

Special Article

Cite this article: Munder T, Flückiger C, Leichsenring F, Abbass AA, Hilsenroth MJ, Luyten P, Rabung S, Steinert C, Wampold BE (2019). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences* **28**, 268–274. <https://doi.org/10.1017/S2045796018000355>

Received: 26 April 2018

Revised: 30 May 2018

Accepted: 10 June 2018

First published online: 30 July 2018

Key words:

Depression; Eysenck; meta-analysis; natural history; psychotherapy effectiveness

Author for correspondence:

Bruce E. Wampold, E-mail: wampold@education.wisc.edu

Is psychotherapy effective? A re-analysis of treatments for depression

T. Munder¹, C. Flückiger², F. Leichsenring³, A. A. Abbass⁴, M. J. Hilsenroth⁵, P. Luyten^{6,7}, S. Rabung⁸, C. Steinert^{3,9} and B. E. Wampold^{10,11}

¹Psychologische Hochschule Berlin, Berlin, Germany; ²Department of Psychology, University of Zürich, Zürich, Switzerland; ³Department of Psychosomatics and Psychotherapy, Justus-Liebig-University Giessen, Giessen, Germany; ⁴Department of Psychiatry, Centre for Emotions and Health, Dalhousie University, Halifax, NS, Canada; ⁵The Derner Institute of Advanced Psychological Studies, Adelphi University, Hy Weinberg Center, Garden City, NY, USA; ⁶Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium; ⁷Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; ⁸Department of Psychology, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria; ⁹Department of Psychology, Medical School Berlin, Berlin, Germany; ¹⁰Modum Bad Psychiatric Center, Modum Bad, Vikersund, Norway and ¹¹University of Wisconsin–Madison, Madison, Wisconsin, USA

Abstract

Aims. The aim of this study was to reanalyse the data from Cuijpers *et al.*'s (2018) meta-analysis, to examine Eysenck's claim that psychotherapy is not effective. Cuijpers *et al.*, after correcting for bias, concluded that the effect of psychotherapy for depression was small (standardised mean difference, SMD, between 0.20 and 0.30), providing evidence that psychotherapy is not as effective as generally accepted.

Methods. The data for this study were the effect sizes included in Cuijpers *et al.* (2018). We removed outliers from the data set of effects, corrected for publication bias and segregated psychotherapy from other interventions. In our study, we considered wait-list (WL) controls as the most appropriate estimate of the natural history of depression without intervention.

Results. The SMD for all interventions and for psychotherapy compared to WL controls was approximately 0.70, a value consistent with past estimates of the effectiveness of psychotherapy. Psychotherapy was also more effective than care-as-usual (SMD = 0.31) and other control groups (SMD = 0.43).

Conclusions. The re-analysis reveals that psychotherapy for adult patients diagnosed with depression is effective.

Those who cannot remember the past are condemned to repeat it.

–George Santayana

Literature and philosophy both allow past idols to be resurrected with a frequency which would be truly distressing to a sober scientist.

–Morris Raphael Cohen

In the 1950s and 1960s, Eysenck made some claims about the effectiveness of psychotherapy (Eysenck, 1952, 1961, 1966). Our collective memories of the specific claims made by Eysenck have diminished over time and we seem to be left with the simple conclusion that Eysenck claimed that psychotherapy was ineffective (Wampold, 2013; Wampold and Imel, 2015). Recently, Cuijpers *et al.* (2018) summarised Eysenck's claims by noting, 'He [Eysenck] suggested that psychotherapies are not effective in the treatment of mental disorders (Eysenck, 1952)' (p. 1). It is important to know whether psychotherapy is effective or not. However, to make any statement about Eysenck and his claims, one has to understand exactly what he claimed and the bases on which he made his claims. We begin by reviewing what Eysenck had to say about the effects of psychotherapy.

Based on a review of research available at that time, Eysenck indeed did conclude that psychotherapy was not effective:

A survey was made of reports on the improvement of neurotic patients after psychotherapy, and the results compared with the best available estimates of recovery *without benefit of such therapy*. The figures fail to support the hypothesis that psychotherapy facilitates recovery from neurotic disorder (emphasis added; Eysenck, 1952, p. 323)

When *untreated* neurotic control groups are compared with the experimental groups of neurotic patients treated by means of psychotherapy, both groups recover to approximately the same extent (emphasis added, Eysenck, 1961, p. 719).

To be clear about Eysenck's claims about the ineffectiveness of psychotherapy, he compared the effects of psychotherapy with those patients who did not receive any treatment.

It is important to note that Eysenck was not simply impugning the absolute effectiveness of psychotherapy, he was at the same time concluding that one form of psychotherapy was effective and that other therapies were unscientific and ineffective (*viz.*, behaviour therapy; Eysenck, 1961; see Wampold, 2013; Wampold and Imel, 2015).

Given the distinction among various psychotherapies, any examination of Eysenck's claims must consider what is and what is not psychotherapy. Eysenck was very careful to define psychotherapy:

- (1) There is an interpersonal relationship of a prolonged kind between two or more people.
- (2) One of the participants has had special experience and/or has received special training in the handling of human relationships.
- (3) One or more of the participants have entered the relationship because of a felt dissatisfaction with their emotional and/or interpersonal adjustment.
- (4) The methods used are of a psychological nature, i.e. involve such mechanisms as explanation, suggestion, persuasion and so forth.
- (5) The procedure of the therapist is based upon some formal theory regarding mental disorder in general, and the specific disorder of the patient in particular.
- (6) The aim of the process is the amelioration of the difficulties which cause the patient to seek the help of the therapist (Eysenck, 1961, p. 698).

Eysenck's definition of psychotherapy is in accord with most definitions of psychotherapy, which emphasize that an interpersonal relationship is at the heart of the endeavor (e.g. Wampold and Imel, 2015).

Eysenck's claims created controversy as well as angst, among mental health professionals as well as the public. There were articles rebutting Eysenck's conclusions and rejoinders, creating a contentious interchange (for a summary see Glass and Kliegl, 1983; Wampold, 2013; Glass, 2015; Wampold and Imel, 2015). The debate about Eysenck's claims led to a proliferation of randomised clinical trials examining both the absolute efficacy of psychotherapy (i.e. the effects of psychotherapy *v.* natural history) and the relative efficacy of various treatments (i.e. the relative effects of different therapies; Wampold, 2013; Wampold and Imel, 2015). In the late 1970s, Mary Lee Smith and Gene Glass (Smith and Glass, 1977; Smith *et al.*, 1980) conducted a comprehensive meta-analysis of controlled studies of psychotherapy and found that psychotherapy was indeed effective, with a standardised mean difference (SMD) between treated and untreated patients of approximately 0.70, a relatively large effect. Of course, Eysenck disputed these results by suggesting that meta-analyses ignored problems with the primary studies, such as the heterogeneity of included studies ('apples and oranges' problem) (Eysenck, 1978, 1984, 1995). However, several re-analyses of Smith and Glass and additional meta-analyses have established an SMD of approximately 0.70, although this varies somewhat depending on the problem being treated (see Wampold and Imel, 2015).

Recently, Cuijpers *et al.* (2018) have addressed several of the problems mentioned by Eysenck (and others) by examining the effects of interventions for a particular disorder, namely depression, considering various factors that might bias the estimates.

In their article, a reassessment of the effects of psychotherapy for adult depression, they claimed, much in the way that Eysenck did, that there is insufficient evidence to declare that psychotherapy is effective:

These results suggest that the effects of psychotherapy for depression are small, above the threshold that has been suggested as the minimal important difference in the treatment of depression, and Eysenck was probably wrong. However, this is still not certain because we could not adjust for all types of bias (p. 1)... [and] the possibility that psychotherapies do not have effects that are larger than spontaneous recovery cannot be excluded (p. 7).

In this article, we address Cuijpers *et al.*'s (2018) claims and show that a different understanding of Eysenck's conjectures produces an estimate of effectiveness for psychotherapy for depression that closely approximates what has been found previously.

A re-analysis of Cuijpers *et al.* (2018)

The goal of Cuijpers *et al.* (2018) was to revisit Eysenck's conclusion that psychotherapy was not effective by meta-analytically examining the corpus of studies comparing an intervention for adults with depression to a control group and correcting obtained effects for bias of various types. Of course, as meta-analytic methods improve, it is commendable to scrutinise prior conclusions in light of the best available methods.

Cuijpers *et al.* (2018) examined 369 effects produced by studies that compared interventions for depression with a control group. The overall effect for these interventions was an SMD of 0.70 suggesting that Eysenck's conclusions were in fact incorrect Q.E.D. But Cuijpers *et al.* claimed that this estimate was biased and when the effects were corrected for these biases the 'true' effect is between 0.2 and 0.3, casting some doubts on whether Eysenck's conclusions were truly incorrect. However, Cuijpers *et al.*'s conclusions depend on several methodological decisions that need to be re-examined. In this article, we examine several of their decisions and then reanalyse their data with decisions that we contend are more in line with Eysenck's conjectures.

Choice of control group

The corpus of studies in Cuijpers *et al.* (2018) included three types of control groups: waiting list (WL, $k = 159$), care-as-usual (CAU, $k = 144$), and 'other control' ($k = 66$).¹ Unfortunately no definition of these types of control groups was presented and no methods for making this determination were provided (e.g. coding procedures, interrater agreement, etc.). What is most important is that Cuijpers *et al.* considered WL controls as *biased* and excluded studies using WL when estimating the *true* effect of psychotherapy. This is a decision that results in a significant decrease in the estimates of psychotherapy effectiveness, and one which is questionable. We examine each of these types of control groups, noting the questions that each is able to address.

Waiting-list controls

WL controls contain patients who are told that during the treatment phase they will receive no treatment (NT) as part of the

¹There was a discrepancy between Table 1 of Cuijpers *et al.*'s (2018) manuscript and Appendix C in Supplemental materials for WL ($k = 159$ *v.* 150, respectively) and CAU ($k = 144$ *v.* 153, respectively), which has been resolved by P. Cuijpers (personal communication, April, 20th, 2018).

study but that after the treatment period, if they choose to, they will receive one of the experimental treatments. To be clear, no treatment is provided to the patients and this type of control group is thought to be a means to estimate the natural history of the disorder (Wampold *et al.*, 2005; Stegenga *et al.*, 2012). If we consider that Eysenck was focusing on the effects of psychotherapy compared with recovery without psychotherapy, it would seem that WL is an appropriate control group as it compares the outcome of psychotherapy with an estimate of natural course of the disorder.

There may well be methodological problems with WL controls. WL patients may actually improve during the study period because they became demoralised by anticipation of being included in a state-of-the-art treatment, which might not be obtainable elsewhere, in 15 or so weeks (Frank and Frank, 1991). There is evidence that patients improve from when they make an appointment to receive services and when they present for such services (Frank and Frank, 1991). Indeed, WL patients in clinical trials for depression improve quite dramatically during the waiting period; the effect for patients on WL in randomised controlled trials for depression from beginning of the waiting period to the end of the waiting period is approximately 0.40 (Minami *et al.*, 2007; see also Posternack and Miller, 2001). WL patients may improve as a function of being included in the trial and therefore the use of WL controls may underestimate the effects of psychotherapy.

On the other hand, patients might feel demoralised by not being selected to receive treatment immediately: ‘Nothing good ever happens to me. I can’t even get selected to receive treatment now.’ This is the resentful demoralisation threat to validity (Shadish *et al.*, 2002). However, there is little evidence that patients on the WL in clinical trials suffer from resentful demoralisation. Of course, many patients in routine clinical care are placed on waiting lists until services are available. Ahola *et al.* (2017) studied patients on waitlists and concluded, ‘scheduled waiting should be regarded as a preparatory treatment and not as an inert non-treatment control’ (p. 611).

Cuijpers *et al.* (2018) chose to question WL as a control group and exclude studies that used WL controls to estimate the ‘true’ effects of psychotherapy:

Waiting list control groups may stimulate patients to *do nothing* about their problems because they will get a treatment after the waiting period. Recent meta-analyses suggest that waiting lists may be a nocebo, and artificially inflate the effect sizes of therapies (Furukawa *et al.*, 2014) (emphasis added, p. 2).²

To be clear, Cuijpers *et al.* (2018) claim is that WL is inappropriate because it might induce patients not to seek help. That is, patients on WL are purported to avoid seeking therapy or any other type of external help. Consequently, according to this view, WL patients represent the population of depressed patients not receiving treatment, which is exactly the control that should be used to determine whether a treatment is superior to spontaneous recovery without treatment. The Eysenckian conjecture that psychotherapy is not more effective than NT would suggest that WL control is a suitable, if not the suitable, hypothesis-driven control group.

²A nocebo is treatment without active ingredients (e.g. inert pill, sham procedures) that results in increased symptoms due to expectations created that the nocebo will be harmful, usually through instructions (Miller *et al.*, 2009; Benedetti, 2014). Clearly, patients on WL are not induced to expect deterioration under this condition, so even if patients deteriorate as a result of being on the WL, WL is not a nocebo. There is a difference between something that is harmful and a nocebo.

What is the best way to empirically determine whether WL is biased? Logically one could compare WL control patients with NT controls. But how would that work? First, NT controls are unethical as one cannot deny patients with mental disorder treatment and that is the reason WL controls are used in lieu of NT controls. Second, NT patients would most likely experience effects of being included in a trial but be denied any treatment at all. They might be discouraged and deteriorate as a result or they might seek alternate treatment and improve – who knows?

Cuijpers *et al.*'s (2018) attribution of bias for WL controls rests on purported evidence that WL cause patients to deteriorate relative to NT patients. But how is that known given that it is unethical to deny treatment to patients with mental health disorders who are seeking treatment? The meta-analysis cited by Cuijpers *et al.* that claimed WL artefactually causes deterioration (*viz.*, Furukawa *et al.*, 2014) is a network meta-analysis that involved clinical trials of cognitive behaviour therapy (CBT) against various controls in the treatment of depression. A closer look showed that this meta-analysis contained 13 comparisons (from only six separate trials) with NT controls. However, none of the studies using an NT control involved patients seeking treatment for depression. The patients in studies with NT controls were college students selected for study (not seeking help for depression) or community members identified through screening. Most, although not all, were mildly to moderately depressed and all were not seeking treatment for depression. It is well established that seeking relief for distress is a vital factor for response to placebo (Price *et al.*, 2008). Interventions in NT trials were more similar to prevention programs than treatment programs and many NT studies did not involve psychotherapy according to Eysenck's definition. Prevention programs and programs for those not seeking treatment typically are ineffective (Lilienfeld, 2007; Wampold and Imel, 2015). Consequently, it is understandable that the effects of CBT *v.* NT would be rather small. In the NT study that contributed more than half of all participants in NT comparisons in the Furukawa *et al.* meta-analysis (*viz.*, Dowrick *et al.*, 2000), the difference of intervention *v.* NT was only SMD = 0.169, suggesting that the treatments employed were only marginally helpful for participants. Of course, in the framework of Furukawa *et al.*'s network meta-analysis, these small treatment effects contribute to the impression of more change for participants in NT than in WL.

On the other hand, the CBT *v.* WL in the Furukawa *et al.* (2014) meta-analysis included studies of patients seeking treatment for depression and it is not surprising that there were larger effects in these studies. The conclusion that WL is a ‘nocebo’ is due to the fact that the CBT *v.* NT prevention studies showed smaller effects than the CBT *v.* WL treatment studies, despite that the studies in these two comparisons were markedly different. Making inferences about the relative effectiveness of treatments or control groups (here WL *v.* NT) from network meta-analyses in lieu of examining direct comparisons often leads to erroneous conclusions (see e.g. Del Re *et al.*, 2013; Jansen and Naci, 2013; Wampold and Serlin, 2014; Wartolowska *et al.*, 2014; Wampold *et al.*, 2017), especially when, as in Furukawa *et al.*, the consistency of indirect estimates with direct estimates cannot be assured (none of the trials directly compared NT and WL controls). Thus, the results of this meta-analysis do not provide persuasive evidence that WL is an inappropriate control. Moreover, Furukawa *et al.* reported that the difference between NT and WL was not significantly different when publication bias was considered.

Given the problematic nature of the Furukawa *et al.* (2014) meta-analysis and other evidence, we contend that WL is indeed an appropriate control group to address Eysenck's conjecture that psychotherapy is not more effective than NT.

Care as usual

To estimate the 'true' effects of psychotherapy, Cuijpers *et al.* (2018) included CAU as appropriate controls. CAU is an appropriate control group if one is estimating whether psychotherapy is more effective than the various mental health treatments being given in routine care. However, CAU typically contains a wide array of treatments (Spielmanns *et al.*, 2010; Wampold *et al.*, 2011), which was noted by Cuijpers *et al.*: '[CAU] is problematic since (sic) this varies considerably across settings and health care systems, making comparisons very heterogeneous' (p. 3). Eysenck was making a claim that psychotherapy was not more effective than NT, not that it was not more effective than the usual care patients were receiving, which might well be psychotherapy or other mental health services. Indeed in Cuijpers *et al.* CAU included credible treatments such as supportive psychotherapy or pharmacotherapy delivered by experienced therapists (Saloheimo *et al.*, 2016); combination of psychotherapy and pharmacotherapy according to Dutch Depression Guidelines (Wiersma *et al.*, 2015), or antidepressant medication (Power and Freeman, 2012). CAU is often nearly as effective as first-line psychotherapeutic treatments for several disorders, including borderline personality disorder, anxiety and depression (Wampold *et al.*, 2011; Cristea *et al.*, 2017). Comparisons of these relatively active treatments produce effects irrelevant to Eysenck's claims about the effectiveness of psychotherapy *vis-à-vis* NT.

Other control groups

Cuijpers *et al.* (2018) included 'other control groups' when estimating the 'true' effect of psychotherapy. They did not define what 'other controls' were but we examined these studies and found that 'other controls' included pill placebos (e.g. Elkin *et al.*, 1989; Dimidjian *et al.*, 2006; Hegerl *et al.*, 2010), or 'so called' psychological placebos (e.g. Watt and Cappeliez, 2000; Spinelli and Endicott, 2003; Armento, 2012; Losada *et al.*, 2015). We know that pill placebo with clinical management is often quite effective, often as effective or nearly as effective as antidepressant medication, particularly for depression (Kirsch, 2002, 2009, 2010; Kirsch *et al.*, 2008). Furthermore, psychological placebos are often quite effective (Baskin *et al.*, 2003; Smits and Hofmann, 2009; Honyashiki *et al.*, 2014). In any event, the use of pill placebo and psychological placebos addresses questions about the relative efficacy of psychotherapy compared to some relatively active controls, but does not address Eysenck's claims about the effectiveness of psychotherapy in comparison with NT.

Definition of psychotherapy

Cuijpers *et al.* (2018) made conclusions about psychotherapy, as evidenced by the subtitle of their article: 'A reassessment of the effects of psychotherapy for adult depression.' If one is to assess the effects of psychotherapy *vis-à-vis* the claims of Eysenck, then it is incumbent to include only studies of psychotherapy. However, Cuijpers *et al.* did not provide any description of inclusion and exclusion criteria for psychotherapy and many of the studies in Cuijpers *et al.* were clearly not psychotherapy. For example, in one study depressed patients were given a copy of a self-help book based on cognitive therapy and were 'asked to

read the book and to complete all the homework exercises in the book within 1 month' (Floyd *et al.*, 2004, p. 305). In a similar study (van Bastelaar *et al.*, 2011), patients with diabetes and elevated depression symptoms were given access to a website with 'eight lessons' (p. 51) on depression and diabetes. Lamers *et al.* (2010) investigated a 'minimal psychological intervention' for elderly depressed patients delivered by 'four nurses with no specific mental health expertise' (p. 219). In Cuijpers *et al.* (2018) we counted at least 61 effects derived from interventions that did not meet common definitions of psychotherapy, including Eysenck's (1961) definition, recreating the 'apples and oranges' problem about which Eysenck was concerned. At the very least, conclusions about psychotherapy are unjustified when interventions that are not psychotherapy are lumped with psychotherapeutic treatments.

Western v. non-Western studies

Cuijpers *et al.* (2018) excluded non-Western studies (*viz.*, those from Africa, Asia and Latin America) based on the finding that the effects of psychotherapy were greater in non-Western countries. Cuijpers *et al.* (2018) did not define 'Western' in a transparent manner (Latin America is in the Western hemisphere and Chile and Argentina are typically classified as 'Western') and provided no hypothesis-driven or theoretical reason to exclude evidence from some countries. Excluding non-Western evidence created smaller effects. Our re-analyses suggested the Western/non-Western effect is at least partially due to outliers, which we omitted in our re-analysis (see below).

Risk of bias

Cuijpers *et al.* (2018) further reduced the number of studies by excluding studies with 'possible systematic errors ... or deviations from the true or actual outcomes' (p. 3). It seems to us, however, that this reduction was conducted in a manner that discards relevant research studies. Specifically, 'four items of the Cochrane risk of bias assessment tool' (p. 3) were used to define risk of bias and studies were excluded if any one of these four criteria were coded as negative or unclear.

Using only four of the six domains of the Cochrane risk of bias (RoB) tool, Cuijpers *et al.*'s (2018) definition did not cover a number of methodological aspects that are especially relevant for psychotherapy, possibly leading to the inclusion of studies with important deficits and to the exclusion of studies with appropriate methodology. In short, Cuijpers *et al.* excluded the RoB domains 'blinding of participants and personnel' and 'selective outcome reporting.' Coding the first domain was considered 'not possible' (p. 4) in the included studies and coding the latter was feared to result in 'very few trials ... with low risk of bias' (p. 4) – that is to say, all psychotherapy studies have significant risk of bias. Clearly, patients and therapists are always cognisant of the psychotherapy they are receiving (or not receiving) and therefore blinding is not possible. However, there is broad consensus in the Cochrane and the psychotherapy research communities that exactly because of this deviation from the ideal experiment it is important to pay attention to methodological dimensions capturing quality of care and expectations (e.g. treatment credibility, therapist allegiance and treatment integrity), which were ignored in Cuijpers *et al.*'s (2018) study (see Baskin *et al.*, 2003; Higgins and Green, 2011; Laird *et al.*, 2017; Munder and Barth, 2018).

One of the studies excluded by Cuijpers *et al.*'s (2018) definition of risk of bias is the NIMH Treatment of Depression Collaborative Research Program (Elkin *et al.*, 1989), even though this was considered the most sophisticated and methodologically rigorous clinical trial of psychotherapy ever conducted. In contrast, Cuijpers *et al.* included other studies that have important methodological shortcomings, including those with few therapists (Milgrom *et al.*, 2005, with two therapists, and Burns *et al.*, 2007 with one therapist) and several that did not monitor or assess adherence (e.g. Burns *et al.*, 2007). Allegiance, an important aspect in psychotherapy studies (Munder *et al.*, 2011; Munder *et al.*, 2012; Munder *et al.*, 2013) was ignored even though many of the studies in this data set were conducted by advocates of one of the treatments.

There are other problems with the Cuijpers *et al.* RoB determination. There are major discrepancies between the number of studies assigned to each risk category reported in Table 1 and Appendix C of Cuijpers *et al.* (2018). Also, no coding procedure or interrater agreement was reported. Given these problems in Cuijpers *et al.*'s RoB determination, we did not use their ratings in our analysis.

Our estimate of the effects of psychotherapy

Professor Cuijpers, upon our request, provided the effect size estimators as well as their standard errors for all 369 comparisons. First, we examined the three types of control groups separately using standard random-effects meta-analysis using the 'metafor' package of 'R' statistical software (Viechtbauer, 2010). In each case, we omitted the outliers (13 WL, 2 CAU and 1 'other control') based on thresholds determined by visual inspection of the effect size distribution for each type of control and omitted comparisons for which $g > 2.00$. Removing such outliers reduces the estimate of the effectiveness of psychotherapy, compared with other procedures, such as Winsorization (Tukey, 1962), in which data are adjusted for outliers rather than eliminated entirely. Then, we also adjusted the effects for publication bias using trim and fill R0-estimates within the 'metafor' package.

The results are shown in Fig. 1, for all comparisons and those that involved psychotherapy. As we have discussed here, the WL is the most appropriate control group for estimating the effects of psychotherapy compared with NT. As can be seen in Fig. 1, the effect of treatment *v.* WL is 0.71 (s.e. = 0.03), a statistically conservative estimate, given elimination of outliers and correcting for publication bias, and one which is similar to that determined by Smith and Glass (1977) and many others (see Wampold and Imel, 2015).

Because we wanted to restrict conclusions to psychotherapy, as defined by Eysenck, for the treatment of adult patients diagnosed with depression, we trimmed the data set accordingly. There were 270 comparisons that met definition of psychotherapy (either individual or group), of these 112 contained adults (excluding elderly, students, patients with general medical conditions or women with post-partum depression) and finally 71 comparisons that involved a diagnosis of depression. The effects for these 71 comparisons (30 WL, 29 CAU and 12 'other control'), after correcting for publication bias, are also presented in Fig. 1 (the standard errors are larger for the psychotherapy studies due to smaller sample sizes). The effect for psychotherapy *v.* WL in this set of comparisons was 0.75 (s.e. = 0.09), again confirming that psychotherapy is effective compared with NT, with a magnitude in the neighbourhood of what Smith and Glass (1977) found. Note, as

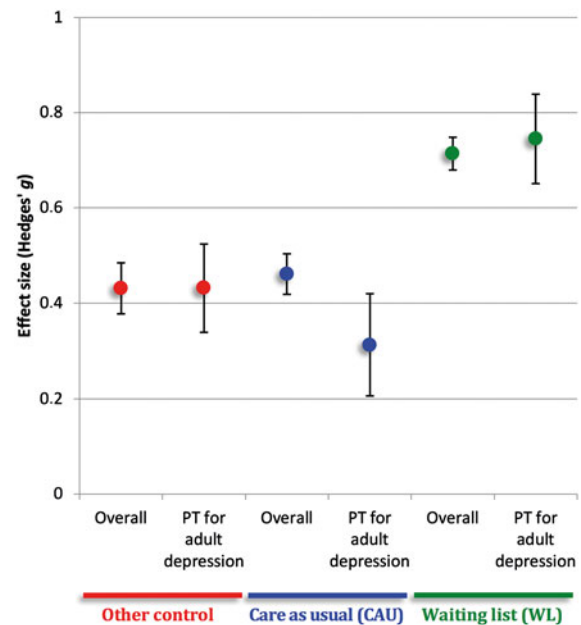


Fig. 1. Effect sizes for psychological interventions for depression. Error bars represent standard errors. PT = psychotherapy. Overall is based on all effect sizes without outliers and corrected for publication bias ($k = 146$ contrasts with WL, $k = 142$ contrasts with CAU, $k = 65$ contrasts with 'other' controls). PT for adult depression only includes (individual or group) psychotherapy for adults with a diagnosis of depression ($k = 30$ contrasts with WL, $k = 29$ contrasts with CAU, $k = 12$ contrasts with 'other control').

well, that in this set of comparisons of treatments that were actually psychotherapy for adult patients diagnosed with depression, psychotherapy was significantly superior to CAU (Hedges' $g = 0.31$, s.e. = 0.11) and 'other control' (Hedges' $g = 0.43$, s.e. = 0.09).

We also tested the set of psychotherapy comparisons to see if there were differences among treatments. We used Cuijpers *et al.*'s (2018) coding and found that there were no statistically significant differences among different types of psychotherapy for adult patients diagnosed with depression (adding type of treatment to a meta-regression model with the type of control did not significantly increase model fit, likelihood ratio test = 9.888, $p = 0.195$). This result is consistent with Cuijpers *et al.* and contradicts Eysenck's claims about the superiority of behavioural treatments.

There are some methodological limitations that may or may not impact the results of the present meta-analysis. First, face-to-face interventional studies are conducted in super-nested designs, randomisation procedures (as one of the key methods to handle risk of biases for internal validity) usually randomise patients to treatment conditions but therapists are neither randomly selected nor randomised to conditions, which may impact the generalisability of the study results to therapists that are not investigated under the study conditions (e.g. Wampold and Imel, 2015). Second, there were unsolved discrepancies between the main text and the Appendix of Cuijpers *et al.* (2018) with regard to RoB criteria, which call into question the reliabilities of the RoB evaluations, further compounding the fact that rater procedures and rater agreement were not reported (see Hartling *et al.*, 2012; Armijo-Olivo *et al.*, 2014). Thus, quality of studies was not considered in our re-analysis. Third, although more statistical driven outlier definitions could have been applied (e.g. Viechtbauer and Cheung, 2010), we opted to exclude outliers based on visual inspection of effect size distribution. This had the advantage of being consistent with Cuijpers *et al.* who used

the same definition of outliers. Fourth, we did not independently calculate effect sizes but used instead the effects provided by Cuijpers *et al.*

Conclusion

After removing outliers, correcting for publication bias, using WL control groups, and restricting analysis to psychotherapy studies, the results of our analyses reveal that psychotherapy for depression is demonstrably effective compared with NT. Indeed, the effect size for psychotherapy compared with natural history, as estimated using WL controls, is about the same size as is generally accepted (i.e. in the neighbourhood of 0.70).

The discrepancy between our results and Cuijpers *et al.* (2018) is due in large part to what is considered an appropriate control group for determining the effectiveness of psychotherapy. Eysenck's claims were about the effectiveness of psychotherapy related to the natural history of the disorder. Determining natural history within the context of randomised clinical trials of psychotherapy is impossible but we have made a case that WL controls are the best possible solution for testing the particular conjecture put forth by Eysenck. Furthermore, dismissing WL conditions as biased is not supported by evidence. In any case, psychotherapy, as defined by Eysenck, is more effective than CAU, even when such care is quite credible, and is more effective than 'other control' as defined by Cuijpers *et al.*

Given these results, as well as a considerable corpus of evidence consistent with these results (Wampold and Imel, 2015), we argue that the field should accept the general conclusion that psychotherapy is an effective practice and give our attention to ways that psychotherapy could be improved.

Acknowledgement. None.

Financial support. None.

Conflict of Interest. None.

References

- Ahola P, Joensuu M, Knekt P, Lindfors O, Saarinen P, Tolmunen T, Valkonen-Korhonen M, Jääskeläinen T, Virtala E, Tiihonen J and Lehtonen J (2017) Effects of scheduled waiting for psychotherapy in patients with major depression. *Journal of Nervous and Mental Disease* 205, 611–617.
- Armento ME (2012) Behavioral activation of religious behaviors: treating depressed college students with a randomized controlled trial (Doctoral Dissertation). University of Tennessee, Knoxville.
- Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes J and Cummings GG (2014) Poor reliability between cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 9, e96920.
- Baskin TW, Tierney SC, Minami T and Wampold BE (2003) Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology* 71, 973–979.
- Benedetti F (2014) *Placebo effects: Understanding the mechanisms in health and disease*, 2nd Edn. New York, NY, US: Oxford University Press.
- Burns A, Banerjee S, Morris J, Woodward Y, Baldwin R, Proctor R, Tarriner N, Pendleton N, Sutherland D and Andrew G (2007) Treatment and prevention of depression after surgery for hip fracture in older people: randomized, controlled trials. *Journal of the American Geriatric Society* 55, 75–80.
- Cristea IA, Gentili C, Cotet CD, Palomba D, Barbui C and Cuijpers P (2017) Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *JAMA Psychiatry* 74, 319–328.
- Cuijpers P, Karyotaki E, Reijnders M and Ebert DD (2018) Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, 1–10. doi: 10.1017/S2045796018000057.
- Del Re AC, Spielmanns GI, Flückiger C and Wampold BE (2013) Efficacy of new generation antidepressants: differences seem illusory. *PLoS ONE* 8, e63509.
- Dimidjian S, Hollon SD, Dobson KS, Schmalzing KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL and Jacobson NS (2006) Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* 74, 658–670.
- Dowrick C, Dunn G, Ayuso-Mateos JL, Dalgard OS, Page H, Lehtinen V, Casey P, Wilkinson C, Vazquez-Barquero JL and Wilkinson G (2000) Problem solving treatment and group psychoeducation for depression: multicentre randomised controlled trial. *BMJ* 321, 1450.
- Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ and Parloff MB (1989) National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry* 46, 971–982.
- Eysenck HJ (1952) The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology* 16, 319–324.
- Eysenck HJ (1961) The effects of psychotherapy. In Eysenck HJ (ed.) *Handbook of Abnormal Psychology*. New York: Basic Books, pp. 697–725.
- Eysenck HJ (1966) *The effects of psychotherapy*. New York: International Science Press.
- Eysenck HJ (1978) An exercise in meta-silliness. *American Psychologist* 33, 517.
- Eysenck HJ (1984) Meta-analysis: an abuse of research integration. *The Journal of Special Education* 18, 41–59.
- Eysenck HJ (1995) Meta-analysis squared—does it make sense? *American Psychologist* 50, 110–111.
- Floyd M, Scogin F, McKendree-Smith NL, Floyd DL and Rokke PD (2004) Cognitive therapy for depression: a comparison of individual psychotherapy and bibliotherapy for depressed older adults. *Behavior Modification* 28, 297–318.
- Frank JD and Frank JB (1991) *Persuasion and healing: A comparative study of psychotherapy*, 3rd Edn. Baltimore: Johns Hopkins University Press.
- Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V and Churchill R (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatrica Scandinavica* 130, 181–192.
- Glass GV (2015) Meta-analysis at middle age: a personal history. *Research Synthesis Methods* 6, 221–231.
- Glass GV and Kliegl RM (1983) An apology for research integration in the study of psychotherapy. *Journal of Consulting and Clinical Psychology* 51, 28–41.
- Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P and Dryden DM (2012) *Validity and inter-rater reliability testing of quality assessment instruments*. Rockville, MD: AHRQ Publication, No. 12-EHC039-EF.
- Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schütze M, Scheunemann W, Altgaier A, Coyne J and Henkel V (2010) Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *International Journal of Neuropsychopharmacology* 13, 31–44.
- Higgins JPT and Green S (2011) *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). Available at <http://handbook-5-1.cochrane.org/>.
- Honyashiki M, Furukawa TA, Noma H, Tanaka S, Chen P, Ichikawa K, Ono M, Churchill R, Hunot V and Caldwell DM (2014) Specificity of CBT for depression: a contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research* 38, 249–260.
- Jansen JP and Naci H (2013) Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Medicine* 11, 159–159.
- Kirsch I (2002) Yes, there is a placebo effect, but is there a powerful antidepressant drug effect? *Prevention & Treatment* 5. doi: 10.1037/1522-3736.5.1.522i.
- Kirsch I (2009) Antidepressants and the placebo response. *Epidemiology and Psychiatric Sciences* 18, 318–322.

- Kirsch I (2010) *The emperor's new drugs: Exploding the antidepressant myth*. New York: Basic Books.
- Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ and Johnson BT (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Medicine* 5, 260–268.
- Laird KT, Tanner-Smith EE, Russell AC, Hollon SD and Walker LS (2017) Comparative efficacy of psychological therapies for improving mental health and daily functioning in irritable bowel syndrome: a systematic review and meta-analysis. *Clinical Psychology Review* 51, 142–152.
- Lamers F, Jonkers CC, Bosma H, Kempen GI, Meijer JA, Penninx BW, Knottnerus JA and van Eijk JT (2010) A minimal psychological intervention in chronically ill elderly patients with depression: a randomized trial. *Psychotherapy and Psychosomatics* 79, 217–226.
- Lilienfeld SO (2007) Psychological treatments that cause harm. *Perspectives on Psychological Science* 2, 53–70.
- Losada A, Márquez-González M, Romero-Moreno R, Mausbach BT, López J, Fernández-Fernández V and Nogales-González C (2015) Cognitive-behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for dementia family caregivers with significant depressive symptoms: results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology* 83, 760–772.
- Milgrom J, Negri LM, Gemmill AW, McNeil M and Martin PR (2005) A randomized controlled trial of psychological interventions for postnatal depression. *British Journal of Clinical Psychology* 44, 529–542.
- Miller FG, Colloca L and Kapchuk TJ (2009) The placebo effect: illness and interpersonal healing. *Perspectives in Biology and Medicine* 52, 518–539.
- Minami T, Wampold BE, Serlin RC, Kircher JC and Brown GSJ (2007) Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology* 75, 232–243.
- Munder T and Barth J (2018) Cochrane's risk of bias tool in the context of psychotherapy outcome research. *Psychotherapy Research* 28, 347–355.
- Munder T, Gerger H, Trelle S and Barth J (2011) Testing the allegiance bias hypothesis: a meta-analysis. *Psychotherapy Research* 21, 670–684.
- Munder T, Flückiger C, Gerger H, Wampold BE and Barth J (2012) Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? A meta-analysis. *Journal of Counseling Psychology* 59, 631–637.
- Munder T, Brüttsch O, Leonhart R, Gerger H and Barth J (2013) Researcher allegiance in psychotherapy outcome research: an overview of reviews. *Clinical Psychology Review* 33, 501–511.
- Posternack MA and Miller I (2001) Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders* 66, 139–146.
- Power MJ and Freeman C (2012) A randomized controlled trial of IPT versus CBT in primary care: with some cautionary notes about handling missing values in clinical trials. *Clinical Psychology and Psychotherapy* 19, 159–169.
- Price DP, Finniss DG and Benedetti F (2008) A comprehensive review of the placebo effect: recent advances and current thought. *Annual Review of Psychology* 59, 565–590.
- Saloheimo HP, Markowitz J, Saloheimo TH, Laitinen JJ, Sundell J, Huttunen MO, Aro TA, Mikkonen TN and Katila HO (2016) Psychotherapy effectiveness for major depression: a randomized trial in a Finnish community. *BMC Psychiatry* 16, 131.
- Shadish WR, Cook TD and Campbell DT (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Smith ML and Glass GV (1977) Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32, 752–760.
- Smith ML, Glass GV and Miller TI (1980) *The benefits of psychotherapy*. Baltimore: The Johns Hopkins University Press.
- Smits JAJ and Hofmann SG (2009) A meta-analytic review of the effects of psychotherapy control conditions for anxiety disorders. *Psychological Medicine* 39, 229–239.
- Spielmanns GI, Gatlin ET and McFall JP (2010) The efficacy of evidence-based psychotherapies versus usual care for youths: controlling confounds in a meta-reanalysis. *Psychotherapy Research* 20, 234–246.
- Spinelli MG and Endicott J (2003) Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. *American Journal of Psychiatry* 160, 555–562.
- Stegenga BT, Kamphuis MH, King M, Nazareth I and Geerlings MI (2012) The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study. *Social Psychiatry and Psychiatric Epidemiology* 47, 87–95.
- Tukey JW (1962) The future of data analysis. *The Annals of Mathematical Statistics* 33, 1–67.
- van Bastelaar KM, Cuijpers P, Pouwer F, Riper H and Snoek FJ (2011) Development and reach of a web-based cognitive behavioural therapy programme to reduce symptoms of depression and diabetes-specific distress. *Patient Education and Counseling* 84, 49–55.
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36, 1–48.
- Viechtbauer W and Cheung MW-L (2010) Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* 1, 112–125.
- Wampold BE (2013) The good, the bad, and the ugly: a 50-year perspective on the outcome problem. *Psychotherapy* 50, 16–24.
- Wampold BE, Budge SL, Laska KM, Del Re AC, Baardseth TP, Flückiger C, Minami T, Kivlighan II DM and Gunn W (2011) Evidence-based treatments for depression and anxiety versus treatment-as-usual: a meta-analysis of direct comparisons. *Clinical Psychology Review* 31, 1304–1312.
- Wampold BE, Flückiger C, Del Re AC, Yulish NE, Frost ND, Pace BT, Goldberg SB, Miller SD, Baardseth TP, Laska KM and Hilsenroth MJ (2017) In pursuit of truth: a critical examination of meta-analyses of cognitive behavior therapy. *Psychotherapy Research* 27, 14–32.
- Wampold BE and Imel ZE (2015) *The great psychotherapy debate: The research evidence for what works in psychotherapy*, 2nd Edn. New York: Routledge.
- Wampold BE, Minami T, Tierney SC, Baskin TW and Bhati KS (2005) The placebo is powerful: estimating placebo effects in medicine and psychotherapy from clinical trials. *Journal of Clinical Psychology* 61, 835–854.
- Wampold BE and Serlin RC (2014) Meta-analytic methods to test relative efficacy. *Quality and Quantity* 48, 755–765.
- Watt LM and Cappeliez P (2000) Integrative and instrumental reminiscence therapies for depression in older adults: intervention strategies and treatment effectiveness. *Aging and Mental Health* 4, 166–177.
- Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJE, Rombach I, Brindley D, Uehiro S, Beard DJ and Carr AJ (2014) Use of placebo controls in the evaluation of surgery: systematic review. *British Medical Journal* 348, 11.
- Wiersma JE, Van Schaik DJ, Hoogendorn AW, Dekker JJ, Van HL, Schoevers RA, Blom MJB, Maas K, Smit JH, McCullough Jr. JP, Beekman ATF, and Beekman AT (2014) The effectiveness of the cognitive behavioral analysis system of psychotherapy for chronic depression: a randomized controlled trial. *Psychotherapy and Psychosomatics* 83, 263–269.