specific setting, and that they did not receive any treatment in the context of the trial but were allowed to seek treatment anywhere. We based this operationalization of care-as-usual on a previous meta-analysis in which we analyzed these different operationalizations and found no significant differences between them (Cuijpers, Quero, Papola, et al., 2021). We included trials regardless of age group or specific target group. We included trials using any treatment format (including individual, group, digital or nondigital guided self-help, telephone) but did not include unguided psychological interventions because we previously found that these are significantly less effective than other formats (Cuijpers et al., 2019). We did not include trials on inpatients (Cuijpers, Ciharova, et al., 2021). We also did not include trials comparing psychotherapies with other psychotherapies or with antidepressants because the trials comparing psychotherapies with control groups are by far the largest cluster of trials, and the other groups of comparisons are relatively small. The comparisons of therapies with control groups also allow to provide one overall pooled effect size and specific effect sizes across types of therapy and across different age- and other-specific target groups. Depression could be defined as meeting criteria for a depressive disorder according to a diagnostic interview or as a score above the cutoff on a validated self-report depression measure.

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 (J. P. T. 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; self-report measures were rated as low risk of bias because they have been found to be more conservative than clinician-rated instruments; Cuijpers et al., 2010); 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.

Clara Miguel

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 on the website of the project (https://docs.metapsy.org/databases/depression-psyctr/).

Outcome Measures

For each comparison between a psychological treatment and a control condition, the effect size indicating the difference between the two groups in terms of depressive symptoms at posttest was calculated (Hedges' *g*). Effect sizes were calculated by subtracting (at posttest) 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 for the depression measures 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.

Historical Overview

In order to examine the historic development of the studies in this field, we first conducted a Poisson regression analysis to examine if the number of trials published each year has increased over time. Then, we added each of the characteristics of the included participants, interventions, and studies as predictors of the number of published studies per year. Because time was expected to be the strongest predictor, we kept this as a coefficient in all models. Each of the characteristics of the participants, interventions, and studies was then added to separate models. In each model, we included an interaction term with publication year to examine if the scientific output in some fields or subgroups has been growing more rapidly than in others. All characteristics were examined as categorical variables.

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). The metapsyTools package was specifically developed for the meta-analytic project of which this study is a part. This package imports the functionality of the meta (Balduzzi et al., 2019), metafor (Viechtbauer, 2010), and dmetar (Harrer et al., 2019) packages.

We calculated the pooled effect sizes in several different ways, as implemented in the metapsyTools package, to explore if different pooling methods resulted in different outcomes. In our main model, all effect size data available for a comparison in a specific study were aggregated within that comparison first. These aggregated effects were then pooled across studies and comparisons. An intrastudy correlation coefficient of $\rho = .5$ was assumed to aggregate effects within comparisons.

We conducted several other analyses to examine whether these main outcomes were robust. First, we estimated the pooled effect using a three-level "correlated and hierarchical effects" (CHE) model, which was recently proposed by Pustejovsky and Tipton (2022), parameter tests and confidence intervals of which were also calculated using Robust variance estimation to guard against model misspecification. We assumed an intra-study correlation of $\rho = .5$ for this model. Second, we pooled effects while excluding outliers, using the "nonoverlapping 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., 2021). Third, we pooled effects while excluding influential cases as defined by the diagnostics in Viechtbauer and Cheung (2010). Fourth, we calculated the effect when only the smallest or largest effect in each study was considered. Fifth, 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 & Tweedie, 2000), Rücker's "limit meta-analysis method" (Rücker et al., 2011), and a step function selection model (Carter et al., 2019; McShane et al., 2016).

A random-effects model was assumed for all analyses. Between-study heterogeneity variance (components) was estimated using restricted maximum likelihood. For models not fitted using RVE, we applied the Knapp–Hartung method to obtain robust confidence intervals and significance tests of the

Marketa Ciharova

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. A value of 0% indicates no observed heterogeneity, and larger values indicate increasing heterogeneity, with 25% as low, 50% as moderate, and 75% as high heterogeneity (J. P. Higgins et al., 2003). For the three-level model, we calculated a multilevel extension of I^2, which describes the amount of total variability attributable to heterogeneity within studies (Level 2) and heterogeneity between studies (Level 3; Cheung, 2014; Harrer et al., 2019). Because I^2 cannot be interpreted as an absolute measure of the between-study heterogeneity, we also added the prediction interval (PI) to the main analyses, which indicates the range in which the true effect size of 95% of all populations will fall (Borenstein et al., 2009, 2017). 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 17%; Cuijpers, Karyotaki, et al., 2021).

We conducted a series of subgroup analyses for all extracted characteristics of the studies, using a mixed-effects model in which studies were pooled with a random effects model within the subgroups and with a fixed-effects model across subgroups. We also conducted a multivariate metaregression analysis in which we added all characteristics of the participants, the interventions, and the studies in order to examine which characteristics remained significant after adjustment for the others.

Results

Selection and Inclusion of Studies

After examining a total of 32,290 records (22,496 after the removal of duplicates), we retrieved 3,816 full-text articles for further consideration. We excluded 3,254 of the retrieved articles. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart describing the inclusion process, including the reasons for exclusion, is presented in Supplemental File A. A total of 562 randomized controlled trials (with 669 comparisons between a treatment and a control group) met the inclusion criteria for this meta-analysis. The references to the included studies are given in Supplemental File B.

A summary of key characteristics of the 562 included studies and 669 comparisons is presented in Supplemental File C. An overview of aggregated characteristics is given in Supplemental File D. This table also gives the number of studies across the different time periods. In the trials, 66,361 patients participated, 35,170 in the intervention, and 31,191 in the control conditions.

Characteristics of Participants and Changes Over Time

We first conducted a Poisson regression analysis to examine whether the number of studies published each year increased over time. This was found to be the case. The average number of studies published per year was roughly 8.6 (antilog of the intercept 2.15; $SE = 0.06$, $p < .001$), and the change over time was significant ($b = 0.98$; $SE = 0.06$; $p < .001$). Next, we examined if the interaction between publication year and each of the characteristics of the participants, studies, and interventions was significant predictor of the number of published studies per year. Supplemental File D gives the absolute number of trials across different time periods. Table 1 gives the results of the Poisson regression analyses.

Target Groups

As can be seen from Supplemental File D, most studies were conducted in adults in general (201; 36%). Smaller numbers of studies were conducted in children and adolescents (54; 10%), patients with comorbid general medical disorders (111; 20%), women with perinatal depression (62; 12%), older adults (47; 8%), college students (22; 4%), and other-specific target groups (65; 12%). The Poisson model (Table 1) showed that compared to the studies in unselected adults (reference group), the proportion of studies in children and adolescents (interaction term: $b = -0.39$; $SE = 0.17$; $p = .02$) and the proportion of studies in older adults ($b = -0.37$; $SE = 0.18$; $p = .04$) have significantly decreased over time. The proportion of studies in patients with general medical disorders has significantly increased over time ($b = 0.36$; $SE = 0.17$; $p = .04$). Figure 1a gives a graphical representation of the predicted number of studies per year by target group.

Eirini Karyotaki

Diagnosis

In 292 trials (52%), participants met the criteria for a depressive disorder according to a diagnostic interview, 234 trials used a cutoff score on a self-report measure to include participants (42%), and 36 trials focused on subthreshold depression (clinically relevant symptoms but no depressive disorder; 6%). Over time, the proportion of trials using a cutoff score for inclusion of participants increased significantly compared to the proportion of trials using a diagnostic interview ($b = 0.25$; $SE = 0.11$; $p = .02$). See Supplemental File E.1 for a graphical representation.

Recruitment

In 155 trials, participants were recruited through clinical referrals (28%), 214 conducted recruitment through the community (38%), and 193 used other recruitment methods, such as screening or through specific settings (34%). With clinical recruitment as reference category, no significant change in proportion of studies occurred over time for community recruitment, but the proportion of trials using other recruitment strategies did increase significantly over time ($b = 0.32$; $SE = 0.15$; $p = .03$; Supplemental Figure E.2).

Proportion of Women

A total of 107 trials (21%) were exclusively aimed at women, 319 (62%) included between 51 and 99% women, 72 (14%) included 1%–50% women, and 17 (3%) were exclusively aimed at men. We found no indication (using the category with 100% women as reference group) that the proportion of studies in one of these categories changed significantly over time (Table 1; Supplemental File E.3) when compared to the reference category.

Characteristics of Therapies

In the 562 trials, 669 psychological treatments were compared with a control condition. We conducted a Poisson regression analysis to examine whether the number of

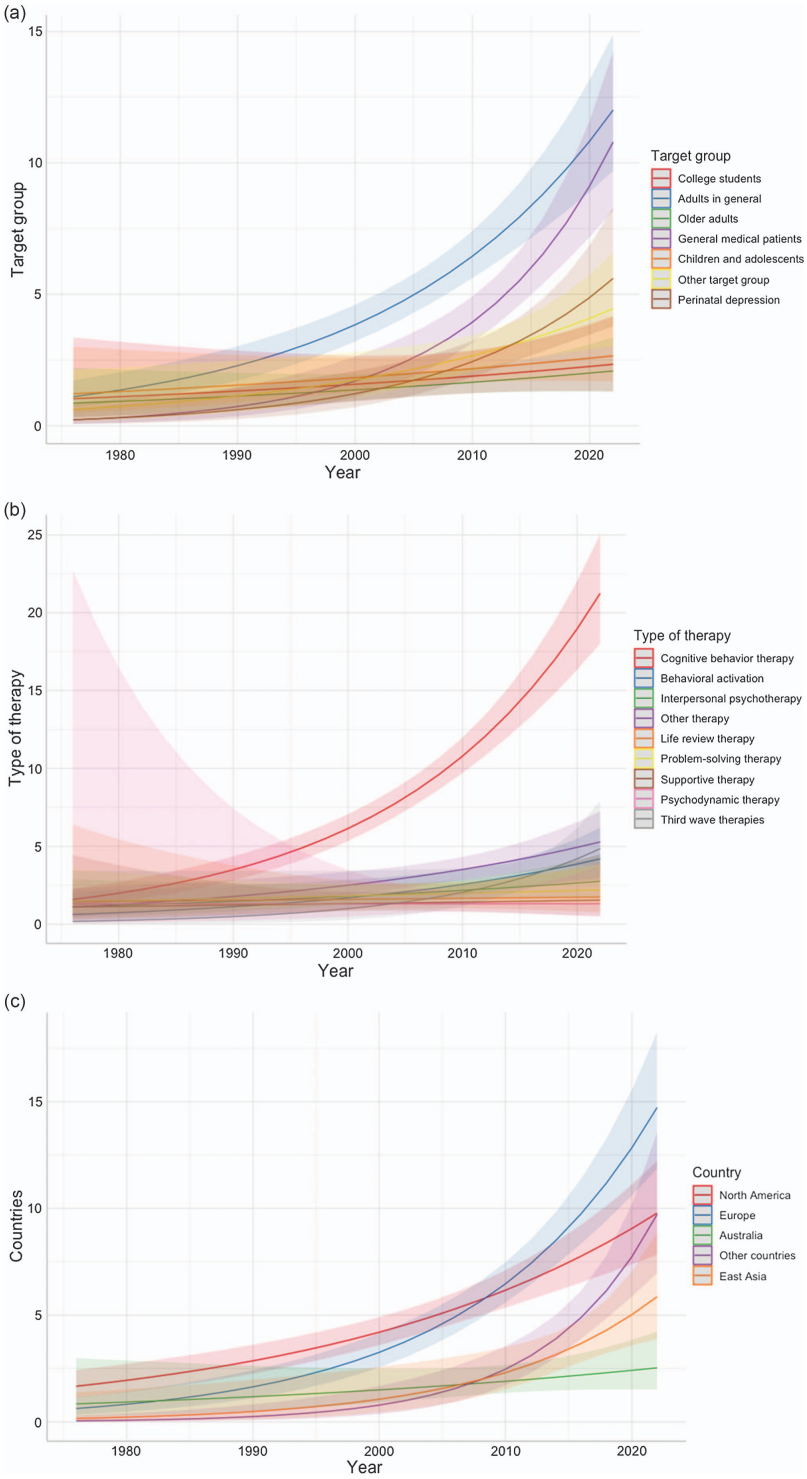
Table 1
Increase of Publications per Year Over Time (Poisson Regression Analyses)

Characteristic	Category	*b*	*SE*	*p*
Population characteristics (all studies)				
Target group	Adults	Ref.		
	Children/adolescents	−0.39	0.17	*.02*
	General medical	0.36	0.17	*.04*
	Older adults	−0.37	0.18	*.04*
	Other	−0.10	0.20	.61
	Perinatal depression	0.20	0.22	.36
	College students	−0.38	0.20	.06
Diagnosis	Clinical interview	Ref.		
	Cutoff	0.25	0.11	*.02*
	Subthreshold	−0.50	0.27	.07
Recruitment	Clinical	Ref.		
	Community	0.07	0.13	.61
	Other	0.32	0.15	*.03*
Proportion women	100%	Ref.		
	51%–99%	0.10	0.15	.53
	1%–50%	−0.06	0.16	.70
	0%	−0.00	0.15	.98
All comparisons				
Type of therapy	CBT	Ref.		
	Behavioral activation	−0.18	0.19	.36
	Psychodynamic therapy	−0.66	0.50	.18
	Interpersonal	−0.43	0.18	*.02*
	Third wave therapies	0.18	0.44	.68
	Problem solving	−0.56	0.17	*.001*
	Supportive	−0.58	0.31	.06
	Life review	−0.61	0.28	*.03*
	Other psychotherapy	−0.26	0.14	.06
Format	Individual	Ref.		
	Group	0.05	0.10	.65
	Guided self-help	0.38	0.15	*.01*
	Telephone	−0.21	0.36	.57
	Other/mixed	0.01	0.20	.95
Number of sessions	<8	Ref.		
	8–12	0.06	0.11	.58
	>12	−0.53	0.14	*<.001*
Study characteristics				
Countries	North America	Ref.		
	Europe	0.36	0.12	*.003*
	Australia	−0.18	0.23	.44
	East Asia	0.47	0.33	.15
	Other	0.91	0.28	*.001*
Control groups	Care-as-usual	Ref.		
	Waitlist	−0.45	0.13	*<.001*
	Pharmacotherapy[a]	−0.93	0.19	*<.001*
	Other control	−0.42	0.16	*.01*
Low risk of bias	Low	Ref.		
	Higher	−0.37	0.15	*.02*
Number of participants	<25	Ref.		
	26–50	0.72	0.12	*<.001*
	51–100	1.00	0.14	*<.001*
	101–200	0.93	0.17	*<.001*
	>200	1.25	0.24	*<.001*

Note. *SE* = standard error; Ref. = reference group; CBT = cognitive behavior therapy. The italic values indicate significance ($p < 0.05$).
[a] These are the studies comparing combined treatment with pharmacotherapy alone.

Figure 1

Predicted Studies Published by Target Group, by Type of Therapy, and by Country



Note. (a) Predicted studies published per year, by target group. (b) Predicted studies published per year, by type of therapy. (c) Predicted studies published per year, by country.

comparisons increased over time. This was found to be the case (intercept 2.44; $SE = 0.05$, $p < .001$; change over time: $b = 0.82$; $SE = 0.05$; $p < .001$).

Type of Therapy

A total of 346 of the intervention arms (52%) examined cognitive behavior therapy (CBT), 51 (8%) interpersonal psychotherapy, 44 (7%) third wave therapies, 43 (6%) behavioral activation, 40 (6%) problem-solving therapy, 20 (3%) nondirective counseling, 18 (3%) life review therapy, 17 (3%) psychodynamic therapy, and 90 (14%) other therapies.

We found that over time, the proportion of studies on interpersonal psychotherapy ($b = -0.43$; $SE = 0.18$; $p = .02$), on problem-solving therapy ($b = -0.56$; $SE = 0.17$; $p = .001$), and life review therapy ($b = -0.61$; $SE = 0.28$; $p = .03$) was significantly reduced, compared with CBT (the reference group; Figure 1b).

Treatment Format

A total of 251 interventions (38%) used an individual format, 211 (32%) had a group format, 125 (19%) a guided self-help format, 23 (3%) delivered the intervention through the telephone, and the remaining 59 interventions (9%) had a mixed format. Over time, we found that when the individual format was used as reference category, only the proportion of studies using guided self-help format had significantly increased ($b = 0.38$; $SE = 0.15$; $p = .01$; Supplemental Figure F.1).

Number of Sessions

The number of sessions was smaller than eight in 236 interventions (36%), 312 (47%) had 8–12 sessions, and 111 (17%) had more than 12 sessions. With less than eight sessions as reference category, the proportion of interventions with more than 12 sessions had significantly decreased over time ($b = -0.53$; $SE = 0.14$; $p = <.001$; Supplemental Figure F.2). Because the smaller number of sessions could be related to the increasing number of trials on guided self-help, we conducted post hoc analyses, excluding studies on guided self-help. We found, however, that the proportion of interventions with more than 12 sessions still had significantly decreased over time compared to the eight sessions as reference category, although the association was less strong ($b = -0.36$; $SE = 0.14$; $p = .01$).

Characteristics of the Studies

Countries

Overall, 208 studies (37%) were conducted in North America, 192 (34%) in Europe, 54 (10%) in East Asia, 34 (6%) in Australia, and the remaining 74 (13%) in other countries. With North America (the United States and Canada) as the reference category, the proportion of trials in Europe (including the United Kingdom) has increased significantly ($b = 0.36$; $SE = 0.12$; $p = .003$), as well as the proportion of trials in other countries (not in North America, Europe, Australia, or East Asia; $b = 0.91$; $SE = -0.28$; $p = .001$; see also Figure 1c).

Type of Control Groups

In 246 studies (44%), usual care was used as the control group, 172 (31%) used a waitlist control group, in 53 studies (9%) all participants received pharmacotherapy (also in the control group), and the 91 remaining studies (16%) used another inactive control group. With usual care as reference category, it can be seen from Table 1 and Supplemental File G.1 that over time, the proportion of studies using any other types of control condition was reduced, and that was true for waitlist control groups ($b = -0.45$; $SE = 0.13$; $p < .001$), pharmacotherapy ($b = -0.93$; $SE = 0.19$; $p < .001$), and other control groups ($b = -0.42$; $SE = 0.16$; $p = .01$).

Low Risk of Bias

Three hundred forty-eight of the 562 studies (62%) reported an adequate sequence generation; 272 (48%) reported allocation to conditions by an independent party; 155 (28%) reported using blinded outcome assessors, while 371 (66%) used only self-report outcomes. In 369 studies (66%) intent-to-treat analyses were conducted. One hundred ninety studies (34%) met all criteria for low risk of bias, 247 studies (44%) met two or three criteria, and 125 (22%) met only one or none of the criteria. In the historical analyses, we looked at the total proportion of studies with low risk of bias (meeting all criteria). With the studies with low risk of bias as reference category, we found that the proportion of trials with at least some risk of bias significantly decreased over time ($b = -0.37$; $SE = 0.15$; $p = .02$; Table 1; Supplemental File G.2).

Number of Participants

We examined the number of participants in the comparisons (not in the trials because that would result in higher numbers for trials with three or more arms). The number of participants was 25 or smaller in 85 (13%), between 26 and 50 in 186 comparisons (28%), between 51 and 100 in 204 (31%), between 101 and 200 in 111 (17%), and larger than 200 in 83 comparisons (12%). With the proportion of studies including less than 25 participants as reference category, we found that all other categories increased over time (Table 1, Supplemental File G.3; 26–50 participants: $b = 0.72$; $SE = 0.12$; $p = <.001$; 51–100 participants: $b = 1.00$; $SE = 0.14$; $p = <.001$; 101–200 participants: $b = 0.93$; $SE = 0.17$; $p = <.001$; >200 participants: $b = 1.25$; $SE = 0.24$; $p < .001$).

Overall Effect Sizes and Effect Sizes in Subgroups

The pooled effects of all psychotherapy conditions compared with the control conditions can be found in Table 2. The primary analyses, in which effect sizes were pooled within a study before pooling across studies, were $g = 0.67$ (95% CI [0.62, 0.72]) with high heterogeneity ($I^2 = 83$; 95% CI [82, 84]) and a wide PI [−0.41, 1.75]. The effect size corresponded with an NNT of 4.57. All sensitivity analyses resulted in a significant effect, although the effects were smaller when only comparisons with low risk of bias were included ($g = 0.49$; 95% CI [0.42, 0.55]). The effect sizes were also considerably smaller after adjustment for publication bias (g ranged from 0.26 to 0.57). Heterogeneity was high in all analyses, except when 202 outliers were removed ($I^2 = 19$; 95% CI [9, 29]). The prediction intervals were very wide in all analyses.

The effect sizes in the subgroups of studies are reported in Table 3. We found significant differences between subgroups for target group, proportion of women, therapy type, treatment format, type of control group, country, risk of bias, and publication year, but not for type of diagnosis, recruitment, and number of sessions. We conducted a multivariate metaregression analysis in which we included all variables to examine which characteristics remained significant after adjustment for the other characteristics (Table 4). We found that studies in which participants met criteria for a mood disorder had a larger effect size compared to studies in which participants scored above a cutoff on a self-rating instrument ($b = 0.17$; $SE = 0.05$; $p = .002$). We also found that studies in children and adolescents had significantly smaller effect sizes compared to studies in unselected adults ($b = -0.27$; $SE = 0.10$; $p = .007$), and studies in other-specific target groups had larger effect sizes than in unselected adults ($b = 0.22$; $SE = 0.09$; $p = .01$). Studies on interpersonal psychotherapy had significantly smaller effect sizes than CBT ($b = -0.29$; $SE = 0.10$; $p = .003$), as had studies on other therapies than the main eight types of therapy ($b = -0.18$; $SE = 0.07$; $p = .02$), while studies on life review therapy had significantly larger effect sizes compared to CBT ($b = 0.57$; $SE = 0.16$; $p < .001$). Studies using a waitlist control group had larger effect sizes than care-as-usual control groups ($b = 0.30$; $SE = 0.07$; $p < .001$). Studies in East Asia ($b = 0.28$; $SE = 0.10$; $p = .005$) and other countries ($b = 0.61$; $SE = 0.09$; $p < .001$) had larger effect sizes compared to those in North America. Studies with low risk of bias had significantly smaller effect sizes than studies with higher risk of bias ($b = -0.21$; $SE = 0.05$; $p < .001$). We found no indication that year of publication was significantly associated with the effect size.

Discussion

We examined major characteristics of the 562 randomized trials comparing psychotherapies for depression with control conditions that have been conducted in the past 5 decades. We found that since the first trials were conducted in the 1970s, every year more and more trials were conducted. However, this growth of the number of trials was not evenly divided across types, format and length of the therapies, the target groups, and the characteristics of the studies.

Compared to the proportion of studies in unselected adults, we found that over time, the proportion of studies in children and adolescents became significantly smaller. This is worrying because we also found that the effects of therapies in this group are significantly smaller than in other age groups (see also Cuijpers, Karyotaki, et al., 2020). At the same time, there are indications that the prevalence of depression is increasing in young people (Mojtabai et al., 2016). The lower proportion of research in older adults compared to unselected adults is also worrying because of the aging populations in many Western countries (Mather et al., 2015). A positive development is that the proportion of trials in general medical patients is increasing compared to the proportion of trials in unselected adults. This is an important acknowledgment of the co-occurrence of medical and mental health problems and the clinical relevance of these problems.

Table 2
Overall Pooled Effect Sizes

Effect sizes	k	g	95% CI	I^2	95% CI	95% PI	NNT
Combined	669	0.67	[0.62, 0.72]	83	[82, 84]	[−0.41, 1.75]	4.57
One ES/study (lowest)	562	0.69	[0.63, 0.75]	84	[83, 85]	[−0.45, 1.82]	4.44
One ES/study (highest)	562	0.70	[0.64, 0.76]	83	[82, 85]	[−0.39, 1.79]	4.35
Outliers removed	467	0.60	[0.58, 0.63]	19	[9, 29]	[0.41, 0.79]	5.18
Influence analysis	651	0.59	[0.56, 0.63]	78	[76, 79]	[−0.18, 1.37]	5.28
Only low risk of bias	218	0.49	[0.42, 0.55]	79	[76, 81]	[−0.27, 1.25]	6.67
Three-level model (CHE)	1,079	0.68	[0.62, 0.73]	88	—	[−0.53, 1.88]	4.52
Publication bias correction							
Trim-and-fill method	855	0.38	[0.32, 0.45]	89	[88, 89]	[−1.27, 2.04]	8.76
Limit meta-analysis	669	0.28	[0.21, 0.35]	87	—	[−0.8, 1.36]	12.49
Selection model	669	0.57	[0.50, 0.65]	91	[89, 93]	[−0.73, 1.87]	5.52

Note. CI = confidence interval; PI = prediction interval; NNT = number-needed-to-treat; ES = effect size; CHE = correlated and hierarchical effects.

Table 3
Subgroup Analyses

Characteristic	Level	*k*	*g*	95% CI	I^2	95% CI	NNT	*p*
Diagnosis	Mood disorder	345	0.69	[0.62, 0.76]	83	[81.2, 84.4]	4.17	.750
	Cutoff	282	0.66	[0.57, 0.74]	82	[80.5, 84.2]	4.39	
	Subthreshold	42	0.63	[0.45, 0.81]	88	[84.4, 90.4]	4.63	
Recruitment	Clinical	175	0.64	[0.53, 0.75]	84	[81.2, 85.5]	4.54	.255
	Community	278	0.63	[0.57, 0.69]	73	[70, 76.3]	4.63	
	Other	216	0.74	[0.63, 0.85]	88	[87, 89.5]	3.85	
Target group	Adults in general	332	0.69	[0.62, 0.76]	81	[78.6, 82.4]	4.17	*<.001*
	Children/adolescents	60	0.35	[0.25, 0.45]	69	[60.1, 76.4]	8.97	
	College students	29	1.00	[0.76, 1.23]	77	[66.5, 83.5]	2.76	
	General medical	123	0.67	[0.52, 0.82]	86	[83.8, 87.9]	4.31	
	Perinatal depression	70	0.66	[0.51, 0.81]	87	[84.5, 89.4]	4.39	
	Older adults	55	0.78	[0.6, 0.96]	82	[77, 85.7]	3.63	
Proportion women	0%	22	0.86	[0.21, 1.51]	88	[83.5, 91.6]	3.25	*<.001*
	1%–50%	85	0.45	[0.35, 0.55]	74	[67.3, 78.6]	6.77	
	51%–99%	386	0.68	[0.62, 0.74]	81	[78.7, 82.3]	4.24	
	100%	123	0.75	[0.62, 0.89]	86	[83.9, 88]	3.80	
Therapy type	CBT	347	0.71	[0.63, 0.79]	85	[83.5, 86.2]	4.04	*<.001*
	IPT	51	0.43	[0.3, 0.56]	78	[70.8, 82.7]	7.13	
	Third wave therapy	44	0.92	[0.7, 1.14]	77	[68.8, 82.4]	3.02	
	Behavioral activation	43	0.72	[0.58, 0.85]	68	[55.6, 76.4]	3.97	
	Problem solving	39	0.63	[0.43, 0.83]	82	[76, 86.4]	4.63	
	Supportive therapy	20	0.52	[0.32, 0.71]	57	[29, 73.8]	5.75	
	Life review therapy	18	1.25	[0.85, 1.65]	90	[85.5, 92.9]	2.18	
	Psychodynamic	17	0.58	[0.16, 1.01]	93	[89.7, 94.7]	5.08	
	Other therapy	90	0.46	[0.36, 0.55]	74	[68.6, 79.1]	6.61	
Format	Individual	251	0.63	[0.54, 0.72]	81	[78.6, 82.9]	4.63	*.003*
	Group	211	0.83	[0.72, 0.95]	87	[85.2, 88.1]	3.39	
	Guided self-help	123	0.60	[0.53, 0.67]	74	[69.4, 78.4]	4.89	
	Telephone	23	0.63	[0.38, 0.87]	86	[80.2, 90.1]	4.63	
	Other/mixed	61	0.52	[0.4, 0.65]	84	[80.1, 87.1]	5.75	
No. of sessions	<8	234	0.61	[0.54, 0.68]	81	[79, 83.3]	4.80	.100
	8–12	305	0.73	[0.65, 0.82]	84	[82.7, 85.7]	3.91	
	>12	108	0.64	[0.5, 0.78]	84	[81.5, 86.7]	4.54	
Control group	Care-as-usual	269	0.59	[0.52, 0.67]	84	[82.5, 85.7]	4.98	*<.001*
	Waiting list	235	0.83	[0.75, 0.91]	76	[72.3, 78.4]	3.39	
	Pharmacotherapy[a]	56	0.51	[0.31, 0.72]	86	[82.9, 88.9]	5.88	
	Other control	109	0.61	[0.45, 0.76]	85	[82.9, 87.5]	4.80	
Country	North America	261	0.56	[0.49, 0.62]	73	[68.9, 75.7]	5.29	*<.001*
	Europe (including the United Kingdom)	222	0.52	[0.46, 0.59]	76	[73.2, 79.2]	5.75	
	East Asia	57	0.86	[0.65, 1.08]	89	[87, 91.3]	3.25	
	Australia	43	0.74	[0.54, 0.94]	75	[66.6, 81.5]	3.85	
	Other	86	1.27	[1, 1.53]	92	[90.3, 92.8]	2.15	
Risk of bias	Low	217	0.49	[0.42, 0.55]	79	[76.2, 81.5]	6.15	*<.001*
	Other	452	0.77	[0.7, 0.84]	84	[82.5, 85.1]	3.68	
Year	<1985	34	1.06	[0.8, 1.32]	76	[66.4, 82.6]	2.59	*<.001*
	1986–1995	53	0.69	[0.55, 0.83]	55	[38.5, 66.9]	4.17	
	1996–2005	90	0.53	[0.42, 0.63]	70	[63.3, 76]	5.63	
	2006–2015	246	0.62	[0.54, 0.7]	83	[81.3, 85]	4.71	
	>2016	246	0.72	[0.62, 0.82]	87	[85.4, 88.1]	3.97	

Note. CI = confidence interval; NNT = number-needed-to-treat; CBT = cognitive behavior therapy; IPT = interpersonal psychotherapy. The italic values indicate significance ($p < 0.05$).
[a] These are the studies comparing combined treatment with pharmacotherapy alone.

The proportion of trials using a cutoff on a self-report depression scale has increased significantly over time compared to the proportion of trials using a diagnostic interview to include participants. At the same time, we found in the metaregression analysis that studies in participants with a mood disorder resulted in somewhat larger effect sizes. Previous meta-analyses have not found significant differences between the effect sizes of these two subgroups of studies (e.g., Cuijpers, Karyotaki, Reijnders, & Huibers, 2018). It is important, however, to examine this finding further because a diagnosis may indicate more severe depressive states, and larger effects would be good news for patients with a diagnosis. The increase of trials using a cutoff on a self-report depression as an inclusion criterion is logistically less challenging compared to a clinical diagnosis and could point at a trend to conduct simpler and "easier" trials (also reflected in the decreasing number of sessions in the trials over time; see below). It can also reflect a development to step away from clinical diagnoses toward a

Table 4
Metaregression Analyses

Characteristic	Level	*b*	*SE*	*p*
Year (continuous)		−0.00	0.00	.17
Diagnosis	Cutoff	Ref.		
	Mood disorder	0.17	0.05	*.002*
	Subthreshold depression	0.13	0.10	.17
Recruitment	Clinical	Ref.		
	Community	0.01	0.07	.91
	Other	0.05	0.08	.56
Target group	Adults in general	Ref.		
	Children/adolescents	−0.27	0.10	*.007*
	College students	0.15	0.13	.25
	General medical patients	0.06	0.08	.48
	Perinatal depression	0.02	0.11	.84
	Older adults	−0.00	0.10	1.00
	Other specific groups	0.22	0.09	*.01*
Proportion women (continuous)		0.00	0.00	.18
Therapy type	CBT	Ref.		
	Third wave therapies	0.07	0.10	.48
	Interpersonal psychotherapy	−0.29	0.10	*.003*
	Behavioral activation	0.01	0.10	.92
	Problem-solving therapy	−0.05	0.10	.63
	Supportive therapy	−0.04	0.14	.80
	Life review therapy	0.57	0.16	*<.001*
	Psychodynamic therapy	−0.01	0.15	.93
	Other therapies	−0.18	0.07	*.02*
Format	Individual	Ref.		
	Group	0.04	0.06	.49
	Guided self-help	−0.04	0.08	.59
	Telephone	0.11	0.13	.40
	Other/mixed	0.05	0.09	.62
No. of sessions (continuous)		−0.00	0.00	.95
Control group	Care-as-usual	Ref.		
	Waiting list	0.30	0.07	*<.001*
	Pharmacotherapy[a]	−0.12	0.10	.20
	Other control	0.04	0.07	.58
Country	North America	Ref.		
	Australia	0.18	0.11	.08
	Europe (including the United Kingdom)	0.03	0.06	.63
	East Asia	0.28	0.10	*.005*
	Other	0.61	0.09	*<.001*
Risk of bias (low vs. other)		−0.21	0.05	*<.001*

Note. *SE* = standard error; Ref. = reference group; CBT = cognitive behavior therapy. The italic values indicate significance ($p < 0.05$).
[a] These are the studies comparing combined treatment with pharmacotherapy alone.

more dimensional approach where anyone suffering from symptoms can potentially benefit from an intervention. The increased use of self-report measures for inclusion in trials may also be one of the causes of the high level of heterogeneity in our meta-analysis because this reflects less stringent criteria and more heterogeneous participants. More research on these issues is, therefore, very important.

More than half of all trials examined the effects of CBT, which makes it by far the best examined type of therapy. None of the other therapies were examined in more than 8% of the trials. This may be related to our broad definition of CBT (any therapy in which cognitive restructuring is a core element), but it still shows the dominance of CBT in research. In a previous meta-analytic study (Cuijpers et al., 2019), we also examined the effects of specific subtypes of CBT (CBT according to Beck, Beck et al., 1979; the "Coping with Depression course," Cuijpers et al., 2009; and guided self-help based on David Burns' self-help book, Burns, 1980). We did not find significant differences between these subtypes of CBT in a multivariate metaregression analysis. We also saw that compared to CBT the proportion of trials on several other therapies, including interpersonal psychotherapy, problem-solving therapy, and life review therapy, is significantly decreasing over time. In the meta-analysis, we found that the effects of therapies significantly differ from each other, with life review being more effective and interpersonal and problem-solving therapies being less effective than CBT. It should be noted that these results should be considered with caution because life review is a sample-specific therapy, which is only used in older adults.

In a previous network meta-analysis (including direct and indirect evidence), we found no indication that these therapies are more or less effective than each other (Cuijpers, Quero, Noma, et al., 2021). This may imply that the increasing number of trials does point to significant differences, but more research is needed to confirm that. However, regardless of this question, it is important to examine other therapies than CBT alone, because several of these therapies have been found to have comparable effects and may benefit other types of participants.

The large majority of trials have examined the effects of individual and group therapies of depression. However, over time, we found a significant increase of the proportion of trials examining guided self-help. This probably refers to the increasing number of trials on digital interventions, which usually use the guided self-help format. We found that format was not associated with large differences in effect sizes, which is in line with a previous network meta-analysis (Cuijpers et al., 2019). This is encouraging because digital interventions typically require fewer resources.

We also found that the proportion of trials examining interventions with more sessions has decreased significantly over time, with the majority of therapies having 12 or fewer sessions. This was partly explained by the increase of trials on (brief) guided self-help interventions, but this finding was still significant after excluding these trials. It makes sense to examine briefer interventions because we found no significant association between the effects number of sessions and the effect size (see also Cuijpers et al., 2013). On the other hand, one should be careful not to focus exclusively on brief interventions and ignore longer treatments that may be needed for some subgroups of participants.

Most trials have been conducted in the United States. However, over time, the proportion of trials in Europe has significantly increased compared to the proportion in the United States. In recent years, the number of trials in Europe is larger than those in the United States. Over time, also the proportion of trials in countries outside the United States, Europe, Australia, and East Asia has increased significantly, compared to the United States. We found that the effects of therapy were comparable in the United States and Europe but significantly larger in East Asia and other countries, which is in line with previous research (Cuijpers, Karyotaki, Reijnders, Purgato, & Barbui, 2018). This larger effect size may indicate that therapies are indeed better in non-Western countries, or they may be related to the low level of usual care in many non-Western countries. It may also be an artifact of the low quality of many trials across the world. However, it can be safely assumed that therapies are effective across the world. Considering the enormous disease burden of depression across the world, more research on how these therapies can be disseminated broadly across non-Western countries is certainly warranted.

Care-as-usual is currently the most important control group for psychotherapies, and the proportion of trials using any other type of control group has decreased significantly over time compared to those using care-as-usual control groups. This is important because especially waiting list control groups have larger effect sizes than other types of control conditions and may overestimate the true effects of therapies (see also Furukawa et al., 2014).

The proportion of studies with low risk of bias has significantly increased over time, which is encouraging. This is also important because studies with low risk of bias have significantly smaller effect sizes than other studies. However, there is also still much room for improvement because, even in recent years, less than half of the trials meet all four basic criteria for low risk of bias. It is well established that trials which are not conducted well may overestimate the effects of an intervention (Sterne et al., 2019). It is therefore very important that researchers avoid bias arising from the randomization process, deviations from intended interventions, bias due to missing outcome data, bias in measurement of the outcome, and bias in selection of the reported result (Sterne et al., 2019). This reduction of bias is a joint responsibility of researchers, journal editors, and funders.

We found few indications that the effects of therapies have changed over time. In the subgroup analyses, we found that early studies (before 1985) had larger effect sizes than later studies. But after adjustment for other characteristics of the studies, such as risk of bias, this association was no longer significant.

One general finding of this review is that the number of controlled trials is increasing rapidly. At some point, however, new controlled trials add very little new knowledge to what is already known, and it may be better to focus on other designs to answer clinical questions that are also relevant (Cuijpers, in press). For example, comparative trials without inactive control groups can examine if a new treatment is as effective as existing ones; fractional factorial designs can be used to examine effective components of psychosocial interventions; and stepped wedged designs can be used to examine how to implement interventions in routine care.

This study has several strengths. It is the largest review of a psychological treatment ever; all searches and data extraction procedures were double-rated, and we used sophisticated (meta-)analytic techniques to examine developments over time and the effect sizes of the therapies. However, there are also some limitations that have to be taken into account when interpreting the results. First, we only analyzed characteristics that were extracted as part of our main meta-analytic project, while several other characteristics may also have been important, such as the proportion of participants from racial/ethnic minorities, the educational level of participants, comorbidities, the training of the clinicians, or the adherence to the manual. Second, we only examined short-term effects and depressive symptoms as outcomes. Third, many of the

included trials suffered from at least some risk of bias. Fourth, although all data were extracted by two independent reviewers, we do not have the interrater reliability between raters available because the data have been collected over a period of more than 15 years and the interrater reliability has not been collected systematically over time. Fifth, we only selected the studies comparing psychotherapies with control conditions in our historical analyses, and we excluded other subsets of our database, such as comparisons with other therapies and pharmacotherapy, comparisons with different delivery formats, unguided digital interventions, dismantling studies, and comparisons with too few studies to conduct a meta-analysis (such as comparisons of psychotherapy with bright light therapy and exercise). This may have introduced some selection bias in the historical analyses, but it did reduce heterogeneity of types of studies. Sixth, we found clear indications of publication bias, which has inflated the effect sizes of therapies.

Another important limitation of this large data set of trials is that heterogeneity is very high. This is not surprising, considering the large data set and the use of continuous outcomes (Alba et al., 2016). We also explored potential sources of heterogeneity and identified a large number of outliers, as well as several study characteristics that may cause parts of this heterogeneity. However, despite these findings, we should consider this data set highly heterogeneous, which may indicate underlying differences between trials that we did not identify. The results should be considered with caution.

Despite these limitations, we can conclude that the number of trials examining the effects of psychotherapies for depression is increasing every year, especially in high-income countries but also in other countries. Research in children and adolescents is lagging behind, which is problematic because the effects of therapies are significantly smaller in this group. CBT is by far the best examined type of therapy but not necessarily more effective than other therapies. The quality of trials is increasing over time, with less use of waitlist controls, larger number of participants in the trials, and a better prevention of methodological biases. However, there is still room for improvement. The most important conclusion, however, is that the evidence that psychotherapies are effective in the treatment of depression is very strong and growing rapidly every year.

References

Alba, A. C., Alexander, P. E., Chang, J., MacIsaac, J., DeFry, S., & Guyatt, G. H. (2016). High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. *Journal of Clinical Epidemiology*, *70*, 129–135. https://doi.org/10.1016/j.jclinepi.2015.09.005

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression* (1st ed.). Guilford Press.

Borenstein, M., Hedges, L., Higgins, J. P., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Burns, D. D. (1980). *Feeling good: The new mood therapy* (1st ed.). Morrow.

Campbell, L. F., Norcross, J. C., Vasquez, M. J., & Kaslow, N. J. (2013). Recognition of psychotherapy effectiveness: The APA resolution. *Psychotherapy*, *50*(1), 98–101. https://doi.org/10.1037/a0031817

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Covi, L., Lipman, R. S., Derogatis, L. R., Smith, J. E., III, & Pattison, J. H. (1974). Drugs and group psychotherapy in neurotic depression. *The American Journal of Psychiatry*, *131*(2), 191–198. https://doi.org/10.1176/ajp.131.2.191

Cuijpers, P. (2023, May 4). *Five decades of research on psychological treatments of depression: A historical and meta-analytic overview*. http://osf.io/psda2

Cuijpers, P. (in press). Has the time come to stop using control groups in trials of psychosocial interventions? *World Psychiatry*.

Cuijpers, P., Ciharova, M., Miguel, C., Harrer, M., Ebert, D. D., Brakemeier, E. L., & Karyotaki, E. (2021). Psychological treatment of depression in institutional settings: A meta-analytic review. *Journal of Affective Disorders*, *286*, 340–350. https://doi.org/10.1016/j.jad.2021.03.017

Cuijpers, P., Huibers, M., Ebert, D. D., Koole, S. L., & Andersson, G. (2013). How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders*, *149*(1–3), 1–13. https://doi.org/10.1016/j.jad.2013.02.030

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica*, *144*(3), 288–299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Karyotaki, E., Ciharova, M., Quero, S., Pineda, B., Munoz, R., Struijs, S. Y., Llamas, J., Vieira, P. M., Figueroa, C., & Rosenström, T. (2022, February 18). *A meta-analytic database of randomised trials on psychotherapies for depression*. Open Science Framework. https://doi.org/10.17605/OSF.IO/825C6

Cuijpers, P., Karyotaki, E., Eckshtain, D., Ng, M. Y., Corteselli, K. A., Noma, H., Quero, S., & Weisz, J. R. (2020). Psychotherapy for depression across different age groups: A meta-analysis. *JAMA Psychiatry*, *77*(7), 694–702. https://doi.org/10.1001/jamapsychiatry.2020.0164

Cuijpers, P., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2018). Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. *Cognitive Behaviour Therapy*, *47*(2), 91–106. https://doi.org/10.1080/16506073.2017.1420098

Cuijpers, P., Karyotaki, E., Reijnders, M., Purgato, M., & Barbui, C. (2018). Psychotherapies for depression in low- and middle-income countries: A meta-analysis. *World Psychiatry*, *17*(1), 90–101. https://doi.org/10.1002/wps.20493

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review*, *30*(6), 768–778. https://doi.org/10.1016/j.cpr.2010.06.001

Cuijpers, P., Miguel, C., Ciharova, M., Quero, S., Plessen, C. Y., Ebert, D., Harrer, M., van Straten, A., & Karyotaki, E. (2023). Psychological treatment of depression with other comorbid mental disorders: Systematic

review and meta-analysis. *Cognitive Behaviour Therapy*, *52*(3), 246–268. https://doi.org/10.1080/16506073.2023.2166578

Cuijpers, P., Miguel, C., Harrer, M., Ciharova, M., & Karyotaki, E. (2023). Does the use of pharmacotherapy interact with the effects of psychotherapy? A meta-analytic review. *European Psychiatry*, *66*(1), Article e63. https://doi.org/10.1192/j.eurpsy.2023.2437

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Papola, D., Ebert, D., & Karyotaki, E. (2023a). Psychological treatment of depression: A systematic overview of a 'meta-analytic research domain'. *Journal of Affective Disorders*, *335*, 141–151. https://doi.org/10.1016/j.jad.2023.05.011

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Papola, D., Ebert, D., & Karyotaki, E. (2023b). *Randomised trials on psychotherapy for depression: Methods of a 'meta-analytic research domain'* [Manuscript submitted].

Cuijpers, P., Muñoz, R. F., Clarke, G. N., & Lewinsohn, P. M. (2009). Psychoeducational treatment and prevention of depression: The "Coping with Depression" course thirty years later. *Clinical Psychology Review*, *29*(5), 449–458. https://doi.org/10.1016/j.cpr.2009.04.005

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. https://doi.org/10.1001/jamapsychiatry.2019.0268

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92–107. https://doi.org/10.1002/wps.20701

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021). 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. https://doi.org/10.1002/wps.20860

Cuijpers, P., Quero, S., Papola, D., Cristea, I. A., & Karyotaki, E. (2021). Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: A meta-analysis. *Psychological Medicine*, *51*(4), 634–644. https://doi.org/10.1017/S0033291719003581

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Furukawa, T. A. (1999). From effect size into number needed to treat. *Lancet*, *353*(9165), Article 1680. https://doi.org/10.1016/S0140-6736(99)01163-0

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., Chen, P., Hunot, V., & Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(3), 181–192. https://doi.org/10.1111/acps.12275

Furukawa, T. A., Shinohara, K., Sahker, E., Karyotaki, E., Miguel, C., Ciharova, M., Bockting, C. L. H., Breedvelt, J. J. F., Tajika, A., Imai, H., Ostinelli, E. G., Sakata, M., Toyomoto, R., Kishimoto, S., Ito, M., Furukawa, Y., Cipriani, A., Hollon, S. D., & Cuijpers, P. (2021). Initial treatment choices to achieve sustained response in major depression: A systematic review and network meta-analysis. *World Psychiatry*, *20*(3), 387–396. https://doi.org/10.1002/wps.20906

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. (2019). *dmetar: Companion R package for the guide 'doing meta-analysis in R'* (R package Version 0.0.9000) [Computer software]. http://dmetar.protectlab.org

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman & Hall/CRC Press. https://doi.org/10.1201/9781003107347

Harrer, M., Kuper, P., & Cuijpers, P. (2022). *metapsyTools: Several R helper functions for the "metapsy" database* (R package Version 0.3.2, 2022) [Computer software]. https://tools.metapsy.org

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ*, *327*(7414), Article 557. https://doi.org/10.1136/bmj.327.7414.557

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A. C., the Cochrane Bias Methods Group, & the Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *The 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 Medical Research Methodology*, *14*(1), Article 25. https://doi.org/10.1186/1471-2288-14-25

Klerman, G. L., DiMascio, A., Weissman, M., Prusoff, B., & Paykel, E. S. (1974). Drug and group psychotherapy for neurotic depression. *The American Journal of Psychiatry*, *131*, 186–191. https://doi.org/10.1176/ajp.131.2.186

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 Systematic Reviews*, *18*(3), Article e1256. https://doi.org/10.1002/cl2.1256

Mather, M., Jacobsen, L. A., & Pollard, K. M. (2015). Population bulletin: Aging in the United States. *Population Reference Bureau*, *70*(2), 1–5. https://www.prb.org/wp-content/uploads/2019/07/population-bulletin-2015-70-2-aging-us.pdf

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. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Mojtabai, R., Olfson, M., & Han, B. (2016). National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, *138*(6), Article e20161878. https://doi.org/10.1542/peds.2016-1878

Pustejovsky, J. E., & Tipton, E. (2022). Meta-Analysis with robust variance estimation: Expanding the range of working models. *Prevention Science: The Official Journal of the Society for Prevention Research*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H., & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, *12*(1), 122–142. https://doi.org/10.1093/biostatistics/kxq046

Sterne, J. A. C., Savovic, 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. *The BMJ*, *366*, Article l4898. https://doi.org/10.1136/bmj.l4898

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Received June 10, 2023
Revision received August 21, 2023
Accepted September 18, 2023 ■