Clinical Psychology Review 40 (2015) 156–169

Contents lists available at ScienceDirect

Clinical Psychology Review

Cognitive behavioral treatments of obsessive–compulsive disorder. A systematic review and meta-analysis of studies published 1993–2014

Lars-Göran Öst [a,b,c,d,*], Audun Havnen [c,d], Bjarne Hansen [c,d], Gerd Kvale [c,d]

[a] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[b] Department of Psychology, Stockholm University, Sweden
[c] Department of Clinical Psychology, University of Bergen, Norway
[d] Haukeland University Hospital, OCD-team, 5021 Bergen, Norway

HIGHLIGHTS

- CBT yielded very large effect sizes compared to wait list and placebo.
- CBT was significantly better than antidepressants.
- The addition of antidepressants did not potentiate the effect of CBT.
- There was no significant difference between ERP and cognitive therapy.
- There was no significant difference between individual and group treatment.

ARTICLE INFO

Article history:
Received 14 January 2015
Received in revised form 14 May 2015
Accepted 11 June 2015
Available online 14 June 2015

Keywords:
Obsessive–compulsive disorder
Exposure and response prevention
Cognitive therapy
Meta-analysis
Systematic review

ABSTRACT

Obsessive–compulsive disorder is ranked by the WHO as among the 10 most debilitating disorders and tends to be chronic without adequate treatment. The only psychological treatment that has been found effective is cognitive behavior therapy (CBT). This meta-analysis includes all RCTs (N = 37) of CBT for OCD using the interview-based Yale–Brown Obsessive Compulsive Scale, published 1993 to 2014. The effect sizes for comparisons of CBT with waiting-list (1.31), and placebo conditions (1.33) were very large, whereas those for comparisons between individual and group treatment (0.17), and exposure and response prevention vs. cognitive therapy (0.07) were small and non–significant. CBT was significantly better than antidepressant medication (0.55), but the combination of CBT and medication was not significantly better than CBT plus placebo (0.25). The RCTs have a number of methodological problems and recommendations for improving the methodological rigor are discussed as well as clinical implications of the findings.

Contents

1. Introduction . . . 157
2. Method . . . 158
 2.1. Literature search . . . 158
  2.1.1. Inclusion criteria . . . 158
 2.2. Categorization of background variables and potential moderators . . . 159
  2.2.1. Conditions . . . 159
  2.2.2. Declining participation . . . 159
  2.2.3. Attrition . . . 159
  2.2.4. Type of exposure . . . 159
  2.2.5. Mode of application . . . 159
  2.2.6. Format and amount of therapy . . . 159
  2.2.7. Statistical analysis . . . 159
  2.2.8. Reliability of categorizations . . . 159

* Corresponding author at: Department of Clinical Neuroscience, Section of Psychology, Karolinska Institutet, S-171 77 Stockholm, Sweden.
E-mail address: ost@psychology.su.se (L.-G. Öst).

http://dx.doi.org/10.1016/j.cpr.2015.06.003
0272-7358/

2.3. Methodological quality . . . 159
2.3.1. The psychotherapy outcome study methodology rating scale . . . 159
2.3.2. Psychometric data . . . 159
2.4. Risk of bias . . . 160
2.5. Meta-analysis . . . 160
3. Results . . . 160
3.1. Description of the studies . . . 160
3.1.1. Background data . . . 160
3.1.2. Treatment data . . . 160
3.2. Methodological data . . . 160
3.2.1. Risk of bias . . . 161
3.2.2. Designs and comparisons . . . 161
3.2.3. Specific methodological issues . . . 161
3.3. Meta-analysis . . . 162
3.3.1. Attrition . . . 162
3.3.2. Primary measure . . . 162
3.3.3. CBT vs. Placebo . . . 162
3.3.4. Publication bias . . . 162
3.3.5. Secondary measures . . . 163
3.3.6. Response and clinically significant change . . . 163
3.3.7. Moderator analyses . . . 163
3.3.8. Within-group effect size . . . 163
4. Discussion . . . 164
4.1. Effect sizes . . . 164
4.2. Moderators . . . 165
4.3. Publication bias . . . 166
4.4. Methodological quality . . . 166
4.4.1. Specific methodological issues . . . 166
4.5. Treatment aspects . . . 167
4.6. Declining participation . . . 167
4.7. Attrition . . . 167
4.8. Limitations . . . 167
4.9. Clinical implications . . . 167
4.10. Recommendations for future research . . . 167
Funding . . . 167
Role of funding sources . . . 167
Contributors . . . 167
Conflict of interest . . . 168
Appendix A and B. Supplementary data and studies included in the meta-analysis . . . 168
References . . . 168

1. Introduction

Obsessive–compulsive disorder (OCD) is characterized by anxiety evoking intrusive thoughts, images or urges (obsessions) and repetitive behaviors aimed at reducing the discomfort (compulsions). The lifetime prevalence of OCD has been estimated to approximately 2% (Kessler et al., 2005) and OCD has been ranked by the WHO among the 10 most debilitating disorders. Untreated OCD tends to be chronic, causing significant functional impairment and reduced quality of life (Koran, Thienemann, & Davenport, 1996).

The most common treatment for OCD is pharmacological (Blanco et al., 2006), primarily with selective serotonin reuptake inhibitors (SSRIs). A meta-analysis by Soomro, Altman, Rajagopal, and Oakley-Browne (2008) found that SSRIs were significantly better than placebo and that the weighted mean difference on the Yale–Brown Obsessive Compulsive Scale (Goodman et al., 1989) was 3.21 points in favor of SSRIs.

The recommended treatment of choice for OCD is cognitive-behavioral therapy (CBT; National Collaborating Centre for Mental Health, 2006) which refers to exposure and response prevention (ERP) with or without the inclusion of cognitive therapy strategies. The most recent meta-analysis on the effects of psychological treatments for OCD was carried out by Olatunji, Davis, Powers, and Smits (2013). However, in this analysis only 16 studies were included, and three of these were on the treatment of pediatric OCD. The main reason for the small sample size with only 13 adult treatment studies may be the application of strict exclusion criteria, as the authors report that 21 studies were excluded due to having active treatment as control condition.

With the exception of Olatunji et al. (2013), the last extensive meta-analysis on the treatment of adult OCD was conducted by Rosa-Alcázar, Sánchez-Meca, Gómez-Conesa, and Marín-Martínez (2008), who included 19 controlled studies published in the period 1980–2006. Of the sample, which also included studies with a non-randomized design, 18 studies were published in the period from 1993 to 2006, thus overlapping with the inclusion period of the present meta-analysis. Since 2006 more than 20 controlled trials have been published on the treatment of OCD and it is evident that a significant proportion of the latest research on OCD has yet to be analyzed by meta-analytic procedures, and potentially twice the number of studies analyzed by Rosa-Alcázar et al. (2008) can be evaluated for inclusion in the present study. Eight of the studies included in the present meta-analysis overlap with the Rosa-Alcázar et al. (2008) study. Nine studies included in the Rosa-Alcázar et al. (2008) analysis were excluded from the present analysis as they failed to meet our inclusion criteria of randomized design ($n = 8$) or using Yale–Brown Obsessive–Compulsive Severity Scale (Y-BOCS; Goodman et al., 1989) as the primary outcome measure ($n = 1$). In summary, this warrants an updated meta-analysis on the cognitive behavioral treatment of adult OCD.

Despite randomization, RCTs may vary substantially in methodological stringency. In a recent meta-analysis of 60 RCTs on Acceptance and Commitment Therapy (Öst, 2014) a significant association between low methodological stringency and high effect size was found. In the

previously published meta-analyses on OCD, no systematic efforts were made to evaluate the methodological qualities of the included studies. To investigate whether methodological stringency and outcome are related in the treatment of OCD this issue will be explored in the present meta-analysis.

In the randomized trials on the treatment of OCD a wide variation in outcome measures have been applied and the differences in outcome measures represent a challenge when comparing studies. Although it may be assumed that by standardizing the outcome measures and comparing effect sizes the outcome will not be influenced, it can be argued that the standardization and comparison of multiple outcome measures may bias the calculated effect sizes because the standard deviations (SDs) may vary substantially between measures (Morris & DeShon, 2002). In order to avoid this potential bias we decided to apply a common outcome measure as criterion for inclusion in the present meta-analysis. The Y-BOCS (Goodman et al., 1989), which has been widely used as primary outcome measure in research on the treatment of OCD and has been established as the "gold standard" of OCD symptom measures, was chosen. Research has demonstrated only a moderate relationship between the interview and self-report version of Y-BOCS in a clinical sample of OCD patients (Federici et al., 2010). In the present analysis we therefore decided to only include studies that applied clinician administered Y-BOCS interview as outcome measure.

The calculation of effect sizes is a widely used means when comparing the results of outcome measures across studies. Effect sizes nicely demonstrate statistically significant changes following treatment and thus provide a basis for comparison of treatment outcome between studies; however, the effect sizes do not provide a way of determining *clinically significant improvement*. To overcome this limitation, Jacobson and Truax (1991) have recommended procedures for calculating clinical improvement and classifying patients accordingly in the categories "recovered", "improved", "unchanged" and "deteriorated". To be classified as recovered, the patient must a) show a change that is larger than the measurement error (the reliable change index; RCI) from pre- to post-treatment, and b) be in the range of the non-clinical population after treatment. In the present analysis the treatment outcome of the included studies will be analyzed and compared both in terms of effect size as well as clinically significant change; an approach not previously applied in meta-analyses of OCD.

ERP has often been described as a challenging treatment due to the confrontation to anxiety provoking cues and it has been estimated that approximately 25% of patients refuse the offer of treatment (Franklin & Foa, 1998), and this number is suggested to reflect that patients "find [behavior therapy] too frightening" (p. 353). Many studies have referred to this refusal rate, even though Franklin and Foa (1998) did not provide a reference for their estimate. It is thus unclear whether the estimate is valid and we will investigate this issue empirically by calculating the refusal rates in the included studies in our meta-analysis. A related question is how many patients drop out of treatment prematurely. An attrition rate of 25–30% has often been referred to (Abramowitz, 2006; Kozak, Liebowitz, & Foa, 2000), however, Kozak et al. (2000) based their estimate on only one study and Abramowitz (2006) did not provide a reference for his estimate. In general there is a huge variation in attrition rates across studies and in order to calculate a valid estimate it is necessary to include a large number of research trials. In the present meta-analysis we will therefore provide a calculation of attrition rates from all included randomized controlled trials published from 1993 to 2014.

ERP has been demonstrated as a treatment that can be disseminated in different modes of therapy and a relevant issue is whether exposure assignments are equally effective when self-administered as when administered by a therapist. In addition there is the question if type of exposure (in vivo versus imaginal exposure) has relevance for the outcome. Abramowitz (1996) published a meta-analysis evaluating different variants of ERP for OCD and the influence on outcome. He concluded that therapist assisted ERP was superior to self-administered ERP. Furthermore, he evaluated the type of exposure and concluded that the combination of in-vivo and imaginal exposure produced better outcome than in-vivo exposure alone. In addition the amount of therapy is known to vary substantially across different treatment formats, both with respect to length of sessions, frequency of sessions and length of the course of therapy. ERP has been disseminated across a range of different formats, e.g. group therapy, family-based interventions, etc., and the potential relation between treatment format and outcome will be analyzed in the present meta-analysis.

To sum up, there are many reasons that suggest the need for an update of the empirical basis for the psychological treatment of OCD and potential moderators of treatment. In the present meta-analysis we aim to provide an updated analysis of several important questions which are of relevance for the treatment of OCD in accordance with the following goals:

a) To provide an updated review and meta-analysis regarding the efficacy of cognitive-behavioral treatments of OCD from 1993 until 2014 using meta-analytic procedures.
b) To evaluate potential moderators for treatment outcome.
c) To evaluate the included studies according to the methodological criteria proposed by Öst (2008) and calculate if there is any difference between the first and the second 10 year period.
d) To provide recommendations for enhanced methodological stringency in future research on the basis of the methodological evaluation of OCD studies.

2. Method

2.1. Literature search

PsycINFO and PubMed were searched from 1993 to February 2015 with the following search words: obsessive–compulsive disorder or OCD, and exposure and response prevention or ERP or behavior therapy or cognitive therapy or cognitive behavior therapy, and Randomized controlled trial or RCT or random*. The reason to use 1993 as the starting year is that the first RCTs that used the Y-BOCS as outcome measure was published that year.

All abstracts were read and when there was an indication of a group of patients receiving the particular cognitive-behavioral treatment being compared with another condition in a randomized clinical trial (RCT) the full-text article was retrieved. Studies using single case designs were excluded since there is no consensus yet regarding the calculation of effect sizes for these designs. The reference lists in the retrieved articles were then checked against the database search and any other articles that might fulfill the inclusion criteria were retrieved.

2.1.1. Inclusion criteria

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

- be published, or in press, in an English language journal
- randomly allocate participants to either treatment and control, or to two or more active treatments
- have participants diagnosed with obsessive–compulsive disorder according to DSM or ICD
- use Yale–Brown Obsessive Compulsive Scale (Y-BOCS) clinician version pre- and post-treatment.

Excluded from the review and meta-analysis were studies using the self-report version of Y-BOCS. Fig. 1 shows a flowchart of the inclusion of studies in the present meta-analysis, which was conducted according to the PRISMA criteria (Liberati et al., 2009).

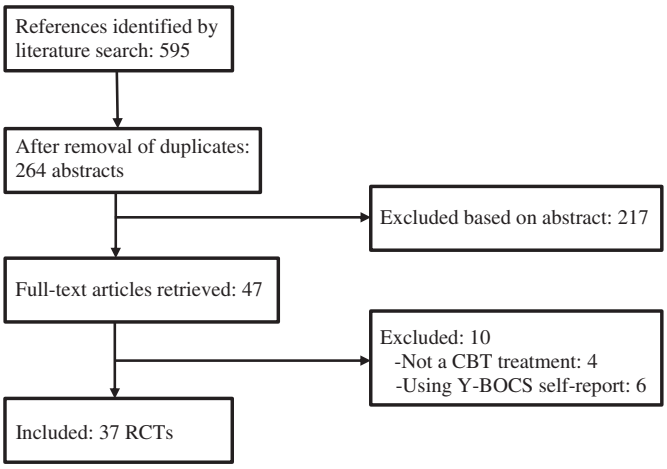

Fig. 1. Flowchart of the inclusion of studies.

2.2. Categorization of background variables and potential moderators

2.2.1. Conditions

Based on the information given in the method section of each article the treatments used for each condition were classified as: 1) exposure and response prevention (ERP), 2) CBT (either described as such or as the combination of ERP and cognitive therapy), 3) cognitive therapy (without any components of ERP), 4) ERP + psychotropic drugs, 5) ERP + placebo drugs, 6) CBT + psychotropic drugs, 7) ERP or CBT + another psychological treatment, 8) drugs alone, 9) placebo (drug or psychological) alone, and 10) waiting-list control (WLC). This classification was done independently by the first author, and by the other three authors yielding agreement in 94% of the cases. Classification of the remaining cases was done after a consensus discussion among the authors.

2.2.2. Declining participation

Declining is defined as fulfilling all criteria for participation but deciding not to participate for any reason, e.g. not wanting drug treatment in a study comparing CBT and drug, or believing that one will not be able to manage going through ERP treatment. However, we also include participants who accept randomization and sign the informed consent but never show up for the first therapy session. These are probably patients who in fact did not accept randomization but "gambled" and when they were randomized to another condition than what they preferred decided to not start the treatment.

2.2.3. Attrition

A patient who participates in at least the first session but then stops before the treatment period used in the study has come to an end is counted as a drop-out. This can occur at any time between the first and the last session and only a few studies have any information about when the attrition occurred. However, we do not count as dropouts people who completed the treatment but failed to show up for the post-treatment assessment.

2.2.4. Type of exposure

In the studies using ERP the type of exposure used was classified as in-vivo, in imagination, or a combination of both types of exposure.

2.2.5. Mode of application

The way the treatment was applied was classified as therapist-administered, self-administered, or a combination of therapist- and self-administered.

2.2.6. Format and amount of therapy

The format of treatment was classified as individual, group or family treatment. Amount of therapy was recorded as number of weeks, number of sessions, total number of hours, and intensity (hours/week).

2.2.7. Statistical analysis

The studies were classified according to the type of analysis used and presented in the article for which effect size could be calculated; either completer or intent-to-treat (ITT) analysis. In some cases ITT-analysis was done and the authors concluded that the results were the same as for completer analysis but no data were presented. In the rare case that both types of analyses were presented we used ITT-analysis.

2.2.8. Reliability of categorizations

The above potential moderators were independently classified by the first and the second author and there was 100% agreement on all variables.

2.3. Methodological quality

In order to assess the quality of the research methodology in RCTs various scales have been developed, e.g. the Jadad criteria (Jadad et al., 1996). They are, however, usually constricted to rather few items rated as present or absent. This means that the range of scores is small (e.g. 2–4 in Cavanagh, Strauss, Forder, & Jones, 2014) with ensuing difficulties of showing a relationship between methodological quality and effect size. Based on previous work by Tolin (1999) Öst developed a scale containing 22 items (Öst, 2008) with a theoretical range of 0–44. When used in the 2008 meta-analysis the total score for the included studies ranged from 10 to 36. Thus, there should not be a problem of "restriction-of-range" with this scale.

2.3.1. The psychotherapy outcome study methodology rating scale

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, and 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.

2.3.2. Psychometric data

The internal consistency of the scale was good with a Cronbach's α of 0.81. In order to assess the inter rater reliability of the scale in the present meta-analysis the following procedure was used. The second, third, and fourth author received 6 h of training in the use of the scale by the first author, with three (randomly selected) of the 37 RCTs included in the meta-analysis as training examples. Then they rated two studies independently and achieved high accuracy (intra-class correlation, ICC(3, 1) = .90). Finally, the second author rated all the remaining 30 studies and conferred with the third and fourth author to reach consensus. This was then compared with the rating done independently by the first author yielding an ICC(3, 1) = .93 for the total score. The *kappa* coefficients on the individual items varied between .50 and 1.00, with a mean of .73, indicating a good inter-rater reliability.

2.4. 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 participants and personnel is not possible in psychotherapy studies. The first author rated all studies and the second author independently rated a random selection of 20% (7 studies). There was perfect agreement between the two authors.

2.5. Meta-analysis

In the current meta-analysis the clinician administered Y-BOCS was used to calculate effect size for each study. This was the primary outcome measure in the included studies. Since patients with OCD often suffer from depression and general anxiety measures of these constructs were also included in the meta-analysis, and separate calculations of pooled effect size were done for each type of measure. We also aimed to look at the ES for quality of life measure but too few studies had any assessment of this concept to make it meaningful. We had to use the data included in each study, which in some studies (mainly older) were completer data and in some studies (mainly more recent ones) were intent-to-treat (ITT) data. When a study presented both sets of data ITT data were used.

The effect size (ES) was calculated as: $(M_{active\ treatment} - M_{comparison}) / SD_{pooled,}$ separately for post- and follow-up assessment. Before pooling the effect sizes we screened for statistical outliers, defined as being outside M ± 2SD. Five (5.9%) of the ESs were outliers. Instead of deleting those ESs from the analysis *winsorising* (Lipsey & Wilson, 2001) was used by reducing outliers to the exact value of M + 2SD. The software *Comprehensive Meta-Analysis, version 2.2.057* (CMA; Biostat Inc., 2006) was used for all analyses and to correct for small sample sizes Hedges's *g* was calculated. Cohen's rule-of-thumb for classification of ES was used; an ES of 0.20–0.49 is considered small, 0.50–0.79 as moderate, and ≥0.80 as large. A random effects model was used since it cannot be assumed that the ESs come from the same population as studies comparing CBT conditions with waitlist, placebo, medication and other forms of CBT are included in the meta-analysis.

Heterogeneity among ES's was assessed with the *Q*-statistic 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, 1997). Moderator analyses of continuous variables were carried out with meta-regression and for categorical variables with sub-group analysis using the mixed effect model.

3. Results

3.1. Description of the studies

3.1.1. Background data

Background data of the studies included in the meta-analysis are displayed in Appendix A, Table A.1. The 37 studies (see Appendix B) originated from USA (n = 10), Canada (n = 7), Brazil (n = 4), Australia (n = 3), Spain (n = 3), Holland (n = 2), Norway (n = 2), and one each from Denmark, France, Germany, Great Britain, Japan, and Sweden. A total of 2414 participants started treatment or control conditions. The proportion of females ranged from 29 to 77% with a mean of 57.9%, and the mean age of the samples varied from 28.6 to 40.0 years with a mean of 34.7. Only 23 of the 37 studies (62%) had information on the mean duration of the participants' OCD and this ranged from 6.0 to 26.3 with a mean of 15.2 years. However, it should be emphasized that very few studies describe whether they count this time period from the emergence of the first OCD-symptom or from when the full OCD diagnostic criteria were fulfilled.

The proportion of patients who declined participation of the study, either not giving informed consent or doing so but not showing up for the first session, varied from 0 to 63% with a mean of 15.0% and a median of 11%. The attrition rate (participating in at least the first session but stopping before the treatment used in the study has come to an end) varied from 0 to 32.3% with a mean of 15.0% and a median of 13.0%. Only 13 (35%) of the studies provided information on what proportion of their patients had received previous treatment for OCD; the figure ranged from 0 to 100% with a mean of 60.1%. Whether the participants were using psychotropic drugs for their OCD concurrently with their CBT was presented by 29 (78%) of the studies; the proportion ranged from 0 to 100% with a mean of 42.4%. The proportion of patients with comorbid axes I or II disorders was described by only 21 (57%) of the studies; the range was 40–83% with a mean of 61.4%.

3.1.2. Treatment data

Various treatment data for the OCD studies are described in Appendix A, Table A.2. The method of CBT in this body of studies was exposure and response prevention (ERP) with no components of cognitive therapy in 23 studies, cognitive therapy (CT) with no components of ERP in 7 studies, and the combination of ERP and CT in 13 studies. However, there are 6 studies in the first two categories comparing ERP and CT.

The type of exposure used was in-vivo in 21 studies, in imagination in 2 studies, and the combination of both in 14 studies. However, it is quite possible that a number of the studies saying that they used in-vivo exposure actually used the combination without explicitly saying so. The mode of application of treatment was therapist-administered in 29 studies, patient self-administered in 1 study, and the combination in 7 studies. The profession of the therapists was reported in 29 (78%) of the studies and the most common profession was psychologist (n = 18, 62%), followed by psychiatrist (n = 3, 10%), social worker (n = 1, 3%), and a mix of professions in 7 (24%) studies. The format of treatment was individual (n = 25, 68%), group (n = 5, 14%), individual versus group (n = 6, 16%), and individual versus individual plus family therapy (n = 1, 3%).

The duration of treatment varied between 3 and 24 weeks with a mean of 12.7 (SD 4.1) and the range of therapy sessions was 8–26 with a mean of 14.7 (SD 5.8). The number of minutes for each therapy session was taken into consideration and total treatment time in hours was calculated. This varied from 6 to 63 h with a mean of 21.5 (SD 12.8) for face-to-face therapy. However, there are two studies using internet-based CBT with much lower therapist time; 1.5 h in Wootton, Dear, Johnston, Terides, and Titov (2013) and 2.1 h in Andersson et al. (2012). In order to obtain the intensity of therapy the total hours of treatment was divided by the number of weeks each treatment lasted. This variable ranged from 0.7 (Belloch, Cabedo, & Carrió, 2008) to 10.0 (Lindsay, Crino, & Andrews, 1997) hours/week with a mean of 3.2 (SD 3.3). Here too the two internet-based CBT studies had the lowest values with 0.2 h, i.e. 12 min/week. It is however important to underscore that the patients who participated in the internet-based studies were almost all patients who were self-referred and wanted this treatment (92% in Andersson et al., 2012; 100% in Wootton et al., 2013). Moreover, patients with a Y-BOCS score as low as 12 were admitted, and patients with a Y-BOCS score of >31 were excluded (Andersson et al., 2012).

3.2. Methodological data

Table A.3 in Appendix A displays the mean scores on the psychotherapy outcome study methodology rating form for all RCTs and divided on the first (1993–2003) and the second time period (2004–2014). The mean sum score on the scale was 23.03 (SD 4.73) for all 37 studies and there was a significant increase ($t(35) = 2.37$, $p < .05$) from the first (M 20.36, SD 3.67) to the second (M 24.15, SD 4.73) time period. Using Bonferroni correction none of the individual items showed a difference between the first and the second time period. Without such a

correction items 13, manualized, replicable, treatment programs, and 18, control of concomitant treatments would have shown a significantly higher mean for the second time period.

3.2.1. Risk of bias

The risk of bias was unclear regarding random sequence generation and allocation concealment since almost all studies just described that the participants were randomly allocated without any further information. A low risk of bias was found in 20 studies (54%) concerning blinding of outcome assessment, in 17 studies (46%) regarding incomplete outcome data (they used intent-to-treat analysis), and 36 studies (97%) when it comes to selective reporting.

3.2.2. Designs and comparisons

The most common design (23 studies) was a comparison of two versions of CBT, e.g. ERP versus CT as in Cottraux et al. (2001) and Whittal, Thordarson, and McLean (2005). Some form of CBT was compared to a waitlist control in 8 studies and a placebo control in 7 studies. A comparison with an antidepressant (ADM) was done in 4 studies and CBT versus the combination of CBT plus ADM in 2 studies. Finally, CBT plus ADM was compared to CBT plus pill placebo in 3 studies. Since a study can combine two of these types of designs, e.g. Whittal, Woody, McLean, Rachman, and Robichaud (2010) which had both a placebo and a waitlist condition to compare cognitive therapy with, the total number of designs is 47.

3.2.3. Specific methodological issues

3.2.3.1. Statistical power. Psychotherapy outcome studies are usually very expensive and it is questionable to start such a study if it is clearly underpowered, i.e. if the chance of detecting a significant difference is markedly lower than the recommended 80%. The sample power table for *t*-test in Kazdin (2003, p. 444) indicates that if a researcher expects to obtain a large effect size ($d = 0.80$) 26 participants per condition is necessary for 80% power. However, if the expected effect size is moderate ($d = 0.50$) it takes 64, and if it is small ($d = 0.20$) the needed number is a staggering 400 per condition. Using the recommended 80% power and an α of 0.05, at randomization 97% of the OCD-studies would only detect a large (or higher) effect size, 3% would detect a moderate or large effect size, and none a small effect size. When taking attrition into consideration and looking at completers all studies could only detect a large effect size, i.e. had a cell size below 64. The mean cell size at the start of the studies (randomization) was 28 (SD 15.7) which means that on average a large effect size could be detected.

3.2.3.2. Reliably diagnosing the participants. In order for treatments evaluated in OCD-studies to be evidence-based it is important that participants are diagnosed, preferably by employing trained interviewers using established interview schedules (or similar instruments) and assessing inter-rater reliability. All 37 studies diagnosed the participants according to some version of the DSM manual and 32 (91%) used an established interview schedule. However, only three studies (Cordioli et al., 2003; McLean et al., 2001; Vogel, Stiles, & Götestam, 2004) tested the inter-rater reliability of the diagnostic procedure by letting independent interviewers blindly rerate a proportion of the video recorded diagnostic interviews.

3.2.3.3. Reliability of the primary outcome measure. The primary outcome measure in the studies of the present meta-analysis is the original interview version of the Yale–Brown Obsessive Compulsive Scale (Y-BOCS), which has been shown to have good psychometric characteristics (e.g. Goodman et al., 1989). However, just because the originator obtained good inter-rater reliability (IRR) with well-trained raters does not mean that RCTs using the scale automatically will get the same IRR figures. This has to be demonstrated in each individual study by letting an independent rater blindly rerate a proportion of the interviews (e.g. 20% randomly selected from each assessment point).

The best procedure from a methodological point of view is to use really independent assessors, i.e. clinicians trained in the use of Y-BOCS, who are not part of the research group, and are blind as to the hypothesis of and treatment conditions in the study. Furthermore, the blindness of the assessors should be tested by asking them to guess, after having done the Y-BOCS interview and ratings, which treatment (if any) the interviewed patient had received. Of the 37 studies in the current meta-analysis only 14 (38%) reported having used independent assessors; however, in most cases these were members of the research team, just not the therapists. A total of 16 (43%) studies reported that the assessors were blind. When the variable independence and blindness were combined we find that 9 (24%) of the studies had assessors who were independent *and* blind, 5 (14%) had independent assessors who were not blind, and 7 (19%) had assessors who were blind without any information about their independence. Only one study (Andersson et al., 2012) tested the blindness of the assessors and found no association between the assessor's guess and treatment condition. Finally, only 7 studies (19%) reported the IRR for their application of Y-BOCS and the coefficients varied between .66 and .98.

3.2.3.4. Number of therapists in the study. If only one therapist is used in a RCT there is a complete confounding between therapist and therapy method, and, consequently, it is not possible to ascribe a certain outcome to the therapy applied. In studies where only one therapist is doing both the compared therapies (e.g. Jaurrieta et al., 2008; Tolin et al., 2007) the therapist factor is controlled to some extent. However, unless adherence and competence ratings are provided in the article (which is not the case in these three studies) it is impossible to conclude that this single therapist carried out both treatments with equal adherence and competence. The OCD-studies are in general not very clear on this point. One study did not mention anything about the number of therapists, in 4 studies we guessed from the plural form therapists that it was more than 1, and in 16 more than 2 therapists. For 16 studies the actual number was mentioned and it varied from 1 to 7, and in 8 of the studies (23%) only one therapist was used. Only 2 therapists were used in 1 study but if we assume that the 4 studies where we guessed more than 1 actually had only 2 this make a total of 5 studies (14%). In total then 37% of the studies had only one or two therapists which means that confounding is quite prevalent in the OCD-studies of this meta-analysis.

3.2.3.5. Adherence and competence ratings. Adherence refers to the extent to which *specified procedures* are used by the therapist during the treatment, whereas competence concerns the degree of *skill and judgment* the therapist displays when carrying out the treatment (Barber, Sharpless, Klostermann, & McCarthy, 2007). These constructs are usually highly correlated (e.g. Barber, Liese, & Abrams, 2003) but cannot replace each other. A therapist can be highly adherent to the procedures in the manual, but not being particularly competent in the therapy situation. The opposite, i.e. a highly competent therapist who is not adherent, is more difficult to envisage. In such a case the therapist is probably doing some other therapy than he/she was supposed to do.

Adherence was assessed in only 5 (14%) of the studies and competence in only 2 (5%) of the studies. This means that it is very difficult to know if the patients actually received the treatment they were supposed to get and how competently it was delivered.

3.2.3.6. Credibility ratings. When two treatments are compared to each other in a RCT the patients' perceived credibility of the respective treatments are important to assess since differences in this respect may be a threat to internal validity of the study. Fully 35 of the 37 studies (95%) were comparisons between two, or more, active treatments. However, only 6 (17%) of these studies included credibility ratings. This issue is

particularly important when it comes to studies using a placebo control condition. Eleven of the studies had a condition with a psychological or pharmacological placebo and only 3 of these (27%) assessed credibility.

3.2.3.7. Control of drug treatment. It is well-known that antidepressive medication (SSRI's or tricyclics) is effective in the treatment of OCD (Soomro et al., 2008). This means that unless the use of these drugs is controlled in some way it may be a serious threat to the internal validity of the study. The way this issue usually is handled in CBT-studies is twofold: (i) the dose has to have been stable before inclusion in the study for the same number of weeks as the coming treatment period, and (ii) the patient agrees not to change dose or type of drug during the treatment period. However, in addition to this procedure a blood sample could be withdrawn to enable plasma concentration analyses of various psychotropic drugs. Only 21 (57%) of the studies used the twofold procedure and none reported plasma concentration analyses. The latter is especially important in the 5 studies comparing a CBT with a medication condition.

3.2.3.8. Drawing conclusion of equivalent effects from superiority designs. A non-significant difference on the primary measure does not allow the conclusion that the two compared treatments are equally good. This requires a noninferiority or an equivalence design (e.g. Walker & Nowacki, 2011). However, equivalence can be tested in a superiority design that yielded a non-significant effect, provided a large enough cell size (at least 30 according to the APA, Division12 Task Force criteria; Chambless et al., 1996; 1998). Out of the 37 RCTs in this review 26 (70%) were comparisons between two or more active treatments. These studies contained 32 comparisons of a form of CBT with another established treatment (e.g. a SSRI drug) or a variant of CBT, and 22 (69%) of the comparisons found no significant difference between the treatments. When reading the abstract and discussion sections of these studies we find that 11 (50%) described CBT and the compared treatment as yielding equivalent outcomes. However, none did an equivalence test, even though three of the studies (Jónsson et al., 2011; Fals-Stewart, Marks, & Schafer, 1993; Whittal et al., 2005) had cell sizes of 30 or more.

3.2.3.9. Assessment of response and clinically significant change. Already in 1984 Jacobson, Follette, and Revensdorf described a way to assess if the response a patient showed after therapy was statistically reliable (reliable change index, RCI) and criteria for considering a patient as recovered (clinically significant change, CSC). The RCI uses the pre-post change of the sample at hand, whereas the CSC is defined in three different ways: a) the cut-off is the mean of the normal population ± 2SD in the direction of dysfunctionality, b) the cut-off is the mean of the patient sample pre-treatment ± 2SD in the direction of functionality, and c) the average of a and b. Other ways of defining response and clinically significant change have also been used.

In the current body of RCTs only 21 studies (57%) reported data on response and 20 (54%) on clinically significant change. Three different criteria were used for treatment response. Most common (9 studies) was a certain percentage of reduction of the pre-treatment score. This varied from 25% (Simpson et al., 2010) to 50% (Cottraux et al., 2001), which in Y-BOCS score corresponded to 7.0 and 14.3 points, respectively. The second most common criterion was RCI, which in these studies varied from 5 to 10 points on Y-BOCS. Finally, three studies (Foa et al., 2005; Nakatani et al., 2006; Storch et al., 2007) used the Clinical Global Impression-Improvement (CGI-I) scale requiring a 1 or 2 on this scale. Concerning clinically significant change 18 of the studies used one of the versions of the Jacobson criteria and only one (Foa et al., 2005) used CGI-I requiring a 1. The studies using the former criteria describe a cut-off score on Y-BOCS that varied from ≤7 to ≤16 points. Thus the range is even larger than that for percent reduction as measure of treatment response. Using 16 is too lenient since many studies have 16 as an inclusion criterion of OCD symptom severity. This review shows that a consensus is warranted when it comes to criteria for treatment response and clinically significant change.

3.3. Meta-analysis

3.3.1. Attrition

Table 1 shows the dropout rate for the different treatments used in this meta-analysis. A subgroup analysis yielded a significant $Q_{between}$ (df 6) = 35.58, $p < 0.0001$. The dropout rates of antidepressants (30.3%) and ERP + antidepressants (32.0%) were significantly higher than for cognitive therapy (11.4%), CBT (15.5%), ERP (19.1%), placebo (18.8%) and waitlist control (13.7%). There were no significant differences between the other cognitive-behavioral treatments or the control conditions.

3.3.2. Primary measure

Table 2 shows the results of the meta-analysis at post-treatment for all comparisons and divided on the different types of comparison conditions at post-treatment. The overall Hedges's *g* was moderate (0.53) but significantly different from zero. Both indices of heterogeneity were also significant. The effect sizes for CBT compared with waiting-list (1.31), all placebo conditions (1.33), and psychological placebo only (1.29) were very large and also significantly heterogeneous, except for WLC. The ES for comparisons between individual and group treatment, and ERP with CT were less than small and non-significant. ERP/CBT was significantly better than ADM (0.55) with significant heterogeneity, whereas ERP + Placebo was nonsignificantly worse (-0.25) than ERP + medications.

Table 3 shows the effect sizes at follow-up assessment, on average 15.1 months after the end of treatment (for the 30 studies reporting follow-up data). The overall ES (0.06) was less than small, which is understandable since WLC conditions (and often placebo conditions) do not continue to follow-up. The individual versus group ES was small (0.21) and of the same magnitude as at post-treatment, which was also the case for the ERP versus CT effect size (0.07). However, the ERP/CBT vs. medication ES (0.38) was reduced from post-treatment and no longer significant, and so was the case for the ERP + Placebo vs. ERP + medication ES (−0.06).

3.3.3. CBT vs. Placebo

There are six studies with a total of eight comparisons of one form of CBT and psychological placebo (Andersson et al., 2012; Fals-Stewart et al., 1993; Lindsey et al., 1997; Nakatani et al., 2005; Whittal et al., 2010) or pill placebo (Foa et al., 2005). For each comparison the difference in pre- to post-treatment change on the Y-BOCS was calculated. The mean difference was 10.15 in favor of CBT, and the only ERP versus pill place comparison yielded a difference of 10.8.

3.3.4. Publication bias

The analyses of possible publication bias used both the trim-and-fill method and Egger's regression intercept. The results are shown in Table 4 and it is evident that publication bias is a potential problem

Table 1
Attrition in the OCD-studies.

Condition	*k*	Dropout	95% CI	*z*-Value	Q-value	I^2
ERP	28	19.1%	16.1–22.7%	13.22[b]	27.7	2.6
CT	8	11.4%	7.4–17.0%	8.55[b]	6.9	0
CBT	19	15.5%	12.5–19.2%	12.92[b]	22.8	21.0
Antidepressants	4	30.3%	23.5–38.3%	4.62[b]	9.0[a]	66.8
ERP/CBT + ADM	7	32.0%	24.2–40.9%	3.83[b]	14.5[a]	57.1
Placebo	6	16.8%	9.3–28.6%	4.58[b]	3.8	0
WLC	8	13.7%	7.9–22.6%	5.88[b]	4.9	0

[a] $p < .05$.
[b] $p < .0001$.

Table 2
Effect sizes (Hedges' *g*) on Y-BOCS for all OCD RCTs and divided on comparison conditions for post-treatment assessments.

Comparison	*k*	*g*-Value	95% CI	*z*-Value	Q-value	I^2
All studies	62	0.57	0.39–0.75	6.20[c]	305.4[c]	80
CBT vs. WLC	15	1.31	1.08–1.55	10.85[c]	22.3	37
CBT vs. placebo: all	8	1.33	0.91–1.76	6.18[c]	24.7[b]	72
CBT vs. placebo: psychological	6	1.29	0.76–1.81	4.81[c]	21.3[b]	77
CBT vs. all active Tx	37	0.09	−0.05–0.22	1.19	70.5[b]	49
Individual vs. Group Tx	6	0.17	−0.06–0.40	1.45	2.6	0
ERP vs. CT	7	0.07	−0.15–0.30	0.64	5.5	0
ERP/CBT vs. Medication	4	0.55	0.05–1.04	2.17[a]	9.7[a]	69
ERP/ERP + Pla. vs. ERP + Med	6	−0.25	−0.46–0.03	1.71	5.1	0

Note: *k* = number of comparisons. A positive *g*-value means that the first treatment in the comparison is better and a negative *g*-value means that the second treatment is better.
[a] $p < 0.05$.
[b] $p < 0.001$.
[c] $p < 0.0001$.

for the OCD RCTs. Regarding the overall ES the trim-and-fill method suggested that 9 studies should be trimmed which would reduce the mean ES from 0.57 to 0.31. Concerning the Placebo-, Individual vs. group-, and ERP/CBT vs. medication two studies each should be trimmed, leading to marked reductions of the ES. However, for the WLC-, ERP vs. CT-, and ERP + Placebo vs. ERP + medication comparisons no studies should be trimmed. For the overall ES, Placebo, and Individual vs. group comparisons Egger's regression intercept also yielded significant *t*-values.

3.3.5. Secondary measures

Even if it is well-known that depression (secondary to OCD) is the most prevalent comorbid disorder in OCD-patients (e.g. LaSalle et al., 2004) only 29 studies (78%) assessed depression pre- and post-treatment. A general measure of anxiety was used in only 17 studies (46%). The results of the meta-analyses are presented in Table 5. Regarding depression the ES was moderate and significant for the WLC-comparisons (0.60), but small and non-significant for all the other comparisons. The ES for anxiety was small but significant for all studies (0.40), and very large for the WLC-comparisons (1.19). None of the other comparisons yielded significant ESs for anxiety.

3.3.6. Response and clinically significant change

Table 6 presents the results for clinical significance. Regarding response the subgroup analysis yielded a significant $Q_{between}$ (df 5) = 45.90, $p < 0.0001$. The three types of CBT gave very similar proportions of response (62–68%), which was markedly higher than that for ADM (33%) and Placebo (27%). The combination of ERP/CT + ADM yielded a very high (80%) response rate but is only based on two studies.

Table 3
Effect sizes (Hedges' *g*) on Y-BOCS for all OCD RCTs and divided on comparison conditions for follow-up assessments.

Comparison	*k*	*g*-Value	95% CI	*z*-Value	Q-value	I^2
All studies	27	0.06	−0.13–0.24	0.60	55.4[b]	53
CBT vs. active Tx	25	0.04	−0.16–0.24	0.41	54.3[c]	56
Individual vs. Group Tx	6	0.21	−0.03–0.45	1.75	1.7	0
ERP vs. CT	4	0.07	−0.27–0.41	0.39	3.7	20
ERP/CBT vs. Medication	2	0.38	−0.81–1.57	0.62	6.1[a]	84
ERP/ERP + Pla. vs. ERP + Med	3	−0.06	−0.66–0.55	0.18	3.3	39

Note: *k* = number of comparisons. A positive *g*-value means that the first treatment in the comparison is better and a negative *g*-value means that the second treatment is better.
[a] $p < 0.05$.
[b] $p < 0.001$.
[c] $p < 0.0001$.

Regarding CSC the subgroup analysis also yielded a significant $Q_{between}$ (df 5) = 12.66, $p < 0.05$. The three types of CBT and the combination with medication gave similar proportions of CSC (43–52%), which were higher than for placebo or medication (12–14%).

As described in the Method section the criterion for response varied from 7.0 to 14.3 points reduction on Y-BOCS, and the cut-off for CSC from 7 to 16 points on Y-BOCS. Meta-regression analysis showed that the proportion of response was higher the lower the criterion was ($z = -4.17$, $p < 0.0001$). For CSC the analysis showed a trend for higher proportion of CSC the more lenient the cut-off criterion was ($z = -1.64$, $p = 0.10$).

3.3.7. Moderator analyses

The following continuous variables, for which at least 75% of the studies provided information, were analyzed with the meta-regression module in the CMA program using fixed effect analysis: proportion of patients declining participation in the study, number of participants starting treatment, proportion of dropouts, proportion of females, mean age of the participants, proportion of participants receiving drug treatments concurrently with the CBT, weeks of treatment, number of sessions, number of therapy hours, intensity of treatment (hours/week), year of publication, and methodological quality of the study. The following variables yielded a significant point estimate of the slope. Studies with higher proportion of women ($z = -2.05$, $p = 0.04$), studies with higher mean age ($z = -2.82$, $p = 0.005$), and studies with higher proportion of patients concurrently receiving antidepressants ($z = -2.65$, $p = 0.008$), all were associated with lower ES. None of the treatment variables were significantly associated with ES.

A specific meta-regression analysis was done on pre-treatment Y-BOCS score for placebo controlled studies. This included six studies and a total of eight comparisons; six of which used psychological and two pill placebo. The mean pre-treatment Y-BOCS score for these studies varied between 17.7 and 29.6, and the post-treatment ES between 0.48 and 3.04. The point estimate of the slope was significant ($z = 4.68$, $p < 0.0001$), which indicates that the higher the OCD-severity at pre-treatment the larger the post-treatment ES favoring CBT over placebo.

For categorical variables sub-group analyses were employed in the CMA program and the results are displayed in Table 7. Four of the six variables yielded significant $Q_{between}$ values. Completer analyses yielded higher ES than intent-to-treat analyses, passive control condition (WLC) gave higher ES than active control, individual treatment was better than group treatment, studies assessing therapist competence yielded higher ES than those that did not, whereas control of drug treatment did not influence ES.

3.3.8. Within-group effect size

The within-group effect sizes are displayed in Table 8. For all conditions combined as well as the individual treatment conditions the ESs were very large and significantly different from zero, both at post-treatment and at follow-up assessment. The difference between post- and follow-up assessment were generally small and non-significant. Placebo and WLC were only assessed at post-treatment since the participants in these conditions obtain treatment afterwards. The placebo ES was moderate and significant but the WLC ES was small.

Sub-group analysis of the post-treatment effect sizes including all sub-groups yielded a significant $Q_{between}$ (df 7) = 264.0, $p < 0.0001$. This was mainly due to the low effect sizes for Placebo and WLC. When only active treatments were compared the Q-value was reduced to 16.46, $p = 0.006$. However, when only CT, ERP, ERP + CT and Medications were compared the Q-value was not significant (6.46, $p = 0.09$). Finally, when the three CBT-conditions were separately compared with Medications only ERP yielded a significantly higher ES than Medications ($Q_{between}$ (df 1) = 5.56, $p = 0.018$).

Meta-regression analysis was used to evaluate if pre-treatment OCD-severity affected the within-group effect size. For all CBT-conditions

Table 4
Publication bias data for the different comparisons in OCD RCTs.

Comparison	Observed ES	Trim-and-fill ES	# of trimmed studies	Egger's regression intercept	*t*-Value
All studies	0.57	0.31	10	3.19	3.27[b]
CBT vs. WLC	1.31	1.26	1	3.17	1.73
CBT vs. placebo: all	1.33	1.03	2	4.58	4.15[b]
CBT vs. Placebo: psychological	1.29	0.91	2	4.41	3.13[a]
CBT vs. all active Tx	0.09	0.02	3	0.44	0.62
Individual vs. Group Tx	0.17	0.09	2	2.58	3.95[a]
ERP vs. CT	0.07	0.07	0	−0.28	0.89
ERP/CBT vs. Medication	0.55	0.17	2	3.98	3.56
ERP/ERP + Pla. vs. ERP + Med	−0.25	−0.25	0	0.87	0.82

[a] $p < 0.05$
[b] $p < 0.01$

($k = 65$) the point estimate of the slope was significant ($z = 5.28$, $p < 0.0001$), indicating larger ES as severity increased. Furthermore, analyses were done for the CBT versus placebo comparisons. These showed that CBT yielded a significant increase in ES as severity increased ($z = 3.68$, $p = 0.0002$), whereas placebo showed a marginally significant decrease in ES as severity increased ($z = -1.81$, $p < 0.07$).

4. Discussion

The primary aims of the present meta-analysis were to provide a methodological review of the randomized controlled trials of cognitive behavioral treatment of OCD in adult patients published from 1993 to 2014, and to assess their efficacy. A total of 37 studies met criteria for inclusion, which makes this the largest meta-analysis of randomized controlled trials for OCD. In accordance with previous, considerably smaller meta-analyses on cognitive behavioral treatment of OCD (e.g. Abramowitz, 1996; Olatunji et al., 2013; Rosa-Alcázar et al., 2008) the results supported the effectiveness of ERP with or without cognitive therapy elements in reducing obsessive–compulsive symptoms.

4.1. Effect sizes

The meta-analysis yielded an overall effect size of 0.57, with large effect sizes for comparisons between CBT and waiting-list, (1.31), all placebo conditions (1.33), and psychological placebo (1.29). These effect sizes are somewhat higher than that reported by Rosa-Alcázar et al. (2008) who included both WLC and placebo controls ($k = 19$) finding a $g = 1.09$. The effect sizes are very similar to that found by Olatunji et al. (2013) who included both adult and pediatric RCTs with WLC or placebo controls ($k = 16$) and reported a Hedges' g of 1.39. Hofmann and Smits (2008) also reported a very similar g-value as we found (1.37) based on three placebo controlled RCTs. It is probably safe to conclude that psychological or pill placebo treatments do not work for OCD-patients.

Studies with passive control conditions had higher ES than active control conditions. This is a trivial finding since decades of research have shown that CBT outperforms passive controls. However, CBT also yielded a significantly higher ES ($g = 0.55$) when compared with medication, albeit this comparison only included four studies since we did not include studies comparing only drug with drug plus CBT treatment without a CBT only condition. This ES was somewhat higher than that obtained in a recent meta-analysis by Romanelli, Wu, Gamba, Mojtabai, and Segal (2014) with four studies reporting a $d = 0.34$, which was marginally significant ($p = 0.073$). An old meta-analysis of Kobak, Greist, Jefferson, Katzelnick, and Henk (1998) also found a significantly higher effect size for ERP than for SRIs. We also found that CBT alone was not inferior to CBT combined with pharmacological treatment ($g = -0.25$). The Romanelli et al. (2014) meta-analysis did not report a mean ES but none of the four studies they included had a d-value significantly different from zero, thus corroborating our result.

There were six studies directly comparing individual and group treatment yielding a nonsignificant ES of 0.17. However, this finding is contradicted by the result of our moderator analysis (using sub-group analysis) which found that individual ($g = 0.50$) did significantly better than group ($g = 0.24$) in the total analysis with all types of comparison conditions (see Table 7). Previous meta-analyses have not focused on head-to-head comparisons but compared how individual and group treatment, respectively, did versus various control conditions. Rosa-Alcázar et al. (2008) reported 0.97 for individual ($k = 14$) and 1.08 for group ($k = 6$) treatment, whereas the corresponding figures in Olatunji et al. (2013) were 1.24 for individual ($k = 10$) and 1.53 for group ($k = 9$) treatment. In addition, Jónsson and Hougaard (2009), focusing only on group CBT found a mean ES of 1.12. This indicates that the format of treatment does not seem to affect the outcome of CBT for OCD-patients.

There were also six studies ($k = 7$) directly comparing ERP with cognitive therapy without systematic exposure, but using behavioral experiments as an integrated component which has been the case since CT was developed for anxiety disorders in the 1980s. The mean ES for these comparisons was 0.07, indicating no significant difference in effect between ERP and CT. Since this result might be unexpected we also analyzed the raw Y-BOCS scores. The ERP conditions had the following means: pre-treatment 23.5, post-treatment 11.9, and follow-up 13.0, and for the CT conditions the corresponding means were 25.2, 12.3, and 11.1. The percent change from pre-treatment was 49% at post and 47% at follow-up for the ERP conditions, whereas it was 51% and 55%, respectively, for the CT conditions. None of these differences were significant. Thus, it can be concluded that modern forms of CT and ERP do not differ significantly in clinical outcome.

Table 5
Effect sizes (Hedges' g) on secondary measures for all OCD RCTs and divided on comparison conditions for post-treatment assessments.

Comparison	Measure	*k*	*g*-Value	95% CI	*z*-Value	Q-value	I^2
All studies	Depression	48	0.12	−0.02–0.26	1.68	103.2[c]	54
	Anxiety	33	0.40	0.17–0.63	3.38[b]	127.4[c]	75
CBT vs. WLC	Depression	14	0.60	0.42–0.78	6.48[c]	6.3	0
	Anxiety	11	1.19	0.97–1.41	10.61[c]	10.8	8
CBT vs. Placebo: psychological	Depression	4	0.09	−0.18–0.37	0.68	2.3	0
	Anxiety	3	0.08	−0.30–0.46	0.41	1.8	0
Individual vs. Group Tx	Depression	4	0.28	−0.20–0.76	1.14	7.4	59
	Anxiety	3	0.40	−0.23–1.03	1.25	6.2[a]	67
ERP vs. CT	Depression	7	−0.01	−0.29–0.26	0.10	8.2	27
	Anxiety	4	−0.28	−0.77–0.21	1.14	4.8	38
ERP/CBT vs. Medication	Depression	2	−0.47	−1.06–0.13	1.54	1.4	30
	Anxiety	2	−0.13	−0.59–0.33	0.55	1.0	0
ERP/ERP + Pla. vs. ERP + Med	Depression	4	−0.33	−0.69–0.04	1.73	3.3	10
	Anxiety	3	−0.27	−0.71–0.17	1.19	3.0	34

Note: k = number of comparisons. A positive g-value means that the first treatment in the comparison is better and a negative g-value means that the second treatment is better.
[a] $p < 0.05$.
[b] $p < 0.001$.
[c] $p < 0.0001$.

Table 6
Proportions of response and clinically significant change (CSC) at post-treatment.

Comparison	Measure	*k*	Proportion	95% CI	*z*-Value*	Q-value	I^2
ERP	Resp.	15	65.0%	59.3–70.2%	5.02[d]	12.1	0
	CSC	13	50.0%	44.0–56.0%	0.01	7.3	0
CT	Resp.	5	67.7%	52.0–80.2%	2.20[a]	9.4	57
	CSC	4	51.6%	27.8–74.8%	0.13	17.5[c]	83
ERP + CT	Resp.	13	62.4%	51.3–72.3%	2.18[a]	46.0[d]	74
	CSC	13	43.4%	34.5–42.8%	−1.38	34.0[c]	65
Antidepressants	Resp.	4	32.9%	25.8–40.9%	−4.06[d]	2.6	0
	CSC	1	14.0%	5.2–32.6%	−3.27[c]		
ERP/CT + ADM	Resp.	2	79.5%	56.0–92.2%	2.38[a]	2.0	51
	CSC	1	43.0%	23.3–56.2%	−0.61		
Placebo	Resp.	3	26.8%	2.5–84.1%	−0.74	10.3[b]	81
	CSC	3	12.0%	1.4–56.7%	−1.73	15.2[c]	87

Note: *k* = number of comparisons.
[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.
[d] $p < 0.0001$.
* Test if significantly different from 50%.

Finally, the six studies that compared CBT with placebo found that there was a large difference in change scores between the conditions, a mean of 10.18 Y-BOCS scores at post-treatment. This is more than three times the difference between SSRIs and pill placebo (3.21) reported in the meta-analysis by Soomro et al. (2008). It is, of course, difficult to compare across meta-analyses but the only study in the present meta-analysis that used ERP, an SRI (clomipramine) and pill placebo found that the ERP-pill placebo difference was 10.8, whereas the SRI-pill placebo difference was 5.3. A tentative conclusion is that CBT leads to larger improvements on the severity scale than SSRIs do.

4.2. Moderators

The meta-analysis showed significant heterogeneity among the effect sizes in the overall analysis with all comparisons. We investigated a number of potential moderators and found three moderators significantly influencing the ES. Studies with higher proportion of women, studies with higher mean age, and studies with higher proportion of patients having antidepressant medication were associated with lower ES. It is difficult to explain why a majority of women in the sample is related to lower ES. Neither Rosa-Alcázar et al. (2008), nor Olatunji et al. (2013) found that gender affected the ES significantly. The finding that older age was associated with lower ES is not corroborated by the meta-analyses of Rosa-Alcázar et al. (2008) and Olatunji et al. (2013) which both found that mean age of the sample was unrelated to ES. One explanation to this difference in outcome might be that we included all 59 comparisons in the 35 studies, whereas the previous meta-analyses only compared CBT with control conditions.

The negative association between proportion of patients having antidepressant medication and ES is probably not due to higher initial OCD-severity in those samples since we found that the higher the initial severity the larger the pre-post change on Y-BOCS. It might be due to less motivation to adhere to the treatment components in those samples. This is, however, difficult to evaluate since motivation is rarely assessed in these RCTs.

We found a strong positive association between initial mean Y-BOCS score and post-treatment ES for the placebo controlled studies. Also, the within-group ES increased with higher severity for the CBT conditions, whereas it remained stable for the placebo conditions. This finding corroborates that of Olatunji et al. (2013) on OCD. Also, Kirsch et al. (2008) in a meta-analysis on antidepressants for major depression found that only from a pre-treatment mean of 27 on the Hamilton Depression Rating Scale was there a significant difference between active drug and placebo.

Table 7
Subgroup analyses of the overall effect size of OCD RCTs at post-treatment.

Variable	*k*	*g*	95% CI	Q_b-value	*p*-Value
Type of data analysis				10.51	0.001
Completer	46	0.55	0.45–0.64		
Intent-to-treat	16	0.27	0.14–0.41		
Type of comparison				58.46	0.0001
Active treatment	46	0.30	0.22–0.39		
Passive control	16	1.08	0.90–1.25		
Treatment format				5.98	0.014
Individual	56	0.50	0.44–0.56		
Group	6	0.24	0.06–0.43		
Control of drug Tx				3.87	0.049
Yes	47	0.41	0.32–0.50		
No	15	0.60	0.43–0.77		
Assessment of adherence				2.40	0.121
Yes	14	0.57	0.40–0.74		
No	48	0.42	0.33–0.51		
Assessment of competence				10.86	0.001
Yes	4	0.90	0.62–1.18		
No	58	0.41	0.33–0.50		

Note: *k* = number of comparisons, Q_b = Q between subgroups.

Table 8
Within-group effect sizes (Hedges' *g*) for all treatment conditions at post-treatment and follow-up assessments.

Comparison	Time point	*k*	*g*-Value	95% CI	*z*-Value	Q-value	I^2
All conditions	Post	87	1.83	1.62–2.04	17.07[c]	810.5[c]	89
	F-up	60	2.12	1.90–2.34	19.10[c]	299.7[c]	80
ERP	Post	27	2.06	1.77–2.36	13.74[c]	104.7[c]	75
	F-up	21	1.77	1.55–1.99	15.66[c]	40.6[c]	51
CT	Post	8	2.21	1.59–2.83	6.97[c]	39.7[c]	82
	F-up	6	2.14	1.50–2.78	6.57[c]	26.8[c]	81
ERP + CT	Post	18	1.90	1.52–2.28	9.79[c]	114.5[c]	85
	F-up	15	2.25	1.82–2.69	10.25[c]	74.1[c]	81
Individual Tx	Post	46	2.31	2.04–2.58	16.90[b]	274.3[c]	84
	F-up	36	2.17	1.90–2.44	15.86[c]	172.1[c]	80
Group Tx	Post	6	1.36	0.86–1.86	5.30[b]	22.5[b]	78
	F-up	6	1.46	1.18–1.74	10.15[c]	7.0	29
Medications	Post	4	1.47	1.03–1.90	6.21[c]	8.6[a]	65
	F-up	2	3.03	1.72–4.33	4.53[c]	4.0[a]	75
ERP/CT + Med.	Post	7	2.95	2.02–3.88	6.19[c]	32.0[c]	81
	F-up	5	2.72	1.59–3.84	4.71[c]	20.6[c]	81
ERP + Placebo	Post	5	2.68	2.21–3.14	11.22[c]	2.4	0
	F-up	5	2.45	1.96–2.94	9.82[c]	3.5	0
Placebo	Post	6	0.53	0.25–0.81	3.74[b]	11.8[a]	58
WLC	Post	8	0.10	−0.06–0.26	1.29	1.9	0

Note: *k* = number of comparisons.
[a] $p < 0.05$.
[b] $p < 0.001$.
[c] $p < 0.0001$.

The sub-group analyses showed that studies using completer analysis yielded significantly higher ES than those with intent-to-treat analysis, and studies with passive control conditions had higher ES than those with active treatment comparisons. These are well-known findings from many meta-analyses and point to the necessity of using active control designs and ITT-analysis in future research. Studies that assessed competence resulted in higher ES than studies not including this important methodological factor, which is encouraging. Our finding should be interpreted with caution since only 4 comparisons involved competence. Also, Olatunji et al. (2013) combined adherence and competence to a factor called integrity checks and found no difference between studies having and not having these checks, respectively.

4.3. Publication bias

We conducted a publication bias analysis and from this it is clear that publication bias is a potential problem for the included OCD studies. For the overall ES the trim-and-fill method indicated that 10 studies should be trimmed, which would have reduced the mean ES from 0.57 to 0.31, a reduction of 45.6%. Rosa-Alcázar et al. (2008) used Egger's test which was significant and the fail-safe N which was not significant and concluded that publication bias could be discarded as a serious threat. Olatunji et al. (2013) only used the fail-safe N and also concluded that the obtained ES was robust. However, the fail-safe N is not considered to be sensitive enough to detect publication bias and modern meta-analyses use the trim-and-fill method and Egger's test much more often. Using these methods we have to conclude that publication bias may be a threat to the validity of the obtained overall ES.

4.4. Methodological quality

All studies were evaluated on methodological aspects by using the psychotherapy outcome study methodology rating scale developed by Öst (2008). The 20 year inclusion period from 1993 to 2014 was divided in two 10 year time periods in order to study changes in methodological quality as an effect of time. The results showed significantly higher methodological rating scores in studies published in the latter period compared to the former, which indicates that the methodological stringency in treatment studies of OCD has increased with time. A recent meta-analysis on Acceptance and Commitment Therapy (Öst, 2014) found a significant association between methodological quality and effect size, whereas the present study did not show such a relationship. One difference between the meta-analyses is that the mean was lower (18.8 vs. 23.0) and the range wider (10–34 vs. 15–34) in the ACT compared to the present meta-analysis.

4.4.1. Specific methodological issues

The Y-BOCS was used as the primary outcome measure in the present meta-analysis. It has been shown to have good psychometric properties; however, unless the inter-rater reliability (IRR) actually is measured in a RCT it is not correct to assume that one would have the same IRR as in the original study by Goodman et al. (1989). Only 6 of the included studies reported the IRR and this represents a common methodological problem. We therefore suggest that future studies should use independent raters who are not part of the research team and are blinded to treatment modality, and in addition the IRR should be measured on a random sample of 20% of the interviews (of both included and excluded patients) throughout the study.

To what degree the treatment manual has been adhered to and with what level of competence the treatment was given is of obvious importance for any treatment study. In the present body of RCTs, ratings of adherence and competence were performed in only 14% and 5% of the studies, respectively. The results also showed that studies assessing competence yielded significantly higher ES than studies without this assessment. The low number of studies conducting this assessment is a problem as it is impossible to know if the outcome of the study has been influenced by the extent to which the manual has been followed and it is difficult to detect potential variability in outcome related to therapists' adherence. In addition, the competence with which the treatment was delivered was only evaluated in a small fraction of the studies, which obviously is a potential problem when comparing different treatment modalities.

To be able to ascribe outcome to therapy in a RCT it is necessary to be able to control the influence of the therapist. The number of therapists in the included studies is overall vaguely described. Based on the information given in the studies we found that as many as 37% of the studies had only one or two therapists, which means that there is a high degree of confounding between therapist and therapy method in OCD studies. This limitation could easily be overcome by using a higher number of therapists and analyzing the therapist effect on the outcome.

Pharmacological treatment of OCD has received empirical support (e.g. Soomro et al., 2008) and many patients with OCD have SSRI medication prescribed. Control of concomitant drug treatment must therefore be performed to reduce threats to the internal validity in any treatment study. One common way to handle the problem of concurrent drug treatment is to use the inclusion criterion that a patient who has a prescribed psychotropic drug at inclusion of the study must have been on a stable dosage for the same number of weeks as the treatment takes (e.g. 12 weeks), and that he/she accepts not to change the medication during the treatment. In that way the effect of the psychological treatment can be evaluated over and above the effect of the drug. Only 21 (57%) of the studies used this criterion and none assessed plasma concentration of psychotropic drugs to evaluate if patients adhered to the criterion. This is a limitation of the included studies in general and in particular for the five studies comparing CBT and pharmacological treatment.

Regarding assessment of clinical response and clinically significant change there is a large variability across the included studies. Treatment response was reported either as a percentage of reduction from pre- to post-treatment, reliable change index (RCI), or a certain score the Clinical Global Impression-Improvement (CGI-I) scale. This is also true for the use of cut-off scores on Y-BOCS that varied from ≤7 to ≤16 points. We would argue that a cut-off score of 16 is too lenient since many studies have 16 (or even less) as a criterion for inclusion. For example in the study by Andersson et al. (2012) a Y-BOCS score of 12–31 was an inclusion criterion, which means that the sample included patients who would be fulfilling the cut-off for clinically significant change in other studies. Yet other studies do not explicitly state whether total Y-BOCS score was used as an inclusion criteria.

It is evident from the present meta-analysis that a consensus is warranted with respect to criteria for treatment response and clinically significant change. Fisher and Wells (2005) used the methodology described by Jacobson and Truax (1991) on five CBT RCTs and arrived at a response criterion of at least 10 points reduction from the pre-treatment score, and a post-treatment recovery criterion of ≤14 on the Y-BOCS. Simpson, Huppert, Petkova, Foa, and Liebowitz (2006) analyzed the Foa et al. (2005) RCT in these respects, comparing four criteria for response and three for remission. They recommended a decrease of ≥25% on the Y-BOCS as response criterion and a score of ≤12 as remission criterion. Their response criterion is more lenient than the 10 point reduction suggested by Fisher and Wells, whereas their remission criterion is more stringent.

A major methodological limitation of many of the included studies concerns power analysis. Cohen (1988) recommended that a power analysis is conducted prior to the research and that the sample size is decided accordingly. The methodological rating scale revealed that during the first time period the included studies obtained a mean score of 0.0 on item 11 (Power analysis). The mean score for the second time period was 0.21, which means that conducting a power analysis is rare in OCD treatment research. Furthermore, 97% of the studies had 80% power to detect only a large effect size, 3% would detect a moderate effect size, and none would have detected a small effect size. This is a

problem that OCD RCTs share with most areas of psychotherapy research and researchers could learn from Kazdin's (2003) suggestions on how to increase the power in RCTs.

A related aspect concerns drawing conclusion of equivalent effects from superiority designs. Even though two compared treatments yield non-significant differences on the primary outcome measure, one cannot conclude that the compared treatments are equally good unless an equivalence test (Rogers, Howard, & Vessey, 1993), or a noninferiority design (Walker & Nowacki, 2011) has been employed. The majority of the included studies compared two or more treatments and more than half of these studies concluded that the treatments were equally effective. However, none of the studies did an equivalence test; maybe because in most studies the sample size was too small.

4.5. Treatment aspects

Six studies directly comparing group and individual ERP were included in the present analysis (Anderson & Rees, 2007; Cabedo et al., 2010; Fals-Stewart et al., 1993; Jaurrieta et al., 2008; Jónsson et al., 2011; O'Connor et al., 2005b). A previous meta-analysis (Jónsson & Hougaard, 2009) summarized the literature on group CBT for OCD and concluded that group and individual treatment gave equal outcome; however, also non-randomized studies as well as studies not directly comparing individual to group therapy were included. The results from the present meta-analysis also did not show any statistically significant differences between individual and group treatment in RCTs where these treatment formats were directly compared.

The studies directly comparing ERP and CT only showed a very small and non-significant ES ($g = 0.07$). Also, the comparison on raw Y-BOCS scores showed very small differences between these treatments, and the percentage of change from pre- to post- and follow-up assessment, respectively, was somewhat larger for CT than for ER. The same lack of significant difference between ERP and CT was found in the meta-analyses by Rosa-Alcázar et al. (2008) and Olatunji et al. (2013). Thus, the tentative conclusion that can be drawn is that modern forms of CT yields treatment effects not significantly different from those obtained with ERP.

4.6. Declining participation

A refusal rate of 25% has often been reported when referring to ERP (Franklin & Foa, 1998). However, the results from the present review show that studies vary considerably with respect to number of patients declining participation. On average 15% of patients refused the offer of treatment with a range from 0% (e.g. Cottraux, Note, & Yao, 2001) to 63% (Kushner et al., 2007). Unfortunately, these figures are for the studies as a whole and the method section in the RCTs usually does not specify the reasons for declining participation. This means that treatment refusal may not necessarily be a reflection of the patients' perceived difficulty and frightening nature of exposure therapy per se, but may be due to a range of different factors.

4.7. Attrition

ERP has frequently been described as a challenging treatment with a high number of patients dropping out, and studies have referred to an attrition rate of 25–30%. The results in the present meta-analysis showed a high variability in dropout rates across studies with a mean attrition rate of 15%. CT had the lowest dropout rate with 11.4%, ERP had 19.1%, and the combination of ERP/CBT and antidepressants the highest with 32.0%. Thus, patients complete cognitive behavioral treatments to a higher degree than previously has been assumed.

4.8. Limitations

One limitation of the current meta-analysis is the exclusion of RCTs not using the interview version of Y-BOCS. However, only six studies were excluded for this reason and it is not probable that they would change the outcome had they been included. Another limitation is that multiple comparisons were done without correction, since this is not included in the applied software.

4.9. Clinical implications

Based on our systematic review and meta-analysis the following clinical implications can be offered:

- There is no additional gain in combining ERP and CT. Each leads to good effects on Y-BOCS on their own with no significant difference between them.
- CBT leads to better effects than antidepressant medication (ADM).
- Adding CBT to ADM leads to a better effect than that of ADM alone.
- Adding ADM to CBT does not yield a better effect than that of CBT alone.
- The format (individual or group) does not affect the outcome.

4.10. Recommendations for future research

Based on the methodological review conducted on the RCTs included in the present meta-analysis we offer some suggestions for improved methodological stringency to consider when planning future RCTs.

- Do a proper power analysis before starting a RCT and adjust the design accordingly. The field does not learn much from underpowered RCTs yielding non-significant differences on the primary outcome measures.
- Assessment of severity of obsessive–compulsive symptoms with the Y-BOCS should be conducted by trained, independent, and blinded assessors, and inter-rater reliability should be reported based on a random sample (20%) of all assessed participants.
- Assess comorbidity and report it. We still have to evaluate if certain types of comorbidity might impede outcome of OCD.
- Assess medication at inclusion and report it.
- Use a proper random sequence generation and describe the concealment of allocation.
- Use intent-to-treat analysis and report results on all measures in the study.
- A consensus need to be reached concerning various Y-BOCS variables: minimum score for inclusion in a RCT, cut-off scores for response and remission/recovery. Also, if other measures are better than Y-BOCS in these respects.
- Assess credibility when two, or more, active treatments are compared.
- Videotape therapy sessions and assess therapists' adherence to the manual used and their competence in carrying out the treatment.
- Use three (or more) therapists with proper training and randomize patients, not only to condition, but also to therapist in order to enable analysis of therapist effect.

Funding

No external funding was obtained for this meta-analysis.

Role of funding sources

The authors have not obtained any financial support for the preparation of the present manuscript.

Contributors

LGÖ designed the meta-analysis, wrote the coding schema, rated the studies, meta-analyzed the included studies, and wrote the first draft of method and results. AH did

the literature search, rated the studies, and wrote the first draft of introduction and discussion. BH and GK rated the studies and co-wrote the first draft of introduction and discussion with AH. All authors participated in the revision and approved of the final manuscript.

Conflict of interest

None of the authors have any conflict of interest to report.

Appendix A and B. Supplementary data and studies included in the meta-analysis

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cpr.2015.06.003.

References

Abramowitz, J.S. (1996). Variants of exposure and response prevention in the treatment of obsessive–compulsive disorder: A meta-analysis. *Behavior Therapy*, *27*(4), 583–600. http://dx.doi.org/10.1016/S0005-7894(96)80045-1.

Abramowitz, J.S. (2006). The psychological treatment of obsessive–compulsive disorder. *Canadian Journal of Psychiatry*, *51*(7), 407–416.

Andersson, E., Enander, J., Andren, P., Hedman, E., Ljotsson, B., Hursti, T., et al. (2012). Internet-based cognitive behaviour therapy for obsessive–compulsive disorder: A randomized controlled trial. *Psychological Medicine*, *42*(10), 2193–2203. http://dx.doi.org/10.1017/S0033291712000244.

Barber, J.P., Liese, B.S., & Abrams, M.J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, *13*(2), 205–221. http://dx.doi.org/10.1093/ptr/kpg019.

Barber, J.P., Sharpless, B.A., Klostermann, S., & McCarthy, K.S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice*, *38*(5), 493–500. http://dx.doi.org/10.1037/0735-7028.38.5.493.

Belloch, A., Cabedo, E., & Carrió, C. (2008). Cognitive versus behaviour therapy in the individual treatment of obsessive–compulsive disorder: Changes in cognitions and clinically significant outcomes at post-treatment and one-year follow-up. *Behavioural and Cognitive Psychotherapy*, *36*(05), 521–540. http://dx.doi.org/10.1017/S1352465808004451.

Biostat Inc. (2006). *Comprehensive meta-analysis, version 2. Statistical software.* Englewood, NJ: Biostat Inc.

Blanco, C., Olfson, M., Stein, D.J., Simpson, H.B., Gameroff, M.J., & Narrow, M.H. (2006). Treatment of obsessive–compulsive disorder by U.S. psychiatrists. *Journal of Clinical Psychiatry*, *67*, 946–951.

Cabedo, E., Belloch, A., Carrio, C., Larsson, C., Fernandez-Alvarez, H., & Garcia, F. (2010). Group versus individual cognitive treatment for obsessive–compulsive disorder: Changes in severity at post-treatment and one-year follow-up. *Behavioural and Cognitive Psychotherapy*, *38*(2), 227–232. http://dx.doi.org/10.1017/s135246580999066x.

Cavanagh, K., Strauss, C., Forder, L., & Jones, F. (2014). Can mindfulness and acceptance be learnt by self-help?: A systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. *Clinical Psychology Review*, *34*(2), 118–129. http://dx.doi.org/10.1016/j.cpr.2014.01.001.

Chambless, D.L., Baker, M.J., Baucom, D.H., Beutler, L.E., Calhoun, K.S., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*(1), 3–16.

Chambless, D.L., Sanderson, W.C., Shoham, V., Johnson, S.B., Pope, K.S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49*, 5–18.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cordioli, A.V., Heldt, E., Braga Bochi, D., Margis, R., Basso de Sousa, M., Fonseca Tonello, J., et al. (2003). Cognitive-behavioral group therapy in obsessive–compulsive disorder: A randomized clinical trial. *Psychotherapy and Psychosomatics*, *72*(4), 211–216. http://dx.doi.org/10.1159/000070785.

Cottraux, J., Note, I., Yao, S.N., Lafont, S., Note, B., Mollard, E., et al. (2001). A randomized controlled trial of cognitive therapy versus intensive behavior therapy in obsessive compulsive disorder. *Psychotherapy and Psychosomatics*, *70*(6), 288–297. http://dx.doi.org/10.1159/000056269.

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. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. http://dx.doi.org/10.1136/bmj.316.7129.469.

Fals-Stewart, W., Marks, A.P., & Schafer, J. (1993). A comparison of behavioral group therapy and individual behavior therapy in treating obsessive–compulsive disorder. *Journal of Nervous and Mental Disease*, *181*(3), 189–193. http://dx.doi.org/10.1097/00005053-199303000-00007.

Federici, A., Summerfeldt, L.J., Harrington, J.L., McCabe, R.E., Purdon, C.L., Rowa, K., et al. (2010). Consistency between self-report and clinician-administered versions of the Yale–Brown Obsessive–Compulsive Scale. *Journal of Anxiety Disorders*, *24*(7), 729–733. http://dx.doi.org/10.1016/j.janxdis.2010.05.005.

Fisher, P.L., & Wells, A. (2005). How effective are cognitive and behavioral treatments for obsessive–compulsive disorder? A clinical significance analysis. *Behaviour Research and Therapy*, *43*(12), 1543–1558. http://dx.doi.org/10.1016/j.brat.2004.11.007.

Foa, E.B., Liebowitz, M.R., Kozak, M.J., Davies, S., Campeas, R., Franklin, M.E., et al. (2005). Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive–compulsive disorder. *American Journal of Psychiatry*, *162*(1), 151–161. http://dx.doi.org/10.1176/appi.ajp.162.1.151 (162/1/151 [pii]).

Franklin, M.E., & Foa, E.B. (1998). Cognitive-behavioral treatments for obsessive–compulsive disorder. In P.E. Nathan, & J.M. Gorman (Eds.), *A guide to treatments that work* (pp. 339–357). New York: Oxford University Press.

Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Fleischmann, R.L., Hill, C.L., et al. (1989). The Yale–Brown obsessive compulsive scale. I. Development, use, and reliability. *Archives of General Psychiatry*, *46*(11), 1006–1011. http://dx.doi.org/10.1001/archpsyc.1989.01810110048007.

Higgins, J.P.T., Altman, D.G., & Sterne, J.A.C. (2011). Chapter 8: Assessing risk of bias in included studies. In J.P.T. Higgins, & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]*. The Cochrane Collaboration.

Hofmann, S.G., & Smits, J.A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*, *69*(4), 621–632. http://dx.doi.org/10.4088/JCP.v69n0415.

Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336–352. http://dx.doi.org/10.1016/S0005-7894(84)80002-7.

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. http://dx.doi.org/10.1037/0022-006X.59.1.12.

Jadad, A.R., Moore, R.A., Carroll, D., Jenkinson, C., Reynolds, D.J., Gavaghan, D.J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12. http://dx.doi.org/10.1016/0197-2456(95)00134-4.

Jaurrieta, N., Jimenez-Murcia, S., Menchon, J.M., Del Pino Alonso, M., Segalas, C., Alvarez-Moya, E.M., et al. (2008). Individual versus group cognitive-behavioral treatment for obsessive–compulsive disorder: A controlled pilot study. *Psychotherapy Research*, *18*(5), 604–614. http://dx.doi.org/10.1080/10503300802192141.

Jónsson, H., & Hougaard, E. (2009). Group cognitive behavioural therapy for obsessive–compulsive disorder: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *119*(2), 98–106. http://dx.doi.org/10.1111/j.1600-0447.2008.01270.x.

Jónsson, H., Hougaard, E., & Bennedsen, B.E. (2011). Randomized comparative study of group versus individual cognitive behavioural therapy for obsessive compulsive disorder. *Acta Psychiatrica Scandinavica*, *123*(5), 387–397. http://dx.doi.org/10.1111/j.1600-0447.2010.01613.x.

Kazdin, A.E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 593–602. http://dx.doi.org/10.1001/archpsyc.62.6.593.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., & Johnson, B.T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *Public Library of Science Medicine*, *5*(2), e45. http://dx.doi.org/10.1371/journal.pmed.0050045.

Kobak, K.A., Greist, J.H., Jefferson, J.W., Katzelnick, D.J., & Henk, H.J. (1998). Behavioral versus pharmacological treatments of obsessive compulsive disorder: A meta-analysis. *Psychopharmacology*, *136*(3), 205–216. http://dx.doi.org/10.1007/s002130050558.

Koran, L.M., Thienemann, M.L., & Davenport, R. (1996). Quality of life for patients with obsessive–compulsive disorder. *American Journal of Psychiatry*, *153*(6), 783–788. http://dx.doi.org/10.1176/ajp.153.6.783.

Kozak, M.J., Liebowitz, M.R., & Foa, E.B. (Eds.). (2000). *Cognitive behavior therapy and pharmacotherapy for OCD: The NIMH-sponsored collaborative study*. Mahwah: NJ Erlbaum.

Kushner, M.G., Kim, S.W., Donahue, C., Thuras, P., Adson, D., Kotlyar, M., et al. (2007). D-Cycloserine augmented exposure therapy for obsessive–compulsive disorder. *Biological Psychiatry*, *62*(8), 835–838. http://dx.doi.org/10.1016/j.biopsych.2006.12.020.

LaSalle, V.H., Cromer, K.R., Nelson, K.N., Kazuba, D., Justement, L., & Murphy, D.L. (2004). Diagnostic interview assessed neuropsychiatric disorder comorbidity in 334 individuals with obsessive–compulsive disorder. *Depression and Anxiety*, *19*, 163–173.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotsche, P.C., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*, e1000100.

Lindsay, M., Crino, R., & Andrews, G. (1997). Controlled trial of exposure and response prevention in obsessive–compulsive disorder. *British Journal of Psychiatry*, *171*, 135–139. http://dx.doi.org/10.1192/bjp.171.2.135.

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

McLean, P.D., Whittal, M.L., Thordarson, D.S., Taylor, S., Sochting, I., Koch, W.J., et al. (2001). Cognitive versus behavior therapy in the group treatment of obsessive–compulsive disorder. *Journal of Consulting and Clinical Psychology*, *69*(2), 205–214. http://dx.doi.org/10.1037/0022-006X.69.2.205.

Morris, S.B., & DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105.

Nakatani, E., Nakagawa, A., Nakao, T., Yoshizato, C., Nabeyama, M., Kudo, A., et al. (2005). A randomized controlled trial of Japanese patients with obsessive–compulsive disorder—Effectiveness of behavior therapy and fluvoxamine. *Psychotherapy and Psychosomatics*, *74*(5), 269–276. http://dx.doi.org/10.1159/000086317.

National Collaborating Centre for Mental Health (2006). *Obsessive–compulsive disorder: Core interventions in the treatment of obsessive–compulsive disorder and body dysmorphic disorder.* London: British Psychological Society, Royal College of Psychiatrists (National clinical practice guideline; no. 31).

O'Connor, K.P., Aardema, F., Bouthillier, D., Fournier, S., Guay, S., Robillard, S., et al. (2005). Evaluation of an inference-based approach to treating obsessive–compulsive disorder. *Cognitive Behaviour Therapy*, *34*(3), 148–163. http://dx.doi.org/10.1080/16506070510041211.

Olatunji, B.O., Davis, M.L., Powers, M.B., & Smits, J.A. (2013). Cognitive-behavioral therapy for obsessive–compulsive disorder: A meta-analysis of treatment outcome and moderators. *Journal of Psychiatric Research*, *47*(1), 33–41. http://dx.doi.org/10.1016/j.jpsychires.2012.08.020.

Öst, L.G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *46*(3), 296–321. http://dx.doi.org/10.1016/j.brat.2007.12.005.

Öst, L.G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, *61*, 105–121. http://dx.doi.org/10.1016/j.brat.2014.07.018.

Rogers, J.L., Howard, K.I., & Vessey, J.T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, *113*(3), 553–565. http://dx.doi.org/10.1037/0033-2909.113.3.553.

Romanelli, R.J., Wu, F.M., Gamba, R., Mojtabai, R., & Segal, J.B. (2014). Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive–compulsive disorder: A systematic review and meta-analysis of head-to-head randomized controlled trials. *Depression and Anxiety*, *31*(8), 641–652. http://dx.doi.org/10.1002/da.22232.

Rosa-Alcázar, A.I., Sánchez-Meca, J., Gómez-Conesa, A., & Marín-Martínez, F. (2008). Psychological treatment of obsessive–compulsive disorder: a meta-analysis. *Clinical Psychology Review*, *28*, 1310. http://dx.doi.org/10.1016/j.cpr.2008.07.001.

Simpson, H.B., Huppert, J.D., Petkova, E., Foa, E.B., & Liebowitz, M.R. (2006). Response versus remission in obsessive–compulsive disorder. *Journal of Clinical Psychiatry*, *67*(2), 269–276. http://dx.doi.org/10.4088/JCP.v67n0214.

Simpson, H.B., Zuckoff, A.M., Maher, M.J., Page, J.R., Franklin, M.E., Foa, E.B., et al. (2010). Challenges using motivational interviewing as an adjunct to exposure therapy for obsessive–compulsive disorder. *Behaviour Research and Therapy*, *48*(10), 941–948. http://dx.doi.org/10.1016/j.brat.2010.05.026.

Soomro, G.M., Altman, D., Rajagopal, S., & Oakley-Browne, M. (2008). Selective serotonin re-uptake inhibitors (SSRIs) versus placebo for obsessive compulsive disorder (OCD). *Cochrane database of systematic reviews*, CD001765.

Tolin, D.F. (1999). *A revised meta-analysis of psychosocial treatments for PTSD.* Poster presented at the at AABT, Toronto, November 11–14, 1999.

Tolin, D.F., Hannan, S., Maltby, N., Diefenbach, G.J., Worhunsky, P., & Brady, R.E. (2007). A randomized controlled trial of self-directed versus therapist-directed cognitive-behavioral therapy for obsessive–compulsive disorder patients with prior medication trials. *Behavior Therapy*, *38*(2), 179–191. http://dx.doi.org/10.1016/j.beth.2006.07.001.

Vogel, P.A., Stiles, T.C., & Götestam, K.G. (2004). Adding cognitive therapy elements to exposure therapy for obsessive compulsive disorder: A controlled study. *Behavioural and Cognitive Psychotherapy*, *32*(3), 275–290. http://dx.doi.org/10.1017/S1352465804001353.

Walker, E., & Nowacki, A.S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, *26*(2), 192–196. http://dx.doi.org/10.1007/s11606-010-1513-8.

Whittal, M.L., Thordarson, D.S., & McLean, P.D. (2005). Treatment of obsessive–compulsive disorder: Cognitive behavior therapy vs. exposure and response prevention. *Behaviour Research and Therapy*, *43*(12), 1559–1576. http://dx.doi.org/10.1016/j.brat.2004.11.012.

Whittal, M.L., Woody, S.R., McLean, P.D., Rachman, S.J., & Robichaud, M. (2010). Treatment of obsessions: A randomized controlled trial. *Behaviour Research and Therapy*, *48*(4), 295–303. http://dx.doi.org/10.1016/j.brat.2009.11.010.

Wootton, B.M., Dear, B.F., Johnston, L., Terides, M.D., & Titov, N. (2013). Remote treatment of obsessive–compulsive disorder: A randomized controlled trial. *Journal of Obsessive-Compulsive and Related Disorders*, *2*(4), 375–384. http://dx.doi.org/10.1016/j.jocrd.2013.07.002.