Behaviour Research and Therapy 159 (2022) 104170

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Cognitive behavior therapy for obsessive-compulsive disorder in routine clinical care: A systematic review and meta-analysis

Lars-Göran Öst [a,b,c,*], Pia Enebrink [c], Anna Finnes [c,d], Ata Ghaderi [c], Audun Havnen [e,f], Gerd Kvale [b,g], Sigrid Salomonsson [h], Gro Janne Wergeland [i,j]

[a] Department of Psychology, Stockholm University, Sweden
[b] Bergen Center for Brain Plasticity, Haukeland University Hospital, Bergen, Norway
[c] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[d] Academic Primary Care Center, Region Stockholm, Sweden
[e] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway
[f] Division of Psychiatry, St. Olavs Hospital, Trondheim, Norway
[g] Department of Clinical Psychology, University of Bergen, Norway
[h] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[i] Department of Child and Adolescent Psychiatry, Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
[j] Department of Clinical Medicine, Faculty of Medicine, University of Bergen, Norway

ARTICLE INFO

Keywords:
Obsessive-compulsive disorder
Effectiveness
Cognitive behavior therapy
Adults
Meta-analysis

ABSTRACT

Cognitive behavioral therapy (CBT) has strong research support for obsessive-compulsive disorder (OCD). However, less is known about how CBT performs when delivered in routine clinical care. A systematic review and meta-analysis was conducted of CBT for OCD in adults treated in routine clinical care. Ovid MEDLINE, Embase OVID, and PsycINFO were systematically searched for studies published until July 2021. The effectiveness of CBT, methodological quality, and moderators of treatment outcome were examined, and benchmarked by meta-analytically comparing with efficacy studies for OCD. Twenty-nine studies (8 randomized controlled trials) were included, comprising 1669 participants. Very large within-group effect sizes (ES) were obtained for OCD-severity at post-treatment (2.12), and follow-up (2.30), on average 15 months post-treatment. Remission rates were 59.2% post-treatment and 57.0% at follow-up. Attrition rate was 15.2%. Risk of bias was considerable in the majority of studies. The benchmarking analysis showed that effectiveness studies had almost exactly the same ES as efficacy studies at post-treatment and somewhat higher at follow-up. Furthermore, effectiveness studies had significantly higher remission rates than efficacy studies, both at post- and follow-up assessment. CBT for OCD is an effective treatment when delivered in routine clinical care, with ES comparable to those found in efficacy studies. However, the evidence needs to be interpreted with caution because of the risk of bias in a high proportion of studies.

Prospero registration id: CRD42021228828.

1. Introduction

Obsessive–compulsive disorder (OCD) is characterized by anxiety-evoking intrusive thoughts, images or urges (obsessions) and repetitive behaviors aimed at reducing the discomfort (compulsions). Untreated OCD tends to be chronic, causing significant functional impairment and reduced quality of life (Koran, Thienemann, & Davenport, 1996; Macy et al., 2013).

The lifetime prevalence of OCD was reported at 1.6% for the National Comorbidity Survey Replication study in the USA (Kessler et al., 2005) and 2.3% in a later article from that study (Ruscio, Stein, Chiu, & Kessler, 2010). In a study from the WHO World Mental Health Surveys covering more than 21,000 respondents in 14 countries, Kessler et al. (2011) found a lifetime prevalence of 6.9% in developed countries (one of which was the USA) and 5.8% in developing countries. All these studies used DSM-IV (American Psychiatric Association, 1994) criteria and, to the best of our knowledge, there is not yet a published study of OCD prevalence using DSM-5 (American Psychiatric Association, 2013)

* Corresponding author. Department of Psychology, Stockholm University, 11419, Stockholm, Sweden.
E-mail address: ost@psychology.su.se (L.-G. Öst).

https://doi.org/10.1016/j.brat.2022.104170
Received 8 November 2021; Received in revised form 11 June 2022; Accepted 25 July 2022
Available online 2 August 2022

criteria.

Comorbidity is very common in OCD patients. Ruscio et al. (2010) found that 90% of patients in the US met diagnostic criteria for another mental disorder; most common was another anxiety disorder (75.8%) followed by mood disorders (63.3%). In a study from the Netherlands Hofmeijer-Sevink et al. (2013) reported that 78% had lifetime comorbidity, which was related to more severe OCD and negative life consequences.

In addition to the high rates of comorbidity, OCD is associated with high levels of disability, severe distress, and impact on everyday life situations. Using the Sheehan Disability Scales, Ruscio et al. (2010) found that 78.7% of OCD-patients reported severe or moderate impairment in at least one of the domains home management, work, relationships or social life. In a review of 58 studies Macy et al. (2013) found that the quality of life in patients with OCD was significantly impaired and lower than that of patients with other psychiatric disorders. Investigating the global burden of disease in 2010, Baxter, Vos, Scott, Ferrari, and Whiteford (2014) reported that anxiety disorders, which included OCD, ranked as the 6th leading cause of disability in both high- and low-income countries.

Effective and empirically supported treatments for OCD exist, with cognitive behavior therapies (CBT) being the psychological treatment with the strongest evidence base. Division 12 of the American Psychological Association (https://div12.org/diagnosis/obsessive-compulsive-disorder/) does the following evaluation on its website: Exposure and Response Prevention (ERP) and the combination of cognitive therapy and ERP both have strong research support. The National Institute for Health and Care Excellence (NICE, 2013) in the UK and the Australian Psychological Society (2018) have basically done the same evaluation.

However, most of the evidence for the empirically supported treatments comes from randomized controlled trials (RCTs) often carried out in specialized research clinics, i.e. efficacy trials. As such, an important question for clinicians is to know what may be expected from empirically supported treatments for OCD when delivered in routine clinical care (Tolin, McKay, Forman, Klonsky, & Thombs, 2015; Weisz, Ugueto, Cheron, & Herren, 2013).

Studies conducted for the purpose of establishing efficacy are often designed for extensive control of trial factors to increase internal validity. Participants are usually subject to stringent selection criteria in order to obtain a more homogenous sample of participants, to minimize the influence of external factors on the treatment. They may be more motivated, and special efforts are made to increase adherence to the treatment (Stewart & Chambless, 2009). Efficacy trials often have randomization of participants to treatment and control conditions, independent and masked assessors, and well-trained and carefully selected therapists. Furthermore, the therapists often have a focused caseload, receive more intensive supervision, and have treatment adherence monitoring following a specific treatment manual (Hunsley, 2007). Also, treatment environment is optimized with staff and facilities dedicated to research, and with resources to reschedule missed appointments and treatment monitoring. One concern raised regarding the treatments having strong research support is how well results generalize into routine clinical care (e.g., Gonzales & Chambers, 2002; Westen & Morrison, 2001), which arises from the belief that patients, therapists and treatment context may all differ in important ways between research settings and routine clinical care. As such, it is important to evaluate how empirically supported treatments for OCD perform in routine clinical care and how results compare with outcomes obtained in efficacy trials.

Effectiveness studies aim to maximize external validity while maintaining an adequate level of internal validity. Efforts to increase external validity include placing the treatment study within routine clinical practice, including patients being referred for treatment to the clinics, and using therapists working at these clinics with caseloads consisting of a broad array of disorders and referred problems to deliver the treatment (Hunsley, 2007; Stewart & Chambless, 2009). These studies can include both RCTs and pre-post trials. One strategy to examine the effectiveness is to take a CBT program proven to be effective and evaluate how results generalize to routine clinical care.

Stewart and Chambless (2009) published the first meta-analysis on effectiveness studies in anxiety disorders, where they included 11 OCD-studies. They found a within-group effect size (ES) of 1.32 for OCD-symptoms and 0.89 for depression. Benchmarking was done against three selected efficacy studies having a range of 1.15–1.88. Thus, the ES they found was within the range of the efficacy studies. Later Hans and Hiller (2013), in a similar meta-analysis including 23 OCD-studies, reported a within-group ES of 1.46 for OCD-symptoms and 0.66 for depression.

A recent meta-analysis of effectiveness studies for internalizing disorders in youth (Wergeland, Riise, & Öst, 2021) used a much more comprehensive benchmarking strategy than that of Stewart and Chambless (2009). In a direct meta-analytical statistical comparison, they included all efficacy studies in the most recent meta-analysis of CBT for OCD in youth. The mean post-treatment ES for effectiveness was 2.29 and for efficacy studies 2.50, and at follow-up 3.51 for effectiveness and 2.70 for efficacy studies, indicating that CBT did as well in routine clinical care as in university settings. Stewart and Chambless (2009) included studies up to 2008 and Hans and Hiller (2013) up to 2012. The overlap between the current meta-analysis and the previous ones is limited; only 11 of the 29 studies (38%) we included were also included in those meta-analyses. Thus, an updated meta-analysis of effectiveness studies of CBT for OCD in adults, that includes studies published during the past 10 years is needed.

The present article will contribute to the existing literature by providing a meta-analysis of the effectiveness of CBT for OCD in adults receiving treatment in routine clinical care. We selected effectiveness studies in which patients were referred for treatment through usual clinical routes, and the treatments were delivered in routine clinical practices by therapists for whom provision of service is a substantial part of their job. To be as comprehensive as possible we included both RCTs and pre-post trials to better capture all studies conducted in routine clinical care. Assuming that the mean ES will be significantly heterogeneous, we extracted a number of background and treatment variables, and rated methodological quality and risk-of-bias to investigate potential moderator variables. Finally, using meta-analytical statistical methods, we directly compared effectiveness and efficacy studies of OCD regarding effect size and remission rates, both at post-treatment and follow-up.

Our specific aims were: First, to examine the effectiveness of CBT for OCD in routine clinical care regarding the primary OCD-measure and remission as well as a secondary measure of depression. Second, to evaluate methodological quality and risk-of-bias in the effectiveness studies, and investigate potential moderators of treatment outcome. Third, to examine how CBT delivered in routine clinical care do in comparison with efficacy studies for OCD. Based on the previous meta-analyses in adults (Hans & Hiller, 2013; Stewart & Chambless, 2009) and youth (Wergeland et al., 2021) we predicted that the ESs for effectiveness studies will be comparable to those of efficacy studies.

2. Method

The protocol for this meta-analysis was pre-registered at PROSPERO with ID CRD42021228828. The meta-analysis was conducted according to the PRISMA guidelines (Liberati et al., 2009), and reported according to AMSTAR 2 (Shea et al., 2017), see online Supplement S1 and S2. The meta-analysis was designed according to the PICOS acronym in the following way:

- Population: adults with OCD diagnosis

- Intervention: CBT, CT, or BT evaluated by APA Division 12 as having strong or modest research support and delivered in routine clinical care
- Comparison: within-group change, i.e. pre vs. post-data (and pre vs. follow-up data)
- Outcome: primary (OCD-symptoms) and secondary (depression) continuous measure, and dichotomous measure of remission
- Study design: randomized controlled trials (RCTs) and pre-post trials

2.1. Literature search

Studies were identified by a systematic and comprehensive literature search of electronic databases and scanning of the included articles' reference lists. The search was applied to Ovid MEDLINE, Embase OVID, and PsycINFO from the start of the data bases to June 23, 2020. An updated search was done July 30, 2021. The list of search terms utilized to identify potential studies were generated by all authors in collaboration with a university librarian, who conducted the database searches. We used the following search terms to search the databases: (Cognitive therapy; behav* therapy; cognitive behav* therapy; cognitive behav* treatment; acceptance and commitment therapy; ACT) AND (obsessive compulsive disorder; OCD) AND (open study; clinical study; community trial; intervention study; Pre-post study; randomized controlled trial) AND (outpatient clinics; community mental health services; effectiveness; routine care; regular care, community clinic*) AND adults. For full search strategy for Ovid MEDLINE, Embase OVID and PsychINFO, see the online Supplement, S3.

Three pairs of authors (AF, AG, AH, GK, PE, SS) read the abstracts of all the papers from this initial search to decide whether a study warranted a more detailed reading. We were over-inclusive at this stage and when there was any indication at all of a target group of patients receiving the particular cognitive-behavioral treatment in a routine clinical care setting the full-text was retrieved. The reference lists in the retrieved articles were then checked against the database search and any other articles that might fulfil the inclusion criteria were retrieved. In total, 84 full-text articles were considered for inclusion. The final decision for article inclusion was made using a stricter set of inclusion and exclusion criteria detailed below. The full-text articles were read by pairs of the authors and any disagreements were resolved by consensus discussion among the authors and/or consultation with the first author. It was determined that 29 articles with 38 treatment conditions could be included in the present meta-analysis.

2.1.1. Inclusion criteria

To be included in the review and meta-analysis a study had to:

1. Be published, or in press, in an English language journal.
2. Have participants diagnosed with OCD according to DSM (III and later) or ICD (10 or 11).
3. Be testing a form of CBT, CT or BT that is evaluated as having strong or modest research support by Division 12 of the American Psychological Association (APA).
4. Have participants referred for treatment through usual clinical routes.
5. Be an effectiveness study, i.e., carried out in a routine clinical care setting such as community mental health centres, patients' homes, etc.
6. Have therapists who are practicing clinicians for whom provision of service is a substantial part of their job (Shadish, Matt, Navarro, & Phillips, 2000).
7. Have a treated sample consisting of at least 10 participants.
8. Have a minimum participant age of 18.
9. Provide a continuous or dichotomous measure of the principal disorder treated.

2.1.2. Exclusion criteria

1. The study is a secondary analysis of a previously published study. However, separate follow-up studies to the basic study are included to provide follow-up data.
2. The study is an evaluation of a service where the results for individual disorders *cannot* be extracted.
3. The study is testing a combination of CBT/CT/BT and pharmacological treatment and all participants in that condition receive both treatments.

Fig. 1 shows a flowchart of the inclusion of studies in the present meta-analysis. For references to included studies, see online Supplement S4, and for references to studies excluded in the meta-analyses, see online Supplement S5.

2.2. Potential categorical moderators

To include any potential categorical or continuous moderator in the analysis we required that at least 70% of the studies provided information on that variable. With lower proportions it is questionable if the information extracted is representative of the entire body of studies.

2.2.1. Type of study and statistical analysis

Type of study was either pre-post (when only a CBT-condition was used in the study, or a non-randomized study where at least one condition was CBT) or RCT (when a CBT-condition was compared with some kind of control/comparison condition). Statistical analysis was categorized as completers (if dropouts were deleted) or as intent-to-treat (ITT, if all randomized or starting participants were included in the statistical analysis).

2.2.2. Treatment format and therapist profession

Treatment format could either be individual, group, or family. The latter required that at least one family member of the patients participated in the therapy sessions. Profession was classified according to which profession the majority of the therapists had. In studies where a team of different professionals were working this variable was classified as mixed.

2.2.3. Continent

The country in which the study was carried out was categorized as situated in Africa. Asia, Australia, Europe, North America, or South America.

2.3. Potential continuous moderators

The following continuous measures on which at least 70% of the studies provided information were used as potential moderators: number of participants in the study, percent female patients, mean age, pretreatment severity (calculated as percentage of the maximum score of the rating scale applied), methodology score (see below), number of treatment sessions, hours of treatment, percent on psychotropic medication for their OCD, and percent attrition in the treatment condition. To be counted as a dropout, the patient had to fulfill the inclusion criteria, be offered to receive the treatment, accepted it, and participated in at least the first session but less than the number of sessions defined as completion of treatment. We also extracted information on a number of other categorical and continuous variables but these did not reach the 70% criterion and were excluded. A coding scheme and a scoring manual including the variables of interest were developed. The data extraction and categorizations were done independently by pairs of the authors and any disagreements were solved after consensus discussion.

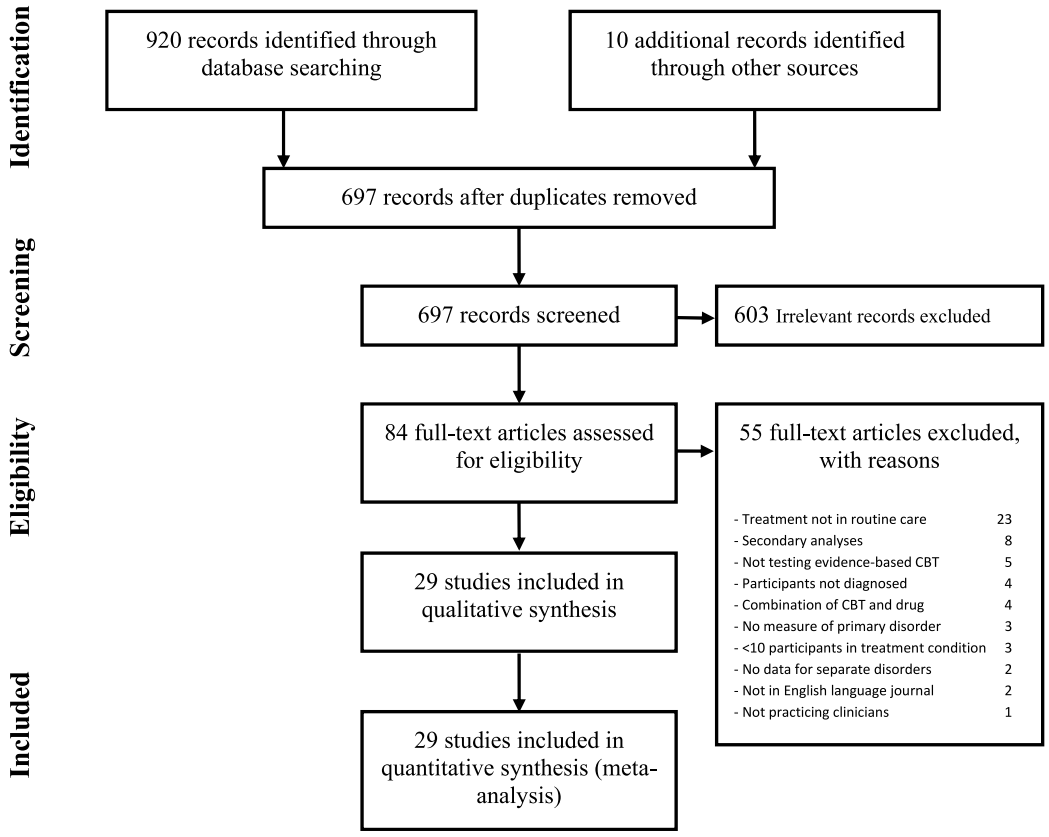

Fig. 1. Flowchart of the inclusion of studies.

2.4. Methodological quality

2.4.1. The psychotherapy outcome study methodology rating scale (POMRS)

The scale consists of 22 items covering various important aspects of the methodology in psychotherapy outcome research (Öst, 2008; Öst and Ollendick, 2017). Each item is rated as 0 = poor, 1 = fair, and 2 = good, and each step has a verbal description of one or more sentences. The total score can vary from 0 to -44 points. Since all items were not applicable to all studies, the total score was recalculated as a percentage of the maximum score possible for the individual study. The internal consistency of the scale was good with a McDonald's ω of 0.80. The inter-rater reliability of the scale (between the first and the last author), based on 20% randomly selected and blindly rated studies was ICC = 0.95 (95% CI 0.72–0.99, $p = 0.0001$), which according to Cicchetti (1994) is excellent.

2.4.2. Risk-of-bias

For RCTs the Cochrane Collaboration tool for assessing risk-of-bias (Sterne et al., 2019) was used, and the following domains were rated: the randomization process, missing outcome data, measurement of the outcome, and selection of the reported result. The domain deviations from intended interventions was not rated since therapists and patients in psychotherapy studies cannot be blind regarding the treatment applied. For non-randomized studies of intervention (NRSI) and pre-post studies the Risk of Bias in Non-randomised Studies - of Interventions (ROBINS-I; Sterne et al., 2016) was applied. The following domains were judged: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcome, and selection of the reported result. An overall classification of the studies was done for RCTs into the categories high, some concerns or low risk of bias. For the NRSI and pre-post studies the categories low, moderate, serious or critical risk-of-bias were used. When the results across these different study designs were judged to be at similar risk-of-bias, these classifications were combined into one: low, moderate (some concerns) and high (serious) risk-of-bias. The rating of the studies was done by two independent researchers and differences were discussed to reach consensus.

2.5. Effect size measures

Patients applying for treatment at clinics in the community are often less interested in whether the treatment is superior to a control condition, and more interested in the degree of improvement that can be expected and the chance of achieving remission following the treatment offered. Thus, in this meta-analysis we used the pre-post and pre-follow-up effect size, as well as the rate of remission at post-treatment and follow-up assessment as outcome measures. We extracted data on both primary and secondary measures in the studies. Since some studies used proportion of remitted participants as their primary outcome measure, whereas other studies used a continuous rating scale, we decided to include both in this meta-analysis.

2.5.1. Continuous rating scales

Most studies used the Yale-Brown Obsessive Compulsive Scale (YBOCS; Goodman et al., 1989); 25 the interviewer version and two the self-report version (Lovell et al., 2006; Warren & Thomas, 2001). One study used the Dimensional Obsessive-Compulsive Scale (Lee, Bistricky, Milam, Wetterneck, & Björgvinsson, 2016), and one the Hamburger Zwangsinventar (Wetzel, Bents, & Florin, 1999).

2.5.2. Remission

Data on remission were provided by 21 of the treatment conditions (55%). Of these, 13 used the Jacobson and Truax (1991) criteria for

clinically significant change, requiring that the pre-post/follow-up change fulfilled the Reliable Change Index and that the post/follow-up score fulfilled a cut-off criterion which made it probable that it belonged to the normal population distribution. Four studies used a modification of the international consensus criteria (Mataix-Cols et al., 2016), requiring at least a 35% pre-post/follow-up change and having a post/follow-up score of ≤12 points on YBOCS. The following cut-off scores on YBOCS were used: ≤ 7 (n = 2), ≤12 (n = 10), ≤14 (n = 2), ≤15 (n = 2), and ≤16 (n = 5). One study used a ≥2 SD change from pre to post and one did not specify any cut-off score.

2.5.3. Secondary outcome measures

Since depressive disorder or symptoms are common comorbid problems in OCD we extracted data on depressive symptoms, which was provided by 69% of the included studies. We also extracted data on general anxiety and quality of life. However, only 38% and 7%, respectively, of the studies provided such data, which precluded their use in the meta-analysis.

2.6. Meta-analysis

In order to obtain as large as possible a body of effectiveness studies we included both RCTs, NRSI, and pre-post trials in the meta-analysis since within-group ES can be calculated from all types of studies. Within-group ES was calculated as $(M_{pre} - M_{post})/SD_{pre}$ according to recommendation by Lakens (2013), since there is good reason to assume that the interventions influence not only the means but also the standard deviations. The mean ES was computed by weighting each ES by the inverse of its variance. When a study presented intent-to-treat data these were used, if not completer data were used.

Before pooling, the effect sizes we screened for statistical outliers, defined as being outside M ± 2SD. At the post-treatment assessment two (5.3%) of the ESs were outliers, and at follow-up assessment there was one (3.6%). For these ESs, *Winsorizing* (Lipsey & Wilson, 2001) was used by reducing outliers to the exact value of M+2SD. The *Comprehensive Meta-Analysis v.3* (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2013) software was used for all analyses and to correct for small sample sizes Hedges' *g* was calculated. A random effects model was used since it cannot be assumed that the ESs come from the same population.

Proportion of remission was calculated in CMA. The values of the individual studies were transformed using logit transformation and the meta-analysis was done on the transformed proportions using the random effects model. Then the pooled proportion and its 95% confidence interval was back transformed to a proportion (according to recommendations by Barendregt, Doi, Lee, Norman, & Vos, 2013; Barker et al., 2021).

Heterogeneity among ESs was assessed with the *Q*- and the I^2- statistic. The possibility of publication bias was analyzed with the trim-and-fill method of Duval and Tweedie (2000) as well as Egger's regression intercept (Egger, Davey Smith, Schneider, & Minder, 1998). Moderator analyses of continuous variables were carried out with meta-regression using the random effects model and for categorical variables with subgroup analysis using the mixed effect model.

2.7. Efficacy studies for comparison

In order to obtain the efficacy studies to be used in comparison of the effect of CBT in effectiveness studies we consulted the most recent comprehensive meta-analyses of CBT for OCD (Leeuwerik, Cavanagh, & Strauss, 2019; Öst, Havnen, Hansen, & Kvale, 2015). From these meta-analyses, we listed the RCTs of cognitive behavioral treatments evaluated as well-established or probably efficacious according to the criteria adopted by Division 12 of APA. Then we deleted those RCTs we had already included in the body of effectiveness studies. This resulted in 32 RCTs for our comparison and the references of these are listed in the online Supplement S6.

As for the effectiveness studies, we extracted data for the primary continuous outcome measure and remission rate, separately at post-treatment and follow-up assessment. In order to compare the two categories of studies on background and treatment variables we also extracted data on mean age, proportion of women, pre-treatment severity (% maximum score on the continuous measure), comorbidity (% of the sample having at least one comorbid disorder), medication (% of the sample that at pre-treatment was prescribed a psychotropic drug for OCD), treatment time (in 60 min units), and attrition rate (% dropping out of those patients who participated in at least one session). Other variables were not reported systematically or not at all in a large enough proportion of studies, which precluded inclusion as a background variable.

Since some of the result tables will entail many statistical tests, e.g., 9 in the meta-regression analysis of continuous moderators we used the Holm-Bonferroni correction to control the family-wise error rate (see Jaccard & Guilamo-Ramos, 2002).

2.8. Power analysis

In the overall comparison of effectiveness and efficacy studies we have the following number of studies and treatment conditions, which is the unit of analysis: effectiveness studies 29/38 and efficacy studies 32/54, for a total number of 61 studies and 92 conditions with an average of 30 participants per condition. According to the formulas for power analysis in meta-analyses by Valentine, Pigott, and Rothstein (2010) we would have 95.8% power to detect a small ES (0.20), when assuming that the heterogeneity of effect sizes will be high.

3. Results

3.1. Description of the studies

3.1.1. Background data

Background data for the included studies are shown in Table 1. There was a total of 29 studies, including 38 treatment conditions, since some studies had two or more CBT interventions. Eight (27.6%) of the studies were RCTs whereas two were NRSI (6.9%) and 19 (65.5%) were pre-post trials. The total number of participants receiving CBT in these studies was 1669. The majority of the 29 studies was done in Europe (n = 18, 62%), followed by North America (n = 10), Asia (n = 1), and both South America and Europe (n = 1). There was a majority of women (60%), and the overall mean age of the samples was 33.7 (*SD* 3.0) years. Comorbidity was reported for only 24 conditions (63.2%) and often unsystematically. The proportion of participants having at least one comorbid disorder was 55.1% (*SD* 17.9%). The mean pre-treatment severity across treatment conditions was 59.5% (*SD* 7.2%). The proportion of the samples taking prescribed psychotropic medications for their OCD at the time of inclusion to the respective studies was reported for 29 conditions (76.3%) and the mean was 57.9% (*SD* 18.6%). Finally, 24 conditions (63.2%) reported the proportion of eligible participants that declined the offer of treatment and the mean was 11.1% (*SD* 9.5%).

3.1.2. Treatment data

Treatment data for the included studies are presented in Table 2. The treatment format was individual in 30 conditions (78%), group in 6, individual + group in 1, and family in 1. The number of therapists was reported in 29 conditions (76.3%), with a range from 1 to 28, and a mean of 5.1 (*SD* 5.7). The most common profession among the therapists was clinical psychologists (n = 22; 70.9%), followed by psychiatrists (n = 3), psychiatric nurses (n = 3), social workers (n = 2), and a mixed team of professions (n = 1). The treatments were carried out over 14.8 (*SD* 10.4) weeks on average, with the mean number of therapy sessions being 14.5 (*SD* 9.3). Calculated as hours of treatment the mean was 19.5 (*SD* 9.7). Attrition rate was reported for 34 conditions (89.5%) and the mean was 15.1% (*SD* 11.4%). Twenty-nine of the conditions (76.3%)

Table 1
Background data of the included studies.

Study	Country	Continent	RCT	Method of CBT	Comparison	N	% declining	Severity	% females	Mean age	% medicated	% comorbidity
Abramowitz, 2005	USA	NA	N	ERP	None	77		0.618	56.8	37.7	74.1	40.3
Andersson, 2011	Sweden	E	N	ICBT	None	23	20.0	0.500	65.0	39.0	34.0	
Belloch, 2008a	Spain	E	Y	CT	Other CBT	17	8.3	0.660	60.0	30.2	80.0	60.0
Belloch, 2008b	Spain	E	Y	ERP	Other CBT	16	8.3	0.618	61.5	34.2	92.3	53.9
Belloch, 2010a	Spain	E	N	ERP-autogen.	None	19		0.583	33.3	31.0	53.3	40.0
Belloch, 2010b	Spain	E	N	ERP-reactive	None	49		0.658	57.6	34.6	84.4	39.4
Cabedo, 2010a	Spain	E	Y	CBT-Ind	Other CBT	18	8.3	0.645		30.4		
Cabedo, 2010b	Argentina	SA	N	CBT-group	None	24	45.0	0.625		37.1		
Grøtte, 2018	Norway	E	N	ERP	None	187	16.9	0.650	63.6	34.2	54.0	
Hansen, 2018	Norway	E	N	cERP	None	65	13.9	0.645	70.8	32.1	46.2	60.0
Havnen, 2014	Norway	E	N	cERP	None	35	5.4	0.653	70.3	32.4	60.0	60.0
Havnen, 2017	Norway	E	N	cERP	None	42		0.643	67.0	32.6	26.2	61.9
Houghton, 2010	England	E	N	CBT	None	37	14.0	0.605	49.0	36.8		86.5
Håland, 2010	Norway	E	N	CBT	None	54		0.598	74.0	35.9	52.0	64.0
Launes, 2019a	Norway	E	N	cERP	None	36	12.8	0.668	71.4	30.4	40.0	68.6
Launes, 2019b	Norway	E	Y	cERP	Other CBT	48	4.0	0.670	79.2	30.4	43.8	87.5
Lee, 2016	USA	NA	N	ERP	None	49		0.433	42.9	30.8		55.1
Lovell, 2006a	England	E	Y	ERP: FtF	Other CBT	36	6.5	0.638	61.0	30.4	42.0	
Lovell, 2006b	England	E	Y	ERP: Tel.	Other CBT	36	6.5	0.648	56.0	33.4	61.0	
Norberg, 2008	USA	NA	N	ERP	None	95		0.560	44.7	31.8		
Olsen 2008	Norway	E	N	ERP	None	25	3.8	0.543	56.0	35.6	68.0	64.0
Papageorgiou, 2018	England	E	N	CBT	None	125	10.8	0.623	52.8	35.0	78.4	76.0
Rector, 2018a	Canada	NA	Y	ERP	Other CBT	62	9.3	0.610	69.1	32.8	64.3	31.5
Rector, 2018b	Canada	NA	Y	ERP + CT	Other CBT	65	9.3	0.580	50.0	31.6	69.2	31.5
Rosqvist, 2001	USA	NA	N	ERP	None	11		0.560	45.5	40.0		27.0
Rothbaum, 2000	USA	NA	N	ERP	None	39		0.608	56.5	32.8	47.8	
Sarichloo, 2020	Iran	Asia	Y	ERP	Other CBT	30		0.533	42.1			
Tolin, 2007a	USA	NA	Y	ERP: Therapist	Other CBT	21	21.2	0.600	52.4	34.1	57.1	28.6
Tolin, 2007b	USA	NA	Y	ERP: Self	Other CBT	20	21.2	0.568	75.0	40.3	55.0	55.0
Tundo, 2007	Italy	E	N	CBT	None	36	0.0	0.705	36.1	31.0	88.8	58.3
van Noppen, 1997a	USA	NA	N	Group BT	None	17	0.0	0.598	59.0	35.1		
van Noppen, 1997b	USA	NA	N	Family BT	None	19	0.0	0.598	74.0	31.1		
van Noppen, 1998a	USA	NA	N	CBGT + med.	None	72		0.548	70.0		73.0	
van Noppen, 1998b	USA	NA	N	CBGT	None	18		0.528	70.0		73.0	
Warren, 2001	USA	NA	N	CBT	None	26		0.575	37.0	30.1	68.0	32.0
Wetzel, 1999	Germany	E	N	ERP	None	85		0.300	77.6	34.1	24.9	
Vogel, 2004a	Norway	E	Y	ERP + CT	Other CBT	16	10.3	0.623	56.0	31.4	34.3	62.0
Vogel, 2004b	Norway	E	Y	ERP + REL	Other CBT	19	10.3	0.593	84.0	39.3	34.3	79.0

Note: Empty cells within each column mean that the information on this variable was not provided. RCT: Y = yes, N = no. Method of CBT: BT = behavior therapy, CBT = cognitive behavior therapy, CBGT = cognitive behavioral group therapy, cERP = concentrated ERP, CT = cognitive therapy, ERP = exposure and response prevention, FtF = face to face, ICBT = Internet-based CBT.

Table 2
Treatment data for the included studies.

Study	Format	# of therapists	Profession	Weeks	Sessions	Hours	Percent attrition	F-up months	Statistical analysis
Abramowitz, 2005	I								Compl.
Andersson, 2011	I	1	Psychol.	15	15.0	1.5	13.0		Compl.
Belloch, 2008a	I	2	Psychol.	26	18.0	16.0	5.9	12	Compl.
Belloch, 2008b	I	2	Psychol.	26	18.0	20.0	18.8	12	Compl.
Belloch, 2010a	I	4	Psychol.	26	18.0	18.0	18.6	12	Compl.
Belloch, 2010b	I	4	Psychol.	26	18.0	18.0	18.6	12	Compl.
Cabedo, 2010a	I	6	Psychol.	26	18.0	18.0	11.1	12	Compl.
Cabedo, 2010b	G	6	Psychol.	18	16.0	32.0	8.3	12	Compl.
Grøtte, 2018	I	18	Nurse	3			11.2	6	ITT
Hansen, 2018	I	11	Psychol.	1	4.0	22.0	0.0	12	ITT
Havnen, 2014	I	8	Psychol.	1	4.0	22.0	2.9	6	ITT
Havnen, 2017	I	6	Psychol.	1	4.0	22.0	0.0	6	ITT
Houghton, 2010	I	8	Nurse		13.0	10.8	24.3		Compl.
Håland, 2010	G	2	Mixed	12	12.0	30.0	7.4	12	ITT
Launes, 2019a	I	5	Psychol.	1	4.0	22.0	2.8	3	ITT
Launes, 2019b	I	7	Psychol.	1	4.0	22.0	0.0	6	ITT
Lee, 2016	I + G		Psychol.						ITT
Lovell, 2006a	I	2		10	10.0	10.0	8.3	6	ITT
Lovell, 2006b	I	2		10	10.0	5.0	2.8	6	ITT
Norberg, 2008	I		Psychol.	20	22.4	22.4	5.3		Compl.
Olsen, 2008	I	2	Nurse	32			4.0	42	Compl.
Papageorgiou, 2018	G			16	12.0	24.0	9.6		ITT
Rector, 2018a	I	4	Psychiat.	16	16.0	16.0	25.8	6	Compl.
Rector, 2018b	I	4	Psychiat.	16	16.0	16.0	13.8	6	Compl.
Rosqvist, 2001	I			12	24.0	36.0	27.3	6	Compl.
Rothbaum, 2000	I	1	Psychiat.		7.7	11.6	41.0		Compl.
Sarichloo, 2020	I	1	Psychol.	12	12.0	16.5	36.7	3	Compl.
Tolin, 2007a	I	2	Psychol.	8	15.0	12.5	19.0	6	ITT
Tolin, 2007b	I	1	Psychol.	8	2.0	1.0	15.0	6	ITT
Tundo, 2007	I	4	Psychol.		30.4	30.4	33.3		Compl.
van Noppen, 1997a	G		Counsellor	36	11.0	22.0	22.7	12	Compl.
van Noppen, 1997b	F		Counsellor	36	11.0	22.0	5.0	12	Compl.
van Noppen, 1998a	G			10	10.0	15.0	19.4	25	Compl.
van Noppen, 1998b	G			10	10.0	15.0	16.6	25	Compl.
Warren, 2001	I	1	Psychol.	16	16.4	16.4	26.9		Compl.
Wetzel, 1999	I	28	Psychol.	3	52.0	52.0		12	Compl.
Vogel, 2004a	I	3	Psychol.	10	20.0	24.0	6.3	12	ITT
Vogel, 2004b	I	3	Psychol.	10	20.0	24.0	36.8	12	ITT

Note: Empty cells within each column mean that the information on this variable was not provided. Treatment format: G = group, I = individual. Profession: Mixed = different professions in the treatment team, Psychol. = psychologists. Statistical analysis: Compl. = completers only, ITT = intention to treat analysis.

Table 3
Within-group effect size (Hedges' *g*) of the primary (OCD) and secondary (depression) effect measures with treatment condition as unit of analysis.

Time point	*k*	*g*-value	95% CI	*z*-value	Q-value	I^2 (%)
Primary measure						
Post	38	2.12	1.85–2.39	15.55[c]	323.5[c]	88.6
Follow-up	28	2.30	1.94–2.66	12.49[c]	243.2[c]	88.9
Secondary measure						
Post	25	0.76	0.65–0.86	14.07[c]	49.8[b]	51.8
Follow-up	18	0.73	0.60–0.86	10.97[c]	29.7[a]	42.8

Note: k = number of treatment conditions. a = $p < 0.05$, b = $p < 0.001$, c = $p < 0.0001$.

provided follow-up data and the mean number of months since post-assessment for these were 15.3 (*SD* 18.7) with a range of 3–96 months. Finally, 39.5% of the conditions were evaluated statistically using intent-to-treat and 60.5% using completer analysis.

3.2. Methodological data

3.2.1. Methodology ratings

The research methodology score (% of maximum possible score for the individual study) had a mean of 52.1 (*SD* 10.8), which corresponds to a raw score of 22.9 points. Restricting the analysis to RCTs only yielded a mean of 59.0 (*SD* 9.1).

3.2.2. Risk-of-bias

The risk-of-bias (RoB) classification is presented in the supplement S7. Among the RCTs 1 study had a low RoB, 4 had some concerns, and 3 (38%) had a high RoB. Regarding the NRSI and pre-post studies 12 had a moderate and 10 (45%) had a high RoB. It is commonly recommended to restrict the analysis to studies with low or moderate risk-of-bias (e.g., Sterne et al., 2019). However, due to the small number of studies with low risk-of bias we determined to run sub-group analyses on high and low/moderate risk of bias and examine whether these made a difference to the outcomes.

3.3. Meta-analysis

3.3.1. Attrition

With treatment condition (k = 34) as the unit of analysis the attrition rate was 15.2% (95% CI 11.7–19.4, $z = 11.44$, $p < 0.0001$), which was significantly heterogeneous (Q = 89.1, $p = 0.0001$, $I^2 = 63\%$). There was no significant difference between RCTs (13.9%) and pre-post trials

Table 4
Rates of remission at post and follow-up with treatment condition as unit of analysis.

Time point	*k*	Percent	95% CI	*z*-value†	Q-value	I^2 (%)
Post	24	59.2	52.3–65.7	2.62[a]	71.8[b]	67.9
Follow-up	19	56.5	47.6–65.0	1.44	68.6[b]	73.8

Note: k = number of treatment conditions. a = $p < 0.01$, b = $p < 0.0001$, † Test if significantly different from 50%.

Table 5
Subgroup analysis of the effect size and remission rate at post-treatment.

Variable	Effect size			Remission		
	k	*g*	95% CI	*k*	*g*	95% CI
Type of study (Qb = 1.42, $p = 0.23$)*				(Qb = 0.03, $p = 0.87$)*		
RCT	13	2.33	1.93–2.74	10	58.3	51.3–67.4
Open trial	25	2.01	1.68–2.34	14	59.6	45.1–64.3
Statistical analysis (Qb = 5.74 $p = 0.017$)*				(Qb = 0.12, $p = 0.73$)*		
Intent-to-treat	15	2.50	2.06–2.94	13	59.7	49.6–69.0
Treatment completers	23	1.85	1.55–2.15	11	57.3	48.6–65.7
Risk-of-Bias (Qb = 0.41, $p = 0.52$)*				(Qb = 1.37, $p = 0.24$)*		
Moderate	25	2.14	1.78–2.50	16	61.5	52.2–70.0
High	12	1.97	1.58–2.36	7	53.0	42.0–63.7
Format (Qb = 3.29, $p = 0.07$)*				(Qb = 5.37, $p = 0.020$)*		
Individual	30	2.33	1.92–2.74	20	62.6	55.8–68.9
Group	6	1.58	0.87–2.30	3	40.5	25.1–58.0
Therapist profession (Qb = 14.27, $p = 0.003$)*						
Clinical psychologist	22	2.46	2.03–2.88			
Psychiatrist	3	2.11	1.55–2.68			
Nurse	3	1.76	1.23–2.28			
Continent (Qb = 15.33, $p = 0.0001$)*				(Qb = 6.80, $p = 0.009$)*		
Europe	22	2.51	2.12–2.90	17	64.2	56.4–71.4
North America	14	1.51	1.19–1.82	6	44.1	31.9–57.1

Note: *k* = number of treatment conditions, Qb = Q between subgroups. * The statistic in parenthesis tests if the subgroups within the individual category differ significantly from each other. Remission rate for the category therapist profession could not be calculated as only one subgroup had ≥3 studies.

(15.9%) in this respect.

3.3.2. Primary continuous measure

Table 3 presents the ESs for the primary continuous measures across all studies at post-treatment and follow-up assessment, which was carried out on average 15 months after post-treatment assessment. At post-treatment the mean ES was very large (2.12) and significantly heterogeneous, as indicated by the Q- and I^2-values. At follow-up, the ES (2.30) had increased somewhat ($Q_{between}$ (1 df) = 0.64, $p = 0.43$) compared to post-assessment, and was significantly heterogeneous.

Publication bias. Egger's regression intercept yielded a significant *t*-value (3.98, $p = 0.001$). The Duval and Tweedie trim-and-fill method suggested trimming 14 studies, which would have reduced the *g*-value to 1.57 (95% CI 1.29–1.85). Thus, publication bias is a problem regarding within-group ES for these effectiveness studies.

3.3.3. Remission

The remission rates at post-treatment and follow-up assessment are shown in Table 4. At post-treatment, the overall remission rate was 59.2%, which was significantly heterogeneous. At follow-up the overall remission rate was 56.5%, which did not differ significantly from the post-treatment rate.

Publication bias. Egger's regression intercept yielded a non-significant *t*-value (0.22, $p = 0.83$). The trim-and-fill method suggested trimming 1 study, which would have reduced the mean remission rate to 58.2% (95% CI 51.0–65.1). Thus, publication bias does not seem to be a problem for the remission rate.

3.3.4. Moderator analyses

Since the mean ES (Table 3) and mean remission rate (Table 4) were significantly heterogeneous, moderator analyses were carried out. Table 5 presents the results for the categorical variables with ES in the left and remission in the right column. RCTs and the NRSI and pre-post trials yielded similar mean ESs and remission rates. Regarding statistical analysis, studies using ITT gave higher ES than studies using completer analysis, which is encouraging, even if it does not reach the Holm-

Table 6
Meta-regression analysis of the effect size and remission rate at post-treatment.

Variable	Effect size				Remission			
	k	Point	*z*-value	*p*-value	*k*	Point	*z*-value	*p*-value
Pre-treatment severity	38	6.430	4.22	0.00001*	23	15.72	3.67	0.0002*
Methodology score	38	0.0368	3.03	0.0024*	23	−0.0068	−0.35	0.729
Percent attrition	34	−0.0312	−2.56	0.0106	22	−0.0312	−2.09	0.0366
Number of sessions	34	−0.0374	−2.48	0.0133	21	−0.0146	−0.49	0.624
Mean age	35	−0.1071	−2.28	0.0225	23	−0.1585	−3.55	0.0004*
Percent females	33	0.0153	1.33	0.183	22	−0.0082	−0.63	0.530
Percent medicated	29	−0.0083	−0.89	0.375	19	−0.0026	−0.29	0.775
Hours of treatment	34	−0.0080	−0.51	0.607	21	0.0072	0.26	0.796
# of participants	38	0.000	0.01	0.99	23	0.0007	0.18	0.856

Note: k = number of treatment conditions, Point = point estimate. * Significant using the Holm-Bonferroni orrection.

Table 7
Some background and treatment data (M and SD) for effectiveness and efficacy studies.

Study type	k	Age (years)	% females	Severity %	% comorbidity	% medicated	Tx time	% attrition
		$p = 0.64$	$p = 0.47$	$p = 0.15$	$p = 0.45$	$p = 0.36$	$p = 0.63$	$p = 0.73$
Effectiveness	39	33.7 (3.0)	59.6 (13.0)	61.0 (7.2)	55.1 (17.9)	57.9 (18.6)	19.6 (9.7)	15.1 (11.4)
Efficacy	53	34.1 (4.0)	57.5 (14.2)	63.2 (7.5)	58.0 (15.1)	51.8 (20.2)	20.7 (11.2)	14.3 (9.6)

Note: k = number of treatment conditions, Severity = percentage of the maximum score on the primary outcome measure. % Comorbidity = proportion having any psychiatric comorbid disorder at inclusion, % Medicated = proportion on any psychotropic medication at inclusion, Tx time = number of 60 min therapy hours, % Attrition = proportion dropping out of those participating in at least one therapy session.

Table 8
Effect sizes (Hedges' g) for effectiveness and efficacy studies at post and follow-up assessment.

Time point	Study type	*k*	*g*-value	95% CI	*z*-value	Qb†	*p*-value
Post	Effectiveness	38	2.12	1.85–2.39	15.55[a]	0.001	0.98
	Efficacy	54	2.13	1.89–2.36	17.78[a]		
Follow-up	Effectiveness	28	2.30	1.94–2.66	12.49[a]	0.74	0.39
	Efficacy	42	2.11	1.87–2.35	17.23[a]		

Note: k = number of comparisons. a = $p < 0.0001$. Qb = Q between, † Comparison Effectiveness vs. Efficacy studies.

Table 9
Remission rates for effectiveness and efficacy studies at post and follow-up assessment.

Time point	Study type	*k*	Percent	95% CI	*z*-value*	Qb†	*p*-value
Post	Effectiveness	25	59.2	52.8–65.4	2.78[b]	11.78	0.001
	Efficacy	27	44.0	38.4–49.8	−2.04[a]		
Follow-up	Effectiveness	20	57.0	48.6–65.0	1.64	5.06	0.024
	Efficacy	18	43.9	36.5–51.6	−1.55		

Note: k = number of comparisons. a = $p < 0.05$, b = $p < 0.01$. * Test if significantly different from 50%. Qb = Q between, † Comparison Effectiveness vs. Efficacy studies.

Bonferroni correction. Regarding risk-of-bias there was no significant difference between studies with high and moderate RoB, neither on ES nor on remission rate. When it comes to treatment format, individual therapy yielded marginally higher ES and remission rate than group therapy. Therapist profession also showed a significant difference for ES. When this was followed up by pairwise comparisons we found that clinical psychologists yielded significantly higher ES than social workers ($Q_{between}$ (1 df) = 13.45, $p < 0.0001$) and psychiatric nurses ($Q_{between}$ (1 df) = 4.08, $p < 0.043$). Finally, the continent at which the study was done also affected ES and remission rate significantly. For both measures, studies done in Europe gave significantly better effects than North American studies.

Table 10
Effect sizes and remission rates for randomized controlled studies only at post and follow-up assessment in RCTs only.

Study type	*k*	ES	95% CI	*z*-value	Qb†	*p*-value
g-value at post-treatment						
Effectiveness	13	2.33	1.92–2.74	11.20[b]	0.72	0.40
Efficacy	54	2.13	1.89–2.36	17.78[b]		
g-value at follow-up						
Effectiveness	11	2.59	1.98–3.19	8.39[b]	2.04	0.15
Efficacy	42	2.11	1.87–2.35	17.23[b]		
Remission rate at post-treatment						
Effectiveness	12	58.7	48.6–68.1	1.69	6.12	0.013
Efficacy	27	44.0	38.4–49.8	−2.04[a]		
Remission rate at follow-up						
Effectiveness	10	56.7	45.8–67.0	1.21	3.54	0.06
Efficacy	18	43.9	36.5–51.6	−1.55		

Note: k = number of comparisons. a = $p < 0.05$, b = $p < 0.0001$. Qb = Q between, † Comparison Effectiveness vs. Efficacy studies.

Continuous variables on which at least 70% of the studies provided information were analyzed with the meta-regression module in the CMA program using the random effects analyses and the results are presented in Table 6. For both ES (left column) and remission rate (right column), pre-treatment severity was a significant positive moderator, i.e., higher severity before the start of treatment was associated with larger ES and higher remission rate. Regarding ES, methodology score was also a significant positive moderator, i.e., more stringent research methodology was associated with higher ES, which is encouraging. For remission, mean age of the sample was a significant negative moderator, i.e. studies with higher mean age were associated with lower remission rates.

Since the cut-off score for remission varied from 7 to -16 on YBOCS we tested this as moderator and found that it was significantly negative ($k = 21$, point estimate = −0.1847, $z = -3.19$., $p = 0.0014$), i.e., more lenient cut-off scores were associated with lower remission rates. To shed more light on this counter-intuitive outcome we classified cut-off scores of 7–12 as stringent and 14–16 as lenient, and did a subgroup analysis. This showed a significant difference in remission rate ($Q_{between}$ (1 df) = 21.61, $p < 0.0001$) between the studies with stringent (68.9%) and those with lenient cut-off scores (41.8%).

3.3.5. Secondary continuous measure

Table 3 (lower part) presents the ESs for the most commonly reported secondary measure, i.e. depressive symptoms, at post-treatment and follow-up. The mean ES (0.76) was close to Cohen's rule-of-thumb for a large effect (0.80) and similar at follow-up.

3.4. Efficacy-effectiveness comparison

3.4.1. Background and treatment variables

Table 7 displays a comparison of effectiveness and efficacy studies on some background and treatment variables. There were no significant differences between the two types of studies for any of the variables.

Thus, the samples of patients in the effectiveness studies are quite similar to those in efficacy studies on the variables for which enough studies provided information to make a statistical comparison relevant.

3.4.2. Effect size on primary outcome measure

Table 8 shows the ES for the two types of studies. At post-treatment as well as follow-up assessment, the mean ESs were very large for both effectiveness and efficacy studies, with no significant difference between them.

3.4.3. Remission

Table 9 displays the remission rates for the two categories of studies. At post-treatment, effectiveness studies (59.2%) had a significantly higher mean remission rate than efficacy studies (44.0%). At follow-up, the remission rates were basically maintained for both categories of studies and the difference between them was still significant.

3.4.4. Comparison of RCTs only

It is possible that the results presented in Tables 8 and 9 may have been unduly affected by NRSI and pre-post trials. In order to test this possibility, we repeated the analyses with only RCTs, and the results are presented in Table 10. There was still no significant difference between effectiveness and efficacy studies for ES. Regarding remission, the effectiveness studies yielded a significantly higher rate than the efficacy studies at post-assessment, whereas the difference at follow-up was marginally significant in favor of effectiveness studies. However, these analyses were based on eight RCTs only so results need to be interpreted with caution.

We also calculated the between group effect size of CBT compared to another treatment in RCTs. Since the effectiveness studies in this meta-analysis only compared CBT with another active psychological treatment we only included efficacy studies with such comparison conditions. Thus, studies having drug treatment, combination of drug and CBT, waitlist, or self-help conditions were excluded in the calculation of between-group effect size. At post-assessment effectiveness studies yielded a small ES ($k = 6$, $g = 0.29$, 95% CI 0.03–0.54, $p = 0.03$) and efficacy studies an even smaller ES ($k = 18$, $g = 0.15$, 95% CI -0.01–0.31, $p = 0.06$). The difference was not significant (Qb = 0.71, $p = 0.40$). At follow-up assessment the effect sizes were even smaller; effectiveness $g = 0.14$ and efficacy $g = 0.02$, with no significant difference between the categories (Qb = 0.38, $p = 0.54$).

4. Discussion

The primary aim of this meta-analysis was to examine the effectiveness of CBT for OCD in routine clinical care regarding the primary anxiety measures and remission, as well as a secondary measure of depression. On the measure of OCD-severity (Table 3), the ES was very large (2.12) at post-treatment and was maintained at follow-up (2.30). This result corroborates the finding of 1.32 by Stewart and Chambless (2009) and the 1.46 reported by Hans and Hiller (2013). The effect sizes are also very similar to the ones found by Öst et al. (2015) in a meta-analysis of primarily efficacy studies in OCD: ERP 2.06, CT 2.21, and CBT 1.90.

Regarding remission (Table 4), we found a post-treatment rate of 59.2% and a follow-up rate of 56.5%; thus, a small non-significant deterioration. Remission rates were not reported by Stewart and Chambless (2009) or Hans and Hiller (2013). However, Öst et al. (2015) found the following remission rates: ERP 50.0%, CT 51.6%, and CBT 43.4%. Thus, the remission rates of effectiveness studies in the current meta-analysis are at least as good as these for mainly efficacy studies.

Concerning the secondary measure, depression (Table 3), we found an almost large ES of 0.76 at post-assessment and 0.73 at follow-up. This finding corroborates the 0.89 reported by Stewart and Chambless (2009) and the 0.66 found by Hans and Hiller (2013). Thus, all three meta-analyses found lower ES for change in depression than in OCD-severity. This is logical because the variants of CBT tested in the included studies focus completely on the OCD problem behaviors and not on comorbid disorders like depression. Since not all OCD patients are depressed before treatment, the samples' mean pre-treatment scores on depression measures are not as elevated as the scores on OCD, which means that there is much less room for improvement, and the ES becomes smaller. However, it is encouraging to see an effect on depression without it being targeted in the treatments.

The second aim was to evaluate the methodological quality and risk-of-bias in the effectiveness studies, and investigate potential moderators of treatment outcome. The mean POMRS score was 52.1%, corresponding to a raw score of 22.9 points, which is virtually identical to the 23.0 we found in our meta-analysis on efficacy studies in OCD (Öst et al., 2015). Regarding risk-of-bias we found that 38% of the RCTs and 45% of the NRSI and pre-post studies had a high risk of bias. When high and low/moderate RoB-studies were compared in a sub-group analysis there was no significant difference between these categories on effect size or remission rate (high risk studies even had nominally lower effects than low/moderate studies). However, the relatively high proportion of high risk studies means that the results should be interpreted with caution. With this in mind, we further discuss the results from the moderator analyses.

Moderator analyses of categorical variables were done with sub-group analysis (Table 5). Encouragingly, from a methodological point of view, there was no significant difference in ES or remission rate between NRSI and pre-post trials and RCTs on the one hand, or completer and intent-to-treat analysis on the other. The same lack of differences was found in a meta-analysis of effectiveness studies of CBT for internalizing (Wergeland et al., 2021) and one of externalizing (Riise, Wergeland, Njardvík, & Öst, 2021) disorders in youth, as well as one on anxiety disorders in adults (Öst et al., submitted). Regarding treatment format, we did not find a significant difference in ES between individual (2.33) and group (1.58) treatment, albeit with a trend favoring individual therapy. In their meta-analysis, Hans and Hiller (2013) found that individual (1.87) was significantly better than group (1.09) treatment in the subset of OCD studies. Finally, the continent where the study was done was a significant moderator; both ES and remission rate showed better outcome for European than North American studies. One possible explanation to this result is that ERP was developed by Victor Meyer at Middlesex hospital in London in the 1960's (Meyer, 1966) and the UK as well as other European countries have a long tradition of OCD-treatment.

Moderator analyses of continuous variables were done with meta-regression (Table 6). For both ES and remission rate, pre-treatment severity was a positive moderator, which is logical since there is more room for improvement for studies having higher OCD-symptom scores before the start of treatment. The same result was found in the Öst et al., (submitted) meta-analysis of effectiveness studies in adult anxiety disorders, in the Wergeland et al. (2021) article on internalizing disorders, and the Riise et al. (2021) paper on externalizing disorders in youth. Another encouraging finding is that the methodology score was a significant positive moderator of ES. The same finding of methodology score as a positive moderator was obtained by Öst, Riise, Wergeland, Hansen, and Kvale (2016), Finnes et al. (2019), Temple et al. (2020), and Öst et al., (submitted). Regarding remission rate, the mean age of the sample was a negative moderator, which corroborates the findings of Riise et al. (2021) for externalizing disorders in children and adolescents. However, mean age was a positive predictor in the Wergeland et al. (2021) meta-analysis of internalizing disorders, so age does not seem to be a consistent moderator.

We also found that the cut-off score on YBOCS to define remission varied from 7 to -16 and was a significant negative moderator of remission rate, i.e., more stringent cut-off scores were associated with higher remission rates. A further subgroup analysis showed that studies with stringent cut-off scores (7–12) had a significantly higher remission rate (69%) than those with lenient (14–16) cut-off scores (42%). This

seems counter-intuitive since it is easier to achieve a more lenient (higher) score. It is also astonishing that two studies used the same score (16), which denotes moderate OCD-severity and most studies nowadays use as inclusion criterion in a treatment outcome study. To avoid this problem future studies should adopt the international consensus criteria described by Mataix-Cols et al. (2016).

The mean proportion of patients declining to participate in the treatment they were offered was 11.1% on average, but should be interpreted with caution since 37% of the studies did not provide information on this issue. The attrition rate was 15.2%. These proportions are similar to what Öst et al. (2015) reported; a declining rate of 15.0% and the following attrition rates: ERP 19.1%, CT 11.4%, and CBT 15.5%. In comparison, a very large meta-analysis of 669 psychotherapy studies (Swift & Greenberg, 2012) found a mean dropout rate of 19.7% but somewhat lower for CBT studies, 18.4%.

The third aim was to examine how CBT for OCD delivered in routine clinical care fare in comparison to efficacy studies. As an initial step we compared effectiveness and efficacy studies on seven background and treatment variables (Table 7) and there were no significant differences. This means that the samples of patients in effectiveness studies were not less, or more, severe than the samples of efficacy studies, which makes for a fair comparison of treatment effects. On the ES measure the two types of studies had very large and almost the same means, 2.12 for effectiveness and 2.13 for efficacy studies at post-treatment. At the follow-up assessment, effectiveness studies showed a small improvement and efficacy studies had basically the same mean as at post. These equal effects for ES corroborate the findings in Stewart and Chambless (2009), Hans and Hiller (2013) in adults with OCD, Öst et al., (submitted) for anxiety disorders in adults, and Wergeland et al. (2021) in youth with internalizing disorders. Regarding remission rate effectiveness studies had a significantly higher proportion of remitted patients (59.2%) than efficacy studies (44.0%) at post-treatment and follow-up assessment, 57.0% vs. 43.9%, respectively. Previous meta-analyses in youth with OCD (Wergeland et al., 2021) and externalizing disorders (Riise et al., 2021) found no significant differences between effectiveness and efficacy studies.

We repeated these analyses with only RCTs for both types of studies and found the same non-significant differences for ES and a significantly higher post-treatment remission rate for effectiveness studies (57.8% vs. 44.0%) and a marginal difference at follow-up (56.7% vs. 43.9%). However, our meta-analysis included eight RCTs only, and more meta-analyses on this subject are needed before firm conclusions can be drawn.

The present meta-analysis has several strong methodological components. A moderate number of studies/treatment conditions (61/92) meant that a power analysis indicated a very high power (95.8%) to detect a small effect size of 0.20. Pairs of researchers screened abstracts, read full-text articles, and extracted information from the included studies, and disparities were solved in consensus discussions. One of the authors rated the included studies for methodological quality and risk-of-bias, and 20% (randomly selected) studies were rated independently by another.

There are also some limitations to consider. Only peer-reviewed published or in press studies in English language journals were included. Studies published in other languages could have provided additional information about the effectiveness of CBT for OCD in adults. However, Hans and Hiller (2013) did not use a language restriction and included studies in Dutch, English, French, German, Italian, Norwegian, and Spanish, and got basically the same results as we did. Furthermore, including only published studies could be viewed as a limitation but our pool of studies spanned 25 years. Including unpublished studies could have introduced bias as it could have been easier to identify unpublished studies from more recent compared to earlier years. Furthermore, we cannot rule out that there may be differences between the effectiveness and efficacy studies on other background variables that may moderate treatment outcome, since we used the criterion that at least 70% of the effectiveness studies in our meta-analysis had to provide information on a variable to be included in the analyses. Other limitations include the considerable risk-of-bias in a high proportion of the studies. RCTs are the gold standard for assessing psychotherapy interventions However, they are more difficult and resource demanding to conduct. Non-randomized studies of interventions have proliferated in recent years and can provide important information on “real world” effectiveness. However, both NRSI and pre-post studies have a higher risk of bias compared to RCT, which can influence effect sizes and remission rates. Thus, the results of this meta-analyses need to be interpreted with the considerable risk-of-bias in mind. Also, the use of pre-post standardized mean difference to indicate treatment effects in meta-analyses has been problematized, as it can contribute to biased outcomes. However, for evaluation of improvement found in routine clinical care compared with improvement found in efficacy studies, these analyses are still considered informative (Cuijpers, Weitz, Cristea, & Twisk, 2017).

Future research should focus on pragmatic trials with large sample sizes to test CBT interventions in routine clinical settings in order to maximize applicability and generalizability. This could be done as multi-site studies or even multi-country studies as the Scandinavian Nord-LOTS study of CBT for OCD in youth which was done in Denmark, Sweden, and Norway (Torp et al., 2015).

Moreover, this type of meta-analysis comparing CBT effectiveness and efficacy studies should be done for depression, bipolar disorder, schizophrenia, eating disorders, alcohol abuse, sleep disorders, and personality disorder, and for post-traumatic stress disorder, which belonged to the anxiety disorders in DSM-IV but was moved to its own chapter in DSM-5.

In conclusion, our findings demonstrate that when therapists trained in CBT, working in routine clinical care settings, apply cognitive behavioral interventions which are evaluated as well-established or probably efficacious for OCD, the within-group effect sizes and remission rates are at least equal to the effects obtained in university settings, and the effects are maintained at follow-up on average 15 months later.

Declaration of competing interest

All authors have declared that they have no competing or potential conflicts of interest.

CRediT authorship contribution statement

Lars-Göran Öst: designed the study, wrote the protocol and the coding schemes, rated the studies, meta-analyzed the included effectiveness and comparison efficacy studies, wrote the first draft of the manuscript, and reviewed and edited the manuscript. **Pia Enebrink:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Anna Finnes:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Ata Ghaderi:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Audun Havnen:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Gerd Kvale:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Sigrid Salomonsson:** screened the studies, read the full-text articles, extracted data from the included studies, and reviewed and edited the manuscript. **Gro Janne Wergeland:** conducted the literature searches in collaboration with an academic librarian, rated the studies for methodological quality and risk-of-bias, and reviewed and edited the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104170.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Press.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th ed.). Washington, DC: American Psychiatric Publishing.

Australian Psychological Society (APS). (2018). *Evidence-based psychological interventions in the treatment of mental disorders: A literature review* (4th ed.) Retrieved from https://www.psychology.org.au.

Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E., & Vos, T. (2013). Meta-analysis of prevalence. *Journal of Epidemiology & Community Health, 67*, 974–978. https://doi.org/10.1136/jech-2013-203104

Barker, T. H., Migliavaca, C. B., Stein, C., Colpani, V., Falavigna, M., Aromataris, E., et al. (2021). Conducting proportional meta-analysis in different types of systematic reviews: A guide for synthesisers of evidence. *BMC Medical Research Methodology, 21*, 189. https://doi.org/10.1186/s12874-021-01381-z

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine, 44*, 2363–2374. https://doi.org/10.1017/S0033291713003243

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis*. Englewood, NJ: Biostat. version 3.3.070.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences, 26*, 364–368. https://doi.org/10.1017/S2045796016000809

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

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1998). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 316*(7129), 469. https://doi.org/10.1136/bmj.316.7129.469

Finnes, A., Enebrink, P., Ghaderi, A., Dahl, J., Nager, A., & Öst, L.-G. (2019). Psychological treatments for return to work in individuals on sickness absence due to common mental disorders or musculoskeletal disorders: A systematic review and meta-analysis of randomized controlled trials. *International Archives of Occupational and Environmental Health, 92*, 273–293. https://doi.org/10.1007/s00420-018-1380-x

Gonzales, J. J., & Chambers, D. A. (2002). The tangled and thorny path of science to practice: Tensions in interpreting and applying "evidence. *Clinical Psychology: Science and Practice, 9*, 204–209. https://doi.org/10.1093/clipsy.9.2.204

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., et al. (1989). The Yale-Brown obsessive compulsive scale. I. Development, use, and reliability. *Archives of General Psychiatry, 46*, 1006–1011. https://doi.org/10.1001/archpsyc.1989.01810110048007

Hans, E., & Hiller, W. (2013). A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clinical Psychology Review, 33*, 954–964. https://doi.org/10.1016/j.cpr.2013.07.003

Hofmeijer-Sevink, M. K., van Oppen, P., van Megen, H. J., Batelaan, N. M., Cath, D. C., van der Wee, N., et al. (2013). Clinical relevance of comorbidity in obsessive compulsive disorder: The Netherlands OCD association study. *Journal of Affective Disorders, 150*, 847–854. https://doi.org/10.1016/j.jad.2013.03.014

Hunsley, J. (2007). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice, 38*(2), 113–121. https://doi.org/10.1037/0735-7028.38.2.113

Jaccard, J., & Guilamo-Ramos, V. (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Issues and recommendations. *Journal of Clinical Child and Adolescent Psychology, 31*, 130–146. https://doi.org/10.1207/S15374424JCCP310115

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of *DSM-IV* disorders in the national comorbidity Survey replication. *Archives of General Psychiatry, 62*, 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., et al. (2011). Development of lifetime comorbidity in the world health organization world mental health Surveys. *Archives of General Psychiatry, 68*, 90–100. https://doi.org/10.1001/archgenpsychiatry.2010.180

Koran, L. M., Thienemann, M. L., & Davenport, R. (1996). Quality of life for patients with obsessive–compulsive disorder. *American Journal of Psychiatry, 153*, 783–788. https://doi.org/10.1176/ajp.153.6.783

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lee, E. B., Bistricky, S., Milam, A., Wetterneck, C. T., & Björgvinsson, T. (2016). Thought control strategies and symptom dimensions in obsessive-compulsive disorder: Associations with treatment outcome. *Journal of Cognitive Psychotherapy, 30*, 177–189. https://doi.org/10.1891/0889-8391.30.3.177

Leeuwerik, T., Cavanagh, K., & Strauss, C. (2019). Patient adherence to cognitive behavioural therapy for obsessive compulsive disorder: A systematic review and meta-analysis. *Journal of Anxiety Disorders, 68*, Article 103135. https://doi.org/10.1016/j.janxdis.2019.102135

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *British Medical Journal, 339*, b2700. https://doi.org/10.1136/bmj.b2700

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc.

Lovell, K., Cox, D., Haddock, G., Jones, C., Raines, D., Garvey, R., et al. (2006). Telephone administered cognitive behaviour therapy for treatment of obsessive compulsive disorder: Randomised controlled non-inferiority trial. *BMJ*. https://doi.org/10.1136/bmj.38940.355602.80

Macy, A. S., Theo, J. N., Kaufmann, S., Ghazzaoui, R. B., Pawlowski, P. A., Fakhry, H. I., et al. (2013). Quality of life in obsessive compulsive disorder. *CNS Spectrums, 18*, 21–33. https://doi.org/10.1017/S1092852912000697

Mataix-Cols, D., Fernandez de la Cruz, L., Nordsletten, A. E., Lenhard, F., Isomura, K., & Simpson, H. B. (2016). Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. *World Psychiatry, 15*, 80–81. https://doi.org/10.1002/wps.20299

Meyer, V. (1966). Modification of expectations in cases with obsessional rituals. *Behaviour Research and Therapy, 4*, 273–280.

National Institute for Health and Care Excellence. (2013). *Obsessive-compulsive disorder. Evidence update*. London: The British Psychological Society and The Royal College of Psychiatrists.

Öst, L.-G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy, 46*, 296–321. https://doi.org/10.1016/j.brat.2007.12.005

Öst, L-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., et al (submitted). Cognitive behavior therapy for adult anxiety disorders in routine clinical care: A systematic review and meta-analysis.

Öst, L.-G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clinical Psychology Review, 40*, 156–169. https://doi.org/10.1016/j.cpr.2015.06.003

Öst, L.-G., & Ollendick, T. H. (2017). Brief, intensive and concentrated cognitive behavioral treatments for anxiety disorders in children: A systematic review and meta-analysis. *Behaviour Research and Therapy, 97*, 134–145. https://doi.org/10.1016/j.brat.2017.07.008

Öst, L.-G., Riise, E. N., Wergeland, G. J., Hansen, B., & Kvale, G. (2016). Cognitive behavioral and pharmacological treatments of OCD in children: A systematic review and meta-analysis. *Journal of Anxiety Disorders, 43*, 58–69. https://doi.org/10.1016/j.janxdis.2016.08.003

Riise, E. N., Wergeland, G. J. H., Njardvík, U., & Öst, L.-G. (2021). Cognitive behavior therapy for externalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology Review, 83*, Article 101954. https://doi.org/10.1016/j.cpr.2020.101954

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the national comorbidity Survey replication. *Molecular Psychiatry, 15*, 53–63. https://doi.org/10.1038/mp.2008.94

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126*, 512–529. https://doi.org/10.1037/0033-2909.126.4.512

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., & Henry, D. A. (2017). Amstar 2: A critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both. *British Medical Journal, 358*, j4008. https://doi.org/10.1136/bmj.j4008

Sterne, J. A. C., Hernan, M. A., Reeves, B. C., Savovic, J., Berkman, N. D., Viswanathan, M., et al. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised atudies of intervenyion. *British Medical Journal, 355*, i4919. https://doi.org/10.1136/bmj.i4919

Sterne, J. A. C., Savovic, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., et al. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *British Medical Journal, 366*, l4898. https://doi.org/10.1136/bmj.l4898

Stewart, R. E., & Chambless, D. L. (2009). Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 77*, 595–606. https://doi.org/10.1037/a0016032

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*, 547–559. https://doi.org/10.1037/a0028226

Temple, J., Salmon, P., Smith, C. T., Huntley, C. D., Byrne, A., & Fisher, P. L. (2020). The questionable efficacy of manualized psychological treatments for distressed breast cancer patients: An individual patient data meta-analysis. *Clinical Psychology Review, 80*, Article 101883. https://doi.org/10.1016/j.cpr.2020.101883

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice, 22*, 317–338. doi.org/10.1037/h0101729.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*, 215–247. https://doi.org/10.3102/1076998609346961

Warren, R., & Thomas, J. C. (2001). Cognitive-behavior therapy for obsessive-compulsive disorder in private practice: An effectiveness study. *Journal of Anxiety Disorders, 15*, 277–285.

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *Journal of Clinical Child and Adolescent Psychology, 42*(2), 274–286. https://doi.org/10.1080/15374416.2013.764824

Wergeland, G., Riise, E. N., & Öst, L.-G. (2021). Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology Review, 83*. https://doi.org/10.1016/j.cpr.2020.101918, 101918.

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*, 875–899. https://doi.org/10.1037//0022-006X.69.6.875

Wetzel, C., Bents, H., & Florin, I. (1999). High-density exposure therapy for obsessive-compulsive inpatients: A 1-year follow-up. *Psychotherapy and Psychosomatics, 68*, 186–192. https://doi.org/10.1159/000012331