Fundamental Problems with Child and Adolescent Depression Treatments

A Quantitative Reappraisal of the Evidence Base

Argyris Stringaris

Charlotte Burman

Dayna Bhudia

Despina Miliou

Giannis Rokas

Lucy Foulkes

Carmen Moreno

Samuele Cortese

Georgina Krebs

2024-04-22

# Introduction

There are two principal treatment modalities for adolescent depression: psychotherapy and medication. Where does one turn to find the evidence that will inform their treatment options? This question is relevant for patients and their carers, for clinicians and for policy makers. However, the question is particularly difficult to answer for adolescent depression, where there are limited data from head-to-head trials of medication and psychotherapy, and where recommendations must therefore be derived from indirect comparisons of treatment efficacy.

Guidelines, such as the ones that the internationally influential UK National Institute of Health and Care Excellence (NICE) have produced for adolescent depression, are a principal source of such information. NICE recommend that youth are first offered psychological therapy (specifically cognitive behaviour therapy and interpersonal therapy) over medication for most presentations of depression (NICE 2023). This conclusion is in keeping with two sources of evidence: medication meta-analyses that cast doubt on the efficacy of most antidepressants, with the exception of fluoxetine, for child and adolescent depression (Cipriani et al. 2016); and psychotherapy meta-analyses that conclude psychotherapy to be efficacious for child and adolescent depression (Cuijpers et al. 2021). However, such conclusions seem at odds with those of a recent network meta-analysis (Zhou et al. 2020), an established method of comparing treatments with each other using both direct (head-to-head) and indirect (indirectly treatment A with treatment C, via studies that directly compare treatments A with B and B with C) evidence. Indeed, the network meta-analysis concluded that only fluoxetine alone and fluoxetine administered together with CBT were significantly more effective than medication control (i.e. pill placebo) or psychotherapy controls. Given this confusing evidence base, how should patients, carers, clinicians and policy makers make treatment decisions for adolescent depression?

In this paper, we examine whether the existing evidence base for adolescent depression treatment can offer valid answers to such questions. Below, we provide a conceptual framework for answering such questions and test a series of hypotheses using data from existing trials. Two points are crucial to indirect comparisons of treatment modalities with each other.

First, whether the participants of trials in one modality are comparable to those in another modality. Second, whether key conditions of the trial, such as the effects of control conditions or the number of sites involved in a trial, are comparable.

Starting with the first point, to be able to compare between different trials one must assume that these trials sample from the same population. If they do not, then the validity of any comparisons, including those conducted through network metanalysis (which rests on the principal of transitivity, i.e. the requirement that the different sets of randomized trials are similar on average (Chaimani et al. 2023)) are questionable (we provide a more formal account of this in the Supplement).

Indeed, comparing outcomes (Y) between psychotherapy and medication trials requires us to contrast what is called the sample average treatment effect of each, defined in the following way:

where the operator E denotes the expectation over the differences in outcomes between those who received the intervention (T = 1) and those who received the control condition (T = 0), for each sample, , where psy and med stand for psychotherapy and medication respectively.

Obviously, this comparison rests on the assumption that trials in both modalities sample from the same population, , of patients. Formally, this can be expressed as follows:

signifying that the effect found in the population would be expected to be found in the same population for each treatment.

This assumption, that medication and psychotherapy trials sample from the same population, may be hard to meet. Clinical experience and empirical evidence indicate that patients and parents often have preferences between psychotherapy and medication (Jaycox et al. 2006; Langer et al. 2021; McHugh et al. 2013), meaning that there is likely to be a self-selection bias with respect to who participates in psychotherapy versus medication trials. Moreover, these treatment preferences correlate with clinically-relevant characteristics of the participants, including severity, sex and comorbidity. Some of these characteristics, particularly baseline severity, may moderate treatment response (Courtney et al. 2022; Lorenzo-Luaces, Rodriguez-Quintana, and Bailey 2020) and therefore, if they differ across psychotherapy and medication trials, they may confound comparisons.

In terms of the second point, differences in trial design may impact outcomes in a differential way between antidepressant and psychotherapy trials (Del Giovane, Cortese, and Cipriani 2019). Perhaps the most obvious way in which this happens is the fact that participants in psychotherapy trials are unblinded in terms of treatment allocation, with the exception perhaps of trials that compare two equally plausible treatment arms (Calvo et al. 2014). By contrast, in new antidepressant trials, patients (and raters) were found to be unlikely to be able to judge treatment allocation (Lin et al. 2022). This creates differential expectations between medication and psychotherapy trials. Importantly, pill-placebo has been developed as the standard medication control in order to match as closely as possible the intervention condition, so as to minimise differences in expectancy between conditions. By contrast, psychotherapy control arms vary across waitlist controls, treatment as usual and active controls. These differences in expectation favour the psychotherapy active condition and disadvantage the psychotherapy control (thus potentially magnifying the difference between them). This is because of what has been aptly described as “disappointment bias” (the SPIN investigators et al. 2017), where participants in the active condition are content for receiving the “cutting edge treatment” whilst those in the control are dissatisfied for having missed out on “proper treatment”. Expectancy is substantially associated with treatment outcomes (Constantino et al. 2011); i.e., it is an effect modifier. If expectancy effects are different for psychotherapy controls and pill placebo, comparisons between them, including in network metanalysis, become questionable.

Another difference in design that has potential implications for which patients are selected into the trial is the number of sites in a trial. Previous research has shown that in medication trials the number of sites is positively related to the magnitude of the placebo response (Bridge et al. 2009; Dechartres et al. 2011). This phenomenon may be due to lower quality of phenotyping in such studies, with higher rates of classification errors, entailing higher rates of spontaneous remission or regression to the mean. It is therefore important to examine whether this basic trial design feature differs between medication and psychotherapy trials.

An inter-related issue concerns the nature of control interventions being used. Control conditions in trials are meant to generate counterfactual conditions (Guo and Fraser 2014) to the intervention: what would have been the outcome in an individual had they not received the intervention, with all else being equal. A control condition such as pill placebo, where the appearance of the drug is faithfully emulated, is an effort for all else to be equal. The question is whether psychotherapy trials make a similar effort: are for example the number of hours of contact with a human comparable between active and control arms in psychotherapy trials?

Understanding the nature of controls for each treatment modality is important for a similar reason. Often, in the public domain, psychotherapy and medication are compared to each other on the basis of their respective effect sizes. However, these effect sizes represent differences between the active intervention and the control condition (placebo or psychotherapy control). For these effect sizes to be comparable, placebo and psychotherapy controls ought to be the equal. Otherwise, misleading conclusions could be drawn, e.g. two effect sizes of 40% would be considered equal, even if one arose from a difference of 100% - 60% and another arose from a difference of 40% - 0%.

In this paper, we examine randomised controlled trials (RCTs) of psychotherapy and medication for child and adolescent depression in the following ways. First, we compare key baseline characteristics of medication vs psychotherapy trials, specifically the extent to which they are comparable in a) baseline severity of depression; b) percentage females vs males; c) mean age; and d) number of sites involved. Second, we determine the standardised mean differences of psychotherapy controls vs medication controls using random effects metanalysis. Third, we examine the quality of psychotherapy control conditions by scrutinising the extent to which they are matched to the active intervention in ways such as number or frequency of sessions, and therefore, whether they represent fair pairings from which to draw valid efficacy inferences.

# Method

This study was registered on the Open Science Framework (OSF); the protocol can be found [here](https://osf.io/bfmc6). In Supplement X we outline deviations from this pre-registered protocol.

## Studies included

We primarily drew upon RCTs included in two recent comprehensive meta-analyses with open data available for each medication and psychotherapy, and supplemented them with an updated systematic review, described below and in extensive detail in the supplement. Please refer to these original meta-analyses for a detailed description of their search strategy and study selection criteria. Psychotherapy studies were drawn from a systematic review and meta-analysis of randomised trials comparing psychotherapy for youth depression against control conditions (Cuijpers et al. 2021). Cuijpers et al. made available a full dataset of psychotherapy trials (via https://www.metapsy.org/), which we used for the current study. Whilst Cuijpers et al. (2023) excluded those studies for which the primary outcome variable could not be calculated due to missing data, we included these studies and performed the imputations outlined below; hence we have more psychotherapy studies included in this review compared to the original meta-analysis. Whilst the online database is regularly updated, we chose to exclude studies published after the final date of Cuijpers et al.’s (2023) literature search.

Medication studies were drawn from a network meta-analysis examining the efficacy and tolerability of a range of antidepressants and placebo for major depressive disorder in children and adolescents (Cipriani et al. 2016). A dataset was made available online though did not include means or standard deviations at baseline or post-test. We were unable to access the full dataset used in this meta-analysis, and hence completed extraction from the included studies ourselves. We excluded three studies because they did not include a control arm. We were unable to locate and therefore complete extraction for two papers. Many studies did not report complete data; we contacted all corresponding authors to request missing data, though did not receive any responses.

We conducted a systematic search for medication studies published after the final search date of Cipriani’s (2016) review up to the final search date of Cuijpers et al’s (2023) review (May 31 2015 up to Jan 1 2021) to ensure we analysed an equivalently up-to-date database of medication trials. We searched PubMed, the Cochrane Central Register of Controlled Trials, Embase, Web of Science, CINAHL, PsycINFO and LiLACS for randomised controlled trials (RCTs) comparing any antidepressant with placebo in the treatment of children and adolescents with a primary diagnosis of major depressive disorder. We also searched trial registers for published and unpublished studies. We used the same search terms as Cipriani (2016) with one additional search term to include only placebo-controlled trials (see Supplement X for further details). We used Covidence, an online software tool, to manage our systematic review. In total, our search produced 538 studies, 88 of which were duplicates and subsequently removed. Two study authors independently completed title and abstract screenings for 450 studies, and full text screening for 38 studies. 7 studies met inclusion criteria and data extraction was completed for these papers. Clinical trial registers were also systematically searched however all RCTs meeting inclusion criteria had already been identified from the database search outlined above.

## Statistical Analysis

### Trial and sample characteristics

We conducted a series of random effects meta-analyses and tested for subgroup differences between psychotherapy versus medication trials in sample characteristics including sex, age, and severity of depressive symptoms at baseline. Meta-analyses were implemented using R’s Meta package.

In order to compare depression severity across the variety of instruments the studies used, we performed a min-max normalisation to turn each study arm mean score at baseline into a percentage using the following formalism:

where,

is the mean score for each study arm on the primary outcome questionnaire, and and are the minimum and maximum possible values of the scale in question, respectively. The standard deviation is calculated thus:

where is the original standard deviation of the mean at baseline.

We also conducted a t-test to compare mean number of trial sites between psychotherapy and medication trials.

### Measures of Effect

As the measure of effect of each individual study, we used the within-group Standardised Mean Difference (SMD), which we defined following Cumming (2013) as:

where, and refer to the means of the main outcome score at the end and beginning of the intervention respectively and and to the respective standard deviations. Where individual studies did not report all data required to calculate the SMD, we imputed missing data according to the methods summarised in this Cochrane Handbook (Higgins, Li, and Deeks 2023), in the following order. If a study reported the standard error of the mean, the SD was obtained simply by multiplying the SE by the square root of the sample size. For conditions where the SD was missing at one time point, the baseline SD was substituted by the post-test SD, and vice versa. If the SD was not available at either time point, missing values were replaced by the mean of the SDs available for comparable cases (defined as same trial type (psy or med), same instrument, same timepoint (pre or post), and same arm (control or active)). Where there were missing means at either baseline or post-test, missing values were calculated using mean change scores, preferring the change scores reported in the paper itself, though where this was unavailable, using the change scores reported in the dataset from Cipriani’s meta-analysis.

For the purposes of metanalysis, it is necessary to estimate a standard error of the SMD. This is calculated according to:

where refers to the study sample size and refers to the correlation between the outcome score obtained at baseline and at the end point. This correlation is typically not reported in studies and is often imputed using previously reported correlations for the instruments used. However, this practice has given rise to concerns about misestimation. Whilst such misestimation is possible, there is no reason to expect that it would be systematic, i.e. bias estimation of the effects for the control group of medication compared to those of psychotherapy. Still, to alleviate such concerns we have used a simulations.

In particular, we simulated one thousand truncated distribution of standard errors with the following general characteristics:

for which we chose the mean to be , the standard deviation to be , and the upper and lower bounds to be and , respectively. We then used these simulated datasets in the subsequent metanalyses.

### Random Effects Metaregression

We estimated the pooled standardised mean difference for each arm by using a random effects metanalysis implemented in R’s metafor package. The main underlying assumption of random effects metanalysis is that each study’s true effect size is affected not only by sampling error , but also by which represents heterogeneity between studies, allowing each study’s estimate to vary along a distribution of effects, and the distribution of true effect sizes termed . Therefore, we can estimate a two stage model with:

where is the estimated effect size for study i, has a normal distribution with as its true mean effect and sampling error . Whereas is a study-specific instantiation of the distribution of effect sizes, with representing heterogeneity.

This then gives rise to:

where,

describes the deviation of each study from the mean of the distribution, and,

describes the sampling error.

We can then specify the following model to obtain the means of each arm of the trials as follows:

where to obtain the mean of each level is the sum of , the intercept for the reference category of medication control, with the coefficient of each level, e.g. for level 3, the psychotherapy controls. The confidence intervals of the means are constructed in the standard way using the standard errors of the mean. Similarly, each coefficient represents the contrast between the reference category and each level, for an example and of main interest to us represents the contrast between psychotherapy and medication control arms. Inference on the contrasts is done as follows:

We used maximum likelihood (ML) to estimate model and applied Hartung-Knapp adjustment to reduce the chance of false positives (IntHout, Ioannidis, and Borm 2014).

We present the SMDs of each level of the dummy variable of the four treatment arms (medication control, medication active, psychotherapy control, psychotherapy active) under investigation. The SMDs are the means across the 1000 simulated datasets.

### Sensitivity Analyses

We then conducted a series of sensitivity analyses of our results. For each of the metanalyses we excluded 1) studies which recruited participants with subclinical levels of depression and 2) studies that used waitlist as their control condition.

Next, we compared the control and active arms of studies where either the CDRS or the HAM-D where used as outcome instruments.

Further, we tested whether the simulated values for the standard error had a substantial influence on the estimation of the differences between the medication vs psychotherapy control condition. To inspect whether this is the case, we plotted the z-value of the difference between the two coefficients against the number of simulations. We make inference on the stability of the difference, by counting the proportion of times that the z-value is above the critical value of z = 1.645 corresponding to an alpha = 0.05.

### Comparing the control versus active arms of psychotherapy trials

We ran t-tests to compare the active versus control arms of psychotherapy trials on key variables of interest regarding the nature and intensity of the interventions. We extracted data pertaining to the number, duration and intensity of sessions, and the total cumulative hours and period of the intervention. Where a range was provided, the maximum was encoded (e.g. if a paper reported that an intervention involved 8-10 sessions lasting 50-60 minutes, we encoded the number and duration of sessions as 10 and 60, respectively). If sessions varied in frequency across an intervention, we calculated an average by dividing total number of sessions by length of intervention period. Similarly, if the length of sessions varied across the course of the intervention, we calculated a weighted average. Phone call, web-chat and online sessions were encoded as sessions, however guided self-help components were not.

# Results

## Included studies

The data for the studies included in this metanalysis are summarised in Supplementary Table 1 and are also available as a csv dataframe on [<https://github.com/transatlantic-comppsych/apples_oranges>].

In total, there were 88 RCTs which included 44 active arms and 32 control arms of antidepressant trials; and 55 active arms and 52 control arms from psychotherapy RCTs Note that the number of active and control arms does not exactly match because some studies feature more than one control or active arm. There were also missing data for 4, 4, 6, and 6 trial arms for medication active, medication control, psychotherapy active, and psychotherapy control conditions respectively, as the data needed to calculate the SMD were missing and could not be imputed by any of the methods outlined above.

Placebo pill was the control condition for all medication trials. In psychotherapy trials, the control arm included 14 WL controls, 25 care as usual and several other conditions such as 4 attention control conditions. All included trials and the types of treatment controls can be found in Supplementary Table 1.

## Sample characteristics at baseline in medication versus psychotherapy trials

Table 1 summarises the results from each of the meta-analyses examining sample characteristics at baseline. The summary statistics are provided for each subgroup (i.e. for medication and psychotherapy trials) and the p-value derives from the test for subgroup differences. Full results for each of the sensitivity analyses are included in the Supplementary Materials (Tables S1 - S3).

| **Table 1** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Sample characteristics at baseline across medication and psychotherapy studies* | | | | | | | |
| **Subgroup** | **K** | **Mean** | **SE** | **Lower CI** | **Upper CI** | **T2** | **p-value** |
| **Baseline Severity** | | | | | | | |
| **Overall** |  |  |  |  |  |  | **0.011** |
| Psychotherapy | 47 | 0.36 | 0.02 | 0.32 | 0.4 | 0.02 |  |
| Medication | 31 | 0.42 | 0.01 | 0.39 | 0.44 | 0 |  |
| **Excluding subclinical** |  |  |  |  |  |  | **0.132** |
| Psychotherapy | 39 | 0.38 | 0.02 | 0.35 | 0.42 | 0.01 |  |
| Medication | 31 | 0.42 | 0.01 | 0.39 | 0.44 | 0 |  |
| **Excluding waitlist** |  |  |  |  |  |  | **0.027** |
| Psychotherapy | 39 | 0.36 | 0.02 | 0.32 | 0.41 | 0.02 |  |
| Medication | 31 | 0.42 | 0.01 | 0.39 | 0.44 | 0 |  |
| **Percent Female** | | | | | | | |
| **Overall** |  |  |  |  |  |  | **0.031** |
| Psychotherapy | 46 | 60.9 | 2.38 | 56.11 | 65.69 | 260.19 |  |
| Medication | 28 | 53.72 | 2.33 | 48.94 | 58.51 | 152.15 |  |
| **Excluding subclinical** |  |  |  |  |  |  | **0.037** |
| Psychotherapy | 39 | 61.21 | 2.73 | 55.67 | 66.74 | 291.49 |  |
| Medication | 28 | 53.72 | 2.33 | 48.94 | 58.51 | 152.15 |  |
| **Excluding waitlist** |  |  |  |  |  |  | **0.047** |
| Psychotherapy | 38 | 60.83 | 2.7 | 55.35 | 66.31 | 277.94 |  |
| Medication | 28 | 53.72 | 2.33 | 48.94 | 58.51 | 152.15 |  |
| **Age** | | | | | | | |
| **Overall** |  |  |  |  |  |  | **0.276** |
| Psychotherapy | 51 | 14.24 | 0.34 | 13.56 | 14.92 | 5.85 |  |
| Medication | 28 | 13.69 | 0.37 | 12.95 | 14.44 | 3.7 |  |
| **Excluding subclinical** |  |  |  |  |  |  | **0.321** |
| Psychotherapy | 42 | 14.22 | 0.38 | 13.44 | 15 | 6.17 |  |
| Medication | 28 | 13.69 | 0.37 | 12.95 | 14.44 | 3.7 |  |
| **Excluding waitlist** |  |  |  |  |  |  | **0.319** |
| Psychotherapy | 43 | 14.22 | 0.37 | 13.46 | 14.97 | 6 |  |
| Medication | 28 | 13.69 | 0.37 | 12.95 | 14.44 | 3.7 |  |

### Baseline severity

On average, severity of depressive symptoms at baseline was significantly higher in medication trials when compared to psychotherapy trials. The statistics provided in Table 1 are baseline depression scores transformed to reflect percentage of a scale range. To take an example, the CDRS gives a possible total score from 17 to 113 (i.e. range of 96). From Table 1, mean severity was 0.36 for psychotherapy studies and 0.42 for medication studies, which would translate to 51.56 (17 + 0.36 x 96) and 57.32 (17 + 0.42 x 96), respectively, as equivalent scores on the CDRS.

When excluding RCTs that used waitlist as their control, baseline severity remained significantly higher in medication trials compared to psychotherapy trials. This difference did not reach statistical significance when excluding studies that recruited samples with sub-clinical depression.

### Sex

For this analysis, we excluded the two psychotherapy trials which included entirely female samples. As can be seen in Table 1, psychotherapy trials featured a significantly higher percentage of females when compared to medication trials. On average, samples were 60.9% (*SE* = 2.38) female across psychotherapy trials and 53.72% (*SE* = 2.33) female across medication trials. Excluding sub-clinical and waitlist control studies yielded similar results.

### Age

As can be seen in Table 1, mean age was 14.24 (*SE* = 0.34) across psychotherapy trials and 13.69 (*SE* = 0.37) across medication trials, with no significant between group differences. There were no significant differences in mean age between modalities on further sensitivity analyses.

## Number of trial sites

| **Table 2** | | | |
| --- | --- | --- | --- |
| *Number of sites across medication and psychotherapy studies* | | | |
| **Modality** | **N** | **Mean** | **SD** |
| Medication | 35 | 35.96 | 25.16 |
| Psychotherapy | 53 | 3.07 | 3.16 |
| *t*(27.5) = 6.9; *p* < 0.001; 95% CI: 23.1-42.7. | | | |

There was a significant difference between the number of sites in medication versus psychotherapy trials, as can be seen in Table 2. Average number of trial sites was significantly higher across medication studies (*M* = 35.96, *SD* =25.16) compared to psychotherapy studies (*M* =3.07, *SD* =3.16)(*t* (27.53) = 6.89, *p* =< 0.001). Of those studies for which we had data on number of sites, 26 of 28 (93%) medication trials were multisite, compared to 24 of 45 (54%) psychotherapy studies.

## Standardised mean differences of control conditions in psychotherapy and medication studies

We applied metaregression to obtain the SMDs and confidence intervals of each of the four arms.

As can be seen in [Figure 1](#fig-plot-means-all) there were substantial differences between the four arms of the metanalysis with striking differences between the medication and the psychotherapy control arms. In particular, pill placebo had an SMD = -1.9 (95% CI: -2.1 to -1.7) whereas psychotherapy controls had an SMD = -0.5 (95% CI: -0.75 to -0.25 ) (see Table 3 below).

|  |
| --- |
| Figure 1: Metanalytic estimates of within-group changes: all studies |

[Table 1](#tbl-coefs)

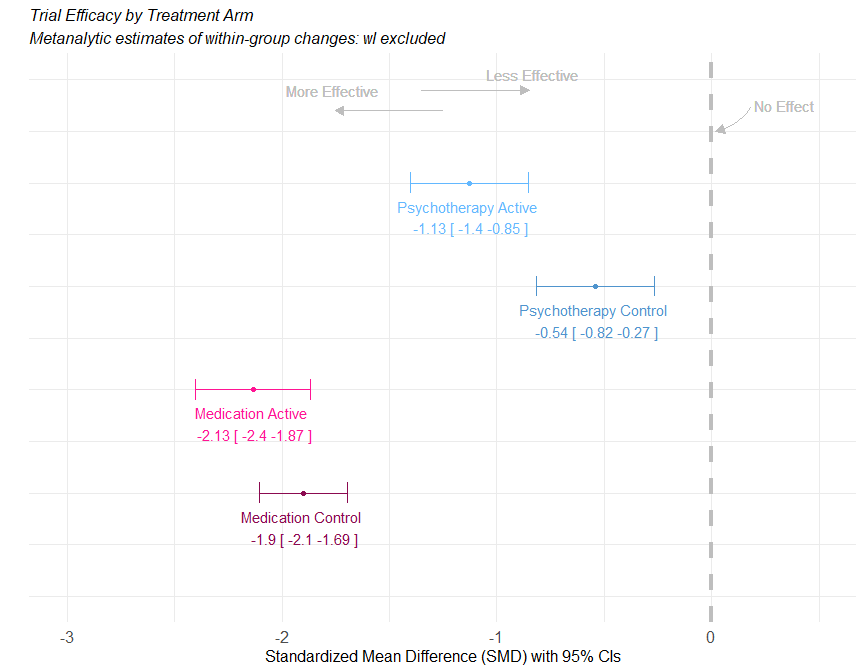
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Summary statistics of the estimated SMDs   | **Table 3** | | | | | | | --- | --- | --- | --- | --- | --- | | *Summary statistics of the estimated SMDs* | | | | | | | **Condition** | **N** | **Coefficient** | **SE** | **Lower CI** | **Upper CI** | | Medication Control | 32 | -1.90 | 0.10 | -2.10 | -1.70 | | Medication Active | 44 | -2.13 | 0.13 | -2.39 | -1.88 | | Psychotherapy Control | 52 | -0.50 | 0.13 | -0.75 | -0.25 | | Psychotherapy Active | 55 | -1.15 | 0.13 | -1.40 | -0.90 | |

In Table 4, we present the regression that tests our hypothesis about differences between medication and psychotherapy controls. In particular, in this metaregression model, medication control is the reference category (termed intercept) to which all other categories of the dummy variable are compared. The strongest difference between arms, as judged by the z-value, is between the psychotherapy and medication controls with a z-value , which yields a very low p-value ( 0.000e+00 ).

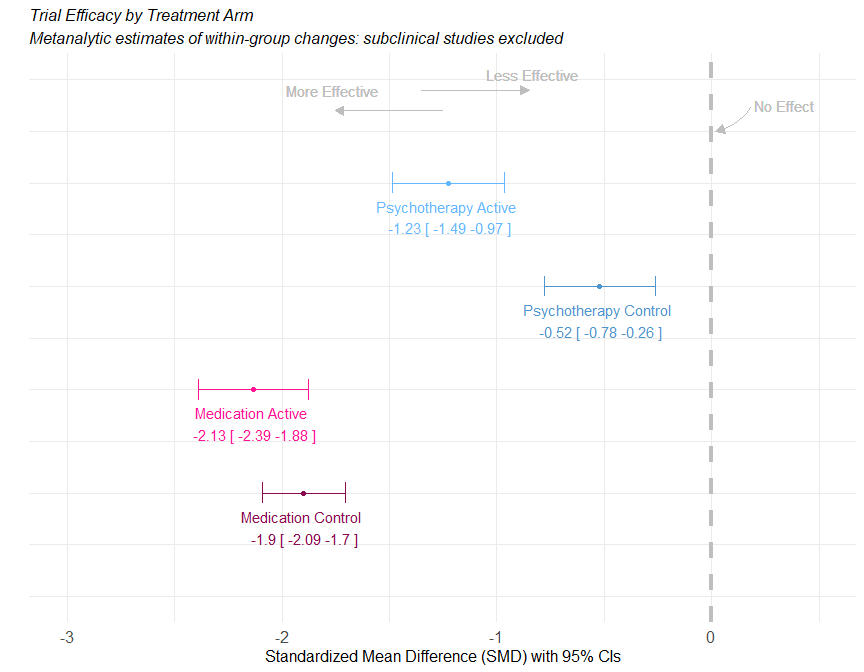
| **Table 4** | | | | | |
| --- | --- | --- | --- | --- | --- |
| *Results from metaregression with overall sample* | | | | | |
| **Condition** | **Coefficient** | **SE** | **z value** | **Lower CI** | **Upper CI** |
| Medication Control | -1.90 | 0.10 | -18.92 | -2.10 | -1.70 |
| Medication Active | -0.24 | 0.13 | -1.79 | -0.50 | 0.02 |
| Psychotherapy Control | 1.40 | 0.13 | 10.94 | 1.14 | 1.65 |
| Psychotherapy Active | 0.75 | 0.13 | 5.82 | 0.49 | 1.00 |
| T2 = 0.2; *I*2 = 92.5; *K* = 183.0; *R*2 = 64.9. | | | | | |

##### Sensitivity analyses

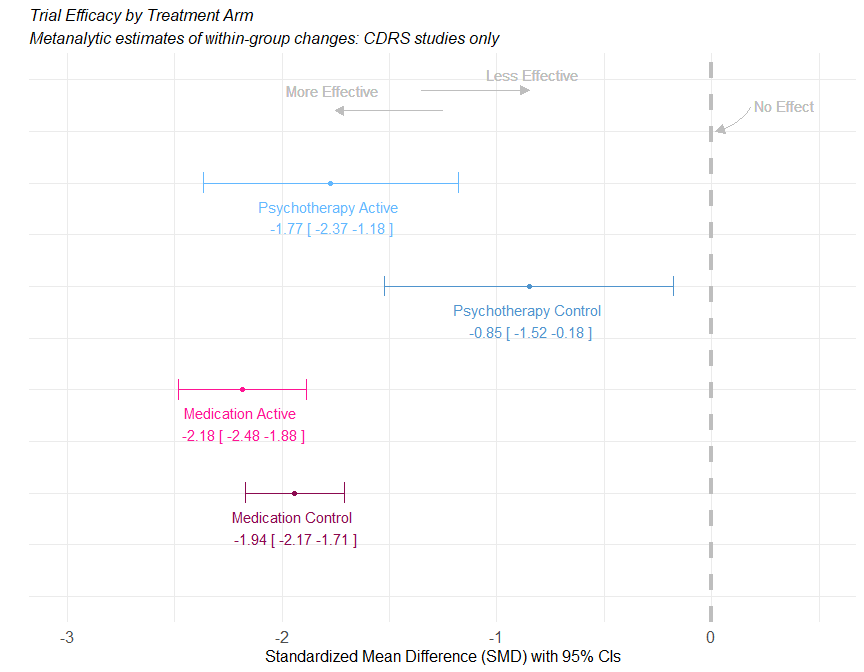
We then conducted a series of sensitivity analyses of our results. First, we analysed the data after excluding waitlist control studies. As can be seen, the pattern of results is very similar to that of the overall analyses.



We next analysed the data after excluding sub-clinical studies. Again, the pattern of results was very similar to that of the overall analyses.

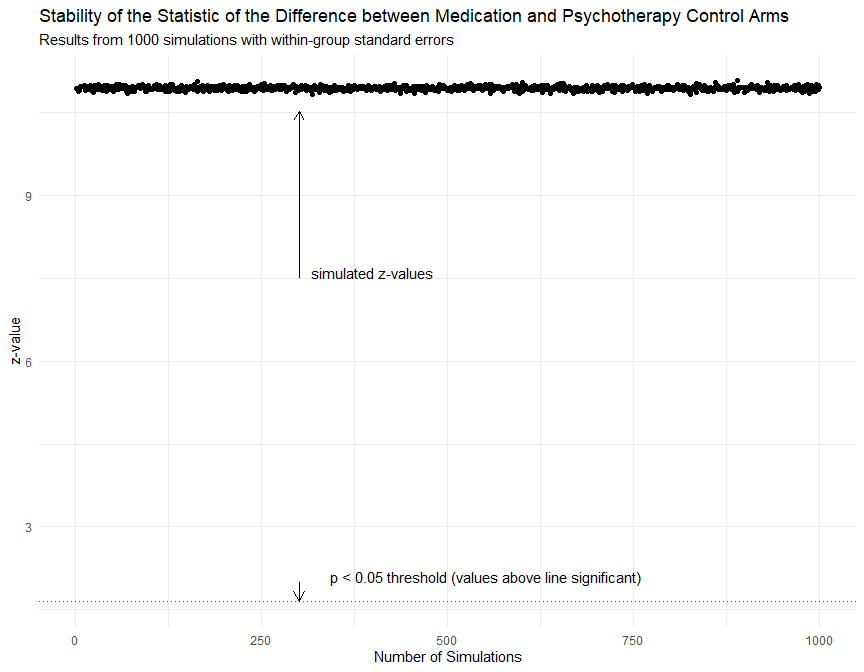


Finally, we examined the data including only those studies that used the CDRS, which yielded a very similar pattern of results.



#### Effect of standard errors of the SMDs

It could be argued that the choice of standard errors of the changes for the calculation of the confidence intervals could have affect the results in one or the other direction. To address such concerns we have simulated 1000 different datasets with SMDs coming from a broad distribution. If standard error distributions were influential, this should show up as substantial variability across simulations. We test this idea in the Figure X which displays across the 1000 simulations the z-value of the contrast between medication and psychotherapy control arms (the mean of which we presented in Table X). As can be seen, the variability in the z-score is minimal and consistently far away from the threshold for significance, i.e. the value of z = 1.645.



#### Addressing Regression to the Mean

We first conducted a meta-analysis to test for differences in means at baseline in the two instruments, CDRS-R and HAM-D, on which there was a sufficient number of studies to metanalyse.

As can be seen in Tables X and Y, the baseline scores in medication studies were on average substantially higher for the CDRS (10.9 points for the control arms) and the HAM-D (7.3 points for the control arms), respectively. The differences were significant at p<0.05 for CDRS, thought they did not reach statistical significance for the HAM-D (p = 0.0573).

We next addressed potential regression to the mean by residualising SMDs by the baseline scores in one regression model.

Table X shows the adjusted SMDs with their confidence intervals that we estimated in metaregression. As can be inferred from the non-overlapping confidence intervals, there were significant differences between the medication control and psychotherapy control arms.

Finally, we tested whether the simulated correlation values we used for simulation of the standard errors had an appreciable effect on the values of the adjusted SMDs. Supplemental Figure X shows that the variability of the estimated z-value was minimal.

## Comparing the nature and intensity of control conditions in psychotherapy trials

| **Table 5** | | | |
| --- | --- | --- | --- |
| *Comparing the intensity of the intervention between active and control arms of psychotherapy studies* | | | |
| **Group** | **N** | **Mean** | **SD** |
| **Number of sessions** | | | |
| Active | 46 | 13.83 | 12.81 |
| Control | 26 | 7.92 | 6.26 |
| **Intensity (sessions per week)** | | | |
| Active | 40 | 1.26 | 0.63 |
| Control | 22 | 0.77 | 0.73 |
| **Session length (mins)** | | | |
| Active | 38 | 66.62 | 28.38 |
| Control | 23 | 38.41 | 35.41 |
| **Total intervention hours** | | | |
| Active | 38 | 13.05 | 8.03 |
| Control | 23 | 7.43 | 8.64 |
| **Intervention duration (weeks)** | | | |
| Active | 45 | 12.02 | 7.74 |
| Control | 26 | 11.08 | 8.27 |

| **Table 6** | | | | | |
| --- | --- | --- | --- | --- | --- |
| *Results for t-tests comparing intervention intensity between active and control arms of psychotherapy trials* | | | | | |
| **Outcome** | **t statistic** | **df** | **p-value** | **Lower CI** | **Upper CI** |
| Number of sessions | 2.62 | 68.92 | 0.011 | 1.41 | 10.40 |
| Intensity (sessions per week) | 2.68 | 38.29 | 0.011 | 0.12 | 0.87 |
| Session length (mins) | 3.24 | 38.93 | 0.002 | 10.61 | 45.81 |
| Total intervention hours | 2.53 | 43.90 | 0.015 | 1.14 | 10.10 |
| Intervention duration (weeks) | 0.48 | 49.50 | 0.637 | -3.05 | 4.94 |

Overall, the active arms of psychotherapy studies were considerably more intensive than the control arms they were compared against (see Table 6 for summary statistics). Active intervention arms featured significantly more sessions than control arms (*t* (68.92) = 2.62, *p* = 0.011). These sessions were also significantly longer (*t* (38.93) = 3.24, *p* = 0.002) and more frequent (*t* (38.29) = 2.68, *p* = 0.011). The total hours involved in an intervention were higher in active vs control arms (*t* (43.9) = 2.53, *p* = 0.015), though the total period of the intervention was similar (*t* () = 0.48, *p* = 0.637).

# Discussion

This paper took as its starting point the question of how anyone, be it a patient, parent, or clinician, should decide whether to opt for medication or psychotherapy for the treatment of adolescent depression. In order to address this question, we asked whether the two treatment modalities — medication and psychotherapy — are comparable on the basis of the existing evidence. Specifically, our paper answered two questions that are fundamental to any attempts at comparing the two treatment modalities.

First, whether the participants of trials in one modality are comparable to those in another modality. Second, whether key conditions of the trial, such as the effects of control conditions or the number of sites involved in a trial, are comparable.

Starting with the first question, we find evidence that participants in the two treatment modalities differ in the following key aspects: those enrolled in medication trials are more likely to be male and have more severe depression compared to those in psychotherapy trials. By contrast, there were no significant differences in age.

Severity is important as it may moderate treatment response, with some evidence suggesting that those with higher baseline scores respond more to antidepressant therapy (Stone et al. 2022) or that their response to pill placebo is lower (Bridge et al. 2009). Some studies argue against severity as a treatment moderator (Tröger et al. 2024; Weitz et al. 2015), however, these are within modality (e.g. within CBT studies), i.e. within those people who have chosen to be in the particular trial and modality. Moreover, there is evidence that severity may represent different subtypes in terms of course of depression and its outcomes in real life settings (Lamers et al. 2016; Simmonds-Buckley, Catarino, and Delgadillo 2021).

Similarly, we found that on average there were about 7% more females in psychotherapy trials. Whilst there is little evidence that sex or sex moderate treatment response within modality in adult trials of depression (Cuijpers et al. 2014), this is yet another indicator that different people enter trials of each modality type, which would violate basic assumptions of comparability between trials.

We then turned to our second basic question, about whether key conditions of the trial design were comparable between modalities. We find a series of critical differences between the two modalities.

First, medication trials are vastly more likely to be multi-site than their psychotherapy counterparts; in the current study, 93% of medication RCTs were multisite compared to 54% of psychotherapy RCTs Multisite trials are associated with higher pill placebo response (Bridge et al. 2009), and are less common in NIH-funded (i.e. non industry) trials which are also more likely to show higher pill placebo efficacy. Also, in single-site trials, principal investigators are often intellectually invested in the treatment (in psychotherapy these are often treatments developed or modified by the PI); this is in stark contrast to the incentive structure in multi-site trials where primacy is given to the number of recruited participants which is the primary unit of reimbursement.

Second, psychotherapy controls have moderate effect size (-0.5) whereas pill-placebo has very large effect size (-1.9). Our analysis could be critiqued as it relies on comparing the within arm symptom change of each trial taken out of randomization, which is generally advised against (CITE). This criticism would apply if our aim were to draw inferences about the efficacy of each arm — in which case preserving the randomization (in order to balance confounders) is critical. However, we note that, our findings are largely in keeping with those of the network metanalysis, which is designed to preserve the randomization structure. More importantly, we do not make the claim with our analyses that these differences are genuinely due to efficacy differences; they may well be because of the fact that the people who attend psychotherapy and medication trials are different and therefore respond differently. In either case (difference in efficacy vs difference in trial participant profile), the vast disparity in the response to control conditions is reason for major concern about our ability to draw inferences concerning the comparison between the two modalities. This is all the more so as clinicians as well as policy makers often resort to effect sizes to summarise findings, with SMD being the one used as per standard. Our findings make obvious that comparing treatment modalities based on the effect size of each modality is misleading.

Moving beyond the comparison of the controls between the two modalities, we examined whether psychotherapy controls are reasonable counterfactuals to receiving the treatment. An optimal control condition is one where the treatment differs (e.g. treatment vs no treatment), but everything else is equal (in technical terms, the ceteris paribus assumption). An obvious disadvantage of psychotherapy trials is that they are typically unblinded (and hard to blind). Yet, our results show that psychotherapy trials are unlikely to fulfill some other very basic conditions of the “all else is equal” assumption. Most (XXX %) of psychotherapy RCT control conditions are either waitlist control or treatment as usual, both of which are very likely to create negative expectations for participants not randomized to treatment; thus, the comparison is not between treatment and no treatment, but rather treatment and the poor luck of being randomized to waiting. But we also find that even psychological controls were unreasonable counterfactuals. In order to test that a psychological treatment is effective per se (e.g. because of the cognitive techniques the therapist deploys) rather than because of generic effects (e.g. pleasant human contact), aspects such as human contact time should be matched. We find that there are vast differences between treatment and control arms in psychotherapy trials (CITE NUMBER) which may lead to an inflation of the true efficacy of psychological treatments.

Given all of the above, the certainty with which guidelines, including UK’s NICE recommend psychotherapy over medication for adolescent depression is surprising. Indeed, we believe that our findings have a number of profound implications for patients, their families, clinicians and policy makers and we list these below.

First, the grounds for a comparison between medication and psychotherapy should be seen as shaky, rather than offering confidence, and there is an urgent need to revisit guidelines and public information in light of the limitations.

Second, the low quality of control conditions of psychotherapy trials for adolescent depression should prompt a rethinking of how these studies are appraised and what needs to be done to create fair and realistic comparators. Indeed, investment of effort and funds should be directed into providing rigorous evidence that established depression psychotherapies are more efficacious than fair controls. There are examples of such psychotherapy RCTS, e.g. in social anxiety (Clark et al. 2006) where such rigor has been applied.

Third, our findings make clear the inherent difficulties of comparing psychotherapy with medication trials (Del Giovane, Cortese, and Cipriani 2019). The first obstacle is the comparability of the populations taking part. Even if a trial were designed to conduct a head-to-head comparison of psychotherapy with medication (as has been done in (“Fluoxetine, Cognitive-Behavioral Therapy, and Their Combination for Adolescents With Depression: Treatment for Adolescents With Depression Study (TADS) Randomized Controlled Trial” 2004)), it might sample the population of those who are indifferent to which one they would receive. And even in such a design, difficulties with blinding of the psychotherapy control would have to be overcome to draw valid inferences.

In summary, our data give cause for consternation about the state of the evidence of youth depression, one of the most common and debilitating disorders in young people. Our data question the state of knowledge about the efficacy of youth psychotherapies and, in particular, the extent to which giving them primacy in the treatment of depression is justified and beneficial for young people. Returning to our motivating question, the stakeholders, including patients and clinicians, deserve better evidence on which to base their choice than currently exists.

# Supplement

Table S1: Included studies.

Figure S1: PRISMA chart summarising the screening process for the systematic review.

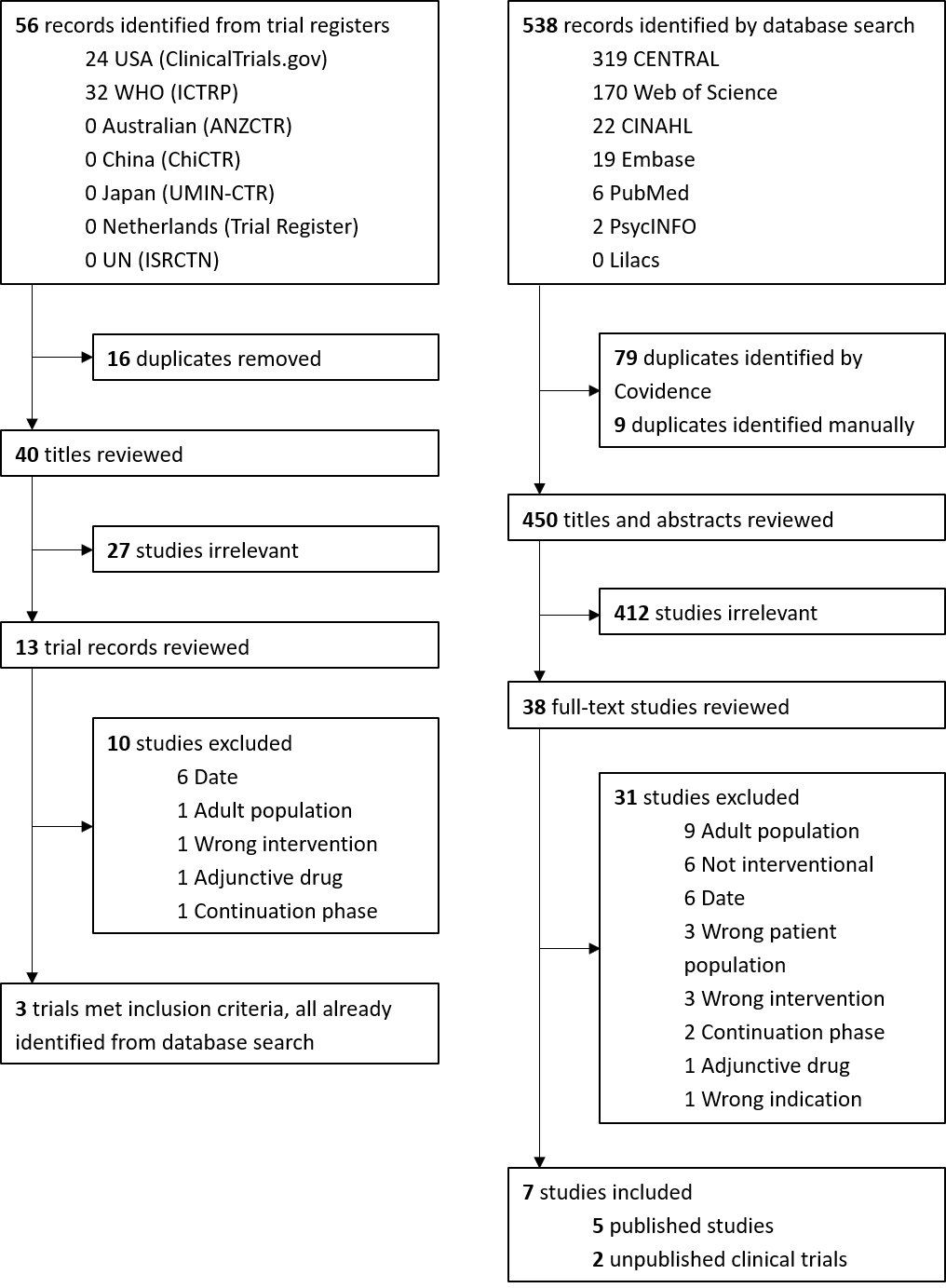
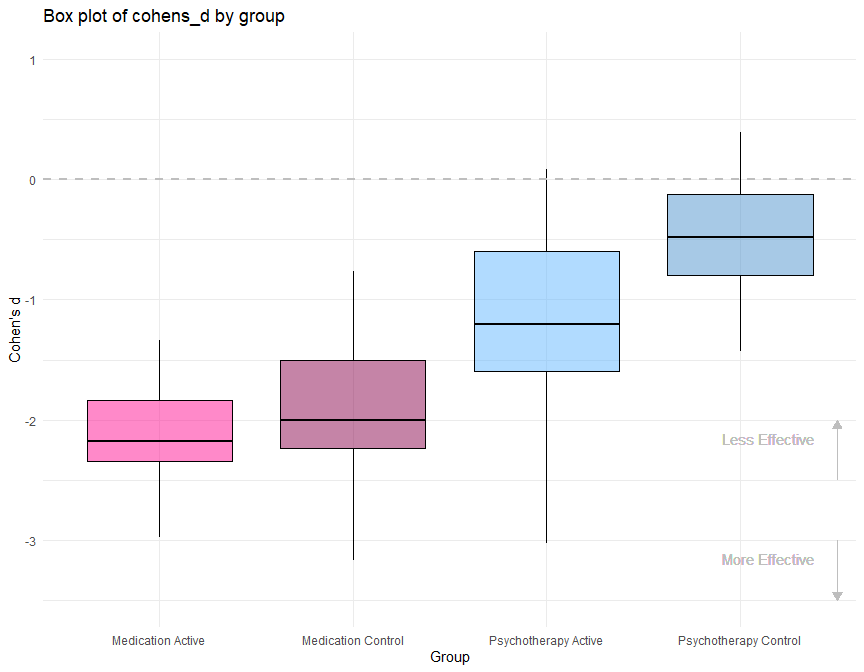


Figure S1



# References

Bridge, Jeffrey A., Boris Birmaher, Satish Iyengar, Rémy P. Barbe, and David A. Brent. 2009. “Placebo Response in Randomized Controlled Trials of Antidepressants for Pediatric Major Depressive Disorder.” *American Journal of Psychiatry* 166 (1): 42–49. <https://doi.org/10.1176/appi.ajp.2008.08020247>.

Calvo, Ana, Miguel Moreno, Ana Ruiz-Sancho, Marta Rapado-Castro, Carmen Moreno, Teresa Sánchez-Gutiérrez, Celso Arango, and María Mayoral. 2014. “Intervention for Adolescents With Early-Onset Psychosis and Their Families: A Randomized Controlled Trial.” *Journal of the American Academy of Child & Adolescent Psychiatry* 53 (6): 688–96. <https://doi.org/10.1016/j.jaac.2014.04.004>.

Chaimani, A, DM Caldwell, T Li, Julian PT Higgins, and G Salanti. 2023. “Chapter 11: Undertaking Network Meta-Analyses.” In *Cochrane Handbook for Systematic Reviews of Interventions Version 6.4*, 6.4 ed. Cochrane.

Cipriani, Andrea, Xinyu Zhou, Cinzia Del Giovane, Sarah E Hetrick, Bin Qin, Craig Whittington, David Coghill, et al. 2016. “Comparative Efficacy and Tolerability of Antidepressants for Major Depressive Disorder in Children and Adolescents: A Network Meta-Analysis.” *The Lancet* 388 (10047): 881–90. <https://doi.org/10.1016/S0140-6736(16)30385-3>.

Clark, David M., Anke Ehlers, Ann Hackmann, Freda McManus, Melanie Fennell, Nick Grey, Louise Waddington, and Jennifer Wild. 2006. “Cognitive Therapy Versus Exposure and Applied Relaxation in Social Phobia: A Randomized Controlled Trial.” *Journal of Consulting and Clinical Psychology* 74 (3): 568–78. <https://doi.org/10.1037/0022-006X.74.3.568>.

Constantino, Michael J., Diane B. Arnkoff, Carol R. Glass, Rebecca M. Ametrano, and JuliAnna Z. Smith. 2011. “Expectations.” *Journal of Clinical Psychology* 67 (2): 184–92. <https://doi.org/10.1002/jclp.20754>.

Courtney, Darren B., Priya Watson, Karolin R. Krause, Benjamin W. C. Chan, Kathryn Bennett, Meredith Gunlicks-Stoessel, Terri Rodak, Kirsten Neprily, Tabitha Zentner, and Peter Szatmari. 2022. “Predictors, Moderators, and Mediators Associated With Treatment Outcome in Randomized Clinical Trials Among Adolescents With Depression: A Scoping Review.” *JAMA Network Open* 5 (2): e2146331. <https://doi.org/10.1001/jamanetworkopen.2021.46331>.

Cuijpers, Pim, Eirini Karyotaki, Marketa Ciharova, Clara Miguel, Hisashi Noma, Yvonne Stikkelbroek, John R. Weisz, and Toshi A. Furukawa. 2021. “The Effects of Psychological Treatments of Depression in Children and Adolescents on Response, Reliable Change, and Deterioration: A Systematic Review and Meta-Analysis.” *European Child & Adolescent Psychiatry*, October. <https://doi.org/10.1007/s00787-021-01884-6>.

Cuijpers, Pim, Erica Weitz, Jos Twisk, Christine Kuehner, Ioana Cristea, Daniel David, Robert J. DeRubeis, et al. 2014. “GENDER AS PREDICTOR AND MODERATOR OF OUTCOME IN COGNITIVE BEHAVIOR THERAPY AND PHARMACOTHERAPY FOR ADULT DEPRESSION: AN ‘INDIVIDUAL PATIENT DATA’ META-ANALYSIS: Research Article: Gender as Moderator of Treatment Outcome.” *Depression and Anxiety* 31 (11): 941–51. <https://doi.org/10.1002/da.22328>.

Cumming, Geoff. 2013. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. 1st ed. Routledge. <https://doi.org/10.4324/9780203807002>.

Dechartres, Agnes, Isabelle Boutron, Ludovic Trinquart, Pierre Charles, and Philippe Ravaud. 2011. “Single-Center Trials Show Larger Treatment Effects Than Multicenter Trials: Evidence From a Meta-epidemiologic Study.” *Annals of Internal Medicine* 155 (1): 39. <https://doi.org/10.7326/0003-4819-155-1-201107050-00006>.

Del Giovane, Cinzia, Samuele Cortese, and Andrea Cipriani. 2019. “Combining Pharmacological and Nonpharmacological Interventions in Network Meta-analysis in Psychiatry.” *JAMA Psychiatry* 76 (8): 867. <https://doi.org/10.1001/jamapsychiatry.2019.0574>.

“Fluoxetine, Cognitive-Behavioral Therapy, and Their Combination for Adolescents With Depression: Treatment for Adolescents With Depression Study (TADS) Randomized Controlled Trial.” 2004. *JAMA* 292 (7): 807. <https://doi.org/10.1001/jama.292.7.807>.

Guo, Shenyang, and Mark W. Fraser. 2014. “Chapter 2: Counterfactual Framework and Assumptions.” In *Propensity Score Analysis: Statistical Methods and Applications*. SAGE Publications.

Higgins, Julian PT, Tianjing Li, and Jonathan J Deeks. 2023. “Chapter 6: Choosing Effect Measures and Computing Estimates of Effect.” In *Cochrane Handbook for Systematic Reviews of Interventions Version 6.4*, 6.4 ed. Cochrane.

IntHout, Joanna, John Pa Ioannidis, and George F Borm. 2014. “The Hartung-Knapp-Sidik-Jonkman Method for Random Effects Meta-Analysis Is Straightforward and Considerably Outperforms the Standard DerSimonian-Laird Method.” *BMC Medical Research Methodology* 14 (1): 25. <https://doi.org/10.1186/1471-2288-14-25>.

Jaycox, Lisa H., Joan Rosenbaum Asarnow, Cathy D. Sherbourne, Margaret M. Rea, Anne P. LaBorde, and Kenneth B. Wells. 2006. “Adolescent Primary Care Patients’ Preferences for Depression Treatment.” *Administration and Policy in Mental Health and Mental Health Services Research* 33 (2): 198–207. <https://doi.org/10.1007/s10488-006-0033-7>.

Lakens, Daniel. 2013. “Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs.” *Frontiers in Psychology* 4.

Lamers, F., A. T. F. Beekman, A. M. Van Hemert, R. A. Schoevers, and B. W. J. H. Penninx. 2016. “Six-Year Longitudinal Course and Outcomes of Subtypes of Depression.” *British Journal of Psychiatry* 208 (1): 62–68. <https://doi.org/10.1192/bjp.bp.114.153098>.

Langer, David A., Tessa K. Kritikos, Joan R. Asarnow, and Martha C. Tompson. 2021. “Parent and Youth Preferences in the Treatment of Youth Depression.” *Child Psychiatry & Human Development* 52 (2): 236–47. <https://doi.org/10.1007/s10578-020-01006-4>.

Lin, Yi-Hsuan, Ethan Sahker, Kiyomi Shinohara, Noboru Horinouchi, Masami Ito, Madoka Lelliott, Andrea Cipriani, Anneka Tomlinson, Christopher Baethge, and Toshi A. Furukawa. 2022. “Assessment of Blinding in Randomized Controlled Trials of Antidepressants for Depressive Disorders 2000–2020: A Systematic Review and Meta-Analysis.” *eClinicalMedicine* 50 (August): 101505. <https://doi.org/10.1016/j.eclinm.2022.101505>.

Lorenzo-Luaces, Lorenzo, Natalie Rodriguez-Quintana, and Allen J. Bailey. 2020. “Double Trouble: Do Symptom Severity and Duration Interact to Predicting Treatment Outcomes in Adolescent Depression?” *Behaviour Research and Therapy* 131 (August): 103637. <https://doi.org/10.1016/j.brat.2020.103637>.

McHugh, R. Kathryn, Sarah W. Whitton, Andrew D. Peckham, Jeffrey A. Welge, and Michael W. Otto. 2013. “Patient Preference for Psychological Vs Pharmacologic Treatment of Psychiatric Disorders: A Meta-Analytic Review.” *The Journal of Clinical Psychiatry* 74 (06): 595–602. <https://doi.org/10.4088/JCP.12r07757>.

NICE. 2023. “Depression in Children and Young People: Identification and Management.”

Simmonds-Buckley, Melanie, Ana Catarino, and Jaime Delgadillo. 2021. “Depression Subtypes and Their Response to Cognitive Behavioral Therapy: A Latent Transition Analysis.” *Depression and Anxiety* 38 (9): 907–16. <https://doi.org/10.1002/da.23161>.

Stone, Marc B, Zimri S Yaseen, Brian J Miller, Kyle Richardville, Shamir N Kalaria, and Irving Kirsch. 2022. “Response to Acute Monotherapy for Major Depressive Disorder in Randomized, Placebo Controlled Trials Submitted to the US Food and Drug Administration: Individual Participant Data Analysis.” *BMJ*, August, e067606. <https://doi.org/10.1136/bmj-2021-067606>.

the SPIN investigators, Clare Relton, Maarten Burbach, Clive Collett, James Flory, Sophie Gerlich, Soren Holm, et al. 2017. “The Ethics of ‘Trials Within Cohorts’ (TwiCs): 2nd International Symposium: London, UK. 7-8 November 2016.” *Trials* 18 (S2): 244, s13063-017-1961-0. <https://doi.org/10.1186/s13063-017-1961-0>.

Tröger, Anna, Clara Miguel, Marketa Ciharova, Nino De Ponti, Güldehan Durman, Pim Cuijpers, and Eirini Karyotaki. 2024. “Baseline Depression Severity as Moderator on Depression Outcomes in Psychotherapy and Pharmacotherapy.” *Journal of Affective Disorders* 344 (January): 86–99. <https://doi.org/10.1016/j.jad.2023.10.047>.

Weitz, Erica S., Steven D. Hollon, Jos Twisk, Annemieke Van Straten, Marcus J. H. Huibers, Daniel David, Robert J. DeRubeis, et al. 2015. “Baseline Depression Severity as Moderator of Depression Outcomes Between Cognitive Behavioral Therapy Vs Pharmacotherapy: An Individual Patient Data Meta-analysis.” *JAMA Psychiatry* 72 (11): 1102. <https://doi.org/10.1001/jamapsychiatry.2015.1516>.

Zhou, Xinyu, Teng Teng, Yuqing Zhang, Cinzia Del Giovane, Toshi A Furukawa, John R Weisz, Xuemei Li, et al. 2020. “Comparative Efficacy and Acceptability of Antidepressants, Psychotherapies, and Their Combination for Acute Treatment of Children and Adolescents with Depressive Disorder: A Systematic Review and Network Meta-Analysis.” *The Lancet Psychiatry* 7 (7): 581–601. <https://doi.org/10.1016/S2215-0366(20)30137-1>.