Clinical Psychology Review 83 (2021) 101918

Contents lists available at ScienceDirect

Clinical Psychology Review

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

Review

Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis

Gro Janne H. Wergeland[a,b,*], Eili N. Riise[c], Lars-Göran Öst[d,e]

[a] Department of Child and Adolescent Psychiatry, Division of Psychiatry, Haukeland University Hospital, Jonas Lies vei 65, N-5021 Bergen, Norway
[b] Department of Clinical Medicine, Faculty of Medicine, University of Bergen, N-5021 Bergen, Norway
[c] Department of Child and Adolescent Psychiatry, District General Hospital of Førde, Svanehaugvegen 2, N-6812 Førde, Norway
[d] Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
[e] OCD-Team, Haukeland University Hospital, Jonas Lies vei 65, N-5021 Bergen, Norway

HIGHLIGHTS

- CBT for internalizing disorders treated in routine clinical care was meta-analyzed.
- Attrition during CBT was low (5–15%).
- CBT was effective in reducing symptoms and increasing remission.
- Outcome was significantly improved from post- to follow-up assessment.
- The outcomes of effectiveness studies were on a par with that of efficacy studies.

ARTICLE INFO

Keywords:
Internalizing disorders
Effectiveness
Cognitive behavior therapy
Children
Meta-analysis

ABSTRACT

Cognitive behavioral therapy (CBT) has received considerable empirical support for internalizing disorders including anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder in children and adolescents. However, there is less knowledge regarding how CBT performs when delivered in routine clinical care. A systematic review and meta-analysis was conducted of CBT for internalizing disorders in children and adolescents in routine clinical care. Ovid MEDLINE, Embase OVID, and PsycINFO were systematically searched for articles published until October 2019. The effectiveness of CBT, methodological quality, and moderators of treatment outcome were examined. The effects of CBT in routine clinical care were benchmarked by comparing with efficacy studies for the same disorders. Fifty-eight studies were included, comprising 4618 participants. Large effect sizes for outcome were detected at post-treatment (g = 1.28–2.54), and follow-up (g = 1.72–3.36). Remission rates across diagnoses ranged from 50.7% - 77.4% post-treatment, to 53.5% -83.3% at follow-up. Attrition rate across the disorders was 12.2%. Quality of the included studies was fair, and heterogeneity was high. Similarities between the effectiveness and efficacy studies were greater than the differences in outcome. CBT delivered in routine clinical care is efficacious in reducing internalizing disorders and symptoms. The outcomes are comparable with results obtained in efficacy studies.

PROSPERO registration: ID CRD42019128709.

1. Introduction

Internalizing disorders such as anxiety disorders, obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), and depressive disorders, represent as a group the most common mental health disorders among children and adolescents (Costello, Egger, & Angold, 2005; Merikangas et al., 2010). These disorders represent partly overlapping variations of emotional distress and symptom presentations in response to life stressors and difficulties (Nigatu et al., 2016). They are associated with numerous negative mental health outcomes, and can lead to psychological, cognitive, social and occupational impairments (Asselmann, Wittchen, Lieb, & Beesdo-Baum, 2018; de Lijster et al., 2018; Kertz, Petersen, & Stevens, 2019; Piacentini, Bergman, Keller, & McCracken, 2003; Trickett, Noll, & Putnam, 2011; Wu et al., 2016).

Effective and empirically supported treatments for these disorders

* Corresponding author at: Department of Child and Adolescent Psychiatry, Haukeland University Hospital, N-5021 Bergen, Norway.
E-mail address: gro.wergeland@uib.no (G.J.H. Wergeland).

https://doi.org/10.1016/j.cpr.2020.101918
Received 23 October 2019; Received in revised form 13 August 2020; Accepted 8 September 2020
Available online 23 September 2020

in children and adolescents exist, with cognitive behavioral therapy (CBT) approaches having empirical support for anxiety disorders, OCD, PTSD and depressive disorders (Dorsey et al., 2017; Freeman et al., 2018; Higa-McMillan, Francis, Rith-Najarian, & Chorpita, 2016; Weersing, Jeffreys, Do, Schwartz, & Bolano, 2017). However, the evidence for the efficacy of empirically supported treatments such as CBT rests mainly on randomized controlled trials conducted in specialized university research settings, i.e. efficacy trials. An important question is how well CBT holds up when transported and delivered in routine clinical practice (Lee, Horvath, & Hunsley, 2013; Weisz, Ugueto, Cheron, & Herren, 2013).

The question of transportability of empirically supported treatment into routine clinical practice is partly grounded in a concern that the methodological rigor of efficacy trials to maximize experimental control may sometimes decrease the external validity and limit the generalizability of the findings (La Greca, Silverman, & Lochman, 2009). Concerns about the generalizability of results from RCTs arises from the assumptions that clients, therapists and treatment contexts may all differ in important ways between research clinics and community clinics (Hunsley, 2007; Lee et al., 2013; Weisz, Ng, Rutt, Lau, & Masland, 2013; Weisz, Ugueto, et al., 2013). Participants in efficacy trials are usually subjects to more rigorous inclusion and exclusion criteria, and may therefore be more homogenous compared with clients in community clinics (Hunsley, 2007; Weisz, Ng, et al., 2013). For example, many inclusion and exclusion criteria may be applied in selecting participants for an efficacy study to minimize the influence of external factors on the treatment. Inclusion criteria are used to ensure that all participants have the condition for which the treatment has been designed (i.e., carefully diagnosed disorders), and exclusion criteria are used to exclude those who have difficulties completing the requirements for the study, e.g., language difficulties, use of psychotropic medication, comorbidity more severe than the targeted condition. Further, client expectations may be higher in specialized university research clinics compared to routine clinical care, which may further enhance motivation and outcome (Stewart & Chambless, 2009). Therapists in university research settings have access to extensive training, supervision, and treatment monitoring with an emphasis on treatment integrity, more often than their colleagues in routine clinical care (Smith et al., 2017). Furthermore, they are more likely to be experts in the delivery of the particular treatment compared with therapists in community clinics. Also, therapists in university settings typically have caseloads focused on particular disorders(s), compared with the caseloads of therapists in community clinics which typically comprise a broad array of disorders and referral problems (Hunsley, 2007; Weisz, Ng, et al., 2013; Weisz, Ugueto, et al., 2013).

Regarding treatment context, there is a discrepancy between university research clinics and community clinics. Staff and facilities are dedicated to research in the university clinic, and there are resources available to do thorough assessment, treatment monitoring, and rescheduling of missed appointments or follow-ups which is seldom the situation in the community clinics. As such, treatment programs developed and evaluated under highly controlled conditions in specialized research settings may not produce similar results when delivered in routine clinical practice (Baker-Ericzen, Hurlburt, Brookman-Frazee, Jenkins, & Hough, 2010; Weisz, Jensen-Doss, & Hawley, 2006; Weisz, Ng, et al., 2013). Thus, it is important for clinicians to know what may be expected from empirically supported treatments of internalizing disorders when delivered in routine clinical practice, and how results compare with outcomes obtained in specialized university research clinics.

Effectiveness studies are typically seen as the best option for evaluating a treatment when delivered in regular clinical practice (Hunsley, 2007). Efforts to increase external validity most commonly involve placing the treatment study within routine clinical care, using regular service providers to deliver the treatment, and including patients who are ordinary referrals to the clinics (Hunsley & Lee, 2007). Such studies can include pretest-posttest designs, quasi-experimental, or experimental designs (Stewart & Chambless, 2009). Importantly, many treatment outcome studies are not easily categorized into the two categories of efficacy or effectiveness research, but can be placed somewhere along a continuum of internal and external validity (La Greca et al., 2009). As such, both efficacy and effectiveness studies are important to better evaluate and understand the potential effect and impact of a treatment (Hunsley, Elliott, & Therrien, 2014).

A few reviews and meta-analyses on effectiveness studies in children and adolescents have been conducted. The most recent review reporting on the effectiveness of empirically supported treatments across various control conditions (e.g., active control, waiting list or usual care) with children and adolescents, included 20 studies, seven of which on internalizing disorders (anxiety, $n = 2$, OCD, n = 2, depression $n = 3$), six of which evaluated CBT (Lee et al., 2013). The review reported treatment completion and improvement rates comparable to those reported in benchmark treatment efficacy studies across the range of included disorders. Thus, the review provided encouraging preliminary evidence of the effectiveness of treatments with established efficacy. Other meta-analyses have either focused on studies comparing empirically supported treatments with a specific control condition (i.e., usual clinical care, Weisz et al., 2006; Weisz et al., 2013), or examined treatment effectiveness across a broad variety of empirically supported treatments for a given disorder (Weisz, Kuppens, et al., 2013). Since these reviews and meta-analyses, the number of effectiveness studies have increased considerably, with a marked increase in studies evaluating CBT across different cultures and countries. An update of this literature is therefore warranted.

The current meta-analysis aims to add information to the existing literature by providing a meta-analysis of the effectiveness of CBT for internalizing disorders in children and adolescents treated in routine clinical care. We selected studies in which patients are referred for treatment through usual clinical routes, treatment is delivered by clinicians for whom provision of service is a substantial part of the job, and treatment is delivered in routine clinical practice. We chose not to limit our meta-analysis to RCTs but include open trials to better capture all research conducted in routine clinical care contexts and be as comprehensive as possible. Background and treatment data, and methodological quality may give important information whether there are systematic differences between effectiveness and efficacy studies that may affect outcome. Thus, systematic evaluations of these characteristics are necessary to inform the field whether differences in effect size are attributable to the actual treatment setting and not to other variables. 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 use 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.

Our specific aims were threefold. First, to examine the effectiveness of CBT for anxiety disorders, obsessive compulsive disorder, posttraumatic stress disorder, and depressive disorders for children and adolescents. Second, to evaluate methodological quality in the effectiveness studies, and investigate potential moderators of treatment outcome. Third, to examine how the different cognitive behavior treatments delivered in routine clinical care fare in comparison with efficacy studies for the same disorders, in order to evaluate if CBT in effectiveness studies are at the same level as found in efficacy studies.

2. Method

The aims and methods of the current meta-analysis have been preregistered in the PROSPERO database with ID CRD42019128709. Two independent raters were involved during the steps of the project, except for the literature screening of title and abstract conducted by one rater

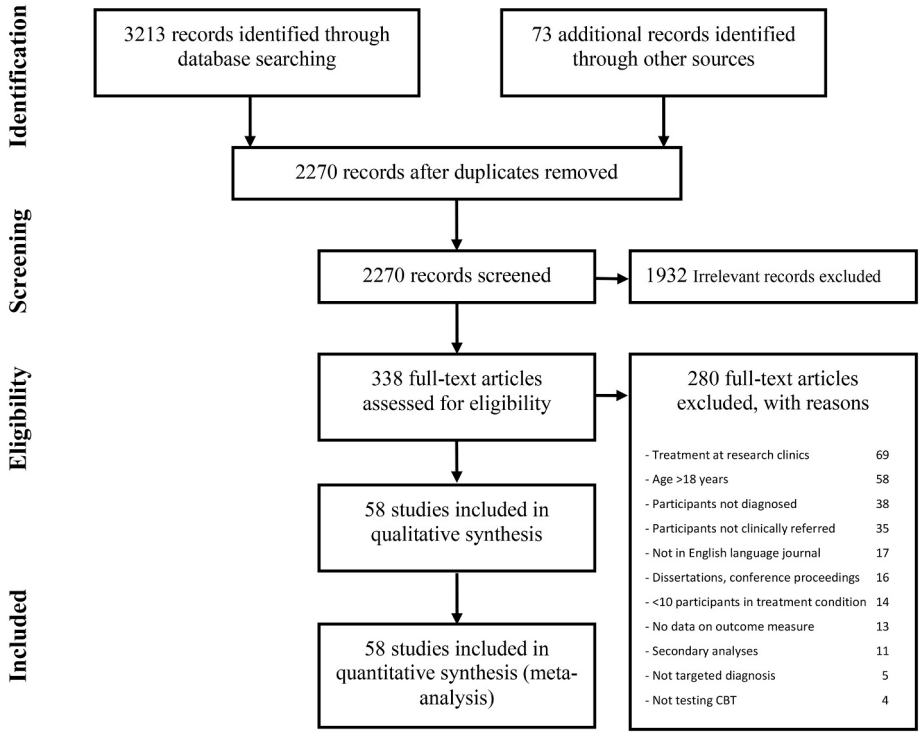

Fig. 1. Flowchart of the inclusion of studies.

only. The meta-analysis was designed according to the PICOS acronym in the following way:

- Population: children and adolescents with anxiety disorders, OCD, PTSD, and depressive disorder
- Intervention: CBT delivered in routine clinical care
- Comparison: within-group change. i.e. pre vs. post-data
- Outcome: primary continuous measure and remission
- Study design: RCTs and open trials

The meta-analysis was conducted according to PRISMA (Liberati et al., 2009) and AMSTAR 2 (Shea et al., 2017), see online Supplement S7 and S8.

2.1. Literature search

Studies were identified by a systematic and comprehensive literature search of electronic databases and scanning reference lists of articles. The search was applied to Ovid MEDLINE, Embase OVID, and PsycINFO from the start of the data bases to September 28th 2018. An updated search was done October 16th 2019. The list of search terms utilized to identify potential studies were generated by all three 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; anxiety (including the different anxiety disorders); depression (including the different depressive disorders); Obsessive compulsive disorder; OCD; Post traumatic stress disorder; PTSD; Outpatient clinics; community mental health services; effectiveness; routine care; regular care, community clinic; youth; child*; adolescent and pediatric. For full search strategy for Ovid MEDLINE, Embase OVID and PsychINFO, see the online Supplement, S1.

The first author read the titles and abstract of all the papers from this initial search to decide whether a study warranted a more detailed reading. When there was an indication of a group of patients receiving the particular cognitive-behavioral treatment in a non-university 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. Although research articles were the target of the search, review articles were also examined for additional references. Key authors were searched in databases for additional publications. Unpublished "grey" literature was not included in the present study. In total, 338 full-text articles were considered for inclusion. The final decision for article inclusion was made using a strict set of inclusion and exclusion criteria detailed below. The full text articles were read by the first and the third author and any disagreements (6% of the articles) were resolved by consensus discussion. It was determined that 58 articles were included in the present meta-analysis. In addition, we found two studies on specific phobia but it was too few to include in the meta-analysis, which was done per disorder.

2.1.1. Inclusion criteria

In order 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 an anxiety or depressive disorder or OCD according to DSM or ICD. For PTSD some proportion of subclinical cases are accepted since research indicates that they often are as severe as those fulfilling full diagnostic criteria (Gutermann et al., 2016).
3. Be testing a form of CBT, cognitive therapy (CT) or behavior therapy (BT).
4. Have participants referred for treatment through usual clinical routes.
5. Be an effectiveness study, i.e. carried out in a non-university setting such as clinical routine care or school health care.
6. Have therapists who are practicing clinicians for whom provision of service is a substantial part of the job.
7. Have a treated sample consisting of at least 10 participants.
8. Have a maximum participant age of 18.
9. Provide a measure of the primary disorder treated.

2.1.2. Exclusion criteria

1. The study is a secondary analysis of a previously published study.
2. The study is an evaluation of a service where the results for individual disorders cannot be extracted.
3. The study is not testing a form of CBT, CT, or BT.
4. The study is testing a combination of CBT and SSRI.

Fig. 1 shows a flowchart of the inclusion of studies in the present meta-analysis. For references to included studies, see online Supplement S2, and for references to studies excluded in the meta-analyses, see online Supplement S3.

2.2. Potential categorical moderators

In order to include any potential categorical or continuous moderators in the analysis we required that at least 75% 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 conditions

Type of study was either RCT or open trial. The various conditions in the RCTs were classified as CBT (various types of cognitive behavioral treatments exclusively), other forms of psychotherapy, SRI (various serotonin reuptake inhibitors), Combo (the combination of CBT and SRI), treatment as usual, and waitlist control (WLC).

2.2.2. Statistical analysis

Statistical analysis was categorized as intent-to-treat (ITT) if all randomized participants were included in the statistical analysis and completers (TC) if only the patients that completed the treatment were included.

2.2.3. Format and parental involvement

Format of therapy was classified as individual, group, family treatment, combinations of these, or Internet-based CBT. Degree of parental involvement was classified as *low* if parents were not present during sessions but informed about progress of therapy, *moderate* if parents were present during some therapy sessions full-time or only part-time of all sessions, and *high* if parents were present full-time during all therapy sessions.

2.2.4. Therapist profession

The profession that the majority of the therapists within a study belonged to was classified as clinical psychologist, child psychiatrist, social worker, nurse, or various professions.

2.2.5. Continent

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

A coding scheme and manual including the variables of interest was developed. The data extraction and categorizations were done independently by the first and the second author and any disagreements (7% of the data items) were solved after consensus discussions.

2.3. Potential continuous moderators

The following continuous measures on which at least 75% of the studies provided information were used as potential moderators: number of participants in the study, percent girls, mean age, pretreatment severity (calculated as percentage of the maximum score of the rating scale applied), number of therapists in the study, methodology score (see 2.4.), risk-of-bias score (see 2.5.), treatment weeks, number of sessions, treatment hours, treatment intensity (hours/week), and percent attrition in the study. In addition, we also extracted information on percent declining treatment, having a comorbid disorder, having a comorbid anxiety disorder, having received treatment for the principal disorder previously, currently on drug treatment, and duration of the disorder, but these variables did not reach the 75% criterion.

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). The scale consists of the following items: 1. Clarity of sample description, 2. Severity/chronicity of the disorder, 3. Representativeness of the sample, 4. Reliability of the diagnosis in question, 5. Specificity of outcome measures, 6. Reliability and validity of outcome measures, 7. Use of blind evaluators, 8. Assessor training, 9. Assignment to treatment, 10. Design, 11. Power analysis, 12. Assessment points, 13. Manualized, replicable, specific treatment programs, 14. Number of therapists, 15. Therapist training/experience, 16. Checks for treatment adherence, 17. Checks for therapist competence, 18. Control of concomitant treatments, 19. Handling of attrition, 20. Statistical analyses and presentation of results, 21. Clinical significance, 22. Equality of therapy hours (for non-WLC designs only). 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. The internal consistency of the scale was good with a McDonald's ω of 0.81. The first author was trained in the POMRS by the developer of the scale. She rated all the studies and inter-rater reliability of the scale (between the first and the third author), based on 20% randomly selected and blindly rated studies was ICC(3,1) = 0.94 (95% CI 0.81–0.98), which according to Cicchetti (1994) is excellent.

2.5. Risk-of-bias

The Cochrane Collaboration tool for assessing risk-of-bias (Higgins, Altman, & Sterne, 2011) was used, and the following domains were rated: random sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data, and selective reporting. Blinding of patients and therapists cannot be used in psychotherapy studies (Borkovec & Sibrava, 2005). A high risk-of-bias in a domain was given 1 point, an unclear risk 0.5, and a low risk 0 point. Summarizing over the five domains the total score could vary between 0 and 5, with higher scores indicating higher risk-of-bias. The first author rated the included studies and inter-rater reliability was assessed between the first and the third author based on 20% randomly selected and blindly rated studies. This yielded an intra-class correlation, ICC(3, 1) = 0.91 (95% CI 0.71–0.97), which also is excellent.

2.6. Effect size 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.

2.6.1. Remission

Below follows a description of the assessment of remission used in the different studies.

Depression: 6 out of 8 studies (75%) provided data and in 5 it was loss of principal diagnosis, whereas 1 used the proportion achieving a normal or borderline score on the Clinical Global Impression-Severity (CGI) scale.

Mixed anxiety: 20 out of 22 studies (90.9%) provided data and in all but one it was loss of principal diagnosis. The remaining study used a combination of CGI-Improvement and Severity scales.

OCD: 7 out of 10 studies (70%) provided data on remission; 3 used a cut-off score (9 or 10 points) on Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS, Scahill et al., 1997), 2 used a ≥ 35% reduction on CY-BOCS plus a post-treatment score of ≤12 points, whereas 2 used Jacobson and Truax (1991) criteria of a Reliable Change Index plus a cut-off score (9 and 12, respectively).

PTSD: Since we included studies where less than 100% fulfilled diagnostic criteria at pre-treatment remission was calculated based on the number of participants who fulfilled diagnostic criteria or the cut-off criteria for clinical range on the respective scale. Fifteen out of 18 studies (83.3%) had data on remission; 13 used loss of principal diagnosis, whereas 2 used scores in the non-clinical range on the applied primary rating scale. Overall 48 out of 58 studies (82.8%) in this meta-analysis provided data on remission.

2.6.2. Continuous rating scales

When a study named its primary outcome measure among rating scales we used that. If none was pinpointed we selected measures in the following order if available: independent assessor rating, self-report scale (for school-aged children/adolescents), and parent report scale (for pre-school children). All studies of depression, OCD, and PTSD provided data on a continuous rating scale, whereas in mixed anxiety 20 of 22 (90.9%) did so. The various rating scales used for the respective studies are described in the online Supplement S4.

2.6.3. Secondary outcome measures

Since depression and general anxiety are common comorbid disorders in the anxiety disorders and general anxiety is a common comorbidity in depression we extracted data on such variables too. However, only 26 of the 50 (52%) anxiety disorder studies had a measure of depression and 31 (62%) had a measure of general anxiety. Among the 8 depression studies only 2 (25%) had a measure of general anxiety. With these low proportions of studies providing the relevant data it is questionable if the outcome of a meta-analysis would be representative so we decided not to carry out one.

2.7. Meta-analysis

In order to obtain as large as possible a body of effectiveness studies we included both RCTs and open trials in the meta-analysis since within-group ES can be calculated from both 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. Rate of remission, with event rate as the effect measure, was analyzed using mixed effect analysis in the subgroup analysis. In this analysis a random effects model is used to combine studies within each subgroup and a fixed effects model is used to compare subgroups and yield the overall effect. When a study presented intent-to-treat (ITT) data these were used, if not completer data were used.

Before pooling the effect sizes we screened for statistical outliers, defined as being above M ± 2SD. At the post-treatment assessment three (4.6%) of the ESs were outliers, and at follow-up assessment there were also three (6.4%). For these ESs *winsorising* (Lipsey & Wilson, 2001) was used by reducing outliers to the exact value of M + 2SD. The software *Comprehensive Meta-Analysis* (Borenstein, Hedges, Higgins, & Rothstein, 2013) was used for all analyses and to correct for small sample sizes Hedges's *g* was calculated. A random effects model was used since it cannot be assumed that the ESs come from the same population.

Heterogeneity among ES's was assessed with the *Q*- and the *I*-square 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 and for categorical variables with subgroup analysis using the mixed effect model.

2.8. 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 evidence base update review of psychosocial treatments published in the *Journal of Clinical Child and Adolescent Psychology* for the respective disorders included in the present meta-analysis. For depression it was Weersing et al. (2017), for mixed anxiety Higa-McMillan et al. (2016), for OCD Freeman et al. (2018), and for PTSD Dorsey et al. (2017). In some of these reviews the authors referred to earlier reviews, which we checked in order to get as comprehensive as possible a list of efficacy RCTs. From each of these reviews we listed the RCTs of some kind of cognitive behavioral treatment and then deleted those RCTs we had already included in the body of effectiveness studies. This resulted in the following number of efficacy RCTs for our comparison: depression 20, mixed anxiety 34, OCD 24, and PTSD 25, for a total of 103 trials. These references are listed in the online Supplement S5.

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 variables we also extracted data on mean age, proportion of girls, pre-treatment severity (calculated as percent of maximum score on the continuous measure), treatment time (60 min. hours), and attrition. Other variables, e.g. comorbidity, were not reported systematically, or not at all in a large proportion of studies, which precluded inclusion as a background variable.

3. Results

3.1. Description of the studies

3.1.1. Background data

Background data for the included studies are presented Table 1. The majority of the 58 studies were done in Europe (n = 29) and North America (n = 18), whereas fewer came from Australia (n = 5), Asia (n = 4), South America (n = 1), and Africa (n = 1). The total number of participants in these studies was 4618 with the following distribution: OCD 560 in 10 studies, PTSD 1266 in 18 studies, Mixed anxiety 1790 in 22 studies, and Depression 1002 in 8 studies. There was an overall majority of girls (58%), and this was due to the uneven distribution in Depression (68.8%) and PTSD (64.3%), whereas it was almost even in Mixed anxiety (49.9%) and OCD (49.3%). The mean age across all studies was 12.5 (SD 2.8) years, but Mixed anxiety had the youngest participants (M = 9.9, SD 1.7) and Depression the oldest (M = 15.0, SD 1.5). Proportion of participants with comorbidity was reported by only 40 studies (69%), and in an unsystematic fashion. With that in mind, 51.9% of the participants had at least one comorbid disorder, and Mixed anxiety had the highest proportion (62.6%), whereas the other disorders varied between 47% and 49%. The mean pre-treatment severity across the studies was 56% (SD 15.5). Since different measures were used for the various disorders it is not meaningful to compare them on this variable. Only 35 studies (60%) reported what proportion of the participants was on psychotropic medication at the inclusion to the respective study. The overall mean was 12.3%, and varied from 3.8% in PTSD and 21.1% in Depression. Finally, 37 studies (64%) reported what proportion of eligible participants declined the offer of treatment and the average was 17.4% (range 14.0–19.1%).

3.1.2. Treatment data

Treatment data for the included studies are presented in Table 2. The number of therapists per study was on average 13.4 (range 11–18),

Table 1
Background data of the included studies.

Disorder and study	Country	Continent	RCT	Comparison	N	% girls	Age (M)	% Comorbidity	% Severity	% declining Tx	% current drug Tx
OCD											
Benazon (2002)	US	NA	N		16	50.0	12.5	50.0	56.4		0
de Haan (1998)	NL	E	Y	SSRI	23	50.0	13.7	18.0	53.8	13.8	0
Farrell (2010)	AU	A	N		35	45.7	12.3	54.0	58.9	0	17.1
Reynolds (2013)	GB	E	Y	Other CBT	50	52.0	14.5	66.0	60.2	41.9	18
Riise (2016)	NO	E	N		27	32.0	13.3	40.9	70.0	21.4	13.6
Riise (2018)	NO	E	N		41	71.0	15.0	63.0	64.3	32.7	12.2
Skarphedinsson (2015)	NO, SE, DK	E	Y	SSRI	50	52.0	14.0	46.0	53.3	30.6	0
Torp (2015)	NO, SE, DK	E	N		269	51.3	12.8	40.5	61.5	0	0
Valderhaug (2007)	NO	E	N		28	50.0	13.3	62.5	57.8		8.3
Williams (2010)	GB	E	Y	WLC	21	38.1	13.6	47.6	57.7	0	33.3
PTSD											
Bicanic (2014)	NL	E	N		55	100.0	16.0			32.1	0
Catani (2009)	LK	Asia	Y	Other Tx	31	45.2	11.9		55.8	0	
Chemtob (2002)	US	NA	Y	WLC	34	68.8	8.4		53.7	12.8	
Cohen (2011)	US	NA	Y	TAU	124	50.8	9.6		48.7	14.5	
de Roos (2017)	NL	E	Y	Other CBT	103	57.3	13.1	54.4	51.3	21.1	0
Giannopolou (2006)	GR	E	Y	WLC	20	55.0	9.6		75.9		
Goldbeck (2016)	DE	E	Y	WLC	159	71.7	13.0	34.0	45.7	7.6	15.1
Habigzang (2016)	BR	SA	N		103	100.0	11.8		33.1		
Jensen (2014)	NO	E	Y	TAU	156	79.5	15.1		47.0	24.5	
Kameoka (2015)	JP	Asia	N		35	74.3	10.9		42.7		
Nixon (2012)	AU	A	Y	Other CBT	34	36.4	10.8		42.4	55.4	
Rossouw (2018)	ZA	Africa	Y	Other PT	63	87.3	15.4	55.6	67.6		
Ruf (2010)	DE	E	Y	WLC	26	46.2	11.5	38.5	63.7		
Salloum (2016)	US	NA	Y	Other CBT	53	49.0	5.0		50.4	29.3	
Scheeringa (2011)	US	NA	Y	WLC	75	33.8	5.3			35.4	
Shein-Szydlo (2016)	MX	NA	Y	WLC	100	64.0	14.9	28.0	64.6	7.4	0
Van der Oord (2010)	NL	E	N		23	73.9	11.9	69.6	62.5	0	
Webb (2014)	US	NA	N		72	64.0	12.4			8.7	
Mixed anxiety											
Barrett (2015)	AU	A	N		31	38.7	5.7		29.3		
Barrington (2005)	AU	A	Y	TAU	62	64.8	10.0	37.0	55.0		
Bodden (2008)	NL	E	Y	Other CBT	134	59.0	12.4	81.0	84.8	0	2.3
Crawford (2013)	US	NA	N		17	29.4	10.1	84.2	54.7	0	41.2
Creswell (2015)	GB	E	Y	Other CBT	211	52.1	10.2	9.5	35.3		25.1
Creswell (2020)	GB	E	Y	Other PT	136	52.9	9.2		63.8	40.4	0
Ginsburg (2012	US	NA	Y	TAU	32	62.5	10.3	63.0	66.9	0	
Jolstedt (2018)	SE	E	N		19	63.0	10.5		60.6		
Jonson (2015)	DK	E	N		87	52.9	11.2	77.0	80.0		20.2
Lau (2010)	HK	Asia	Y	WLC	51	46.7	8.7	23.0	30.1	32.9	
Martinsen (2009)	NO	E	N		21	48.4	9.8	83.0			
Monga (2009)	CA	NA	N		34	59.4	6.5	62.5	41.2		3.1
Monga (2015)	CA	NA	Y	Other CBT	77	62.3	6.8	85.7	70.1	19.8	0
Nauta (2001)	NL	E	Y	Other CBT	18	44.4	10.2	78.0	38.9		17
Nauta (2003)	NL	E	N	Other CBT	79	50.6	11.0	70.0	76.3	11.2	5
Southam-Gerow (2010)	US	NA	Y	TAU	48	56.2	10.9	72.9	57.2		
Storch (2015)	US	NA	Y	TAU	100	44.0	9.8		55.0	2.9	21
Tobon (2011)	CA	NA	N		38	38.0	10.3	68.0	70.9		35
Van Steensel (2015)	NL	E	N		174	40.2	12.4	49.4		13	19
Wergeland (2014)	NO	E	Y	Other CBT	182	53.0	11.5	77.5	87.1	8.7	6
Villabö (2018)	NO	E	Y	Other CBT	165	45.5	10.5	67.9	37.9	4.6	0
Özyurt (2018)	TR	E	Y	WLC	74	34.0	9.7	36.4	40.1	34.5	
Depression											
Charkhandeh (2016)	IR	Asia	Y	Other Tx	188	53.7	14.7		72.2		0
Goodyear (2017)	GB	E	Y	Other PT	470	74.8	15.6	48.0	70.0		27.2
Kobak (2015)	US	NA	Y	TAU	76	66.0	15.4				
Melvin (2006)	AU	A	Y	SSRI	73	65.8	15.3	69.0	69.8	9.9	0
Shirk (2014)	US	NA	Y	TAU	43	83.7	15.5	46.0	47.4	14	39
Straub (2014)	DE	E	N		15	73.3	16.4	20.0	47.6	21	6.7
Weersing (2006)	US	NA	N		80	77.0	15.6	53.0	38.1		65
Weisz (2009)	US	NA	Y	TAU	57	56.0	11.8	60.0	20.1	31.3	9.7

Note: Country: Codes from the International Organization for Standardization. Codes for the representation of names of countries and their subdivisions. Part 1: Country codes. Geneva (Switzerland): The Organization; 1997. (ISO 3166-1: 1997), have been used. RCT = Randomized controlled trial, N = No, Y = Yes, SSRI = Selective Serotonin reuptake inhibitors, WLC = Waitlist control, TAU = Treatment as usual. Blanks = Data not provided. % Severity = the mean of the sample divided by the maximum severity score of the primary outcome measure.

which indicates the number of participating therapists working at the routine clinical sites where the studies were done. Treatments were carried out over 11.7 weeks on average (range 9.5 for PTSD to 15.0 for Depression) and the mean number of sessions was 11.2. Most studies had one session per week. The total hours of treatment was 14.1 (range 12.9–15.0) and the intensity (hours/week) was on average 2.0. The Depression studies had a mean of 1.1, Mixed anxiety 1.2, PTSD 1.5, and OCD 4.4. However, the latter mean was completely carried by the two

Table 2
Treatment data of the included studies.

Disorder and study	Method of CBT	Format	Therapist (N)	Parental involvement	Duration (weeks)	# of Sessions	Tx time (hrs)	Intensity (hrs/ week)	Attrition (total, %)	Attrition (CBT, %)	Attrition (comp., %)	F-up (months)	Analysis
OCD													
Benazon (2002)	CBT + ERP	I	1	Moderate	12	12	12	1.0	0.0	0.0		0	ITT
de Haan (1998)	ERP + CT	I	8	Low	12	12	12	1.0	4.4	7.7	0.0	0	TC
Farrell (2010)	CBT+ ERP	I,G	4	Moderate	12	12	16.1	1.3	5.7			0	ITT
Reynolds (2013)	ERP + CT	I	6	Moderate	14	14	14	1.0	6.4	0.0	12.5	6	ITT
Riise (2016)	ERP	G	4	High	0.6	4	18	18.0	9.1			0	ITT
Riise (2018)	ERP	G	9	High	0.6	4	18	18.0	0.0			6	ITT
Skarphedinsson (2015)	ERP + CT	I	19	Moderate	16	10	15	0.9	16.3	25.0	0.0	6	ITT
Torp (2015)	ERP + CT	I	44	Moderate	14	14	17.5	1.3	10.4	10.4		36	ITT
Valderhaug (2007)	ERP + CT	I	8	Moderate	14	12	12	0.9	14.3	14.3		6	TC
Willians (2010)	CBT	I	3	Low	12	10	10	0.8	9.5	9.1	10.0	3	ITT
PTSD													
Bicanic (2014)	CBT	G		Moderate	9	9	18	2.0	1.8			12	TC
Catani (2009)	KIDNET	I	6	Low	2	6	9	4.5	0.0	0.0	0.0	6	ITT
Chemtob (2002)	EMDR	I	4	Low	3	3	3	1.0	5.9			6	TC
Cohen (2011)	TF CBT	I	3	Moderate	8	8	6	0.8	29.2	28.3	30.4	0	ITT
de Roos (2017)	EMDR/CBWT	I	21	Moderate	6	6	4.5	0.8	2.4	2.4	2.3	12	ITT
Giannopolou (2006)	CBT	G	2	Moderate	7	7	14	2.0	15.0	0.0	30.0	48	TC
Goldbeck (2016)	TF CBT	I	26	Moderate	16	12	18	1.5	12.2	13.7	10.8	12	ITT
Habigzang (2016)	CBT	G		Low	16	16	24	1.5	1.0			0	TC
Jensen (2014)	TF CBT	I	71	Moderate	15	15	15	1.0	15.2	14.9	15.6	18	ITT
Kameoka (2015)	TF CBT	I	7	Moderate	14	14	21	1.5	8.6			0	ITT
Nixon (2012)	CBT/CT	I		Moderate	9	9	13.5	1.5	24.1	26.7	21.4	12	ITT
Rossouw (2018)	PE-A	I		Low	9	9	7	1.0	17.5	19.4	15.6	6	ITT
Ruf (2010)	KIDNET	I	8	Low	8	8	16	2.0	3.8	7.7	0.0	10	ITT
Salloum (2016)	TF CBT	I	4	Moderate	12	14	14	1.0	7.5	11.4	0.0	3	ITT
Scheeringa (2011)	TF CBT	I	4	Moderate	12	12	12	1.0	42.9	34.6	52.2	6	TC
Shein-Szydlo (2016)	CBT	I	2	Low	12		12	1.0	1.0	2.0	0.0	6	TC
Van der Oord (2010)	CBWT	I		Low	11	6	5.5	0.5	13.0			12	TC
Webb (2014)	TF CBT	I	12	Moderate	10	10	10	1.0	12.5			10	
Mixed anxiety													
Barrett (2015)	CBT	G	3	Moderate	10	10	15	1.5	0.0			12	TC
Barrington (2005)	CBT	I	18	Moderate	12	12	12	1.0	12.9			12	TC
Bodden (2008)	CBT child	I		Moderate	13	13	19.5	1.5	10.9	3.1	18.8	3	ITT
Crawford (2013)	Comp. assisted	iCBT	3	Low	12	12	10	0.8	11.8			0	TC
Creswell (2015)	CBT child	I	10	Moderate	8	8	8	1.0	15.6	21.1	12.6	12	ITT
Creswell (2020)	CBT parent	P	19	High	8	8	5	0.6	9.7	16.7	3.2	6	ITT
Ginsburg (2012	CBT	I	11	Moderate	12	8	6	0.8	9.4	5.9	13.3	1	ITT
Jolstedt (2018)	CBT	iCBT	3	Moderate	12	12	4	0.3	11.0			3	ITT
Jonson (2015)	CBT	G	16	Moderate	12	10	20	1.7	4.7			3	ITT
Lau (2010)	CBT	G	6	High	10	9	18	1.8	16.0	4.2	14.3	6	TC
Martinsen (2009)	CBT	G		Moderate	12	10	15	1.5	14.3			0	TC
Monga (2009)	CBT	G		High	12	12	12	1.0	5.9	5.9		0	TC
Monga (2015)	CBT fam	G		High	12	12	12	1.0	0.0	0.0	0.0	12	ITT
Nauta (2001)	ICBT	I	6	Moderate	12	12	12	1.0	5.6			15	TC
Nauta (2003)	ICBT	I	27	Moderate	12	12	12	1.0	4.1	10.3	0.0	3	ITT
Southam-Gerow (2010)	ICBT	I	39	Moderate	22	18	18	1.0	22.9	25.0	20.8	0	ITT
Storch (2015)	CBT	iCBT	6	Moderate	12	12	12	1.0	8.0	8.1	7.8	1	ITT
Tobon (2011)	CBT	G	6	Moderate	12	12	18	1.5	10.5			4	TC
Van Steensel (2015)	CBT	I, F		Moderate	12	12	18	1.5	18.1			24	ITT
Wergeland (2014)	CBT	I	17	Moderate	10	10	10	1.0	13.4		5.7	47	ITT
Villabö (2018)	CBT	I	32	Moderate	12	10	15	1.5	12.7	29.1	7.3	24	ITT
Özyurt (2018)	CBT	G	2	High	8	8	11.5	1.4	12.2	7.1	17.2	4	TC
Depression													
Charkhandeh (2016)	CBT				12	12	24	2.0	0.0	0.0	0.0	0	TC
Goodyear (2017)	CBT	I	145	Moderate	20	20	20	1.0	9.6	8.4	10.2	12	ITT
Kobak (2015)	CBT	iCBT	18	Low	12	12	12	1.0	14.5	10.2	18.9	0	TC
Melvin (2006)	CBT	I	11	Moderate	12	12	10	0.8	15.6	4.5	19.6	6	ITT
Shirk (2014)	CBT-m	I	4	Low	12	12	12	1.0	5.7	6.3	5.0	0	ITT
Straub (2014)	CBT	G	2	Low	5	6	6	1.2	0.0			1	TC
Weersing (2006)	CBT	I		Low	22	20	20	0.9	0.0			0	ITT
Weisz (2009)	CBT	I	54	Low	25	16	16	0.6	0.0	0.0	0.0	0	ITT

Note: F-up = Follow-up, I = Individual therapy, G = Group therapy, iCBT = internet delivered therapy, F = Family therapy, ITT = intention to treat, TC = Treatment completers. * = Four studies report on follow-up in separate publications, and data from these publications are included in the follow-up (Tutus 2017, Jensen 2017, Nixon 2017, Kodal 2018). Blanks = Data not provided.

Table 3
Within-group effect size (Hedges' *g*) for all studies (RCTs and open trials) divided on disorder with treatment condition as unit of analysis.

Disorder	k	*g*	95% CI	*z*-value	Q_{-value}	p_{-value}	I^2
				Post-treatment			
All disorders	69	1.50	1.32–1.67	16.86[a]	753.78	0.001	91.0
Depression	8	1.28	0.81–1.74	5.40[a]	21.54*	0.001	
Mixed anxiety	29	1.28	1.03–1.53	9.94[a]			
OCD	11	2.54	2.05–3.02	10.25[a]			
PTSD	21	1.43	1.16–1.70	10.28[a]			
				Follow-up			
All disorders	49	2.13	1.85–2.40	14.99[a]	640.84	0.001	92.5
Depression	3	1.72	0.95–2.49	4.37[a]	13.85*	0.003	
Mixed anxiety	23	1.87	1.48–2.26	9.46[a]			
OCD	7	3.36	2.63–4.10	9.01[a]			
PTSD	16	2.05	1.75–2.35	13.39[a]			

Note: k = number of treatment conditions. [a] $p < 0.0001$. * Comparison between the four disorders.

studies by Riise et al. (2016; 2018) with the concentrated exposure and response prevention of 18 h over four days. When removing these studies from the calculation the intensity for OCD was 1.0. Information on continent, therapist and CBT-manual information on the included studies are provided in Supplement S9.

3.2. Methodological data

3.2.1. Methodology ratings

The research methodology score had an overall mean of 22.28 (SD 5.94, 95% CI 20.71–23.84). The studies for the different disorders had the following means (SD): Depression 20.50 (5.73), PTSD 21.11 (6.46), Mixed anxiety 23.36 (6.11), and OCD 23.40 (4.72). A one-way ANOVA yielded a non-significant effect ($F(3) = 0.83$, $p = 0.49$). When the analysis was restricted to the 38 RCTs the overall mean was 24.42 (SD 5.61, 95% CI 22.57–26.28). The means for the different disorders were: Depression 21.17 (6.43), PTSD 23.54 (5.67), Mixed anxiety 25.80 (5.23), and OCD 27.00 (4.97), and the ANOVA was non-significant ($F(3) = 1.39$, $p = 0.26$).

3.2.2. Risk of bias

The risk of bias categorization is presented in the online Supplement, Table S6. If we delete the 21 open trials from the evaluation of the first two domains we find the following proportions of a low risk-of-bias: Random sequence 78%, allocation concealment 59%, blinding of assessors 47%, incomplete data 64%, and selective reporting 97%. A high risk-of-bias was found in: Random sequence 0%, allocation concealment 3%, blinding of assessors 26%, incomplete data 28%, and selective reporting 2%. Thus, it was much more common that these studies had a low than a high risk-of-bias regarding the evaluated domains.

In order to score the risk-of-bias a low risk was given 0, an unclear risk 0.5, and a high risk 1 point, which means that the total score could vary from 0 to 5 points (see Supplement S6). The total mean score was 1.60 (SD = 1.27) and the studies within each disorder had the following mean (SD): Depression 1.25 (1.13), PTSD 1.44 (1.44), Mixed anxiety 1.71 (1.31), and OCD 1.95 (1.01). A one-way ANOVA yielded a non-significant effect ($F(3) = 0.58$, $p = 0.63$). Restricting the analysis to the RCTs only yielded the following means: Depression 1.25 (1.13), PTSD 0.81 (1.15), Mixed anxiety 1.07 (0.96), and OCD 0.92 (1.11), and the ANOVA was non-significant ($F(3) = 0.15$, $p = 0.93$).

3.3. Meta-analysis

3.3.1. Attrition

Using treatment condition (k = 97) as the unit of analysis the overall attrition rate was 12.2% (95% CI 10.1–14.6, $z = 18.33$, $p < 0.001$). The different disorders had the following attrition rates: Depression 7.1%, Mixed anxiety 12.6%, OCD 13.4%, and PTSD 14.3%, with no significant difference between them ($Q_{between}$ (3 df) = 4.62, $p = 0.20$). When analyzing dropout rate for the different treatment conditions we found the following rates: CBT 11.4%, Other forms of psychotherapy 6.6%, SRIs 25.8%, Treatment as usual 14.3%, Waitlist control 20.1%, with a significant difference between them ($Q_{between}$ (4 df) = 11.59, $p = 0.021$). Subsequent pairwise comparisons showed that CBT had a lower attrition rate than SRI ($Q_{between}$ (1 df) = 7.73, $p = 0.005$) but did not differ significantly from the other conditions.

3.3.2. Primary continuous measure

Table 3 displays the mean effect sizes of the primary continuous measure for all studies at post-treatment and follow-up assessment, which was done on average 10.7 months after the end of therapy. At post-treatment the average ES was very large ($g = 1.50$) and significantly different from zero. Heterogeneity was significant and large as indicated by the Q- and I^2-values. The comparison between disorders was also significant and pairwise comparisons showed that OCD had a significantly higher ES than PTSD ($Q_{between}$ (1 df) = 15.23, $p < 0.001$), Mixed anxiety ($Q_{between}$ (1 df) = 20.33, $p < 0.001$), and Depression ($Q_{between}$ (1 df) < 13.59, $p = 0.001$). The other means did not differ significantly from each other.

At follow-up the mean ES ($g = 2.13$) was significantly higher than at post-treatment ($Q_{between}$ (1 df) = 14.02, $p < 0.001$), but also significantly heterogeneous. Once more, the comparison between disorders was significant and pairwise comparisons indicated that OCD had a significantly higher ES than PTSD ($Q_{between}$ (1 df) = 10.64, $p < 0.001$), Mixed anxiety ($Q_{between}$ (1 df) = 12.54, $p < 0.001$), and Depression ($Q_{between}$ (1 df) = 9.16, $p < 0.001$). The differences between the other means were not significant.

3.3.2.1. Publication bias. The possibility of publication bias was investigated, using Duval and Tweedie's trim-and-fill method and Eggers regression intercept. Regarding the post-treatment data the trim-and-fill method suggested trimming 25 conditions to the left of the mean which would lower the ES from 1.50 to 0.94 (95% CI 0.74–1.14). The regression intercept had a significant *t*-value = 4.78, $p < 0.001$, indicating that publication bias probably is an issue for this body of studies.

3.3.2.2. Moderator analyses. As the mean ES was significantly heterogeneous we followed up with moderator analyses. Using subgroup analysis Table 4 shows the results for categorical variables and there was no significant difference between RCTs and open trials. Also, studies with intent-to-treat analysis yielded nominally higher ES than studies using completer analysis, which is unexpected. There was no significant differences depending on treatment format, degree of parental involvement, or therapist profession. However, the continent at which the study was carried out was associated with a significant difference; studies from Europe had the highest ES.

Continuous variables on which at least 75% of the studies provided information were analyzed with the meta-regression module in the CMA program using the fixed effects analysis (see Table 5). The following variables were positive moderators: number of participants in the study, mean age of the sample, pre-treatment severity, number of therapists in the study, and intensity of the treatment, i.e. higher values on these variables were associated with higher ES. However, regarding intensity there were two outliers with 18 h of therapy during a single week (Riise et al., 2016; Riise, Kvale, Öst, Skjold, & Hansen, 2018); when these were removed intensity no longer was a significant positive moderator. There were two negative moderators: Proportion of girls in the sample and attrition rate, i.e. the higher these proportions the lower the ES. This was, however mainly due to PTSD, which had the highest proportion of girls and the moderator was no longer significant when PTSD-studies were removed. Finally, higher risk-of-bias score was

Table 4
Subgroup analysis of the overall effect size for all studies at post-treatment.

Variable	k	*g*	95% CI	Q_b-value	*p*-value
Type of study				2.60	0.10
RCT	49	1.42	1.23–1.62		
Open trial	20	1.80	1.42–2.18		
Statistical analysis				2.24	0.13
Intent-to-treat	47	1.58	1.36–1.80		
Treatment completers	22	1.31	1.04–1.58		
Format				2.97	0.23
Individual	47	1.40	1.20–1.60		
Group	14	1.72	1.20–2.24		
Internet CBT	4	1.75	1.32–2.18		
Parental involvement				2.45	0.29
Low	15	1.57	1.14–2.01		
Moderate	45	1.41	1.21–1.61		
High	8	2.01	1.24–2.79		
Therapist profession				5.68	0.13
Clinical psychologist	43	1.54	1.32–1.75		
Child psychiatrist	2	1.93	1.41–2.45		
Social worker	7	1.00	0.42–1.58		
Various professions	11	1.41	0.97–1.85		
Continent				11.22	0.011
Europe	37	1.75	1.51–2.00		
North America	20	1.14	0.85–1.42		
Australia	6	1.26	0.81–1.70		
Asia	4	1.35	0.91–1.78		

Note: k = number of treatment conditions, Q_b = Q between subgroups.

associated with higher ES, whereas the methodology score was not significantly associated with ES.

3.3.3. Remission

The remission rates at post-treatment and follow-up are presented in Table 6. At post-treatment 58.7% of the participants had remitted, which was significantly different from 50%, and significantly heterogeneous. A subgroup analysis indicated that PTSD had the highest remission rate (77.4%), which was significantly higher that the rates for Depression ($Q_{between}$ (1 df) = 6.84, p = 0.009), Mixed anxiety ($Q_{between}$ (1 df) = 36.68, $p < 0.001$), and OCD ($Q_{between}$ (1 df) = 7.55, $p < 0.001$). The rates for the other disorders did not differ significantly between each other.

At follow-up assessment the overall remission rate had increased significantly ($Q_{between}$ (1 df) = 15.59, $p < 0.001$), to 71.6%, which was significantly different from 50% and heterogeneous. The subgroup analysis also showed that the disorders differed significantly between each other. Once again, PTSD had the highest remission rate (83.3%), which differed significantly from the rates for Depression ($Q_{between}$ (1 df) = 4.93, p = 0.026), Mixed anxiety ($Q_{between}$ (1 df) = 7.36, p = 0.007), and OCD ($Q_{between}$ (1 df) = 4.77, p = 0.029). The differences between the other rates were not significant.

Table 5
Meta-regression analysis of the overall effect size for all studies at post-treatment.

Variable	k	Point est.	*z*-value	*p*-value
Number of participants	69	0.002	4.99	0.0001
Percent girls	69	−0.004	−2.48	0.013
PTSD removed	48	0.004	1.49	0.135
Mean age	69	0.086	8.35	0.0001
Pre-treatment severity	63	1.828	11.96	0.0001
Number of therapists	55	0.006	5.08	0.0001
1 outlier removed	54	0.004	2.02	0.042
Methodology score	69	0.004	1.24	0.216
Risk-of-bias	69	0.094	5.16	0.0001
Weeks of treatment	69	−0.013	−1.88	0.059
Number of sessions	69	−0.002	−0.21	0.837
Total hours of treatment	69	−0.004	−0.92	0.356
Intensity	69	0.169	7.10	0.0001
2 outliers removed	67	−0.052	−0.80	0.426
Attrition	69	−0.017	−5.07	0.0001

Note: k = number of treatment conditions.

Table 6
Rates of remission for all studies divided on disorder.

Disorder	k	%	95% CI	*z*-value†	Q_{-value}	p_{-value}	I^2
Post-treatment							
All disorders	54	58.6	53.6–63.4	3.34[a]	244.87	0.001	78.4
Depression	5	53.2	34.6–70.9	0.32	37.33*	0.001	
Mixed anxiety	27	50.7	45.3–56.2	0.27			
OCD	7	56.7	41.7–70.6	0.87			
PTSD	15	77.4	71.3–82.6	7.50[b]			
Follow-up							
All disorders	39	71.6	67.3–75.5	8.96[b]	109.94	0.001	65.4
Depression	3	53.5	25.9–79.1	0.23	9.20*	0.027	
Mixed anxiety	22	69.4	64.1–74.3	6.65[b]			
OCD	3	72.6	67.6–77.1	7.92[b]			
PTSD	11	83.3	74.9–89.3	6.11[b]			

Note: k = number of treatment conditions. [a] $p < 0.001$, [b] $p < 0.0001$. † Test if significantly different from 50%, * Comparison between the four disorders.

3.3.3.1. Publication bias. Eggers regression intercept did not yield a significant *t*-value (1.46, p = 0.15). The trim-and-fill method suggested trimming 8 studies, which would have reduced the remission rate from 58.7% to 54.8% (95% CI 49.5–60.0). Thus, publication bias does not seem to be an important problem regarding the remission rate.

3.3.3.2. Moderator analyses. Table 7 displays the results of the subgroup analyses for the categorical variables. The only significant variable was parental involvement where treatment conditions with low involvement from the parents yielded a higher remission rate than conditions with moderate or high parental involvement. As for the primary continuous measure open trials did not have a significantly higher remission rate than RCTs and treatment completer analyses did not differ from intent-to-treat analyses.

The results of the meta-regression analysis of continuous variables are shown in Table 8. The following variables were significant positive moderators: percent girls in the sample (which disappeared when the PTSD-studies were removed), mean age, total hours of therapy, and treatment intensity, i.e. higher values on these variables were associated with higher remission rates. As was the case for the *g*-value (Table 5), the significant effect for intensity was driven by two outliers and when those studies were removed it was no longer significant. Number of therapists in the study was also a positive moderator but it was due to an outlier with an extremely high number of therapists (n = 145), and removing this study yielded a non-significant effect. Only two variables, methodology score and pre-treatment severity, were negative moderators; i.e. more stringent methodology and higher pre-treatment severity were associated with lower remission rate.

3.4. Efficacy-effectiveness comparison

3.4.1. Background and treatment variables

Table 9 presents a comparison between efficacy and effectiveness studies on some background and treatment variables. There were no significant differences between the two types of studies on mean age at the start of treatment, proportion of girls in the studies, or the pre-treatment severity (as assessed by the primary continuous outcome measure). Regarding treatment time (hours of therapy) the effectiveness studies had nominally lower means than efficacy studies for all four disorders, and the difference was significant for mixed anxiety disorders. Finally, the attrition rate was lower for effectiveness than efficacy studies in depression, but did not differ significantly for the other disorders.

Table 7
Subgroup analysis of the remission rate for all studies at post-treatment.

Variable	k	%	95% CI	Q_b-value	*p*-value
Type of study				1.09	0.30
RCT	40	56.8	51.1–62.3		
Open trial	14	63.1	52.4–72.6		
Statistical analysis				1.33	0.25
Intent-to-treat	41	56.7	51.5–61.6		
Treatment completers	13	64.7	51.7–75.9		
Format				0.05	0.98
Individual	37	57.7	51.8–63.3		
Group	10	59.7	42.9–74.5		
Internet CBT	4	58.1	44.6–70.5		
Parental involvement				6.11	0.047
Low	7	72.5	60.4–82.0		
Moderate	38	55.5	49.8–61.1		
High	6	58.8	38.2–76.7		
Therapist profession				4.03	0.13
Clinical psychologist	33	61.2	54.9–67.1		
Social worker	7	64.7	49.4–77.5		
Various professions	11	49.5	38.8–60.2		
Continent				2.08	0.56
Europe	29	56.6	50.8–62.1		
North America	16	57.7	45.4–69.0		
Australia	5	55.5	34.8–74.4		
Asia	2	69.2	52.4–82.1		

Note: k = number of treatment conditions, Q_b = Q between subgroups.

Table 8
Meta-regression analysis of the overall remission rate for all studies at post-treatment.

Variable	k	Point est.	*z*-value	*p*-value
Number of participants	53	−0.0001	−1.10	0.27
Percent girls	53	0.017	4.76	0.0001
PTSD removed	39	0.001	0.13	0.90
Mean age	53	0.056	2.43	0.015
Pre-treatment severity	48	−1.453	−5.66	0.0001
Number of therapists	42	0.003	2.04	0.041
1 outlier removed	41	−0.0006	−0.18	0.86
Methodology score	53	−0.036	−4.98	0.0001
Risk-of-bias	53	0.011	0.33	0.74
Weeks of treatment	53	−0.007	−0.57	0.57
Number of sessions	53	0.007	0.50	0.61
Total hours of treatment	53	0.019	2.01	0.04
Intensity	53	0.062	3.45	0.0006
2 outliers removed	51	0.183	1.62	0.11
Attrition	53	−0.006	−0.92	0.36

Note: k = number of treatment conditions.

3.4.2. Effect size on primary outcome measure

Table 10 presents the subgroup analyses comparing the uncontrolled effect size (Hedges' *g*) for efficacy and effectiveness studies within each disorder. At post-treatment assessment (upper part of Table 10) there were no significant differences between the two types of studies. However, at follow-up assessment effectiveness studies yielded a significantly higher ES than efficacy studies for PTSD, whereas the differences for the remaining disorders were non-significant.

3.4.3. Remission

Table 11 contains subgroup analyses comparing the remission rates at post-treatment and follow-up assessment. The efficacy vs. effectiveness comparison regarding follow-up period was not significant for any disorder, and varied from 6.3 (SD 5.5) months for effectiveness studies in depression to 17.9 (SD 19.9) months for efficacy studies in mixed anxiety. The post-treatment remission rates (upper part of Table 11) did not differ between study types for three of the disorders, but in mixed anxiety disorders the efficacy studies had a significantly higher rate than the effectiveness studies. However, at follow-up assessment (lower part of Table 11) the difference within mixed anxiety disorders was no longer significant. The increase in remission rate from post-treatment was 18.7 percentage points for effectiveness studies and 9.7 for efficacy studies. The difference in remission rates was not significant for the other disorders.

3.4.4. Comparison of RCTs only

Since the outcome presented in Tables 10 and 11 might have been unduly influenced by open trials we repeated the analyses using only RCT effectiveness studies. Table 12 summarizes the results across disorders and there were no significant differences between efficacy and effectiveness studies on remission rates or effect size at post-treatment or follow-up assessment.

4. Discussion

The overall focus of this meta-analysis was to examine whether cognitive behavior therapy for anxiety disorders, obsessive-compulsive disorder, post-traumatic stress disorder, and depressive disorder in children and adolescents works in routine clinical care. Our main finding supports the effectiveness of CBT in reducing internalizing disorders and symptoms, and the outcomes are comparable with how CBT works in university settings, i.e. efficacy studies. Our results provide updated evidence about the degree of improvement which can be expected following the treatment offered in routine clinical care when carried out by therapists having appropriate training.

Across the different disorders, the pre-post treatment effect size for disorder-specific outcome measures was very large and over half of the youth who started treatment achieved remission at the end of treatment. Our results also showed that some disorders yielded more favorable outcomes, with OCD having the highest effect size, and PTSD the highest remission rate. Furthermore, the treatment effects significantly increased at follow-up on average 10.7 months after the end of therapy. Almost two thirds of youth achieved remission at follow-up, indicating that they continued to improve after therapy had ended. Several explanations for this continued improvement may apply. First, the improvements may relate to a delayed treatment effect, which may stem from a prolonged consolidation of acquired skills among youth and their parents (Ishikawa, Okajima, Matsuoka, & Sakano, 2007). Also, the time available to conduct exposure exercises in CBT is limited in most manual-based treatment programs, which were carried over 11.7 weeks on average in the current meta-analysis. However, the mean follow-up period was 10.7 months which meant that the youth had much more time to apply the acquired skills after ending treatment, leading to improved outcome at follow-up. Also, a possible explanation could be maturation and spontaneous recovery among children and adolescents with anxiety disorders (Adler Nevo et al., 2014). Similar findings of continued improvement are found in effectiveness studies (Kodal et al., 2018) and efficacy studies (Gibby, Casline, & Ginsburg, 2017) for mixed anxiety disorders in children and adolescents.

Moreover, the overall attrition rate across the disorders was only 12.2%. This figure is lower than the 20% in the review by Lee et al. (2013). Importantly, there was no significant difference in the attrition rates between the different disorders, indicating that these treatments were acceptable to youth and their caregivers. Taken together, these findings suggest that youth receiving CBT for internalizing disorders in routine clinical care improved significantly and substantially following treatment, and that CBT is effective in clinical settings. We attempted to broaden the view by extracting effect data for the common comorbid disorders (depression for anxiety disorders and anxiety for depressive disorders) and quality of life. However, only 48% of the studies provided data on comorbid disorders and much fewer on quality of life and if we were to meta-analyze them the results would most probably not be representative for the entire body of studies.

Our findings expand the results from a previous review of the effectiveness of CBT for internalizing and externalizing disorders in youth (Lee et al., 2013). Considerable evidence regarding the topic has been

Table 9
Some background and treatment data (M and SD) for efficacy and effectiveness studies in the different disorders.

Disorder	k	Age (years)	% Girls	Severity	% Attrition	Tx time
Depression		$p = 0.35$	$p = 0.10$	$p = 0.08$	$p = 0.001$	$p = 0.30$
Efficacy	24	14.0 (2.8)	57.2 (15.7)	36.6 (10.9)	15.4 (13.3)	19.7 (9.9)
Effectiveness	8	15.0 (1.4)	67.6 (11.2)	52.2 (19.6)	5.1 (3.8)	15.4 (6.4)
Mixed anxiety		$p = 0.87$	$p = 0.72$	$p = 0.95$	$p = 0.97$	$p = 0.01$
Efficacy	59	9.9 (1.8)	48.7 (12.9)	58.0 (18.5)	10.0 (8.6)	15.1 (5.3)
Effectiveness	29	9.9 (1.7)	49.9 (10.0)	57.8 (19.5)	10.1 (9.5)	12.0 (4.1)
OCD		$p = 0.12$	$p = 0.40$	$p = 0.21$	$p = 0.39$	$p = 0.37$
Efficacy	27	12.0 (2.5)	46.1 (9.6)	62.0 (6.1)	6.1 (7.5)	15.4 (4.6)
Effectiveness	11	13.3 (1.3)	49.2 (10.2)	59.4 (4.8)	8.5 (7.4)	14.0 (2.8)
PTSD		$p = 0.64$	$p = 0.46$	$p = 0.57$	$p = 0.81$	$p = 0.24$
Efficacy	33	10.8 (2.7)	68.3 (19.8)	56.0 (16.1)	14.9 (10.5)	14.8 (7.1)
Effectiveness	21	11.2 (3.2)	63.8 (19.9)	52.7 (10.8)	15.5 (14.4)	12.5 (5.8)

Note: k = number of treatment conditions, Severity = percentage of the maximum score on the primary outcome measure. Tx time = number of 60 min therapy hours.

Table 10
Effect sizes (Hedges' g) for efficacy and effectiveness studies within the different disorders.

Disorder	Study type	k	*g*	95% CI	*z*-value	$Q_{between}$†	*p*-value
Post-treatment							
Depression	Efficacy	19	1.31	1.01–1.62	8.46[a]	0.06	0.81
	Effectiveness	7	1.24	0.71–1.77	4.62[a]		
Mixed anxiety	Efficacy	60	1.32	1.13–1.52	13.12[a]	0.002	0.97
	Effectiveness	28	1.32	1.04–1.59	9.45[a]		
OCD	Efficacy	33	2.50	2.25–2.76	19.31[a]	0.86	0.35
	Effectiveness	11	2.29	1.92–2.66	12.05[a]		
PTSD	Efficacy	31	1.18	0.96–1.41	10.28[a]	1.92	0.17
	Effectiveness	20	1.44	1.15–1.73	9.81[a]		
Follow-up							
Depression	Efficacy	12	1.54	1.18–1.90	8.41[a]	0.11	0.74
	Effectiveness	3	1.69	0.90–2.49	4.17[a]		
Mixed anxiety	Efficacy	47	1.84	1.57–2.11	13.39[a]	0.06	0.81
	Effectiveness	22	1.91	1.47–2.34	8.63[a]		
OCD	Efficacy	20	2.70	2.31–3.08	13.67[a]	3.06	0.08
	Effectiveness	7	3.51	2.69–4.33	8.35[a]		
PTSD	Efficacy	26	1.42	1.16–1.69	10.44[a]	8.45	0.004
	Effectiveness	16	2.02	1.72–2.32	13.17[a]		

Note: k = number of comparisons. [a] $p < 0.0001$. † Comparison Efficacy vs. Effectiveness within the respective disorders.

Table 11
Remission rates for efficacy and effectiveness studies for the different disorders.

Disorder	Study type	k	%	95% CI	*z*-value*	$Q_{between}$†	*p*-value
Post-treatment							
Depression	Efficacy	20	55.5	47.3–63.4	1.31	0.15	0.70
	Effectiveness	6	51.7	35.1–68.0	0.85		
Mixed anxiety	Efficacy	51	60.3	56.0–64.4	4.64[c]	7.57	0.006
	Effectiveness	27	50.6	45.2–56.0	0.21		
OCD	Efficacy	26	50.7	43.9–57.5	0.21	0.50	0.48
	Effectiveness	7	56.7	41.7–70.6	0.87		
PTSD	Efficacy	13	72.1	64.6–78.6	5.33[c]	1.36	0.24
	Effectiveness	15	77.4	71.3–82.6	7.50[c]		
Follow-up							
Depression	Efficacy	8	66.7	58.3–74.2	3.77[c]	0.77	0.38
	Effectiveness	3	53.5	25.9–79.1	0.82		
Mixed anxiety	Efficacy	45	70.0	64.8–74.8	6.91[c]	0.04	0.83
	Effectiveness	22	69.3	63.8–74.2	6.52[c]		
OCD	Efficacy	12	65.5	53.6–75.7	2.54[a]	1.41	0.24
	Effectiveness	3	72.6	67.6–77.1	7.92[c]		
PTSD	Efficacy	4	78.7	61.3–89.6	3.02[b]	0.42	0.52
	Effectiveness	13	83.5	75.9–87.6	6.94[c]		

Note: k = number of comparisons. [a] $p < 0.05$, [b] $p < 0.01$, [c] $p < 0.0001$. * Test if significantly different from 50%. † Comparison Efficacy vs. Effectiveness within the respective disorders.

generated since the Lee et al. review was published in 2013. Using similar inclusion criteria but including also PTSD, the number of effectiveness studies for internalizing disorders has almost increased tenfold, with a total of 58 studies included in this meta-analysis. This increase reflects the effort to adopt evidence-based interventions and to evaluate their effectiveness. Consistent with findings from Lee et al.

Table 12
Effect sizes for randomized controlled studies only: all disorders combined.

Study type	k	*ES*	95% CI	z-value	$Q_{between}$†	*p*-value
g-value at post-treatment						
Efficacy	143	1.55	1.41–1.68	22.45[b]	2.49	0.12
Effectiveness	47	1.36	1.17–1.55	13.85[b]		
g-value at follow-up						
Efficacy	105	1.83	1.67–2.00	21.23[b]	0.06	0.80
Effectiveness	37	1.87	1.58–2.16	12.76[b]		
Remission rate at post-treatment						
Efficacy	110	58.6	55.3–61.8	5.12[b]	0.01	0.95
Effectiveness	41	58.8	52.7–64.6	2.83[a]		
Remission rate at follow-up						
Efficacy	69	69.3	65.2–73.1	8.58[b]	1.31	0.25
Effectiveness	33	73.0	67.8–77.7	7.78[b]		

Note: k = number of comparisons. [a] $p < 0.01$, [b] $p < 0.0001$. † Comparison Efficacy vs. Effectiveness.

(2013), the majority of the included studies were conducted outside of North America. However, several more studies from different countries were included in the current meta-analysis, indicating an increased effort to address the transportability and generalizability of interventions developed in North America to other countries and cultures (see Supplement S9).

The meta-analysis revealed a significant heterogeneity among the effect sizes and remission rates in the analysis with all comparisons. Thus, we examined some characteristics of the patient sample and treatment variables as potential moderators influencing treatment outcome. When analyzing all the studies together, older age was associated with better outcome. This finding may be difficult to interpret, as age has been found to moderate treatment outcome differently across the disorders. For example, older age was positively associated with outcome for depressive disorders (Weersing et al., 2017), negatively associated with outcome in OCD (Freeman et al., 2018), whereas mixed finding for age and outcome was found for anxiety disorders (Higa-McMillan et al., 2016; Öst & Ollendick, 2017). Finally, for PTSD, age was found to be unrelated to treatment (Dorsey et al., 2017). Thus, the variability across the disorders challenges the interpretation of the results.

Further, we found that pre-treatment severity moderated the effect size positively and remission rate negatively. This finding is probably due to the greater room for improvement for those with higher pre-treatment symptom levels, resulting in a higher within-group effect size. On the other hand, a higher pre-treatment severity makes remission harder to achieve, and as such these patients end treatment with a higher symptom level. Such findings are in line with those reported in treatment outcome studies for mixed anxiety (Knight, McLellan, Jones, & Hudson, 2014; Lundkvist-Houndoumadi, Thastum, & Hougaard, 2015), OCD (Freeman et al., 2018), and depressive disorders (Weersing et al., 2017).

The significant association between number of therapists and remission was probably due to one study with a high number of therapists; when excluding this study number of therapists no longer moderated remission. Similarly, the association between intensity of treatment and outcome was no longer significant when excluding two studies for OCD with exposure and response prevention of 18 h over four days. When removing these two studies, treatment intensity no longer moderated outcome. Number of participants in the study was also associated with higher effect size, but not remission rate. This is encouraging since small studies do not seem to be driving the high ES for these effectiveness studies. Also, a higher attrition rate was associated with lower effect size, but was unrelated to remission. The lower effect size could be explained by attrition causing a more conservative and less precise estimate of change in the intention-to-treat analyses. Furthermore, we found a significant association between continent and effect size. A possible explanation to this finding is that the organization of routine clinical care might be different in different continents, whereas the organization of care in the more privileged university settings might be more uniform over continents.

We found that gender moderated outcome both negatively and positively, i.e. a higher proportion of girls in the sample was associated with both a lower effect size and a higher remission rate. Across the disorders, gender is most often not found to significantly moderate treatment outcome (Dorsey et al., 2017; Freeman et al., 2018; James, James, Cowdrey, Soler, & Choke, 2015; Weersing et al., 2017). Separate analyses showed that the association between gender and outcome was, however, completely due to the PTSD studies; when excluding these studies gender no longer moderated outcome.

The finding that the degree of parental involvement in treatment was inversely related to outcome is somewhat unexpected. Our result may be related to the treatment of PTSD, with the high remission rate found for this disorder, and in which low parental involvement was common in several of the studies. Similar findings of low parental involvement being positively related to outcome have been reported in a recent meta-analytic review of brief, intensive and concentrated CBT treatments for anxiety in children (Öst & Ollendick, 2017). One possible hypothesis could be that in treatments with low parental involvement the child may feel more free to test various anxiety-arousing behaviors under the guidance of the therapist. On the other hand, a high parental involvement may mean a continuation of old interaction patterns between parent and child, making it more difficult for the child to try new behaviors that will be helpful in reducing anxiety. Mixed results regarding the role of parental involvement in CBT have been found in previous meta-analyses (Dorsey et al., 2017; Thulin, Svirsky, Serlachius, Andersson, & Öst, 2014; Weersing et al., 2017).

Regarding methodological aspects, it is encouraging that the effect size and remission rates were not moderated by type of studies, statistical analysis, or by the treatment variables such as format, therapist profession, number of sessions, or weeks of treatments. These results provide confidence in the overall findings of the meta-analysis.

We found that higher risk-of-bias score was associated with higher ES, but not with remission. Although methodological flaws were noted in some of the studies, the total mean score was 1.6 (SD = 1.27) with a non-significant difference between the disorders. When excluding the uncontrolled trials, our overall results showed that it was more common with a low than a high risk-of-bias. It may be that the more recent RCTs carried out in routine clinical care apply the CONSORT statement (Altman, Moher, & Schulz, 2012) to a larger extent, and thus more information on the methods to evaluate risk-of-bias is provided.

All studies were evaluated on methodological aspects by using the psychotherapy outcome study methodology rating scale developed by Öst (2008). The results showed an overall mean of 22.28 for the different disorders. This result is comparable to a recent meta-analyses of brief, intensive and concentrated CBT treatment for anxiety in children (Öst & Ollendick, 2017), but somewhat lower than in a previous meta-analyses on OCD in children (Öst, Riise, Wergeland, Hansen, & Kvale, 2016). Comparisons between the different disorders yielded a non-significant difference, indicating an overall fair methodological quality of the included studies. However, regarding moderator analyses, a higher methodological score was associated with lower remission rates, but did not moderate effect size of continuous outcomes. This discrepancy makes it difficult to draw firm conclusions about the significance of methodological quality for the outcome of the included effectiveness studies.

We conducted a publication bias analysis and identified publication bias as a potential problem when using effect size as outcome measure for the included studies. For the overall effect size the trim-and-fill method indicated that 25 studies should be trimmed, which would have reduced the mean effect size from 1.50 to 0.94, a reduction of 37.3%. No publication bias was evident when using remission as outcome.

Using the trim-and-fill method and Egger's test, the effects of publication bias cannot be ruled out in evaluation the overall impact of effectiveness studies.

We compared our outcomes to efficacy studies to evaluate whether the magnitude of improvement achieved in routine clinical care is at the same level as randomized controlled trials from specialized research settings. With very few exceptions, this was the case. The effectiveness studies of mixed anxiety, OCD, PTSD, and depression generated post-treatment effect sizes very close to the effect sizes from efficacy studies (*p*-values 0.17–0.97), and the effect sizes achieved in both settings were in the very large range (1.18 to 2.50 at post- and 1.42 to 3.51 at follow-up assessment). Only for PTSD, there was a difference in the effect size at follow-up in favor of studies conducted in routine clinical care. Such a difference favoring PTSD was also found in a meta-analysis of effectiveness studies of CBT for adult anxiety disorders (Stewart & Chambless, 2009). Furthermore, remission rates for OCD, PTSD, and depression were comparable to the efficacy studies. Only in the case of mixed anxiety, the remission rate was lower in effectiveness studies at post-treatment, but not at follow-up. This finding for mixed anxiety at post-treatment may be explained by the difference in number of treatment sessions applied (12.0 for effectiveness and 15.1 for efficacy studies), as this has been related to outcome (Ishikawa et al., 2007). Importantly, there were no differences between the effectiveness and efficacy studies on effect sizes or remission rates when we excluded the uncontrolled effectiveness studies and only analyzed RCTs, providing confidence in the findings. Regarding attrition rates, our results corresponds to the attrition rates of 4.2% found in efficacy studies for depression (Yang et al., 2017), the 10.8% for mixed anxiety (James et al., 2015), the 12.7% for OCD (Öst et al., 2016), and the 14.4% for PTSD (Gutermann et al., 2016). The only difference in the background and treatment variables was fewer treatment sessions for mixed anxiety and lower attrition rate for depression in the effectiveness studies. There were no differences in mean age, proportion of females or pre-treatment severity between the effectiveness and efficacy studies for any of the disorders, indicating more similarities than differences between the samples on these variables. Overall, the results provide positive evidence for the generalizability of treatments developed in university research settings to routine clinical care. Moreover, our results are consistent with those of previous reviews finding that results of CBT for internalizing disorders from university research settings can be generalized to routine clinical care in youth (Lee et al., 2013), and adults (Hans & Hiller, 2013a, 2013b; Stewart & Chambless, 2009). Although CBT in general had positive effects on outcomes, there is still a large proportion of children and adolescents who do not recover after CBT. This finding is in line with recent meta-analyses of the effect of empirically supported youth psychotherapies across a broad variety of disorders, and indicates a need for improved therapies (Weisz et al., 2017; Weisz et al., 2019).

Some limitations warrant consideration. We only included peer-reviewed published or in press studies in English language journals. Studies published in other languages could have provided us with additional information about the effectiveness of CBT for internalizing disorders. Furthermore, the inclusion of only published studies could be viewed as a limitation. However, our pool of studies spanned three decades. Including unpublished studies could have introduced bias as it could have been easier to identify unpublished studies from more recent compared to earlier decades. Also, titles and abstracts were screened by one rater only. According to the PRISMA guideline the use of two independent raters when screening title and abstract may reduce the possibility of rejecting relevant reports. However, by reading 338 full-text articles and including 17.2% of them in the current meta-analyses the risk of missing many relevant studies that would have changed the results is probably low. Another limitation may be our classification of effectiveness studies. Although classification criteria were predefined and assessment could be made reliably by trained raters, studies differed on the quality of reporting the needed information. Thus, judgment was based on the limited and sometimes ambiguous information available, and perhaps some studies are missed that should have been included. There are examples in the literature where authors classify a study as effectiveness when it is classified by us as not (Ale, McCarthy, Rothschild, & Whiteside, 2015; Lee et al., 2013).

Furthermore, we cannot rule out that there may be other differences in the background and treatment variables between the effectiveness and efficacy studies that moderates treatment outcome, since we used the criterion that at least 75% of the effectiveness studies in our meta-analysis had to provide information on a variable to be included in the moderator analyses. The use of pre-post standardized mean difference to indicate treatment effects in meta-analyses has been problematized in a recent study, as it can contribute to biased outcomes and does not provide reliable information about the effects of the intervention (Cuijpers, Weitz, Cristea, & Twisk, 2017). However, for evaluation of improvement found in routine clinical care compared with improvement found in efficacy studies these analyses are still considered informative (Cuijpers et al., 2017; van der Lem, van der Wee, van Veen, & Zitman, 2012). It should also be noted that there were no third-wave CBT studies among the studies included in the meta-analysis, and the findings may therefore not apply to these forms of therapy. Finally, it may be claimed that the effectiveness of a CBT program is demonstrated when it exceeds the effects of the treatment youth and families usually received in the clinic, i.e. usual clinical care (Weisz, Kuppens, et al., 2013; Weisz, Ugueto, et al., 2013). However, our aim was to examine the degree of improvement that can be expected and the chance of achieving remission following CBT for internalizing disorders when provided by trained therapist in routine clinical care. Thus, a comparison with usual care was outside the scope of the current meta-analysis.

Future meta-analyses regarding CBT might want to include studies with single-case designs once consensus on how to calculate effect size has been reached and if such effect sizes can be combined with those from RCTs for an overall effect size. Furthermore, when rating the studies' methodological quality there was a lack of consistency in the reporting of the included studies. Although we selected studies that met our inclusion criteria, there was variability in the background and treatment data provided, e.g. reporting of attrition, intent to treat or completer analyses, and information of sample characteristics such as comorbidity and use of psychopharmacological treatments. Use of the POMRS, or a similar scale, is recommended in future meta-analyses. In addition, evaluation of treatment integrity and researcher allegiance effects might be valuable.

Our findings demonstrate overall encouraging treatment outcomes for internalizing disorders, and suggest that clinicians can be confident about the effectiveness of CBT treatments in routine clinical care. Adequately trained clinicians that provide these treatments in their work with youth in routine clinical care can achieve outcomes comparable to those in university research clinic settings both at post-treatment and at follow-up. At the same time, the results also suggest there is room for improvement. A substantial number of clients do not respond to the treatments currently available. Although treatment effects are not lost when programs are transported from research clinics to routine clinical care, there is a need to further develop and implement effective interventions for children and adolescents with internalizing disorders to improve outcome.

Acknowledgements

The authors wish to thank Regina Küfner Lein, academic librarian at the Medical library, University of Bergen Library, Norway, for assistance with the literature search.

Study funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contribution

GJW and LGÖ designed the study and wrote the protocol. GJW conducted literature searches in collaboration with an academic librarian. LGÖ wrote the coding scheme, rated the studies, meta-analyzed the included studies, and wrote the first draft of methods and results. ER and GJW extracted data and rated the studies. GJW wrote the first draft of the introduction and discussion. All authors contributed to and have approved the final manuscript.

Declaration of Competing Interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2020.101918.

References

Adler Nevo, G. W., Avery, D., Fiksenbaum, L., Kiss, A., Mendlowitz, S., Monga, S., & Manassis, K. (2014). Eight years later: Outcomes of CBT-treated versus untreated anxious children. *Brain and Behavior, 4*(5), 765–774. https://doi.org/10.1002/brb3.274.

Ale, C. M., McCarthy, D. M., Rothschild, L. M., & Whiteside, S. P. (2015). Components of cognitive behavioral therapy related to outcome in childhood anxiety disorders. *Clinical Child and Family Psychology Review, 18*(3), 240–251. https://doi.org/10.1007/s10567-015-0184-8.

Altman, D. G., Moher, D., & Schulz, K. F. (2012). Improving the reporting of randomised trials: The consort statement and beyond. *Statistics in Medicine, 31*(25), 2985–2997. https://doi.org/10.1002/sim.5402.

Asselmann, E., Wittchen, H. U., Lieb, R., & Beesdo-Baum, K. (2018). Sociodemographic, clinical, and functional long-term outcomes in adolescents and young adults with mental disorders. *Acta Psychiatrica Scandinavica, 137*(1), 6–17. https://doi.org/10.1111/acps.12792.

Baker-Ericzen, M. J., Hurlburt, M. S., Brookman-Frazee, L., Jenkins, M. M., & Hough, R. L. (2010). Comparing child, parent, and family characteristics in usual care and empirically supported treatment research samples for children with disruptive behavior disorders. *Journal of Emotional and Behavioral Disorders, 18*(2), https://doi.org/10.1177/1063426609336956.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis (version 2.2.057)*. Englewood, NJ: Biostat.

Borkovec, T. D., & Sibrava, N. J. (2005). Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. *Journal of Clinical Psychology, 61*, 805–818. https://doi.org/10.1002/jclp.20127.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284.

Costello, E. J., Egger, H., & Angold, A. (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: I. methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(10), 972–986. https://doi.org/10.1097/01.chi.0000172552.41596.6f.

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*(4), 364–368. https://doi.org/10.1017/S2045796016000809.

Dorsey, S., McLaughlin, K. A., Kerns, S. E. U., Harrison, J. P., Lambert, H. K., Briggs, E. C., ... Amaya-Jackson, L. (2017). Evidence base update for psychosocial treatments for children and adolescents exposed to traumatic events. *Journal of Clinical Child and Adolescent Psychology, 46*(3), 303–330. https://doi.org/10.1080/15374416.2016.1220309.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

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.

Freeman, J., Benito, K., Herren, J., Kemp, J., Sung, J., Georgiadis, C., ... Garcia, A. (2018). Evidence base update of psychosocial treatments for pediatric obsessive-compulsive disorder: Evaluating, improving, and transporting what works. *Journal of Clinical Child and Adolescent Psychology, 47*(5), 669–698. https://doi.org/10.1080/15374416.2018.1496443.

Gibby, B. A., Casline, E. P., & Ginsburg, G. S. (2017). Long-term outcomes of youth treated for an anxiety disorder: A critical review. *Clinical Child and Family Psychology Review, 20*(2), 201–225. https://doi.org/10.1007/s10567-017-0222-9.

Gutermann, J., Schreiber, F., Matulis, S., Schwartzkopff, L., Deppe, J., & Steil, R. (2016). Psychological treatments for symptoms of posttraumatic stress disorder in children, adolescents, and young adults: A meta-analysis. *Clinical Child and Family Psychology Review, 19*(2), 77–93. https://doi.org/10.1007/s10567-016-0202-5.

Hans, E., & Hiller, W. (2013a). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81*(1), 75–88. https://doi.org/10.1037/a0031080.

Hans, E., & Hiller, W. (2013b). A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clinical Psychology Review, 33*(8), 954–964. https://doi.org/10.1016/j.cpr.2013.07.003.

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology, 45*(2), 91–113.

Higgins, J. P. T., Altman, D. G., & Sterne, J. A. (2011). Assessing risk ofbias in included studies. In J. P. T. Higgins, & S. Green (Eds.). *Cochrane handbook for systematic reviews of interventions v. 5.1.0. S.l.* The Cochrane Collaboration.

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.

Hunsley, J., Elliott, K., & Therrien, Z. (2014). The efficacy and effectiveness of psychological treatments for mood, anxiety, and related disorders. *Canadian Psychology/Psychologie Canadienne, 55*(3), 161–176. https://doi.org/10.1037/a0036933.

Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*(1), 21–33. https://doi.org/10.1037/0735-7028.38.1.21.

Ishikawa, S. I., Okajima, I., Matsuoka, H., & Sakano, Y. (2007). Cognitive behavioural therapy for anxiety disorders in children and adolescents: A meta-analysis. *Child and Adolescent Mental Health, 12*(4), 164–172.

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*(1), 12–19.

James, A. C., James, G., Cowdrey, F. A., Soler, A., & Choke, A. (2015). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *The Cochrane Database of Systematic Reviews, 2*, Article CD004690. https://doi.org/10.1002/14651858.CD004690.pub4.

Kertz, S. J., Petersen, D. R., & Stevens, K. T. (2019). Cognitive and attentional vulnerability to depression in youth: A review. *Clinical Psychology Review, 71*, 63–77. https://doi.org/10.1016/j.cpr.2019.01.004.

Knight, A., McLellan, L., Jones, M., & Hudson, J. (2014). Pre-treatment predictors of outcome in childhood anxiety disorders: A systematic review. *Psychopathology Review, 1*(1), 77–129. https://doi.org/10.5127/pr.034613.

Kodal, A., Fjermestad, K., Bjelland, I., Gjestad, R., Öst, L. G., Bjaastad, J. F., ... Wergeland, G. J. (2018). Long-term effectiveness of cognitive behavioral therapy for youth with anxiety disorders. *Journal of Anxiety Disorders, 53*, 58–67. https://doi.org/10.1016/j.janxdis.2017.11.003.

La Greca, A. M., Silverman, W. K., & Lochman, J. E. (2009). Moving beyond efficacy and effectiveness in child and adolescent intervention research. *Journal of Consulting and Clinical Psychology, 77*(3), 373–382.

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, C. M., Horvath, C., & Hunsley, J. (2013). Does it work in the real world? The effectiveness of treatments for psychological problems in children and adolescents. *Professional Psychology: Research and Practice, 44*(2), 81–88. https://doi.org/10.1037/a0031133.

van der Lem, R., van der Wee, N. J., van Veen, T., & Zitman, F. G. (2012). Efficacy versus effectiveness: A direct comparison of the outcome of treatment for mild to moderate depression in randomized controlled trials and daily practice. *Psychotherapy and Psychosomatics, 81*(4), 226–234. https://doi.org/10.1159/000330890.

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.

de Lijster, J. M., Dieleman, G. C., Utens, E., Dierckx, B., Wierenga, M., Verhulst, F. C., & Legerstee, J. S. (2018). Social and academic functioning in adolescents with anxiety disorders: A systematic review. *Journal of Affective Disorder, 230*, 108–117. https://doi.org/10.1016/j.jad.2018.01.008.

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

Lundkvist-Houndoumadi, I., Thastum, M., & Hougaard, E. (2015). Effectiveness of an individualized case formulation-based cbt for non-responding youths with anxiety disorders. *Journal of Child and Family Studies*. https://doi.org/10.1007/s10826-015-0225-4 No Pagination Specified.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., ... Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the national comorbidity survey replication–adolescent supplement (ncs-a). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(10), 980–989. https://doi.org/10.1016/j.jaac.2010.05.017.

Nigatu, Y. T., Liu, Y., Uppal, M., McKinney, S., Rao, S., Gillis, K., & Wang, J. (2016). Interventions for enhancing return to work in individuals with a common mental illness: Systematic review and meta-analysis of randomized controlled trials. *Psychological Medicine, 46*(16), 3263–3274. https://doi.org/10.1017/S0033291716002269.

Öst, L. G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behavior Research and Therapy, 46*(3), 296–321.

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. *Behavior 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.

Piacentini, J., Bergman, R. L., Keller, M., & McCracken, J. (2003). Functional impairment in children and adolescents with obsessive-compulsive disorder. *Journal of Child and Adolescent Psychopharmacology, 13*(Suppl. 1), S61–S69. https://doi.org/10.1089/104454603322126359.

Riise, E. N., Kvale, G., Öst, L.-G., Skjold, S. H., & Hansen, B. (2018). Concentrated exposure and response prevention for adolescents with obsessive-compulsive disorder: A replication study. *Journal of Obsessive-Compulsive & Related Disorders, 19*, 15–22. https://doi.org/10.1016/j.jocrd.2018.07.002.

Riise, E. N., Kvale, G., Öst, L.-G., Skjold, S. H., Hansen, H., & Hansen, B. (2016). Concentrated exposure and response prevention for adolescents with obsessive-compulsive disorder: An effectiveness study. *Journal of Obsessive-Compulsive and Related Disorders, 11*, 13–21.

Scahill, L., Riddle, M. A., McSwiggin-Hardin, M., Ort, S. I., King, R. A., Goodman, W. K., ... Leckman, J. F. (1997). Children's Yale-brown obsessive compulsive scale: Reliability and validity. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(6), 844–852. https://doi.org/10.1097/00004583-199706000-00023.

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 randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal, 358*, j4008. https://doi.org/10.1136/bmj.j4008.

Smith, M. M., McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Kendall, P. C., & Weisz, J. R. (2017). Does the delivery of CBT for youth anxiety differ across research and practice settings? *Behavior Therapy, 48*(4), 501–516. https://doi.org/10.1016/j.beth.2016.07.004.

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*(4), 595–606. https://doi.org/10.1037/a0016032.

Thulin, U., Svirsky, L., Serlachius, E., Andersson, G., & Öst, L. G. (2014). The effect of parent involvement in the treatment of anxiety disorders in children: A meta-analysis. *Cognitive Behaviour Therapy, 43*(3), 185–200. https://doi.org/10.1080/16506073.2014.923928.

Trickett, P. K., Noll, J. G., & Putnam, F. W. (2011). The impact of sexual abuse on female development: Lessons from a multigenerational, longitudinal research study. *Development and Psychopathology, 23*(2), 453–476. https://doi.org/10.1017/S0954579411000174.

Weersing, V. R., Jeffreys, M., Do, M. T., Schwartz, K. T., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child and Adolescent Psychology, 46*(1), 11–43. https://doi.org/10.1080/15374416.2016.1220310.

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*(7), 671–689. https://doi.org/10.1037/0003-066X.61.7.671.

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry, 70*, 750–761.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., ... Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist, 72*(2), 79–117. https://doi.org/10.1037/a0040360.

Weisz, J. R., Kuppens, S., Yi Ng, M., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science, 14*, 216–237. https://doi.org/10.1177/1745691618805436.

Weisz, J. R., Ng, M. Y., Rutt, C., Lau, N., & Masland, S. (2013). Psychotherapy for children and adolescents. In M. J. Lambert (Ed.). *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 541–586). (6th ed.). Hoboken, N.J: Wiley.

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.

Wu, M. S., McGuire, J. F., Martino, C., Phares, V., Selles, R. R., & Storch, E. A. (2016). A meta-analysis of family accommodation and OCD symptom severity. *Clinical Psychology Review, 45*, 34–44. https://doi.org/10.1016/j.cpr.2016.03.003.

Yang, L., Zhou, X., Zhou, C., Zhang, Y., Pu, J., Liu, L., ... Xie, P. (2017). Efficacy and acceptability of cognitive behavioral therapy for depression in children: A systematic review and meta-analysis. *Academic Pediatrics, 17*(1), 9–16. https://doi.org/10.1016/j.acap.2016.08.002.