Check for updates

ORIGINAL ARTICLE

Acta Psychiatrica Scandinavica WILEY

The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis

Pim Cuijpers[1,2] | Eirini Karyotaki[1,2] | Marketa Ciharova[1] | Clara Miguel[1] |
Hisashi Noma[3] | Toshi A. Furukawa[4]

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

[2]WHO Collaborating Centre for Research and Dissemination of Psychological Interventions, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[3]Department of Data Science, Institute of Statistical Mathematics, Tokyo, Japan

[4]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

Correspondence
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

Abstract

Objective: Meta-analyses of psychotherapies usually report effects sizes, while clinicians and patients need to know the proportion of patients who benefit from therapy. We conducted a meta-analysis of therapies for depression reporting the rates of response (50% symptom reduction), remission (HAM-D <7), clinical significant deterioration for psychotherapy, and control conditions (CAU, waitlist, and pill placebo), as well as the relative risk of these outcomes and the numbers-needed-to-be-treated (NNTs).

Methods: We searched bibliographic databases and included 228 randomized trials comparing psychotherapy for depression against control conditions (75 with low risk of bias). Only therapies with at least 10 trials were included. We extracted outcomes from the studies, and for those studies not reporting the outcomes, we used a validated method to estimate the rates.

Results: The overall response rate in psychotherapies at 2 (±1) months after baseline was 41% (95% CI: 38~43), 17% (15~20) for usual care (CAU), and 16% (95% CI: 14~18) for waitlist. No significant differences between types of therapy were found. The NNT for therapy versus CAU was 5.3 and versus waitlist 3.9. About one third of patients remitted after therapy compared with 7%–13% in control conditions. The rates of deterioration were 5% versus 12%–13%, respectively. Most sensitivity analyses supported the general findings.

Conclusion: Psychotherapies for depression may be effective compared with control conditions, but more than half of patients receiving therapy do not respond and only one third remitted. More effective treatments and treatment strategies such as sequencing and combining treatments are clearly needed.

KEYWORDS
cognitive behavior therapy, depression, major depressive disorder, meta-analysis, psychotherapy

1 | INTRODUCTION

Psychological treatments have been found to be potentially effective in the treatment of depression,[1] probably have comparable effects as antidepressant medication at the short term,[2] and may have superior effects at the longer term.[3] Although cognitive behavior therapy is by far the best studied type of therapy for depression, several other types have also been found to be potentially effective, including interpersonal psychotherapy, problem-solving therapy, and behavioral activation.[4] These therapies have been found to have comparable effects, with no therapy being significantly more effective than another.[1] These therapies have also been found to be effective in several different target groups, including older adults, women with perinatal depression, and people with comorbid general medical disorders such as cancer, diabetes, and heart disease.[5] Although the effects of psychotherapies have probably been overestimated because of publication bias and the low quality of most trials,[6] the effects are still significant after adjustment for these problems and comparable to those of antidepressants.

But what does it mean that these therapies are probably effective? Many meta-analyses examining the effects of psychotherapies report the outcomes in terms of standardized mean differences (SMD), indicating the difference between the therapy and a control group after the treatment in terms of standard deviation. For patients and clinicians, however, it is important to know what are the chances of getting better after a treatment. The SMD is not very informative in this respect and cannot be seen as an indicator of clinical relevance, because it is still a statistical concept.[7–9]

Categorical outcomes such as response and remission are easier to interpret because they indicate how many patients get better when they get a treatment. However, these outcomes are often reported as Relative Risks (RRs) or Odds Ratios (ORs) and indicate the relative benefit of a treatment in comparison to a control condition or another treatment. This is easier to interpret than effect sizes, but these outcomes still do not indicate the chance of getting better when receiving treatment.

For patients and clinicians, it is important to simply know the proportions of patients who get better when they receive treatment and the proportion of patients in the control conditions who get better. Most meta-analyses do not report these proportions, because heterogeneity of the pooled proportions is usually very high and may make the interpretation of these proportions complicated. However, the importance of such straightforward numbers is very high for clinicians and patients. Furthermore, meta-analyses of prevalence rates are often published despite very high levels of heterogeneity.[10,11]

In a previous meta-analysis, we did examine pooled response and remission rates of psychotherapy for unipolar depression and control conditions.[8] We found a response rate

Summations

- 41% of patients respond to psychotherapy for depression, compared with 17% in usual care.
- About one third of patients remit in therapies, compared with 9%–17% in control conditions.
- Most patients do not respond or remit after therapy, and more effective treatments are clearly needed.

Limitations

- Heterogeneity was high in most analyses.
- Most rates were not reported but estimated with a validated method.

of 48% for psychotherapies and 19% in control groups, and remission rates of 43% and 27% respectively for therapy and control conditions. However, the number of studies included in this meta-analysis was relatively small ($N < 20$), the quality of these studies was limited, and the follow-up periods were variable, making these results uncertain.

In the current meta-analysis, we used a method to estimate response and remission rates, as well as rates of clinical significant change in psychotherapy for unipolar depression and control conditions using estimates based on the means at baseline, and the means, standard deviations, and N at post-test.[12] This method estimates how many patients are scoring above or below a cutoff assuming a normal distribution of the outcome. In a previous meta-analysis, we found a correlation of 0.94 between the response and remission rates reported in the paper and the estimated rates using this method.[2] Because the data needed to calculate these rates are reported in most studies on psychotherapy, we could include a large sample of studies and report rates at different follow-up points, for different therapies and for studies with low risk of bias. We could also use this method to estimate clinically significant deterioration, which is usually not reported in studies because these rates are typically low and non-significant.[13] Based on these data, it is also possible to directly calculate numbers-needed-to-be-treated (NNTs), indicating how many patients have to be treated in order to have one more positive outcome compared with the comparison group.[14]

1.1 | Aims of the study

To calculate the pooled rates of response, remission rates, clinical significant improvement, and clinical significant deterioration in psychotherapy for unipolar depression and control conditions in a meta-analysis.

2 | METHODS

2.1 | Search strategy and selection criteria

The protocol for this meta-analysis was registered at the Open Science Framework.[15] We used an existing database of studies on the psychological treatment of unipolar depression.[16] The database is continuously updated and was developed through a comprehensive literature search (up to January 1st 2020). For this database, we searched four major bibliographical databases (PubMed, PsycINFO, Embase, and Cochrane Library) by combining terms (index and text words) indicative of depression and psychotherapies, with filters for randomized controlled trials. Full search strings are available on the project's website (www.metapsy.org) and in Appendix A. We also checked the references of previous meta-analyses. 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 also done by two independent researchers, and disagreements were solved through discussion and consensus.

For the current meta-analysis, we included (a) randomized trials (b) in which a psychological treatment (c) for adult depression (d) was compared with a control group (waitlist, CAU, and pill placebo). We included only studies that either reported response rates or reported sufficient data to calculate the response rate using the method developed by Furukawa.[12] We included only psychotherapies for which at least 10 studies were available, because that allowed to calculate outcomes for each therapy separately. Psychotherapies were defined according to the definitions developed in previous meta-analyses.[1,17,18] Waitlist was defined as a control group in which participants were offered the intervention at some point after participants in the experimental group had finished the intervention. We included studies in which the presence of a depressive disorder was established using a diagnostic interview as well as studies in which participants had to score above a cutoff on a self-report depression scale. Co-morbid mental or somatic disorders were not used as exclusion criteria. Studies with control conditions other than waitlist, CAU, or pill placebo were excluded. Studies with a follow-up of less than 4 weeks were also excluded. No language restrictions were applied.

2.2 | Quality assessment and data extraction

We assessed the validity of included studies using four criteria of the Cochrane "Risk of bias" assessment tool[19]: allocation sequence generation; concealment of allocation to conditions; prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as low risk when intention-to-treat

analyses were conducted). Items were dichotomized as low or high/unclear risk. These assessments were conducted by two independent researchers, and disagreements were solved through discussion.

We also coded participant characteristics, study characteristics, and the time from baseline to outcome. If available, we also extracted baseline severity on the HAM-D.[20]

2.3 | Outcome measures

Treatment response (50% reduction in depressive symptomatology) was the primary outcome. We retrieved all response rates at all time points that were reported in the included studies, but we focused the main analyses on response rates at 2 ($\pm$1) months after baseline. We choose the time point of 2 ($\pm$1 months) as the main outcome for pragmatic reasons, because most studies reported data for this follow-up point and this can therefore be considered as the typical length of the acute phase treatment for depression. When more than one outcome measure was reported, we selected the outcome according to an algorithm that has been used in previous meta-analyses[2] so as to minimize potential selective outcome reporting bias. The main outcome was the response rate at post-test, assuming that all study drop-outs were non-responders, because this was considered to be the most conservative estimate. We also conducted sensitivity analyses, one in which all participants lost to follow-up were considered to be responders, and one in study completers.

We categorized the response rates according to the time between baseline and post-test, and selected the post-test at 2 ($\pm$1) months after baseline as the main outcome. We also examined remission rates, defined as a score of $\leq$7 on the 17-item HAM-D,[21] using the same method of Furukawa. In addition, we calculated the Reliable Change Index, which is a psychometric criterion used to evaluate whether the change between baseline and post-test is considered statistically significant (the difference between baseline and post-test means divided by the standard error of the difference between the two scores is greater than 1.96, conservatively assuming a Cronbach's alpha of 0.75).[22] We used the same method to calculate the Reliable Deterioration Index, indicating whether a patient reliably deteriorated.

2.4 | Meta-analyses

We calculated the pooled response rates at different follow-up times, as well as the remission rates, reliable change, and reliable deterioration, as well the RRs for these outcomes and the NNTs. We first pooled rates for response, remission, reliable change, and reliable deterioration using the "meta" package in R (version 3.6.3). In these analyses, we

—WILEY—

synthesized the binomial outcome data by random-effects pooling models after transforming to a logit scale. The summary results were converted to the raw proportion scale, and the estimates and their 95% confidence intervals (CIs) are presented. Because we expected considerable heterogeneity, we employed a random-effects pooling model in all analyses, according to the DerSimonian-Laird method. As indicator of heterogeneity, we calculated the I^2 statistic and its 95% CI.[23]

We first meta-analyzed response rates for psychotherapies and control conditions separately at 2 ($\pm$1 month) follow-up (our primary outcome). We also pooled response rates assuming that all drop-outs are responders and the rates for the completers of the study. We then examined the risk of small study effects by testing asymmetry through Egger's test and adjusted the rates for the small study effects through Duval and Tweedie trim-and-fill procedure (R0 estimator).[24] We also conducted sensitivity analyses by excluding outliers, defined as studies whose 95% CI of the response rate does not overlap with the response rate of the pooled studies, by limiting the analyses to those studies with low risk of bias, and by limiting the analyses to those studies that reported response rates in the papers.

In the next step, we calculated the Relative Risk (RR) of response, indicating the pooled response rate in the therapy groups divided by the pooled response rate in the control groups. Then, we calculated the NNT using the pooled RR and the response rate in the control group, as recommended by the Cochrane Collaboration[25] because the RR is more stable than risk differences as would be calculated by subtracting pooled event rate in the control groups from the pooled event rate in the intervention groups.[26] Because we found no significant differences between the response rates in the psychotherapy conditions, we conducted subgroup analyses with all psychotherapies together. As predictors, we used the diagnosis at baseline (clinical interview vs. score above a cutoff on a self-report scale), target groups (adults in general vs. specific target group), age group (younger vs. older adults), treatment format (individual, group, other), and country (Western vs. non-Western; Western studies were conducted in North America, Europe, and Australia).

3 | RESULTS

3.1 | Selection and inclusion of studies

After examining 24,771 records (18,217 after removal of duplicates), we retrieved 2914 full-text papers for further consideration and excluded 2686 of these. The PRISMA flowchart, including the reasons for exclusion, is presented in Figure 1. Two hundred twenty-eight studies including 23,574, participants (12,308 in the treatment groups and 11,266 in the control groups) met inclusion criteria. Key characteristics of the included studies are presented in Appendix B and references in Appendix D.

3.2 | Characteristics of included studies

A total of 98 studies (43.0%) were aimed at adults in general, 23 at older adults (10.1%), 18 at women with perinatal depression (7.9%), 54 at patients with comorbid general medical disorders (23.7), and 35 at other specific target groups (15.3%). In 53 studies (23.2%), patients were recruited from clinical samples, 101 studies recruited participants through the community (44.3%), while the other 74 studies (32.5%) used other recruitment strategies. Patients were diagnosed with a mood disorder according to a diagnostic interview in 122 studies (53.5), while the other 106 used a cutoff on a self-report measure to include participants.

One hundred seventeen studies used a CAU control condition, 106 used a waitlist control condition, and eight included a pill placebo control condition (three studies had two control conditions). Seventy-two studies (31.6%) were conducted in the United States, 88 in Europe (38.6%), while the remaining 68 studies (29.8%) were conducted in other countries.

Eighteen studies had two psychotherapy conditions, resulting in a total of 246 therapy conditions, including CBT (148 studies; 60.2%), problem-solving therapy (24; 9.8%), interpersonal psychotherapy (21; 8.5), third wave therapies (21; 8.5), behavioral activation therapy (20; 8.1%), and life review therapy (12; 4.9%). An individual treatment format was used in 82 therapies (33.3%), a group format in 61 therapies (24.8), guided self-help (including internet-based interventions) in 49 therapies (19.9%), 11 used a telephone format (4.5%), and the other 25 therapies used a mixed format (10.2%).

In 25 (11.0%) of the 228 studies, the response rates were reported, and in the other 203 studies (89.0%), the rates were imputed. The response rates in each individual study for the main outcome (at 2 [$\pm$1] months) are reported in Appendix C.

Risk of bias was considerable. One hundred thirty-six of the 228 studies reported an adequate sequence generation (59.6%); 117 reported allocation to conditions by an independent party (51.3%); 62 reported using blinded outcome assessors (27.2%); 148 used only self-report outcomes (64.9%). In 143 studies, intent-to-treat analyses were conducted (62.7%). Seventy-five studies (32.9%) met all quality criteria (17 when self-report measures were rated as high risk of bias), 102 studies (44.7%) met two or three criteria, and 51 met no or only one criterion (22.4%).

3.3 | Response rates in psychotherapy

The response rate at 2 ($\pm$1) months was available for 193 psychotherapy conditions and resulted in an overall response rate

FIGURE 1 PRISMA flowchart
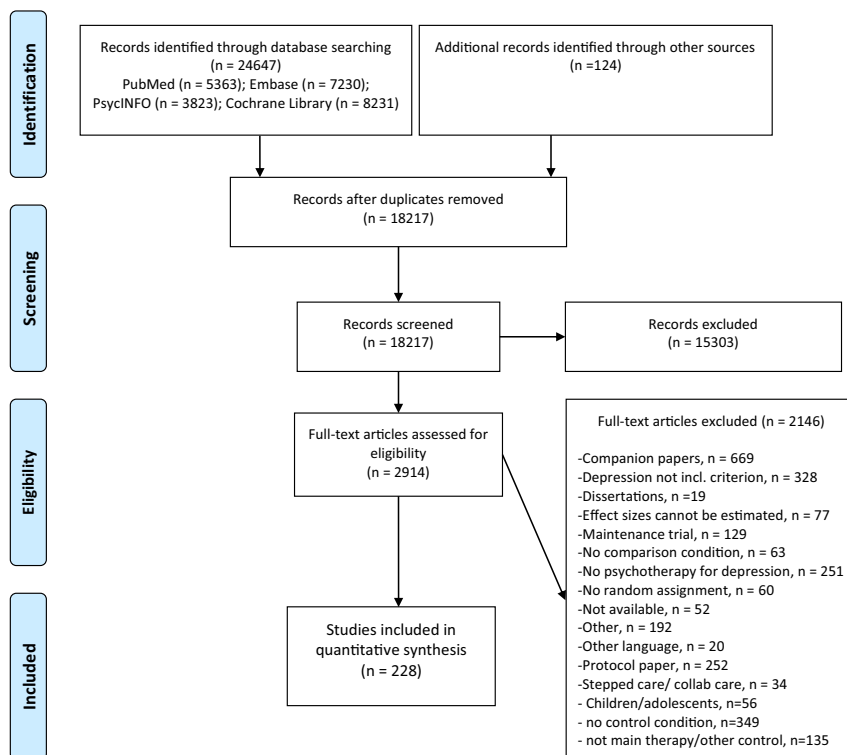
of 0.41 (95% CI: 0.38~0.43) (Table 1). Less conservative estimates resulted in somewhat higher rates (0.45 for completers samples, and 0.51 when all drop-outs were considered to be responders). Excluding outliers and selecting only studies with low risk of bias resulted in comparable response rates. No indication for significant asymmetry of the funnel plot was found ($p = 0.11$), and adjustment for publication bias resulted in a comparable response rate (0.38). Heterogeneity was high (>77%) in all analyses (except when outliers were excluded).

The results of the subgroup analyses are reported in Table 1. The response rates for each of the specific types of psychotherapy varied from 0.37 for life review therapy and interpersonal psychotherapy to 0.46 for behavioral activation therapy. CBT was examined in more than half of the studies and resulted in a response rate of 0.42. Again, heterogeneity was high for all response rates. The differences between the response rates for the different therapies were not significant ($p = 0.39$).

We also found significant differences between the response rates across treatment formats ($p = 0.01$), with the highest rates for individual treatment and the lowest for guided self-help and the category of mixed treatment formats. The rates also differed significantly between countries, with considerably lower response rates in Europe compared with other parts of the world ($p < 0.001$). No significant differences were found for recruitment method, diagnosis, age group, and specific target groups.

The response rates for psychotherapies at longer follow-up (Table 1) ranged from 0.37 at 3- to 6-month follow-up to 0.48 at 13- to 24-month follow-up.

3.4 | Response rates for the control conditions

The response rates for the control conditions are reported in Table 2. As can be seen, the overall response rate for CAU was 0.17, for waitlist 0.16 and for pill placebo 0.31 (p for difference between control conditions: <0.001). Sensitivity analyses resulted in broadly comparable outcomes, with higher rates for completers-only and when all drop-outs were considered responders, and small variations when outliers were excluded, results were adjusted for publication bias, when only studies with low risk of bias were included, and when only studies that directly reported response rates were examined. Heterogeneity was again high in all analyses (except when outliers were excluded). It should be noted that the number of studies using a pill placebo control group was small ($N = 7$). The l'Abbé plots, graphically presenting the rates in the therapy at the vertical axis and the control conditions at the horizontal axis,[27] for psychotherapy versus CAU and versus waitlist are presented in Figure 2A,B.

Response rates for CAU at longer follow-up (Table 2) ranged from 0.21 at 3- to 6-month follow-up to 0.29 at

TABLE 1 Response rates in psychotherapies for adult depression and control conditions at 2 (±1)-month follow-up

	N	Resp rate	95% CI	I²	95% CI	p
Main outcomes at 2 (±1)-month follow-up						
All psychotherapies	193	0.41	0.38~0.43	79	76~82	
Completers only	193	0.45	0.42~0.48	77	74~80	
All drop-outs responders	193	0.51	0.48~0.53	80	77~82	
Outliers excluded	133	0.41	0.39~0.43	28	10~42	
Only low risk of bias	61	0.38	0.34~0.43	85	82~88	
Adjusted for publication bias	206	0.38	0.36~0.41	82	79~84	
Only reported	19	0.50	0.38~0.61	89	84~92	
Subgroup analyses						
Type						
CBT	109	0.42	0.39~0.45	80	77~84	0.39
3rd wave therapies	21	0.38	0.31~0.46	74	60~83	
Problem-solving therapy	20	0.38	0.31~0.45	81	71~87	
Life review therapy	12	0.37	0.24~0.52	84	74~90	
Interpersonal psychotherapy	14	0.37	0.28~0.47	77	61~86	
Behavioral activation	17	0.46	0.37~0.55	72	54~83	
Format						
Individual	57	0.48	0.42~0.53	80	75~84	**0.01**
Group	56	0.41	0.36~0.46	83	78~86	
Guided self-help	49	0.37	0.34~0.41	66	54~75	
Other/mixed	31	0.37	0.30~0.44	82	75~87	
Recruitment						
Community	98	0.42	0.39~0.45	65	57~72	0.37
Only clinical samples	38	0.42	0.36~0.48	83	78~88	
Other recruitment	57	0.37	0.31~0.43	87	84~89	
Diagnosis						
Mood disorder	98	0.42	0.38~0.46	82	79~85	0.41
Cutoff	95	0.40	0.36~0.43	75	70~80	
Age group						
Adults	152	0.42	0.39~0.45	77	74~81	0.07
Older adults	41	0.36	0.30~0.42	84	79~88	
Target group						
Adults in general	114	0.42	0.39~0.45	76	71~80	0.15
Specific target group	79	0.38	0.34~0.43	83	79~86	
Country						
US	63	0.43	0.39~0.47	64	53~73	**<0.001**
Europe	68	0.35	0.31~0.38	79	73~83	
Other western	23	0.47	0.39~0.54	68	50~79	
Other	39	0.47	0.39~0.55	88	84~90	
Response at follow-up						
≥3 and ≤6 months	38	0.37	0.32~0.43	80	73~85	
6–8 months	49	0.38	0.34~0.43	83	78~87	
9–12 months	30	0.42	0.35~0.50	91	88~93	
13–24 months	5	0.48	0.29~0.68	86	68~93	

Bold values indicate significant differences between subgroups.

TABLE 2 Response rates in control conditions, relative risks of psychotherapies, and numbers-needed-to-be-treated and response at longer follow-up

	Response rate in control group					Psychotherapy versus control						
	N	Rate	95% CI	I^2	95% CI	N	RR	95% CI	I^2	95% CI	NNT	95% CI
Care-as-usual (vs. therapy)												
All studies	79	0.17	0.15~0.20	74	68~79	83	2.12	1.80~2.50	62	52~70	5.3	3.9~7.4
Completers only	79	0.19	0.16~0.22	74	68~79	83	2.17	1.85~2.55	61	51~70	4.5	3.4~6.2
All drop-outs responders	79	0.25	0.21~0.28	81	77~85	83	1.94	1.68~2.24	67	59~74	4.3	3.2~5.9
Outliers excluded	66	0.16	0.14~0.18	24	0~44	75	2.14	1.85~2.48	41	22~56	5.5	4.2~7.4
Only low risk of bias	28	0.21	0.16~0.26	81	74~87	29	1.79	1.36~2.34	59	38~73	6.0	3.6~13.2
Adj. for publication bias	108	0.24	0.21~0.28	79	74~82	113	1.43	1.16~1.78	73	67~77	9.7	5.3~26.0
Only reported	10	0.16	0.09~0.27	85	73~91	10	2.17	1.42~3.30	48	0~75	5.3	2.7~14.9
Waitlist (vs. therapy)												
All studies	92	0.16	0.14~0.18	56	44~65	102	2.61	2.29~2.98	40	24~53	3.9	3.2~4.8
Completers only	92	0.17	0.15~0.19	53	41~63	102	2.74	2.40~3.13	45	31~57	3.4	2.8~4.2
All drop-outs responders	92	0.21	0.19~0.24	64	55~71	102	2.42	2.12~2.77	50	37~60	3.4	2.7~4.3
Outliers excluded	83	0.16	0.15~0.18	0	0~25	96	2.65	2.35~2.98	0	0~24	3.8	3.2~4.6
Only low risk of bias	28	0.18	0.15~0.21	61	42~74	29	2.09	1.69~2.58	47	19~66	5.1	3.5~8.1
Adj. for publication bias	125	0.20	0.18~0.23	62	54~69	143	1.87	1.56~2.23	51	41~60	5.7	4.1~8.9
Only reported	6	0.22	0.06~0.53	89	79~94	7	3.13	1.19~8.23	75	47~88	2.1	0.6~23.9
Pill placebo (vs. therapy)[a,b,c]												
All studies	7	0.31	0.23~0.40	70	33~86	8	1.48	1.14~1.93	42	0~74	6.7	3.5~23.0
Completers only	7	0.33	0.26~0.42	61	11~83	8	1.65	1.19~2.28	66	28~84	4.7	2.4~15.9
All drop-outs responders	7	0.37	0.30~0.46	62	14~84	8	1.59	1.18~2.15	69	34~85	4.6	2.4~15.0
Outliers excluded	6	0.28	0.23~0.33	0	0~58	8	1.48	1.14~1.93	42	0~74	7.4	3.8~25.5
Only low risk of bias	3	0.34	0.21~0.52	88	65~96	3	1.12	0.84~1.49	0	0~75	24.5	n.s.
Adj. for publication bias	9	0.35	0.26~0.44	72	46~86	12	1.20	0.89~1.62	62	28~80	14.3	n.s.
Response at follow-up[d]												
≥3 and ≤6 months (WL)	14	0.13	0.09~0.19	49	5~72	14	3.24	2.32~4.53	24	0~59	3.4	2.2~5.8
>3 and <6 months (CAU)	23	0.21	0.17~0.26	60	37~75	23	1.61	1.32~1.96	44	8~66	7.8	5.0~14.9
6–8 months	47	0.26	0.23~0.30	71	60~78	49	1.47	1.29~1.68	44	22~60	8.2	5.7~13.3
9–12 months	28	0.28	0.22~0.35	88	84~91	30	1.38	1.19~1.61	52	27~68	9.4	5.9~18.8
13–24 months	5	0.29	0.18~0.42	69	20~88	5	1.59	0.98~2.60	17	0~83	5.8	n.s

[a] No outliers were detected for the RR in pill placebo (one for the response rate in the placebo group).

[b] Response rates for pill placebo were reported in only one study (with 2 comparisons); results not included in the table.

[c] p for difference between control conditions is <0.001.

[d] Response rates for waitlist were not available after 6-month follow-up. Only two studies reported outcomes at longer than 24-month follow-up.

13–24 months. The response rate for waitlist control groups was 0.13 at 3- to 6-month follow-up (no response rates were available at later follow-up periods).

3.5 | Relative risks and NNTs

The RRs of the therapy versus control conditions and the NNTs are also reported in Table 2. The RR for all psychotherapies compared with CAU was 2.12 (95% CI: 1.80~2.50), compared with waitlist 2.61 (95% CI: 2.29~2.98) and for pill placebo 1.48 (95% CI: 1.14~1.93). All sensitivity analyses indicated broadly comparable outcomes for all three types of control groups, although there was significant publication bias for CAU and waitlist ($p < 0.001$), and the RRs after adjustment for publication bias the RRs were considerably smaller. For pill placebo, the RR was no longer significant.

FIGURE 2   L'Abbé plot. (A)
Psychotherapy versus care-as-usual. (B)
Psychotherapy versus waitlist
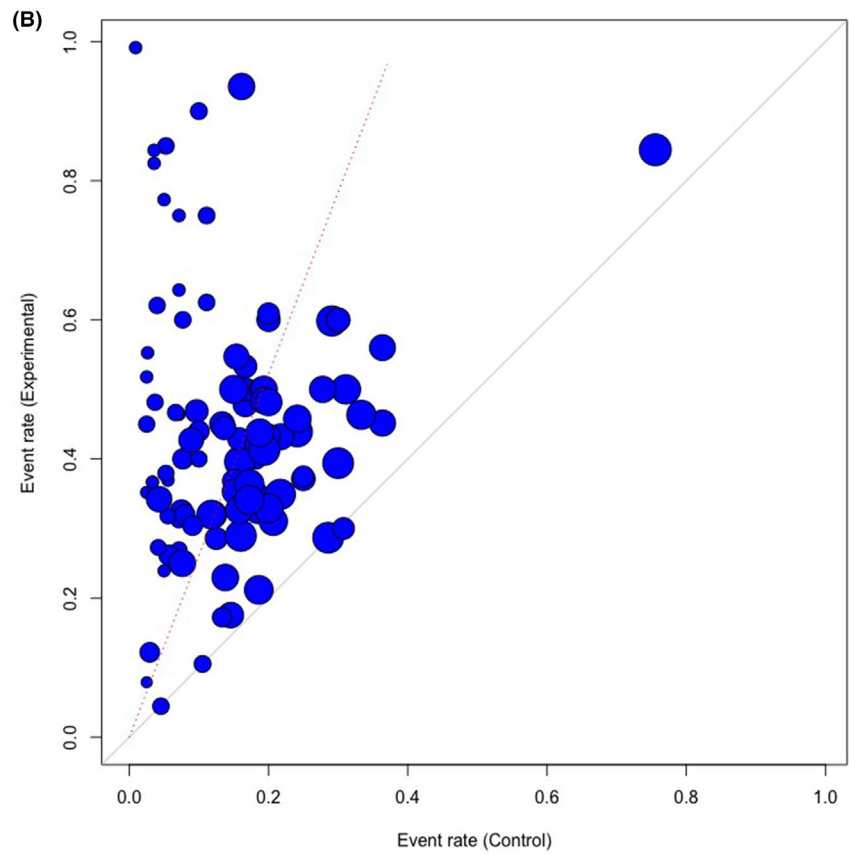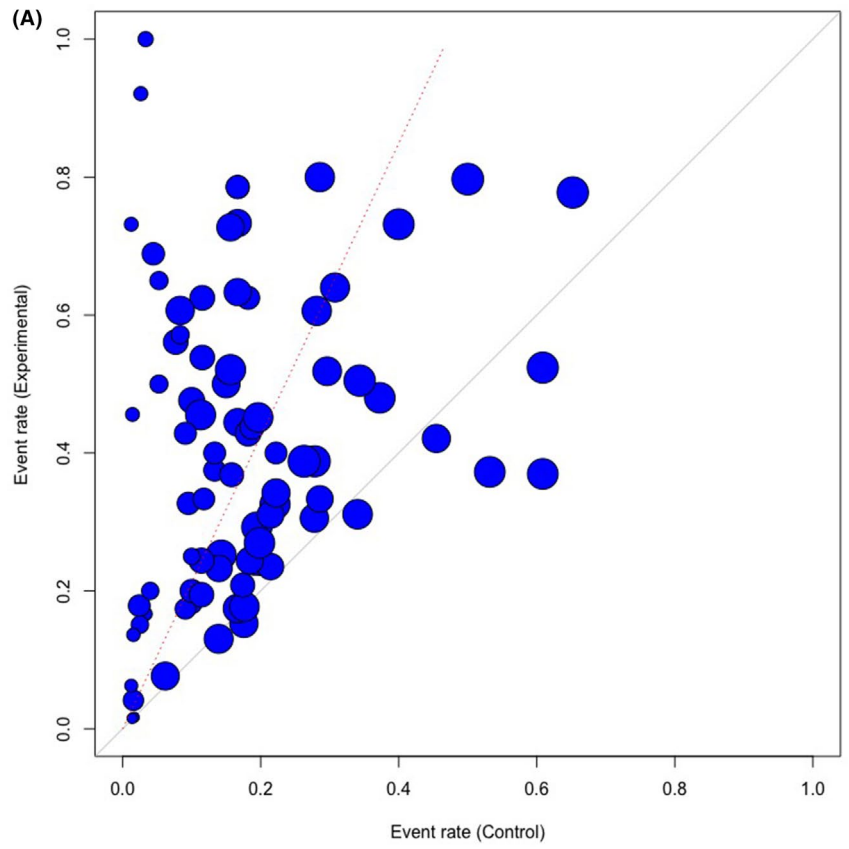
Numbers-needed-to-be-treated were calculated using the response rate in the control condition and the RR indicating the relative benefit of treatment compared with control conditions. The NNT for therapy versus CAU was 5.3 (95% CI: 3.9~7.4), for therapy versus waitlist 3.9 (95% CI: 3.2~4.8) and for therapy versus placebo 6.7 (95% CI: 3.5~23.0). Sensitivity analyses are reported in Table 2. Considerable differences with the main outcomes were found after adjustment for publication bias when the NNT for CAU was 9.7, for waitlist 5.7, and for pill placebo 14.3 (which was no longer significant). Limiting the analyses to studies with low risk of bias also resulted in somewhat higher NNTs (6.0 for CAU, 5.1 for waitlist, and 24.5 for pill placebo).

The RRs for CAU were relatively stable across the different follow-ups and the NNTs ranged from 5.8 to 9.4 at the different follow-up periods. The RR for waitlist was 3.24 and the NNT 3.4.

3.6 | Reliable improvement, reliable deterioration, and remission

The outcomes for reliable improvement, reliable deterioration, and remission at 2 (±1)-month follow-up are reported in Table 3. Reliable improvement rates ranged from 0.59 to 0.68 for psychotherapies, and from 0.30 to 0.57 for control conditions. However, reliable improvement in CAU and waitlist was both 0.30, while the rate for placebo was 0.57. Unfortunately, only four studies gave enough data to calculate the rates for pill placebo. The RRs ranged from 1.14 to 1.91 and the NNTs from 3.7 to 12.5. The rates for reliable deterioration were 0.2 to 0.5 for psychotherapies and 0.07 to 0.13 for control conditions, while the RRs ranged from 0.27 to 0.37 and the NNTs from 2.8 to 7.4. Remission rates ranged from 0.26 to 0.34 for psychotherapies, 0.09 to 0.17 for control conditions, with RRs between 2.11 to 4.39 and NNTs from 3.3 to 7.0.

4 | DISCUSSION

We examined rates for response, remission, reliable improvement and reliable deterioration for psychotherapies for adult depression. We found that the response rates were 0.41 for all psychotherapies together at 2 (±1)-month follow-up, when most therapies had ended. Sensitivity analyses broadly supported these findings, although the rates were somewhat higher depending on whether drop-outs were considered responders or not. Heterogeneity was high, but excluding outliers resulted in a comparable response rate and low heterogeneity.

We found no significant differences between the response rates in the included therapies, suggesting that these

are broadly comparable. This is in line with other meta-analyses indicating that psychotherapies have comparable effects.[1,17,18] However, we also found a significant difference for treatment formats. This *p*-value was not very high, and it may very well be a chance finding. This finding is also not in line with other meta-analyses suggesting that treatment format for therapies is not associated with differential outcomes. We did find a strong difference between response rates across different countries, with the rates in Europe being considerably smaller than in other countries. It is not clear why this would be the case. Because these are the rates within the therapy conditions, this cannot be attributed to for example differences in usual care delivered in these countries. Future research is needed to explore the causes of this difference.

Overall, the response rates are modest, with more than 50% of those receiving therapy not responding. In the control conditions, response was much lower (0.16 for waitlist and 0.17 for CAU), but the additional benefit of therapies above the control condition is still modest. Such modest additional effects of treatments were already shown for antidepressants[28] and a previous, but much smaller meta-analysis also pointed for such modest additional benefits for psychotherapies.[8] The current meta-analysis is much larger but clearly confirms these modest outcomes. The numbers for reliable improvement were somewhat better than those for the response rates, but still about one third of patients did not reach this level of improvement. Remission rates were only realized in about one third of the patients receiving therapies. It is clear that new, more effective strategies are needed to further reduce the disease burden of depression.

One of the strong points of this study was that we could estimate clinical significant deterioration with the same method across all included studies. In a previous meta-analysis, we found that only 6% of all trials comparing psychotherapy for depression with a control condition reported deterioration rates.[13] In that meta-analysis, we also found that the studies reporting deterioration rates used different ways to define deterioration, which made pooling the prevalence rates across treatments and control groups impossible. In the current study, we were able to pool deterioration rates in a large sample of studies using the same method. The results indicate that deterioration rates are below 5% in psychotherapy conditions and between 7% and 13% in control conditions. These numbers are low, but still large enough to warrant further research.

This study does not allow comparisons with the effects of antidepressant medication, because it only included studies in psychotherapy for depression. However, previous meta-analyses of trials directly comparing the two have shown that psychotherapy and pharmacotherapy broadly have comparable effects,[2] although this is complicated by issues like blinding,[29] sponsorship bias,[30] and general differences in the designs, such as the use of self-report measures for inclusion in psychotherapy trials and the impossibility of blinding

TABLE 3 Reliable improvement, reliable deterioration, remission, relative risks (RRs), and Numbers-needed-to-be-treated (NNT) of psychotherapies versus control groups

	Psychotherapy					Control groups					Psychotherapy versus control					
	N	Rate	95% CI	I^2	95% CI	N	Rate	95% CI	I^2	95% CI	RR	95% CI	I^2	95% CI	NNT	95% CI
Reliable improvement at 2 (±1) months																
RCI improvement (CAU)	81	0.59	0.54~0.64	87	85~89	76	0.30	0.26~0.34	82	78~85	1.85	1.62~2.10	83	79~86	3.9	3.0~5.4
RCI improvement (WL)	101	0.60	0.56~0.63	78	73~81	90	0.30	0.27~0.32	54	41~64	1.91	1.75~2.09	72	65~77	3.7	3.1~4.4
RCI improvement (PLA)	5	0.68	0.53~0.80	78	47~91	4	0.57	0.41~0.71	77	37~92	1.14	0.94~1.38	0	0~77	12.5	n.s.
Reliable deterioration at 2 months																
RCI deterioration (CAU)	81	0.05	0.04~0.06	13	0~35	76	0.12	0.10~0.14	55	41~65	0.37	0.29~0.48	0	0~9	4.9	3.4~7.7
RCI deterioration (WL)	101	0.05	0.04~0.05	0	0~7	90	0.13	0.12~0.15	34	14~49	0.27	0.22~0.33	0	0~0	2.8	2.2~3.8
RCI deterioration (PLA)	5	0.02	0.01~0.06	0	0~0	5	0.07	0.04~0.11	0	0~0	0.34	0.13~0.88	0	0~36	7.4	2.1~104.8
Remission (HAM-D <7) at 2 months																
Remission (CAU)	22	0.26	0.20~0.33	74	60~83	21	0.12	0.09~0.18	57	30~74	2.19	1.40~3.42	37	0~62	7.0	3.4~20.8
Remission (WL)	27	0.34	0.27~0.41	67	50~78	22	0.09	0.06~0.12	0	0~23	4.39	3.29~5.86	0	0~0	3.3	2.3~4.9
Remission (PLA)	4	0.34	0.27~0.42	0	0~84	3	0.17	0.11~025	0	0~85	2.11	1.90~2.35	0	0~0	5.3	4.4~6.5

participants and clinicians. Overall, however, it can be hypothesized that the effects of both psychotherapy and pharmacotherapy do not differ substantially from each other at the short term. Pharmacotherapy therefore shows probably comparable response rates as we found in this study, although this assumption would need additional research.

This study has several important limitations that should be take into account when interpreting the effects. One important limitation is that heterogeneity was very high, especially when estimating the response rates (less so for the RRs). However, the estimated rates appeared to be relatively robust in a series of sensitivity analyses. Second, the number of studies directly reporting response rates was low, and in most cases, the rates were estimated. Although the method to estimate these rates has been found to be reliable, it is still an estimate and this may have affected the pooled rates. Third, risk of bias was considerable in the large majority of trials. Although the subsample of studies with low risk of bias resulted in comparable outcomes, this still means that the results should be interpreted with caution. Another limitation of this study was that not all domains of risk of bias were examined. Blinding of participants and clinicians is not possible in psychological interventions, and we also did not rate selective outcome reporting because only a small number of trials in this field have been registered prospectively.[31] This may have resulted in an underestimation of the risk of bias in the included studies. Furthermore, a considerable number of studies included participants scoring high on a self-report scale instead of meeting criteria for a depressive disorder. Although we did not find a significant difference between the response rates in these groups of studies, this may have contributed to the heterogeneity of the studies. The outcomes at follow-up should also be considered with caution, because we examined outcomes at fixed periods of time, while the therapies varied in length.

Despite the limitations, this study showed that psychotherapies for depression may be effective compared with control conditions, but that still more than half of patients receiving therapy did not respond and only one third remitted. Furthermore, 16%–17% also respond when they are in a waitlist or CAU control conditions. More effective treatments and treatments for those not responding to a first treatment are clearly needed.

CONFLICT OF INTEREST

AUTHOR CONTRIBUTION

PC and TAF conceptualized the main idea for this paper. PC and EK included the studies. PC, EK, CM, and MC extracted the data. The analyses were conducted by PC with the support of HN and TAF. PC wrote the first draft of this paper. All authors read and revised the paper critically for important intellectual content and approved the final version of the paper.

PEER REVIEW

DATA AVAILABILITY STATEMENT

ORCID

Pim Cuijpers https://orcid.org/0000-0001-5497-2743
Eirini Karyotaki https://orcid.org/0000-0002-0071-2599
Marketa Ciharova https://orcid.org/0000-0002-7131-1549
Clara Miguel https://orcid.org/0000-0001-5563-5896
Hisashi Noma https://orcid.org/0000-0002-2520-9949
Toshi A. Furukawa https://orcid.org/0000-0003-2159-3776

REFERENCES

1. Cuijpers P, Quero S, Noma H, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry. 2021;20:283-293.

2. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry. 2020;19:92-107.

3. Karyotaki E, Smit Y, de Beurs DP, et al. The long-term efficacy of acute-phase psychotherapy for depression: a meta-analysis of randomized trials. Depress Anxiety. 2016;33:370-383.

4. Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. Psychother Res. 2020;30:279-293.

5. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJH. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cogn Behav Ther. 2018;47:91-106.

6. Cuijpers P, Karyotaki E, Reijnders M, Ebert DD. Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. Epidemiol Psychiatr Sci. 2019;28:21-30.

7. Cummings P. Arguments for and against standardized mean differences (effect sizes). Arch Pediatr Adolesc Med. 2011;165: 592-596.

8. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. J Affect Disord. 2014;159:118-126.

9. Hieronymus F, Jauhar S, Østergaard SD. One (effect) size does not fit at all: interpreting clinical significance and effect sizes in depression treatment trials. J Psychopharmacol. 2020;34:1074-1078.

10. Charlson F, van Ommeren M, Flaxman A, Cornett J, Whiteford H, Saxena S. New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis. Lancet. 2019;394:240-248.

11. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA. 2016;316:2214-2236.

12. Furukawa TA, Cipriani A, Barbui C, et al. Imputing response rates from means and standard deviations in metaanalyses. Psychopharmacol. 2005;20:49-52.

13. Cuijpers P, Reijnders M, Karyotaki E, de Wit L, Ebert DD. Negative effects of psychotherapies for adult depression: a meta-analysis of deterioration rates. J Affect Disord. 2018;239:138-145.

14. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318:1728-1733.

15. Cuijpers P, Karyotaki E. The effects of psychological treatments of depression on response, remission, reliable change, and deterioration: a meta-analysis. Open Science Framework. 2021. https://doi.org/10.17605/OSF.IO/SG274

16. Cuijpers P, Karyotaki E. A meta-analytic database of randomised trials on psychotherapies for depression. 2020. https://doi.org/10.17605/OSF.IO/825C6

17. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol. 2008;76:909-922.

18. Barth J, Munder T, Gerger H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med. 2013;10:e1001454.

19. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

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

21. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry. 1991;48:851-855.

22. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12-19.

23. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-560.

24. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometr. 2000;56:455-463.

25. Higgins JPT, Green S (Eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. London: The Cochrane Collaboration; 2011. www.handbook.cochrane.org

26. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. Int J Epidemiol. 2002;31:72-76.

27. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med. 1987;107:224-233.

28. Levkovitz Y, Tedeschini E, Papakostas GI. Efficacy of antidepressants for dysthymia: a meta-analysis of placebo-controlled trials. J Clin Psychiatry. 2011;72:509-514.

29. Cuijpers P, Karyotaki E, Andersson G, Li J, Mergl R, Hegerl U. The effects of blinding on the outcomes of psychotherapy and pharmacotherapy for adult depression: a meta-analysis. Eur Psychiatr. 2015;30:685-693.

30. Cristea IA, Gentili C, Pietrini P, Cuijpers P. Sponsorship bias in the comparative efficacy of psychotherapy and pharmacotherapy for adult depression: a meta-analysis. Br J Psychiatry. 2017;210:16-23.

31. Miguel C, Karyotaki E, Cuijpers P, Cristea IA. Selective outcome reporting and the effectiveness of psychotherapies for depression. World Psychiatry. 2021.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.