Received: 27 July 2023 | Revised: 16 November 2023 | Accepted: 18 November 2023

DOI: 10.1002/eat.24104

REVIEW

International Journal of EATING DISORDERS WILEY

# Cognitive behavior therapy for adult eating disorders in routine clinical care: A systematic review and meta-analysis

Lars-Göran Öst PhD[1,2,3] | Martin Brattmyr MSc[4] | Anna Finnes PhD[3,5] | Ata Ghaderi PhD[3] | Audun Havnen PhD[4,6] | Maria Hedman-Lagerlöf PhD[7] | Thomas Parling PhD[7,8] | Elisabeth Welch PhD[3,9] | Gro Janne Wergeland MD, PhD[10,11]

[1]Department of Psychology, Stockholm University, Stockholm, Sweden

[2]Bergen Center for Brain Plasticity, Haukeland University Hospital, Bergen, Norway

[3]Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[4]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

[5]Academic Primary Care Center, Stockholm, Sweden

[6]Division of Psychiatry, St. Olavs Hospital, Trondheim, Norway

[7]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[8]Stockholm Health Care Services, Stockholm, Sweden

[9]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[10]Department of Child and Adolescent Psychiatry, Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

[11]Department of Clinical Medicine, Faculty of Medicine, University of Bergen, Bergen, Norway

**Correspondence**
Lars-Göran Öst, Department of Psychology, Stockholm University, 114 19 Stockholm, Sweden.
Email: ost@psychology.su.se

**Action Editor:** Kelly L. Klump

## Abstract

**Objective:** Cognitive behavior therapy (CBT) is a recommended treatment for eating disorders (ED) in adults given its evidence, mainly based on efficacy studies. However, little is known about how CBT works in routine clinical care. The goal of the present meta-analysis is to investigate how CBT works for various ED when carried out in routine clinical settings.

**Method:** Ovid MEDLINE, Embase OVID, and PsycINFO were systematically searched for articles published until June 2023. The outcome of CBT, methodological quality, risk of bias (RoB), and moderators of treatment outcome were examined and benchmarked by meta-analytically comparing with efficacy studies for ED. Fifty studies comprising 4299 participants who received CBT were included.

**Results:** Large within-group effect sizes (ES) were obtained for ED-psychopathology at post-treatment (1.12), and follow-up (1.22), on average 9.9 months post-treatment. Attrition rate was 25.5% and RoB was considerable in the majority of studies. The benchmarking analysis showed that effectiveness studies had very similar ESs as efficacy studies (1.20 at post-treatment and 1.28 at follow-up).

**Conclusion:** CBT for ED is an effective treatment when delivered in routine clinical care, with ESs comparable to those found in efficacy studies. However, the evidence needs to be interpreted with caution due to the RoB in a high proportion of studies.

**Public Significance:** Eating disorders are common in the population and often lead to multiple negative consequences. CBT has been found effective for ED and is recommended in clinical guidelines. Since these recommendations are primarily based on university studies we wanted to investigate how CBT performs in routine clinical care. Our meta-analysis found that CBT worked as well in routine care as in university setting studies.

## Abstracto

**Objetivo:** La terapia cognitivo-conductual (TCC) es un tratamiento recomendado para los trastornos de la conducta alimentaria (TCA) en adultos debido a su evidencia, basada principalmente en estudios de eficacia. Sin embargo, se sabe poco sobre cómo funciona la TCC en la atención clínica rutinaria. El objetivo de este meta-análisis es investigar cómo funciona la TCC para diversos TCA cuando se lleva a cabo en entornos clínicos habituales.

**Método:** Se realizó una búsqueda sistemática en Ovid MEDLINE, Embase OVID y PsycINFO de artículos publicados hasta junio de 2023. Se examinaron el resultado de la TCC, la calidad metodológica, el riesgo de sesgo y los moderadores del resultado del tratamiento, y se compararon metaanalíticamente con estudios de eficacia para TCA. Se incluyeron cincuenta estudios que comprendían a 4299 participantes que recibieron TCC.

**Resultados:** Se obtuvieron tamaños del efecto (TE) grandes dentro del grupo para la patología de los TCA en el post-tratamiento (1.12) y en el seguimiento (1.22), en promedio 9.9 meses después del post-tratamiento. La tasa de abandono fue del 25.5% y el riesgo de sesgo fue considerable en la mayoría de los estudios. El análisis de comparación mostró que los estudios de efectividad tenían TE muy similares a los estudios de eficacia (1.20 en el post-tratamiento y 1.28 en el seguimiento).

**Conclusión:** La TCC para los TCA es un tratamiento efectivo cuando se administra en la atención clínica rutinaria, con TE comparables a los encontrados en estudios de eficacia. Sin embargo, la evidencia debe interpretarse con cautela debido al riesgo de sesgo en una alta proporción de los estudios.



**KEYWORDS**

adults, anorexia nervosa, binge-eating disorder, bulimia nervosa, CBT, eating disorders, effectiveness, meta-analysis


## 1 | INTRODUCTION

During the last 20 years, there has been a strong surge of research on ED treatments, resulting in clinical recommendations that favor cognitive behavioral therapy (CBT), and treatment guidelines have been published by various organizations. In Great Britain, the National Institute for Health and Care Excellence (NICE, 2017) recommends individual CBT-ED for anorexia nervosa (AN), and individual CBT-ED or guided self-help (GSH-CBT) for bulimia nervosa (BN) and binge eating disorder (BED). The Australian Psychological Society (APS) (2018) reviewed the evidence base and reported that for both BN and BED face-to-face CBT was at the highest level and online CBT at the second level, whereas for AN face-to-face CBT and online CBT were at Level II. The American Psychiatric Association (Crone et al., 2023) recommends that AN should be treated with an eating disorder-focused psychotherapy, and both BN and BED with an eating disorder-focused CBT. The clinical recommendations are based on efficacy studies (RCTs) and it is important to assess how CBT performs in routine clinical practice, which is the focus of the present meta-analysis where we review the forms of CBT which are recommended in the treatment guidelines from these organizations.

Cognitive behavior therapy addresses cognitive, physiological, and behavioral components that interact in the development and maintenance of the dysfunctional cognitive processes, emotions, and maladaptive behavior through use of goal-oriented and systematic procedures.

The evidence base for treatments of ED consists to a large extent of randomized controlled trials (RCT) conducted under formal research conditions (efficacy studies), and it has been questioned (e.g., Westen et al., 2004) if the results are transferable to routine clinical care (effectiveness studies). RCTs in efficacy studies are designed to give high internal validity, for example, by randomizing participants to conditions, having highly trained therapists with documented adherence to a manual and competence in carrying out the treatment, and using independent and masked assessors of outcome. Effectiveness studies focus on external validity, for example, by having fewer exclusion criteria to increase generalizability, using therapists with varying degrees of experience and training, and primarily applying self-report measures of outcome. We believe that effectiveness studies complement the results from efficacy studies by examining how empirically supported treatments perform when delivered in routine clinical care.

In addition to open and nonrandomized trials, RCTs can also be used in effectiveness studies (Stewart & Chambless, 2009) provided they include referred participants, are carried out in routine clinical care, and use ordinary therapists working at these services.

To the best of our knowledge, there is only one previous meta-analysis (MA) of CBT for ED that has examined effectiveness studies (Linardon, Messer, & Fuller-Tyszkiewicz, 2018). They included studies of patients with BN, BED, or Other Specified Feeding and Eating Disorder (OSFED), receiving a face-to-face treatment in a nonrandomized design, and reported rates of abstinence from binge eating and/or purging. The authors found 27 studies with mean intent-to-treat abstinence rates for BN of 29.8% and BED of 47.2%. Comparing these data with results from their previous efficacy meta-analyses on BN (Linardon & Wade, 2018) and BED (Linardon, Hindle, & Brennan, 2018), the authors concluded that the effects of CBT in controlled research settings are generalizable to routine clinical settings.

The current MA differs from that of Linardon, Hindle, and Brennan (2018) in several ways. First, we included all common eating disorders in adults. Second, we included self-help CBT since a recent MA by Hedman-Lagerlöf et al. (2023) yielded effect sizes for self-help CBT on par with face-to-face treatments. Third, we included both nonrandomized and randomized designs, since RCTs can be used in effectiveness studies (Stewart & Chambless, 2009). Fourth, we included various effect measures that have been found to significantly predict treatment outcome in eating disorders (Vall & Wade, 2015). Fifth, differences in effect sizes for effectiveness and efficacy studies were directly tested using meta-analytical statistical methods as we have done in previous meta-analyses (e.g., Öst et al., 2022, 2023a, 2023b, 2023c). The overlap in included studies between Linardon, Hindle, and Brennan (2018) and our MA is only 22%. Since the publication of the Linardon meta-analysis 21 new effectiveness studies have emerged which could be included in our meta-analysis. Thus, a new meta-analysis on the effectiveness of CBT for ED routine clinical care is warranted.

In order to enable inclusion of both RCTs and pre–post trials in this meta-analysis, we used the uncontrolled pre–post effect size, while being aware of the problems with this ES measure. Cuijpers et al. (2017) give a detailed description of these problems. The pre–post ES is influenced by other factors than the treatment, for example, spontaneous recovery, regression to the mean, and various patient characteristics. Further, the pre- and post-scores are not independent of each other and the degree of correlation between them influences the ES. Thus, Cuijpers et al. (2017) argue that the pre–post ES should be avoided in general, but consider this ES as useful when comparing improvement in routine care with that in efficacy studies, which is the case in the present meta-analysis.

Previous meta-analyses of psychological treatments have found different moderators of the ES. In the present meta-analysis, we will use four categorical variables. *Statistical analysis*; some studies have found no difference in ES between intent-to-treat (ITT) and completer analysis (e.g., Cuijpers et al., 2012), that ITT analysis yielded higher ES (e.g., Schwartze et al., 2019) or that completer analysis yielded a higher ES (e.g., Öst et al., 2015). Thus, from a methodological point of view, this is an important moderator to assess. *Risk of bias*; high risk of bias (RoB) has been associated with high ES (e.g., Bürkner et al., 2017; Cuijpers et al., 2014), but there are meta-analyses finding that studies with low RoB yielded higher effect (e.g., van den Berg et al., 2019; van Dis et al., 2019), and those that did not find RoB to be a significant moderator (e.g., Carpenter et al., 2018; Cuijpers et al., 2012). Thus, RoB is included as a moderator. *Treatment format*; in some meta-analyses, individual therapy has been better than group therapy (e.g., Hans & Hiller, 2013), whereas others have not found a significant difference (e.g., Wergeland et al., 2021). *Continent*; previous meta-analyses investigating this variable have found different results. For example, Cuijpers et al. (2013) found that studies from North America yielded higher ES than studies from Europe, whereas Öst et al. (2022) and Wergeland et al. (2021) reported that studies from Europe yielded higher ES than studies from other continents.

There are also some continuous variables of interest as potential moderators. *Pretreatment severity* has in a number of meta-analyses using within-group ES been found to positively moderate outcome (e.g., Cuijpers et al., 2014; Öst et al., 2015; Riise et al., 2021). The *methodological quality* of the included studies has in previous meta-analyses been found to be associated with lower ES (e.g., A-Tjak et al., 2015; Öst, 2014), as well as with higher ES (e.g., Finnes et al., 2019; Öst et al., 2016). *Amount of treatment* measured as weeks of therapy and hours of treatment has in some meta-analyses been found to be a positive moderator (e.g., Cuijpers et al., 2014; Hans & Hiller, 2013; Wergeland et al., 2022) but in at least one meta-analysis it was a negative moderator (Öst & Ollendick, 2017). In addition, it would be of interest to analyze if the number of participants in the CBT-condition and its mean age moderate the ES.

The aims of the present meta-analysis were (1) to examine the effectiveness of CBT for ED in routine clinical care on various primary measures of ED and depression as secondary outcome (since it is a common comorbidity), (2) to evaluate methodological quality and RoB in the effectiveness studies, and investigate potential moderators of treatment outcome, and (3) to compare the outcome of CBT delivered in routine clinical care with that reported in efficacy studies for ED. Based on the previous meta-analysis of abstinence rates from binge eating and purging in BN and BED (Linardon, Hindle, & Brennan, 2018), as well as our previous meta-analyses of effectiveness studies in obsessive-compulsive disorder (Öst et al., 2022), post-traumatic stress disorder (Öst et al., 2023b), anxiety disorders (Öst et al., 2023a), and depression (Öst et al., 2023c) we predicted that the ESs for effectiveness studies will be comparable to those of efficacy studies.

## 2 | MATERIALS AND METHODS

This meta-analysis was pre-registered at PROSPERO with ID CRD42023373548 and there was no deviation from the protocol. It was conducted according to the PRISMA guidelines (Page et al., 2021). For more details, see Supporting Information S1. The following PICOS specifies the design of this meta-analysis.

- *Population*: adults with an ED diagnosis.
- *Intervention*: CBT-E (Fairburn, 2008), CBT-BN (Fairburn & Beglin, 1994), CBT (e.g., Agras & Apple, 2007; Garner et al., 1997), CBT-T (Waller et al., 2019), CBT+ (Lammers et al., 2020), applied in various formats (e.g., Individual, Group, Self-help, Guided Self-help), and delivered in routine clinical care. CBT+ is an extension and adaptation of the previous CBT manual (Fairburn & Beglin, 1994) into an intensive outpatient group CBT for eating disorders.
- *Comparison*: within-group change, that is, pre versus post/follow-up data.
- *Outcome*: primary (ED symptoms) and secondary (depression).
- *Study design*: RCTs and pre–post/nonrandomized studies of intervention (NRSI).

## 2.1 | Literature search

We identified relevant studies through a systematic and comprehensive literature search of electronic databases and scanned the reference lists of the included studies. The search strategy was applied to Ovid MEDLINE, Embase OVID, and PsycINFO from the start of the databases to November 22, 2022. An updated search was done on June 22, 2023. The authors in collaboration with a university librarian, who conducted the database searches, generated the list of search items to identify relevant studies. We used both subject headings and free text words for the following search terms to search the databases: CBT and variations of thereof (i.e., CBT OR CBT-E OR cognitive behavio* therap* OR cognitive therap* OR cognitive behavio* treatment*), the different EDs, the design of the study, and adults. For the full search strategy, see Supporting Information S2.

Four pairs of authors (MB, AF, AG, AH, MHL, TP, and EW) read the abstracts independently of each other to decide whether a study warranted a more detailed reading. Full-text articles were retrieved if there was any indication of a target group of patients receiving the particular CBT in a routine clinical care setting. 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. In total, 403 full-text articles were considered for inclusion. The final decision for article inclusion was made using a stricter set of inclusion and exclusion criteria, and any disagreements were resolved by consensus discussion among the authors and/or consultation with the first author.

### 2.1.1 | Inclusion criteria

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

1. Be published, or in press, in an English language journal.
2. Have participants diagnosed with ED according to DSM (III and later) or ICD (10 or 11).
3. Be testing a form of CBT, that received recommendations by the organizations described in the introduction.
4. Have participants referred, or self-referred, for treatment through usual clinical routes.
5. Be an effectiveness study, that is, carried out in a routine care setting.
6. Have therapists who are practicing clinicians for whom provision of service is a substantial part of their job (Shadish et al., 2000).
7. Have a treated sample consisting of at least 10 adult participants.
8. Provide data for a standardized and validated measure of ED symptoms.

Non-English articles excluded at the abstract reading level are listed in Supporting Information S3.

### 2.1.2 | Exclusion criteria

1. Studies where participants are required to fulfill diagnostic criteria for another disorder in addition to ED in order to be included in the study.
2. Studies being a secondary analysis of a previously published study. However, separate follow-up studies to the basic study were included to provide follow-up data.
3. Studies testing a combination of CBT and pharmacological treatment, and where all participants in that condition received both treatments.

## 2.2 | Categorization of studies

To be categorized as an effectiveness study, it had to have participants referred through ordinary clinical channels (or self-referred), the treatment is carried out in routine clinical care settings (or in patients' homes for internet-based CBT), and the therapists are ordinary clinicians who work with a caseload of patients with different diagnoses.

We included studies with adults diagnosed with AN, BN, BED, EDNOS, or OSFED. A few studies had a combination of BN and EDNOS or BED and EDNOS but with a majority of participants having BN and BED, respectively, so these were classified as BN and BED. In addition, we included a number of studies that had a mix of the above eating disorders and presented combined results.

## 2.3 | Potential categorical moderators

An a priori requirement for including any potential categorical or continuous moderator in the analysis was that at least 70% of the studies provided information on that variable, as lower rates would probably lead to questionable representativity. Statistical analysis was categorized as completers (if dropouts were deleted) or as intent-to-treat (ITT, if all randomized or starting participants were included in the statistical analysis). RoB was based on a summary evaluation of the domains rated for the different designs (see below) and the studies were categorized as low, moderate, or high RoB. Treatment format could either be

individual, group, or a combination of individual and group. The country in which the study was carried out was categorized as situated in Africa, Asia, Australia, Europe, North America, or South America.

## 2.4 | Potential continuous moderators

We used the following variables as potential continuous moderators: mean age, pretreatment severity (calculated as a percentage by dividing the sample mean with the maximum score possible of the rating scale applied), methodology score (see below), weeks of therapy, hours of treatment, and number of participants in the CBT-condition. We developed a coding scheme and a scoring manual including the variables of interest. The data extraction and categorizations were done independently by pairs of authors and any disagreements were solved after consensus discussion.

## 2.5 | Methodological quality

### 2.5.1 | The Psychotherapy Outcome Study Methodology Rating Scale

The Psychotherapy Outcome Study Methodology Rating Scale (POMRS) consists of 22 items covering various important aspects of the methodology in psychotherapy outcome research (Öst, 2008). Each item is rated on a 3-point scale (0 = poor, 1 = fair, and 2 = good). Since all items do not apply to all studies, the total score was recalculated as a percentage of the maximum score possible for the individual study. The internal consistency of the scale was good with a McDonald's $\omega$ of .80. The inter-rater reliability of the scale (between GJW and LGÖ), based on 20% randomly selected and blindly rated studies, was ICC = .98 (95% CI .94–.99, $p = .0001$), which according to Cicchetti (1994) is excellent.

### 2.5.2 | Risk of bias

We used the Cochrane Collaboration tool for assessing RoB (Sterne et al., 2019) for RCTs and the RoB in nonrandomized Studies of Interventions (ROBINS-I; Sterne et al., 2016) for NRSI and pre–post studies. An overall classification of the studies was done for RCTs into the categories high, moderate (some concerns), or low RoB. For the NRSI and pre–post studies, the categories low, moderate, and high (serious or critical) RoB were used. The rating of the studies was done by two independent researchers and differences were discussed to reach consensus.

## 2.6 | Effect size measures

When selecting outcome measures for our meta-analysis, we consulted the meta-analysis on predictors of treatment outcomes in ED by Vall and Wade (2015). Among the many predictors in this meta-analysis, we selected those that had been investigated most often and yielded significant predictions of outcome.

### 2.6.1 | Primary outcome measures

The first primary measure is scores on a validated rating scale of eating disorder psychopathology. The Eating Disorders Examination (EDE; Cooper & Fairburn, 1987) is an interview-based assessor rating and different versions were used in 12 studies. The Eating Disorders Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994) was applied in 25 studies. The Eating Disorder Inventory (EDI; Garner & Olmsted, 1984) was used in two studies, the Bulimic Investigatory Test, Edinburgh (BITE, Henderson & Freeman, 1987) in three studies, the Short Evaluation of Eating Disorders Symptoms (SEEDS; Kordy et al., 1999) in two studies, and the Binge Eating Scale (BES; Gormally et al., 1982) in one study.

Body mass index (BMI) defined as kg/m$^2$ is primarily used in AN studies where one goal of the treatment is to normalize the participant's weight. Even if BMI is used in studies of other EDs, we will not use it for these because there is usually no uniform treatment goal of BMI change for these disorders.

Number of binge and purge episodes were either assessed during a 28-day period as part of EDE/EDE-Q (29 studies) or with self-recordings during a 7-day period (eight studies). Abstinence from binge eating and purge episodes was defined as proportion of participants with zero episodes reported within the time period of assessment.

### 2.6.2 | Secondary outcome measure

We extracted data on depressive symptoms, which were provided by 24 (49%) of the included studies. Seventeen studies used the Beck Depression Inventory (BDI; Beck et al., 1961, or BDI-II; Beck & Steer, 1993), three used the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), two used the Depression Anxiety Stress Scales (DASS-Depression subscale; Lovibond & Lovibond, 1995), and one each used the Montgomery-Åsberg Depression Rating Scale (MADRS-S; Svanborg & Åsberg, 1994) and the Center for Epidemiological Studies Depression scale (CES-D; Radloff, 1977).

## 2.7 | Meta-analysis

To obtain as many effectiveness studies as possible, we included both RCTs and open trials in the meta-analysis since within-group ES can be calculated from both types of studies. ES was calculated as $(M_{pre} - M_{post})/SD_{pre}$ according to a recommendation by Lakens (2013), since there is good reason to assume that the interventions influence not only the means but also the standard deviations. The mean ES was computed by weighting each ES by the inverse of its variance. We used intent-to-treat (ITT) data when a study provided those, otherwise completer data were used.

Before pooling, the effect sizes were screened for statistical outliers, defined as being outside M ± 2SD. At the posttreatment assessment, 4.6% of the ESs for primary measures combined were outliers, and at follow-up assessment, there was 5.5%. For these ESs, *winsorizing* (Lipsey & Wilson, 2001) was used by reducing outliers to the exact value of M + 2SD. The *Comprehensive Meta-Analysis v.4* (CMA; Borenstein et al., 2022) software was used for the analyses and Hedges' *g* was calculated to correct for small sample sizes. A random effects model was used since it cannot be assumed that the ESs come from the same population. Lipsey (1990) described an empirically developed rule-of-thumb for considering an ES as small (<.32), moderate (.33–.55), and large (.56–1.20). Also, Sawilowsky (2009) denoted ESs as very large (1.20–1.99) and huge (≥2.00).

Sensitivity analysis was done for the primary outcome measure ED-psychopathology in two ways to test the robustness of the pooled ES. First, the pre–post correlation was varied from .1 to .9 and then the effect of the different ED-psychopathology measures was tested by deleting each of them not being EDE or EDE-Q.

Proportions were analyzed in CMA. The values of the individual studies were transformed using logit transformation and the statistical analysis was done on the transformed proportions using the random effects model. Then the pooled proportion and its 95% confidence interval was back-transformed to a proportion.

Heterogeneity among ESs was assessed with the *Q*-statistic and the prediction interval (the true effect size in 95% of all comparable populations will fall within this interval; Borenstein, 2022), and publication bias with Egger's regression intercept (Egger et al., 1998) and Duval and Tweedie's (2000) trim and fill method. Moderator analyses of categorical variables were done with subgroup analysis using the mixed effect model and of continuous variables with meta-regression using the random effects model.

## 2.8 | Efficacy studies for comparison

We consulted recent comprehensive meta-analyses of CBT for ED (AN: Gan et al., 2022; Solmi et al., 2021; BN: Linardon & Wade, 2018; Svaldi et al., 2019; BED: Hilbert et al., 2020; Linardon, 2018; Mixed ED: Chang et al., 2021; Linardon et al., 2018a) to obtain the efficacy studies to be used in a comparison with effectiveness studies. From these meta-analyses, we listed the RCTs of CBT recommended by the treatment guidelines reviewed in the introduction. Since these meta-analyses included both efficacy and effectiveness studies, we deleted those RCTs we had already included in the body of effectiveness studies. This resulted in 58 RCTs for our comparison and the references are listed in Supporting Information S4. This type of benchmarking in which ES for effectiveness and efficacy studies are statistically compared using a meta-analysis software has previously been done in four similar meta-analyses on effectiveness studies in adults (Öst et al., 2022, 2023a, 2023b, 2023c) and three in children and adolescents (Riise et al., 2021; Wergeland et al., 2021, 2022).

As for the effectiveness studies, we extracted data for the type of primary outcome measure most frequently used in both types of studies (some ED-psychopathology measure), at

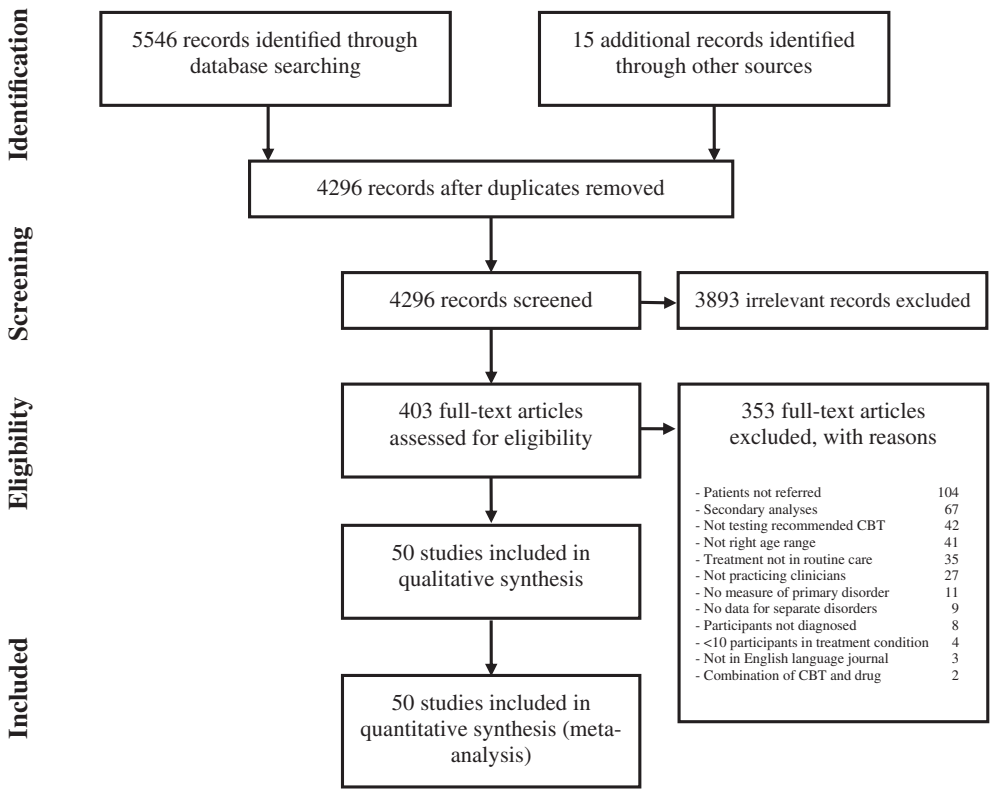


**FIGURE 1** Flowchart of the inclusion of studies.

**TABLE 1** Background data for the included studies.

| Study | Eating disorder | Country | RCT | CBT-method | Race % | Ethnicity % | *N* | % Females | Mean age | % Severity | % Comorbidity | % Medicated |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Allen, 2012 (I) | AN, BN, E | Australia | No | CBT-E | NR | NR | 43 | 100.0 | 26.4 | 70.0 | 34 | 50.0 |
| Allen, 2012 (II) | AN, BN, E | Australia | No | CBT-E + MFT | NR | NR | 52 | 100.0 | 26.5 | 68.3 | 50 | 54.0 |
| Bara-Carril, 2004 | BN | UK | No | SH | NR | NR | 47 | 94.0 | 30.0 | 68.0 | NR | 40.0 |
| Beintner, 2020 | AN, BN, E | Germany | No | CBT | NR | NR | 148 | 98.6 | 25.4 | 61.7 | NR | NR |
| Bell, 2001 | BN, BED, E | UK | No | GSH-CBT | NR | NR | 21 | 90.5 | 30.0 | 28.2 | NR | NR |
| Byrne, 2011 | AN, BN, E | Australia | No | CBT-E | O 6, W 92 | NR | 125 | 97.6 | 26.0 | 66.7 | 31 | 36.0 |
| Calugi, 2017 (I) | AN | Italy | No | CBT-E | NR | NR | 34 | 97.1 | 23.0 | 63.3 | NR | 10.9 |
| Calugi, 2017 (II) | AN | Italy | No | CBT-E | NR | NR | 32 | 96.9 | 29.4 | 63.3 | NR | 10.9 |
| Calugi, 2019 | AN | Italy | No | CBT-E | NR | NR | 90 | 100.0 | 26.3 | 60.0 | 46 | NR |
| Calugi, 2021a | AN | Italy | No | CBT-E | NR | NR | 214 | 95.3 | 25.3 | 65.0 | NR | NR |
| Calugi, 2021b | AN | Italy | No | CBT-E | NR | NR | 30 | 96.7 | 22.4 | 56.7 | NR | NR |
| Carter, 2003 | BN | Canada | Yes | SH | B 2, O 15, W 83 | NR | 85 | 100.0 | 27.0 | 78.3 | NR | NR |
| Chen, 2017 (I) | BED | USA | Yes | GSH-CBT | B 17, W 74 | H 9 | 42 | 100.0 | 38.6 | 51.7 | 50 | NR |
| Chen, 2017 (II) | BED | USA | Yes | CBT | B 19, W 71 | H 10 | 31 | 100.0 | 37.8 | 51.7 | 58 | NR |
| Cooper, 1994 | BN | UK | No | GSH-CBT | NR | NR | 18 | 100.0 | NR | 88.3 | NR | NR |
| Cooper, 1996 | BN | UK | No | GSH-CBT | NR | NR | 82 | 100.0 | 23.8 | 73.3 | 31 | NR |
| Dalle Grave, 2007 (I) | AN | Italy | No | CBT-E | NR | NR | 60 | 91.3 | 24.6 | 61.7 | NR | NR |
| Dalle Grave, 2007 (II) | BN | Italy | No | CBT-E | NR | NR | 25 | 91.3 | 25.6 | 71.7 | NR | NR |
| Dalle Grave, 2007 (III) | EDNOS | Italy | No | CBT-E | NR | NR | 64 | 91.3 | 26.4 | 48.3 | NR | NR |
| Dalle Grave, 2020 | AN | Italy | No | CBT-E | NR | NR | 81 | 93.8 | 30.6 | 65.0 | NR | NR |
| Dalle Grave, 2022 | AN, BN, E | Italy | No | CBT-E | NR | NR | 43 | 86.0 | 28.8 | 70.0 | NR | NR |
| Dalle Grave, 2023 | AN | Italy | No | CBT-E | NR | NR | 57 | 93.0 | 26.4 | 66.7 | NR | NR |
| De Jong, 2020 | AN, BN, E, O | Netherlands | Yes | CBT-E | NR | NR | 71 | 97.2 | 28.9 | 68.3 | NR | NR |
| Duchesne, 2007 | BED | Brazil | No | CBT | NR | NR | 21 | 86.7 | 37.1 | 62.0 | 33 | 33.3 |
| Durand, 2003 | BN | UK | Yes | GSH-CBT | B 9, W 85 | NR | 34 | 100.0 | 28.3 | 50.0 | NR | NR |
| Fernandez-Aranda 2009 (I) | BN | Spain | No | CBT | NR | NR | 150 | 100.0 | 26.7 | 49.6 | NR | NR |
| Fernandez-Aranda 2009 (II) | BN | Spain | No | CBT | NR | NR | 19 | .0 | 22.4 | 31.7 | NR | NR |
| Freeman, 1988 | BN | UK | Yes | CBT | NR | NR | 32 | 100.0 | 24.2 | NR | NR | NR |
| Frostad,2018 | AN | Norway | No | CBT-E | NR | NR | 44 | 97.7 | 23.3 | NR | 60 | NR |
| Garner, 1993 | BN | Canada | Yes | CBT | NR | NR | 30 | 100.0 | 23.7 | 61.7 | NR | NR |
| Garte, 2015 (I) | AN | Norway | No | CBT-E | NR | NR | 12 | 93.5 | 27.6 | 56.7 | 61 | NR |
| Garte, 2015 (II) | BN | Norway | No | CBT-E | NR | NR | 21 | 93.5 | 27.6 | 68.3 | 61 | NR |
| Garte, 2015 (III) | EDNOS | Norway | No | CBT-E | NR | NR | 14 | 93.5 | 27.6 | 65.0 | 61 | NR |

(Continues)

**TABLE 1** (Continued)

| Study | Eating disorder | Country | RCT | CBT-method | Race % | Ethnicity % | *N* | % Females | Mean age | % Severity | % Comorbidity | % Medicated |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Goldbloom 1997 | BN | Canada | Yes | CBT | NR | NR | 24 | 100.0 | 25.8 | 46.7 | 13 | .0 |
| Högdahl 2013 | BN, BED, E | Sweden | No | GSH-CBT | NR | NR | 44 | 95.5 | 27.9 | 66.7 | 75 | NR |
| Högdahl, 2023 | BN | Sweden | Yes | GSH-ICBT | NR | NR | 98 | 100.0 | 27.3 | 61.7 | 62 | NR |
| Ibrahim, 2022 | AN | UK | No | CBT-E | NR | NR | 34 | 100.0 | 26.9 | NR | 40 | NR |
| Jenkins, 2019 | AN | UK | No | CBT | O 16, W 84 | NR | 63 | 98.4 | 23.6 | 61.7 | NR | NR |
| Juarascio 2021 | BN | USA | Yes | CBT-E | B 6, W 89 | H 28, O 72 | 18 | 88.9 | 35.2 | 56.7 | NR | NR |
| Knott, 2015 | BN | UK | No | CBT-E | W 92 | NR | 272 | 97.1 | 28.7 | 76.7 | NR | 67.3 |
| Lammers, 2020 | BED | Netherlands | Yes | CBT + | NR | NR | 33 | 89.2 | 37.3 | 51.7 | NR | NR |
| Lammers 2022 | BED | Netherlands | No | CBT + | NR | NR | 133 | 88.7 | 33.5 | 56.7 | NR | NR |
| Melisse, 2022 (I) | BED | Netherlands | No | CBT-E | NR | NR | 113 | 89.4 | 36.2 | 63.3 | NR | 32.6 |
| Melisse, 2022 (II) | BN | Netherlands | No | CBT-E | NR | NR | 370 | 97.3 | 27.8 | 68.3 | NR | 22.1 |
| Melisse, 2022 (III) | OSFED | Netherlands | No | CBT-E | NR | NR | 142 | 95.1 | 28.3 | 65.0 | NR | 28.1 |
| Melisse, 2023 | BED | Netherlands | Yes | GSH-CBT-E | NR | NR | 90 | 92.1 | 39.2 | 65.0 | 62 | 27.9 |
| Olmsted, 1991 | BN | Canada | No | CBT-E | NR | NR | 25 | 100.0 | 23.7 | NR | NR | .0 |
| Ramklint, 2012 | BN, BED, E | Sweden | No | GSH CBT | NR | NR | 89 | 96.6 | 25.0 | 73.3 | 66 | 40.4 |
| Ricca, 2001 | BED | Italy | Yes | CBT | NR | NR | 20 | 65.0 | 26.3 | 63.3 | 14 | .0 |
| Ricca, 2010 (I) | BED | Italy | Yes | CBT | NR | NR | 72 | 86.1 | 46.5 | 53.3 | 51 | .0 |
| Ricca, 2010 (II) | BED | Italy | Yes | CBT | NR | NR | 72 | 90.3 | 47.4 | 50.0 | 57 | .0 |
| Schmidt, 2006 (I) | BN | UK | Yes | GSH-CBT + FB | NR | NR | 32 | 100.0 | 29.5 | 56.0 | NR | NR |
| Schmidt, 2006 (II) | BN | UK | Yes | GSH-CBT | NR | NR | 29 | 100.0 | 28.1 | 52.0 | NR | NR |
| Schmidt, 2008 | BN | UK | Yes | SH | W 71 | NR | 49 | 96.9 | 25.6 | 58.3 | NR | 27.9 |
| Setsu, 2018 | BN | Japan | No | GSH-CBT | NR | NR | 25 | 100.0 | 25.6 | 68.3 | 36 | 28.0 |
| Shapiro, 2010 | BN | USA | No | CBT-BN | NR | NR | 25 | 100.0 | 26.3 | 38.1 | NR | NR |
| Signorini, 2018 | AN, BN, E | Australia | No | CBT-E | NR | NR | 114 | 100.0 | 26.1 | 66.7 | 42 | NR |
| Tasca, 2019 | BED | Canada | No | SH | NR | NR | 135 | 88.9 | 41.9 | 56.7 | 17 | NR |
| Turner, 2015 | AN, BN, E | UK | No | CBT-E | NR | NR | 94 | 94.7 | 28.4 | 68.3 | NR | NR |
| Waller, 2014 | BN | UK | No | CBT | NR | NR | 78 | 100.0 | 27.8 | 54.5 | 44 | 11.5 |
| Waller, 2018 | BN, BED, O | UK | No | CBT-T | NR | NR | 93 | 96.8 | 27.4 | 68.3 | NR | NR |
| Wiberg, 2022 | BN | Sweden | No | ICBT-E | NR | NR | 41 | 95.1 | 34.5 | 66.7 | NR | NR |

*Note*: Study: Roman letters in parentheses indicate different subgroups within the same study, a and b after the year indicate different articles. No study provided information on socioeconomic status of the sample. Severity—the mean of the sample on the ED measure divided by the maximum score possible for that measure. Comorbidity—proportion having at least one comorbid disorder at inclusion in the study. Medicated—proportion taking a prescribed antidepressant medication at the inclusion.

Abbreviations: NR, not reported. Disorder: AN, anorexia nervosa; BN, bulimia nervosa; BED, binge eating disorder; E, eating disorder not otherwise specified; O, other specified feeding and eating disorder. Method of CBT: CBT, cognitive behavior therapy; CBT+, intensive outpatient CBT; CBT-E, CBT-enhanced; CBT-T, Ten session CBT; FB, feedback; GSH, guided self-help; ICBT-E, internet-based CBT-E; MFT, motivation focused therapy; SH, self-help. Race: B, black; O, other; W, white. Ethnicity: H, Hispanic; O, other.

posttreatment and follow-up assessment separately. To compare the two categories of studies on background and treatment variables, we also extracted data on mean age, proportion of women, pretreatment severity, comorbidity (% of the sample having at least one comorbid disorder), medication (% of the sample that at pretreatment was prescribed a psychotropic drug for ED), treatment time (in 60 min units), and attrition rate (% dropout of patients who participated in at least one session). Other variables were not reported systematically (or not at all) in a large enough proportion of studies, which precluded inclusion as a background variable. Since the result tables will entail many statistical tests, we used the Holm–Bonferroni correction to control the family-wise error rate (see Jaccard & Guilamo-Ramos, 2002).

## 2.9 | Power analysis

The number of studies and treatment conditions, which are the unit of analysis in the overall comparison of effectiveness and efficacy studies, were as follows: effectiveness studies 50/62 and efficacy studies 58/72. This yields a total number of 108 studies and 134 treatment conditions with an average of 54 participants per condition. According to the formulas for power analysis in meta-analyses by Valentine et al. (2010), with these figures we would have a 100% power to detect an ES of .20, assuming a high heterogeneity.

# 3 | RESULTS

## 3.1 | Description of the effectiveness studies

A total of 50 studies comprising 62 treatment conditions were included. Figure 1 presents a flowchart of the study inclusion. References to the included studies are shown in Supporting Information S5.

### 3.1.1 | Background data

Background data for the included studies are displayed in Table 1. The studies came from the following continents: Europe 37, North America 8, Australia 3, South America 1, and Asia 1. The number of conditions for the different ED were: AN 12, BN 24, BED 11, Mixed 12, and EDNOS 3. The total number of participants receiving CBT in the studies was 4299 (range 12–370), with 95.1% on average being females, however, the studies did not provide information on this being sex assigned at birth or gender. Only 16% of the studies reported race, 4% reported ethnicity, and no study reported socioeconomic status of the participants. Mean age across the studies was 29 years. Prevalence of comorbid psychiatric disorders was only reported in 40% of the conditions with a mean of 47%, and use of psychotropic medication in only 34% of the conditions with a mean of 25%. The pretreatment severity on an ED-psychopathology measure could be calculated for 95% of the conditions and the mean was 61%.

### 3.1.2 | Treatment data

Treatment data are presented in Supporting Information S6. The treatment format was individual in 39 conditions, group in 8, and a combination in 15. Treatment was carried out over a mean of 19.4 weeks (range 8–47) and 27.4 sessions (range 5–104). Calculated as hours of treatment the mean was 37.8 (range 2–104). Follow-up assessment was done in 32 conditions (52%) and on average 9.9 months (range 3–36) after the end of treatment. Intention-to-treat statistical analysis was provided for 41 conditions (66%) and completer analysis for 21.

## 3.2 | Methodological data

The research methodology score had a mean of 46.6% (*SD* 9.4), which corresponds to a raw score of 20.5 points. The RoB classification is

**TABLE 2** Results on ED-psychopathology measures for all eating disorders and specific disorders at post-assessment and follow-up assessment.

| Time point | Disorder | *k* | *g* | 95% CI | *z*-value | *Q*-value | 95% PI | Qb[a] | *p*-value |
|---|---|---|---|---|---|---|---|---|---|
| Post | All disorders | 59 | 1.12 | .98–1.25 | 16.00[b] | 575.2[b] | .12–2.11 | 5.23 | .16 |
| | AN | 10 | 1.46 | 1.12–1.79 | 8.48[b] | 45.7[b] | .40–2.51 | | |
| | BN | 23 | .98 | .76–1.21 | 8.59[b] | 231.9[b] | −.04 to 2.00 | | |
| | BED | 11 | 1.11 | .80–1.43 | 6.88[b] | 163.8[b] | .07–2.16 | | |
| | Mix | 12 | 1.15 | .84–1.45 | 7.44[b] | 79.2[b] | .10–2.19 | | |
| Follow-up | All disorders | 31 | 1.22 | 1.04–1.41 | 12.98[b] | 296.1[b] | .21–2.23 | 6.06 | .11 |
| | AN | 6 | 1.59 | 1.13–2.05 | 6.80[b] | 37.6[b] | .41–2.77 | | |
| | BN | 10 | 1.12 | .77–1.47 | 6.28[b] | 71.7[b] | .01–2.24 | | |
| | BED | 9 | .96 | .50–1.32 | 5.21[b] | 97.2[b] | −.18 to 2.10 | | |
| | Mix | 6 | 1.47 | 1.01–1.93 | 6.26[b] | 48.6[b] | .29–2.65 | | |

Abbreviations: CI, confidence interval; *k*, treatment conditions; *g*, Hedges's *g*; PI, prediction interval.
[a]Comparison between the disorders (with EDNOS, $k = 3$, deleted).
[b]$p < .0001$.

**TABLE 3** Results on BMI, binge episodes, and depressive symptoms at post-assessment and follow-up assessment.

| Time point | Measure/disorder | *k* | *g* | 95% CI | *z*-value | Q-value | 95% PI | Qb[a] | *p*-value |
|---|---|---|---|---|---|---|---|---|---|
| *BMI* | | | | | | | | | |
| Post | AN | 11 | 2.30 | 1.87–2.73 | 10.43[e] | 92.8[e] | .83–3.77 | | |
| Follow-up | AN | 6 | 1.86 | 1.39–2.34 | 7.67[e] | 41.6[e] | .59–3.13 | | |
| *Binge episodes* | | | | | | | | | |
| Post | All studies | 46 | .73 | .64–.82 | 15.54[e] | 171.2[e] | .21–1.25 | | |
| | BN | 22 | .69 | .65–.83 | 9.70[e] | 12.2[c] | .13–1.25 | 3.00 | .22 |
| | BED | 11 | .89 | .70–1.09 | 8.98[e] | 71.0[e] | .32–1.47 | | |
| | Mix | 7 | .70 | .46–.95 | 5.58[e] | 34.0[e] | .11–1.30 | | |
| Follow-up | All studies | 23 | .86 | .72–1.00 | 12.04[e] | 102.6[e] | .26–1.45 | | |
| | BN | 9 | .79 | .54–1.03 | 6.35[e] | 21.0[c] | .09–1.48 | 5.19 | .08 |
| | BED | 9 | .80 | .56–1.03 | 6.61[e] | 53.8[e] | .10–1.49 | | |
| | Mix | 4 | 1.29 | .90–1.69 | 6.40[e] | 16.9[d] | .52–2.07 | | |
| *Depression* | | | | | | | | | |
| Post | All studies | 29 | .70 | .55–.85 | 9.08[e] | 192.3[e] | −.07 to 1.46 | | |
| | BN | 14 | .82 | .59–1.06 | 6.83[e] | 62.3[e] | −.04 to 1.69 | 1.70 | .19 |
| | BED | 7 | .56 | .24–.88 | 3.41[d] | 63.7[e] | −.34 to 1.45 | | |
| Follow-up | All studies | 10 | .54 | .31–.78 | 4.55[e] | 52.8[e] | −.26 to 1.35 | | |
| | BN | 4 | .71 | .27–1.15 | 3.14[c] | 3.6 | −.35 to 1.76 | .48 | .49 |
| | BED | 5 | .51 | .14–.88 | 2.70[c] | 43.2[e] | −.51 to 1.52 | | |

Abbreviations: CI, confidence interval; *k*, treatment conditions; *g*, Hedges's *g*; PI, prediction interval.
[a]Comparison between the disorders (with EDNOS, $k = 3$, deleted).
[b]$p < .05$.
[c]$p < .01$.
[d]$p < .001$.
[e]$p < .0001$.

presented in Supporting Information S7. Among the 18 RCT conditions, 6 had a low and 12 had a moderate RoB. Regarding the 42 NRSI/pre–post conditions 38 had a moderate and 6 had a high RoB.

## 3.3 | Meta-analysis

### 3.3.1 | Attrition

Data on attrition were provided for 60 of the conditions (96.8%) and the mean rate was 25.5% (95% CI 22.2–29.0). The ED disorders had the following mean dropout rates: Mixed ED 32.4%, BN 27.5%, AN 21.2%, and BED 18.0%. The Q between studies (Qb; 3 df) was significant (8.99, $p = .029$). Subsequent pairwise comparisons showed that BED had significantly lower attrition rate than Mixed ED (Qb = 6.51, $p = .011$) and BN (Qb = 4.60, $p = .032$).

### 3.3.2 | Primary outcome measures

The primary measure on which most of the conditions (95%) provided data was *ED-psychopathology* and the results are displayed in Table 2. Fully 82% of the studies that provided such data used EDE or EDE-Q, whereas other self-report measures were used in 18% (8 studies). Since they all measure ED psychopathology, we believe that pooling within this category is acceptable. At posttreatment, the mean ES across all disorders was large (1.12) and significantly heterogeneous. A subgroup analysis comparing the types of ED did not yield a significant difference between disorders. At follow-up, the mean ES was still large (1.22) and heterogeneous with no significant differences between the ED disorders. Thus, it seems that the effects of treatment were maintained at follow-up. However, only 53% of the conditions had follow-up data. Regarding publication bias for the ED-psychopathology, Egger's regression intercept yielded a nonsignificant *t*-value (.54, $p = .56$). Thus, publication bias does not seem to be a problem for the ED-psychopathology measure.

The sensitivity analysis varying the pre–post correlation yielded the following results at posttreatment: .1:1.10, .3:1.10, .5:1.12, .7:1.12, and .9:1.12. At follow-up assessment, the ES was .1:1.21, .3:1.21, .5:1.22, .7:1.22, and .9:1.23. Thus, the mean ES changed very little due to the various estimates of the pre–post (pre-follow-up) correlation. Regarding the effect of the various ED psychopathology measures at post the overall ES was 1.12 and when BES was removed 1.10, when BITE was removed 1.12,

**TABLE 4** Subgroup analysis of the effect size for ED psychopathology measures at post-assessment.

| Variable | *k* | *g* | 95% CI | $I^2$ (%) | Qb[a] | *p*-value |
|---|---|---|---|---|---|---|
| *Statistical analysis* | | | | | | |
| Intent-to-treat | 40 | 1.08 | .92–1.25 | 92.0 | .53 | .47 |
| Treatment completers | 19 | 1.19 | .95–1.44 | 75.5 | | |
| *Risk of bias* | | | | | | |
| High | 6 | 1.04 | .64–1.44 | 87.5 | 6.12 | .05 |
| Moderate | 47 | 1.19 | 1.04–1.33 | 95.0 | | |
| Low | 6 | .67 | .29–1.06 | 86.7 | | |
| *Treatment format* | | | | | | |
| Individual | 36 | .99 | .81–1.16 | 91.7 | 7.04 | .03 |
| Group | 7 | 1.19 | .79–1.59 | 85.6 | | |
| Individual + group | 16 | 1.41 | 1.14–1.68 | 78.9 | | |
| *Continent* | | | | | | |
| Europe | 45 | 1.18 | 1.02–1.33 | 90.8 | 4.89 | .09 |
| North America | 8 | .91 | .54–1.28 | 76.4 | | |
| Australia | 4 | .68 | .20–1.16 | 0 | | |

Abbreviations: CI, confidence interval; *k*, treatment conditions; *g*, Hedges's *g*; $I^2$, the variance in observed effects that reflects variance in true effects rather than sampling error.
[a]Comparison between the subcategories.

**TABLE 5** Some background and treatment data (M and SD) for effectiveness and efficacy studies.

| Type of study | *k* | Age (years) | % females | % severity | % comorbid | % medicated | Tx hours | % attrition |
|---|---|---|---|---|---|---|---|---|
| Effectiveness | 62 | 29.0 (5.6) | 95.1 (5.2) | 61.1 (10.5) | 47.1 (16.3) | 24.3 (20.4) | 40.7 (42.7) | 25.1 (13.0) |
| Efficacy | 70 | 32.1 (8.5) | 94.8 (6.4) | 54.5 (10.3) | 42.4 (16.2) | 13.8 (17.5) | 20.2 (18.0) | 21.6 (12.3) |
| *p*-value | | .02 | .76 | <.001* | .29 | .04 | <.001* | .12 |

*Note*: *k*, number of treatment conditions; severity, percentage of the maximum score on the primary outcome measure. % comorbid, proportion having any psychiatric comorbid disorder at inclusion; % medicated, proportion on any psychotropic medication at inclusion; Tx time, number of 60 min therapy hours; % attrition, proportion dropping out of those participating in at least one therapy session.
*Significant using Holm–Bonferroni correction.

when EDI was removed 1.13, and when SEEDS was removed 1.15. At follow-up, the overall ES was 1.22, and when BITE was removed 1.22, when EDI was removed 1.22, and when SEEDS was removed 1.26. Thus, the overall ES was robust at both assessment points.

The results for the remaining primary outcome measures are presented in Table 3. BMI was provided by 92% of the conditions with AN participants and yielded a very large ES (2.30) at postassessment, which was significantly heterogeneous. At follow-up, the ES was somewhat lower (1.86). The analysis of publication bias yielded a nonsignificant Egger's regression intercept ($t = 1.93$, $p = .09$).

Regarding *binge episodes* data were provided for 74% of the conditions. The mean posttreatment ES across ED diagnoses was large (.73) and heterogeneous. At follow-up, the mean ES was still large (.86) and heterogeneous. Subgroup analyses yielded no difference between the EDs, neither at post-assessment nor at follow-up assessment. Regarding publication bias, Egger's regression intercept yielded a significant *t*-value (2.88, $p < .01$), and the Duval and Tweedie's trim and fill method suggested trimming 14 studies, which would have given a *g*-value of .57. To give a perspective on the results for this measure we calculated the percent change from pre-assessment to post-assessment and pre-assessment to follow-up assessment. The mean pre-post change was 66% (*SD* 20%) and the pre-follow-up change was 70% (*SD* 18%).

Purging episodes were reported for only 60% of the conditions, abstinence from binge eating for 47%, and abstinence from purging behaviors for 31% of the conditions. The results of these measures are shown in Supporting information S6.

### 3.3.3 | Secondary outcome measure

Depression was assessed in 47% of the total conditions but in 60% of BN and BED conditions combined. The results are displayed in the lower part of Table 3. At posttreatment the mean ES was large (.70) and heterogeneous, and at follow-up it was moderate (.54) and still heterogeneous. The subgroup analysis yielded no difference between BN and BED conditions, neither at post-assessment nor at follow-up assessment. Regarding publication bias, Egger's regression intercept was not significant ($t = 1.24$, $p = .24$).

**TABLE 6** Effect sizes on ED-psychopathology measures and BMI for effectiveness and efficacy studies at post-assessment and follow-up assessment.

| Time point | Disorder | Study type | *k* | *g* | 95% CI | *z*-value | 95% PI | Qb[a] | *p*-value |
|---|---|---|---|---|---|---|---|---|---|
| Post | All disorders | Effectiveness | 59 | 1.12 | .98–1.25 | 16.21[b] | .15–2.09 | .72 | .40 |
| | | Efficacy | 60 | 1.20 | 1.06–1.34 | 17.32[b] | .23–2.17 | | |
| | AN | Effectiveness | 10 | 1.46 | 1.14–1.78 | 8.89[b] | .40–2.52 | 5.71 | .02 |
| | | Efficacy | 7 | .85 | .48–1.23 | 4.44[b] | −.24 to 1.94 | | |
| | BN | Effectiveness | 23 | .98 | .76–1.20 | 8.71[b] | −.02-1.99 | 4.33 | .04 |
| | | Efficacy | 24 | 1.31 | 1.09–1.54 | 11.62[b] | .30–2.32 | | |
| | BED | Effectiveness | 11 | 1.11 | .80–1.43 | 6.91[b] | .05–2.18 | .07 | .79 |
| | | Efficacy | 19 | 1.06 | .82–1.30 | 8.56[b] | .01–2.10 | | |
| | Mix | Effectiveness | 12 | 1.13 | .88–1.39 | 8.66[b] | .34–2.04 | 2.89 | .09 |
| | | Efficacy | 10 | 1.47 | 1.18–1.76 | 8.91[b] | .56–2.39 | | |
| Follow-up | All disorders | Effectiveness | 31 | 1.22 | 1.04–1.41 | 12.98[b] | .21–2.23 | .15 | .70 |
| | | Efficacy | 49 | 1.28 | 1.13–1.43 | 16.74[b] | .31–2.26 | | |
| | AN | Effectiveness | 6 | 1.59 | 1.10–2.09 | 6.32[b] | .20–2.99 | 2.04 | .16 |
| | | Efficacy | 6 | 1.09 | .60–1.58 | 4.38[b] | −.31 to 2.48 | | |
| | BN | Effectiveness | 10 | 1.12 | .77–1.47 | 6.28[b] | .01–2.24 | 1.08 | .30 |
| | | Efficacy | 21 | 1.32 | 1.10–1.55 | 11.64[b] | .35–2.28 | | |
| | BED | Effectiveness | 9 | .96 | .64–1.29 | 5.80[b] | −.06 to 1.99 | .70 | .41 |
| | | Efficacy | 15 | 1.14 | .88–1.40 | 8.66[b] | .14–2.14 | | |
| | Mix | Effectiveness | 6 | 1.46 | 1.02–1.90 | 6.54[b] | .26–2.66 | .43 | .52 |
| | | Efficacy | 7 | 1.66 | 1.26–2.06 | 8.08[b] | .49–2.85 | | |
| Post | BMI in AN | Effectiveness | 11 | 2.30 | 1.87–2.73 | 10.43[b] | .83–3.77 | 5.02 | .03 |
| | | Efficacy | 7 | 1.52 | .99–2.05 | 5.66[b] | .01–3.03 | | |
| Follow-up | BMI in AN | Effectiveness | 6 | 1.86 | 1.39–2.34 | 7.67[b] | .59–3.13 | 1.51 | .22 |
| | | Efficacy | 6 | 1.45 | .99–1.91 | 6.20[b] | .10–2.71 | | |

Abbreviations: AN, anorexia nervosa; BED, binge eating disorder; BN, bulimia nervosa; CI, confidence interval; *g*, Hedges's *g*; *k*, treatment conditions; Mix, mixed eating disorders; PI, prediction interval.
[a]Comparison between effectiveness and efficacy studies.
[b]$p < .0001$.

### 3.3.4 | Moderator analysis for ED psychopathology

Since moderator analysis divides the total number of studies into subgroups, the power to detect a significant effect is much decreased. Thus, we decided to use the ED psychopathology measure, which is relevant for all ED patients and for which 95% of the conditions provided data. The subgroup analyses of categorical variables are displayed in Table 4. Using Holm–Bonferroni correction none of the variables showed a significant difference between the included categories. Regarding the six continuous variables, the meta-regression analyses did not yield a significant point estimate for any of the variables.

## 3.4 | Effectiveness–efficacy comparison

### 3.4.1 | Background and treatment variables

The comparisons of effectiveness and efficacy studies on some background and treatment variables are displayed in Table 5. Effectiveness studies had a significantly higher mean severity score and more treatment hours than efficacy studies. The latter is partly explained by a long AN treatment used in seven of the effectiveness but only two of the efficacy conditions. When AN studies were deleted from the analysis the *p*-value was .01, which is not significant using the Holm–Bonferroni correction. Thus, the only significant difference is that effectiveness studies had a higher pretreatment severity score.

### 3.4.2 | Effect size on primary outcome measure

The comparison between effectiveness and efficacy studies on eating pathology and BMI is presented in Table 6. On the ED psychopathology measure at posttreatment, there were large ESs for both types of studies with a small difference between them (1.12 vs. 1.20). Comparisons for the different EDs showed that regarding AN there was a tendency for effectiveness studies to yield a higher ES than efficacy studies (1.46 vs. .85), and for BN there was a tendency that efficacy studies gave a higher ES than effectiveness studies (1.31

vs. .98). However, when applying the Holm–Bonferroni correction none of these differences was significant. At follow-up assessment for both types of studies, the mean ES was maintained with a small difference between them (1.22 vs. 1.28). For the individual disorders, there was no significant difference between the types of studies.

The results for BMI in AN conditions are shown in the lower part of Table 6. At posttreatment, effectiveness studies had a significantly higher ES than efficacy studies (2.30 vs. 1.52). However, there was a significant outlier among the effectiveness studies and when that study was deleted the mean ES was reduced to 2.15 and the difference was no longer significant (Qb = 3.75, $p = .053$). At follow-up assessment, the types of studies did not differ significantly, and deleting the outlier reduced the ES from 1.86 to 1.60.

## 4 | DISCUSSION

### 4.1 | Effectiveness in routine care

The first aim of this meta-analysis was to examine the effectiveness of CBT for ED in routine clinical care on various primary measures of ED and a secondary measure of depression. On the ED-psychopathology measure, BMI for AN, and binge eating episodes, the ESs were large–very large for all disorders combined and for the individual disorders, and the effects were maintained at follow-up. These results corroborate the pre–post ES of 1.06 on ED-psychopathology measures of CBT-E studies (Dahlenburg et al., 2019), and the ES of 1.49 reported by Keegan et al. (2022) for 10-session CBT.

On the measure of depression, the ES was moderate and lower than the ES for the ED-psychopathology. For the studies in the present meta-analysis that provided data on both types of measures, we compared the pretreatment severity. The mean severity for depression scores (39.8%, *SD* 8.4) was significantly lower ($t(28) = 9.53$, $p < .001$) than that for ED scores (60.0%, *SD* 11.5), which means that there is less room for improvement, and the pre–post ES becomes smaller. A similar finding was obtained by Keegan et al. (2022).

### 4.2 | Methodological quality and RoB

The second aim was to evaluate methodological quality and RoB in the effectiveness studies and investigate potential moderators of treatment outcome. The mean POMRS score was 46.6%, which is somewhat lower than we have found in previous meta-analyses of effectiveness studies of anxiety disorders (50.6%; Öst et al., 2023c), depression (51.7%; Öst et al., 2023b), OCD (52.1%; Öst et al., 2022), and PTSD (53.7%: Öst et al., 2023b). Effectiveness studies on ED could improve their research methodology in various ways, for example, by using a semi-structured interview schedule for diagnosis, having independent, masked, and properly trained evaluators of outcome, controlling any concomitant treatments, and assessing clinical significance (remission). Regarding RoB six of the conditions (9.7%) had a high RoB, 50 (80.6%) had a moderate, and six (9.7%) had a low RoB. The proportion of conditions with a low RoB was too low to enable a sensitivity analysis on these studies only, and future effectiveness studies on ED should improve their methodology to reduce different risks of bias, for example, by registration of the study, analyzing all measures applied, and using intention-to-treat analysis.

The moderator analysis of categorical variables showed no significant difference between ITT and completer analysis, between levels of RoB, and between different treatment formats, which corroborate the results from our three previous meta-analyses on child disorders (Riise et al., 2021; Wergeland et al., 2021, 2022) and four on disorders among adults (Öst et al., 2022, 2023a, 2023b). There was also no significant difference in ES between continents, which runs contrary to the finding for OCD in adults (Öst et al., 2022) and internalizing disorders in children (Wergeland et al., 2021), in which Europe had a higher ES than North America. None of the six continuous variables were significantly moderating the ES. The finding that pretreatment severity was not a significant moderator was unexpected since we found it to be significant in OCD (Öst et al., 2022), anxiety disorders (Öst et al., 2023c), internalizing disorders (Wergeland et al., 2021), and externalizing disorders in children (Riise et al., 2021).

### 4.3 | Comparison with efficacy studies

The third aim was to examine how CBT was delivered in routine clinical care performed in comparison with efficacy studies for ED. We first compared effectiveness and efficacy studies on some background and treatment variables. The only significant difference was that effectiveness studies had a higher ED-psychopathology severity score than efficacy studies, which makes for a fair comparison. On this primary outcome measure, effectiveness and efficacy studies had large and very similar ESs, both at post-assessment (1.12 vs. 1.20) and follow-up assessment (1.22 vs. 1.28). This small difference in *g*-value (.08) corroborates the findings of our previous meta-analyses of effectiveness studies in adults, which varied between .01 for OCD and .20 for depression. Our results on abstinence from binge eating (41.6% for BN and 42.5% for BED, Supporting Information S6) also corroborate the 29.8% for BN and 47.2% for BED reported by Linardon, Messer, and Fuller-Tyszkiewicz (2018). Altogether, the present meta-analysis is the eighth using the same statistical comparison method and the result is the same: CBT does as well in routine clinical care as in university research settings.

### 4.4 | Strengths and limitations

This meta-analysis has a number of strong points. We could include enough effectiveness and efficacy studies for the comparison between them to have 100% power to detect a small ES, and the comparison was done statistically using meta-analytic software. Screening of abstracts and reading of full-text studies were done independently in pairs of researchers and any disagreements were solved

in consensus discussion. Ratings of methodological quality and RoB were done by one of the authors and 20% of the studies were independently rated by another to assess inter-rater reliability, which was excellent.

However, the meta-analysis also has some limitations. We only included published studies in English language journals, which means that we cannot be sure that unpublished studies, and studies in other languages, have obtained the same results. The uncontrolled pre–post ES we used is influenced by other factors than the treatment, for example, spontaneous recovery, regression to the mean, and various patient characteristics. Despite these problems, Cuijpers et al. (2017) consider this ES to be useful when comparing improvement in routine care with that in efficacy studies, which is the case in the present meta-analysis. In the included studies, different measures for assessing specific psychopathology of eating disorders have been used. Pooling the outcome based on different measures is a limitation that calls for efforts to establish international standards in the assessment of eating disorders. Reporting of the study samples' race and ethnicity was only done by 16% and 4%, respectively, and none of the studies in this meta-analysis reported socioeconomic characteristics. This is a major limitation of this research area, and researchers of therapy outcome studies are encouraged to report this information. The overwhelming majority of studies had a moderate or high RoB, which prevented an analysis of low RoB studies only. However, the moderator analysis (Table 4) of RoB categories indicated that low RoB studies had a nominally, but not significantly, lower ES than moderate and high RoB studies.

## 5 | CONCLUSION

Our findings demonstrate encouraging treatment outcomes for ED among adults in routine clinical care using CBT methods that are recommended in the clinical guidelines by NICE, APA, and APS. Clinicians trained in the CBT-ED methods can achieve outcomes comparable to those in university research settings, indicating that treatment effects are not lost when programs developed in research settings are implemented in routine clinical care. Future research on CBT effectiveness studies should investigate studies of children/adolescents with ED.

### AUTHOR CONTRIBUTIONS

**Lars-Göran Öst:** Conceptualization; data curation; formal analysis; methodology; project administration; writing – original draft; writing – review and editing. **Martin Brattmyr:** Data curation; writing – review and editing. **Anna Finnes:** Data curation; writing – review and editing. **Ata Ghaderi:** Data curation; writing – review and editing. **Audun Havnen:** Data curation; writing – review and editing. **Maria Hedman-Lagerlöf:** Data curation; writing – review and editing. **Thomas Parling:** Data curation; writing – review and editing. **Elisabeth Welch:** Data curation; writing – review and editing. **Gro Janne Wergeland:** Data curation; formal analysis; methodology; writing – review and editing.

### FUNDING INFORMATION

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

### CONFLICT OF INTEREST STATEMENT

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

### DATA AVAILABILITY STATEMENT

Data covered in this systematic review and meta-analysis are available from the corresponding author, upon reasonable request.

### ORCID

*Lars-Göran Öst* https://orcid.org/0000-0002-4351-2810
*Martin Brattmyr* https://orcid.org/0000-0003-2919-7698
*Anna Finnes* https://orcid.org/0000-0002-5615-4266
*Ata Ghaderi* https://orcid.org/0000-0001-8483-7964
*Audun Havnen* https://orcid.org/0000-0002-4610-6842
*Maria Hedman-Lagerlöf* https://orcid.org/0000-0002-3581-099X
*Thomas Parling* https://orcid.org/0000-0002-6159-598X
*Elisabeth Welch* https://orcid.org/0000-0003-3042-6187
*Gro Janne Wergeland* https://orcid.org/0000-0003-0913-5473

### REFERENCES

Agras, W. S., & Apple, R. F. (2007). *Overcoming eating disorders: A cognitive-behavioral therapy approach for bulimia nervosa and binge-eating disorder, therapist guide* (2nd ed.). Oxford University Press.

A-Tjak, J. G. L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics*, *84*, 30–36.

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

Beck, A. T., & Steer, R. A. (1993). *Beck depression inventory—Manual*. The Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Borenstein, M. (2022). In a meta-analysis, the I-squared statistic does not tell us how much the effect size varies. *Journal of Clinical Epidemiology*, *152*, 281–284.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2022). *Comprehensive meta-analysis (version 4.0)*. Biostat.

Bürkner, P.-C., Bittner, N., Holling, H., & Buhlmann, U. (2017). D-cycloserine augmentation of behavior therapy for anxiety and obsessive-compulsive disorders: A meta-analysis. *PLoS One*, *12*, e0173660.

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety*, *35*, 502–514.

Chang, P., Delgadillo, J., & Waller, G. (2021). Early response to psychological treatment for eating disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *86*, 102032.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cooper, Z., & Fairburn, C. G. (1987). The eating disorder examination. *International Journal of Eating Disorders*, *6*, 1–8.

Crone, C., Fochtmann, L. J., Attia, E., Boland, R., Escobar, J., Fornari, V., Golden, N., Guarda, A., Jackson-Triche, M., Manzo, L., Mascolo, M., Pierce, K., Riddle, M., Seritan, A., Uniacke, B., Zucker, N., Yager, J., Craig, T. J., Hong, S. H., & Medicus, J. (2023). The American Psychiatric Association practice guideline for the treatment of patients with eating disorders. *American Journal of Psychiatry*, *180*, 167–171.

Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-analysis. *Clinical Psychology Review*, *32*, 280–291.

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders*, *159*, 118–126.

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F., III. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons. *Canadian Journal of Psychiatry*, *58*, 376–385.

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

Dahlenburg, S. C., Gleaves, D. H., & Hutchinson, A. D. (2019). Treatment outcome research of enhanced cognitive behaviour therapy for eating disorders: A systematic review with narrative and meta-analytic synthesis. *Eating Disorders*, *27*, 482–502.

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

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1998). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *316*, 469.

Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*, 363–370.

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

Gan, J. K. E., Wub, V. X., Chowa, G., Chanc, J. K. Y., & Klainin-Yobasb, P. (2022). Effectiveness of non-pharmacological interventions on individuals with anorexia nervosa: A systematic review and meta-analysis. *Patient Education and Counseling*, *105*, 44–55.

Garner, D. M., & Olmsted, M. P. (1984). *The eating disorder inventory manual*. Psychological Assessment Resources.

Garner, D. M., Vitousek, K. M., & Pike, K. M. (1997). Cognitive-behavioral therapy for anorexia nervosa. In D. M. Garner & P. E. Garfinkel (Eds.), *Handbook of treatment for eating disorders*. Guilford Press.

Gormally, J., Black, S., Daston, S., & Rardin, D. (1982). The assessment of binge eating severity among obese persons. *Addictive Behaviors*, *7*, 47–55.

Hans, E., & Hiller, W. (2013). A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clinical Psychology Review*, *33*, 954–964.

Hedman-Lagerlöf, E., Carlbring, P., Svärdman, F., Riper, H., Cuijpers, P., & Andersson, G. (2023). Therapist-supported internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *World Psychiatry*, *22*, 305–314.

Henderson, M., & Freeman, C. L. P. (1987). A self-rating scale for bulimia. The BITE. *British Journal of Psychiatry*, *150*, 18–24.

Hilbert, A., Petroff, D., Herpertz, S., Pietrowsky, R., Tuschen-Caffier, B., Vocks, S., & Schmidt, R. (2020). Meta-analysis on the long-term effectiveness of psychological and medical treatments for binge-eating disorder. *International Journal of Eating Disorders*, *53*, 1353–1376.

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

Keegan, E., Waller, G., & Wade, T. D. (2022). A systematic review and meta-analysis of a 10-session cognitive behavioural therapy for non-underweight eating disorders. *Clinical Psychologist*, *26*, 241–254.

Kordy, H., Richard, M., Herrmann, A., Murphy, F., Treasure, J., & Charpentier, P. (1999). A computer assisted eating disorder specific quality management system: EQUAL-TREAT. *European Eating Disorders Review*, *7*, 239–258.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*, 606–613.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lammers, M. W., Vroling, M. S., Crosby, R. D., & van Strien, T. (2020). Dialectical behavior therapy adapted for binge eating compared to cognitive behavior therapy in obese adults with binge eating disorder: A controlled study. *The Journal of Eating Disorders*, *8*, 27.

Lammers, M. W., Vroling, M. S., Crosby, R. D., & van Strien, T. (2022). Dialectical behavior therapy compared to cognitive behavior therapy in binge-eating disorder: An effectiveness study with 6-month follow-up. *International Journal of Eating Disorders*, *55*, 902–913.

Linardon, J. (2018). Rates of abstinence following psychological or behavioral treatments for binge-eating disorder: Meta-analysis. *International Journal of Eating Disorders*, *51*, 785–797.

Linardon, J., Hindle, A., & Brennan, L. (2018). Dropout from cognitive-behavioral therapy for eating disorders: A meta-analysis of randomized, controlled trials. *International Journal of Eating Disorders*, *51*, 381–391.

Linardon, J., Messer, M., & Fuller-Tyszkiewicz, M. (2018). Meta-analysis of the effects of cognitive-behavioral therapy for binge-eating-type disorders on abstinence rates in nonrandomized effectiveness studies: Comparable outcomes to randomized, controlled trials? *International Journal of Eating Disorders*, *51*, 1303–1311.

Linardon, J., & Wade, T. (2018). How many individuals achieve symptom abstinence following psychological treatments for bulimia nervosa? A metanalytic review. *International Journal of Eating Disorders*, *51*, 287–294.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Sage.

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

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales*. The Psychology Foundation of Australia.

National Institute for Health and Care Excellence. (2017). *Eating disorders: recognition and treatment*. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/ng69

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

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.

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2022). Cognitive behavior therapy for obsessive-compulsive disorder in routine clinical care: A

systematic review and meta-analysis. *Behaviour Research and Therapy*, *159*, 104170.

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023a). Cognitive behavior therapy for adult post-traumatic stress disorder in routine clinical care: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *166*, 104323.

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023b). Cognitive behavior therapy for adult depressive disorders in routine clinical care: A systematic review and meta-analysis. *Journal of Affective Disorders*, *331*, 322–333.

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023c). Cognitive behavior therapy for adult anxiety disorders in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology: Science and Practice*, *30*(3), 272–290.

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

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

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401.

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

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*, 597–599.

Schwartze, D., Barkowski, S., Strauss, B., Knaevelsrud, C., & Rosendahl, J. (2019). Efficacy of group psychotherapy for posttraumatic stress disorder: Systematic review and meta-analysis of randomized controlled trials. *Psychotherapy Research*, *29*, 415–431.

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

Solmi, M., Wade, T. D., Byrne, S., del Giovane, C., Fairburn, C. G., Ostinelli, E. G., de Crescenzo, F., Johnson, C., Schmidt, U., Treasure, J., Favaro, A., Zipfel, S., & Cipriani, A. (2021). Comparative efficacy and acceptability of psychological interventions for the treatment of adult outpatients with anorexia nervosa: A systematic review and network meta-analysis. *Lancet Psychiatry*, *8*, 215–224.

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of intervention. *British Medical Journal*, *355*, i4919.

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *British Medical Journal*, *366*, l4898.

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

Svaldi, J., Schmitz, F., Baur, J., Hartmann, A. S., Legenbauer, T., Thaler, C., von Wietersheim, J., de Zwaan, M., & Tuschen-Caffier, B. (2019). Efficacy of psychotherapies and pharmacotherapies for bulimia nervosa. *Psychological Medicine*, *49*, 898–910.

Svanborg, P., & Åsberg, M. (1994). A new self-rating scale for depression and anxiety states based on the comprehensive psychopathological rating scale. *Acta Psychiatrica Scandinavica*, *89*, 21–28.

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

Vall, E., & Wade, T. D. (2015). Predictors of treatment outcome in individuals with eating disorders: A systematic review and meta-analysis. *International Journal of Eating Disorders*, *48*, 946–971.

van den Berg, E., Houtzager, L., de Vos, J., Daemen, I., Katsaragaki, G., Karyotaki, E., Cuijpers, P., & Dekker, J. (2019). Meta-analysis on the efficacy of psychological treatments for anorexia nervosa. *European Eating Disorders Review*, *27*, 331–351.

van Dis, E. A. M., van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L. H., van den Heuvel, R. M., Cuijpers, P., & Engelhard, I. M. (2019). Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders. A systematic review and meta-analysis. *JAMA Psychiatry*, *77*(3), 265–273.

Waller, G., Turner, H., Tatham, M., Moutford, V., & Wade, T. (2019). *Brief cognitive Behavioural therapy for non-underweight patients—CBT-T for eating disorders*. Routledge.

Wergeland, G. J. H., Posserud, M.-B., Fjermestad, K., Njardvik, U., & Öst, L.-G. (2022). Early behavioral interventions for children with autism spectrum disorder in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology: Science and Practice*, *29*, 400–414.

Wergeland, G. J. H., Riise, E. N., & Öst, L.-G. (2021). Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology Review*, *83*, 101918.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*, 631–663.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Öst, L.-G., Brattmyr, M., Finnes, A., Ghaderi, A., Havnen, A., Hedman-Lagerlöf, M., Parling, T., Welch, E., & Wergeland, G. J. (2023). Cognitive behavior therapy for adult eating disorders in routine clinical care: A systematic review and meta-analysis. *International Journal of Eating Disorders*, 1–16. https://doi.org/10.1002/eat.24104