Psychotherapy for Eating Disorders: Two Meta-Analyses of Direct Comparisons and an Evaluation of the Quality of Randomized Controlled Trials

Renee Grenon

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

© Renee Grenon, Ottawa, Canada, 2020

Abstract

Introduction: Treatment guidelines for eating disorders (EDs) recommend individual and group cognitive behavioural therapy (CBT) and individual interpersonal psychotherapy (IPT) as first-line treatments for bulimia nervosa (BN) and binge-eating disorder (BED). This suggests that CBT and IPT have sufficient research evidence and possess specific therapeutic ingredients that will result in a reduction of ED symptoms. There is limited understanding about how different psychotherapies for EDs compare in the treatment of EDs, and whether individual and group psychotherapy modalities are equally effective. Additionally, given that the findings of randomized controlled trials (RCTs) of psychotherapy for EDs are used to develop national treatment guidelines which influence clinician's practice, and that the quality of such studies can influence the magnitude of treatment effects, it is imperative to evaluate the quality of RCTs in order to better interpret the findings. Thus, the overall goals of this dissertation were to systematically and comprehensively evaluate the efficacy of psychotherapy for EDs, and to evaluate and describe the quality of said RCTs. **Method:** I conducted two meta analyses of direct comparison RCTs of psychotherapy for EDs, including both individual and group psychotherapy. In the third study, I assessed the quality of these RCTs using a tool specifically designed for psychotherapy trials. **Results:** Results indicated a clear advantage of bona fide psychotherapy over wait-list controls. Bona fide psychotherapy was superior to non-bona fide treatment (treatments that were meant to be effective but were not psychotherapy). These findings were not stable, meaning that when a single study was removed from analyses, results became non-significant. There were no significant differences between bona fide CBT and bona fide non-CBT, with the exception of bona fide CBT resulting in greater reductions in ED psychopathology that was defined by a CBT-model. Group psychotherapy was significantly

more effective than wait-list controls. The effects of group psychotherapy and other active treatments (treatments that were meant to be effective, but were not psychotherapy) did not differ on any outcome at post-treatment or at follow-ups. Group CBT and other forms of group psychotherapy did not differ significantly on outcomes at any time point. The mean total quality score for all included RCTs was in line with those reported for RCTs of psychotherapy for depression and anxiety disorders. Several standards of quality were unfulfilled by over half of the included RCTs. More recent RCTs were of higher quality, and higher quality was moderately associated with lower effect sizes. **Discussion:** Overall, results indicate that there are no differences between the included bona fide non-CBT psychotherapies and bona fide CBT in the treatment of EDs. Whereas the number of trials remains modest, I hope that as more research is conducted, treatment guidelines can be updated, and more evidence-based psychotherapies will be available for EDs. Additional research is needed to evaluate other group psychotherapy approaches, along with CBT, in order to provide more evidence-based treatment options for individuals with an ED. Group psychotherapy appears as effective as other common treatments and is perhaps more cost-effective than the most popular treatment, individual psychotherapy. To improve the quality of RCTs of psychotherapy for EDs, I recommend that researchers address the criteria assessed in quality checklists. Psychotherapy trials should be registered, have a published protocol, and be reported following the Consolidated Standards of Reporting Trials (CONSORT) guidelines. Results from the three manuscripts included in this dissertation will give readers a better understanding of the efficacy of psychotherapy for EDs so they can be better informed when interpreting the findings of studies and recommendations in treatment guidelines. Further, authors of treatment guidelines for EDs, and clinicians' who treat individuals

with EDs can consider these findings when updating treatment guidelines and when developing treatment plans for patients.

Acknowledgments

I would like to thank my committee members Drs. Louise Balfour, John Hunsley, and Stuart Hammond for your commitment, time, and feedback throughout the dissertation process. I would also like to thank Dr. Gary Burlingame from Brigham Young University for acting as my external examiner and for providing your feedback and recommendations.

I would like to thank my friends and family for supporting me throughout my graduate studies. Hilary Maxwell, Alicia McMullan, and Cathy Sawyer, this process would not have been nearly as enjoyable without your friendship, support, and humour. To my parents Susan and Randy, my sister Liza, and my brother Jesse, words cannot express my gratitude for your support and encouragement throughout my studies.

I would like to thank Pre, my husband, for his dedication to me throughout this process, for encouraging me when I needed a boost, and for celebrating with me every step along the way. We welcomed our son, Milo, into the world on September 7th, 2018. Our love for this little boy could not be any greater. He has brought so much happiness into our lives. Our two Pomeranians, Muffin and Bubbles are okay too.

Finally, I would like to thank my supervisor, Dr. Giorgio Tasca. I feel extremely lucky that I have had the opportunity to work side by side with you and to learn from you over the past eleven years. I have had many amazing opportunities throughout graduate school that would not have been possible without you. While the research projects over the years were a lot of work, I always looked forward to group meetings over delicious dinners and homemade bread. I would like to express my deepest gratitude for your mentorship and support throughout my Master's and Doctoral degrees. Thank you.

This thesis was funded by a Doctoral Award – Frederick Banting and Charles Best Canada Graduate Scholarship from the Canadian Institutes of Health Research.

Content of Dissertation and Contribution of Authors

This dissertation contains three studies. The three studies were prepared in collaboration with Dr. Giorgio A. Tasca, my supervisor. Study one and study two of this dissertation are meta-analyses. Prior to beginning my dissertation, I completed a graduate level Meta-Analysis course at Carleton University. Study three is a comprehensive and systematic evaluation of the quality of randomized controlled trials of psychotherapy for eating disorders that includes the primary studies included in study one and study two.

For study one and study two, I was responsible for completing the literature review, systematic literature search, formulating the hypotheses, coding and extracting effect size information, preparing databases in SPSS and Comprehensive Meta-Analysis (CMA), completing data analyses, and writing the introduction, methods, results, and discussion. Both studies were accepted for publication in *Psychotherapy Research* and in the *International Journal of Eating Disorders*, respectively. For both manuscripts, I am the first author. Dominique Schwartze is a doctoral student who aided in the literature search, coding and extracting effect size information, and data analyses, and she is co-author for all three manuscripts. Nicole Hammond is also a doctoral student who aided in the literature search, and coding and extracting effect size information, and she is included as a co-author for all three manuscripts. Samantha Carlucci and Agostino Brugnera are both doctoral students who aided in the literature search, and coding and extracting effect size information for manuscript one. Samantha and Agostino are included as co-authors for manuscripts one and three.

For study three, I was responsible for completing the literature review, formulating the hypotheses, rating the quality of all primary studies using the Randomized Controlled Trials-Psychotherapy Quality Rating Scale (RCT-PQRS), preparing databases in SPSS, completing data

analyses, and writing the introduction, methods, results, and discussion. The study was accepted for publication in the *International Journal of Methods in Psychiatric Research*. Alena McKenna is an undergraduate student who aided in collecting data, conducting a literature review, developing an SPSS database, and data analysis, and she is included as second author. Hilary Maxwell is a doctoral student who rated the quality of half the primary studies using the RCT-PQRS, for reliability purposes, and she is included as a co-author.

The remaining co-authors of all three manuscripts (Iryna Ivanova, Nancy Mcquaid, and Genevieve Proulx) aided in the literature search, and coding of primary studies for manuscript two.

Table of Contents

Abstract………… ii
Acknowledgments………… v
Content of the Dissertation and Contribution of Authors………… vii
Table of Contents………… ix
General Introduction………… 1
Introduction Overview………… 1
Overview of Eating Disorders………… 2
Anorexia Nervosa………… 3
Bulimia Nervosa………… 3
Binge-Eating Disorder………… 4
Eating Disorder Not Otherwise Specified………… 4
Bona Fide Psychotherapy vs. Non-Bona Fide Treatment………… 5
Group Psychotherapy vs. Individual Psychotherapy………… 7
Treatment Guidelines and Recommendations………… 8
Testing Equivalence of Treatments………… 10
Study Quality in Psychotherapy Trials………… 12
Cochrane Risk of Bias Tool………… 13
Randomized Controlled Trials-Psychotherapy Quality Rating Scale………… 15
The Current Dissertation………… 16
Study One………… 18
Abstract………… 19
Psychotherapy for Eating Disorders: A Meta-Analysis of Direct Comparisons………… 20
Method………… 22
Results………… 27
Discussion………… 33
References………… 41
Table 1………… 48
Table 2………… 52
Table 3………… 53
Table 4………… 55
Study Two………… 57
Abstract………… 58
Group Psychotherapy for Eating Disorders: A Meta-Analysis………… 59
Method………… 63
Results………… 70
Discussion………… 78
References………… 87
Table 1………… 99
Table 2………… 102
Table 3………… 104
Table 4………… 105
Figure 1………… 107
Figure 2………… 108
Figure 3………… 109
Study Three………… 110

Abstract………………………………………………………………………………………. 111
The Quality of Randomized Controlled Trials of Psychotherapy for Eating Disorders… 112
Method………………………………………………………………………………… 116
Results…………………………………………………………………...…..………... 121
Discussion…………………………………………………………………………….. 124
References……………………………………………………………………..……. 130
Table 1…………………………………………………………..……….……. 137
Figure 1………………………………………………………………..…....………. 141
Figure 2……………………………………………………………..…………… 143
General Discussion……………………………….…………………………….……….. 144
Review of the Findings ……………………………….…………………….……… 145
Implications for Treatment Guidelines for Eating Disorders………………….. 148
Implications for Group Psychotherapy for Eating Disorders………………….…... 150
Implications for the Equivalence of Treatments……………………………………… 151
Implications for Considering the Quality of Trials of Psychotherapy for Eating Disorders………………………………………………………………………... 153
Future Directions ……………………………………………………………………….. 153
Conclusions……………………………………………………………………………... 155
References……………………………………………………………………………… 157
Appendix A – Study One……………………………………………………………….... 167
Appendix B – Study Two...……………………………………………………………... 191
Appendix C – Study Three……………………………………………………………... 196

General Introduction

Introduction Overview

Eating disorders (EDs) are prevalent and result in both psychological and physical impairments (National Institute for Health and Care Excellence [NICE], 2004). Understanding the state of the research on psychotherapy for EDs has important implications for treating this burdensome disorder. For the current dissertation, I conducted a systematic and comprehensive review of the efficacy of psychotherapy for EDs with two meta-analyses of randomized controlled trials (RCTs). Additionally, I evaluated and described the quality of the RCTs and identified whether study quality was related to the magnitude of the reported treatment effects. Together, the three studies, address key issues that arise when broadly evaluating the efficacy of psychotherapy for EDs. The overarching goal of this dissertation is to give readers a better understanding of the efficacy of psychotherapy for EDs, so they can be better informed when interpreting the findings of studies and recommendations in various treatment guidelines.

The most recent American Psychiatric Association treatment guidelines (Yager et al., 2014) suggest certain psychotherapies and not others are more efficacious and have sufficient evidence to be considered evidence-based. For example, enhanced cognitive behavioural therapy (CBT) is recommended for patients with anorexia nervosa (AN) during acute refeeding and weight gain. Cognitive behavioural therapy is also recommended as the most effective and best-studied intervention for patients with bulimia nervosa (BN) and binge-eating disorder (BED). Examining whether treatment specificity can be demonstrated for EDs has clinical implications. If some psychotherapies are more efficacious than others, then only some treatments should be used for EDs. However, if many bona fide psychotherapies are equally efficacious, more treatment options would be available for individuals with an ED. In the current dissertation, I

examine whether the available research of psychotherapies for EDs supports the hypothesis of treatment specificity. Authors of treatment guidelines for EDs, and clinicians' who treat individuals with EDs, can consider the findings from each of the three studies when updating guidelines, and developing treatment plans. Results from the current dissertation will also inform future research that examines the efficacy of psychotherapy for EDs.

In this general introduction, I will begin with a brief review of EDs including the prevalence rates and diagnostic criteria of each diagnosis. I then review different treatment options for EDs including bona fide psychotherapy (individual and group) and non-bona fide treatments, or other active treatments. Following this, I will describe how the American Psychiatric Association treatment guidelines and recommendations are developed, what they include, and why they exist. I then describe the most recent American Psychiatric Association treatment guidelines and recommendations for EDs (Yager et al., 2014), and how they imply that there is treatment specificity for EDs. Lastly, I will discuss how study quality is measured in psychotherapy trials and why quality is important to consider when interpreting results of individual studies and meta-analyses for the development of treatment guidelines, theory, and future research methodology.

Overview of Eating Disorders

The most recent Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013) includes five ED diagnoses: AN, BN, BED, specified feeding and ED, and unspecified feeding and ED. The DSM-5 no longer includes the diagnostic category of ED not otherwise specified (EDNOS) that was included in the previous versions of the manual. For the purposes of this dissertation, I will be referring to the diagnostic categories of AN, BN, BED, and EDNOS, because authors of the primary treatment studies used these

categories derived from previous versions of the DSM (DSM-III; American Psychiatric Association, 1980; DSM-IV; American Psychiatric Association, 1994; DSM-IV-TR; American Psychiatric Association, 2000) to specify the study sample.

Anorexia Nervosa

Lifetime prevalence estimates for adults are approximately 0.6% in the United States with an average age of onset of 18.9 years (*SD* = 0.8; Hudson, Hiripi, Pope, and Kessler, 2007). Individuals with AN tend to over-evaluate (excessively evaluate and overestimate) their weight and shape, and their self-worth is unduly dependent upon the evaluation of their appearance (Fairburn and Harrison, 2003). One of the key characteristics of AN is the refusal to maintain a healthy body weight. Individuals with AN severely restrict their food intake which may lead to malnutrition and malfunctioning of the major organ systems. Of all the ED diagnoses, AN has the highest mortality rates at 1.2% to 12.82% (American Psychiatric Association, 2013).

Bulimia Nervosa

The lifetime prevalence estimates of BN for adults is approximately 1% in the United States with an average age of onset of 19.7 years (*SD* = 1.3; Hudson et al., 2007). Similar to individuals with AN, individuals with BN also over-evaluate their weight and shape. In addition to this cognitive component, individuals with BN engage in binge eating. Binge eating is defined as eating a large amount of food within a short amount of time while feeling a loss of control (American Psychiatric Association, 2013). A binge-eating episode may include some of the following criteria: eating faster than usual; eating not to satisfy hunger; eating in seclusion due to embarrassment; eating until painfully full; and feeling depressed or guilty following the binge (Fairburn and Wilson, 1993). To meet diagnostic criteria for BN, binge-eating episodes must be followed by inappropriate compensatory behaviours such as dietary restriction, self-induced

vomiting, excessive exercise, or laxative misuse in order to prevent weight gain (American Psychiatric Association, 2013). Although there are negative physical and mental health issues that result from binge eating and purging, fatal complications are rare (American Psychiatric Association, 2013).

Binge-Eating Disorder

Binge-eating disorder (BED) was a provisional diagnosis in the DSM-IV and is now a full category of ED in the DSM-5. Binge-eating disorder is one of the most common EDs with a lifetime prevalence estimate of approximately 3% for adults in the United States and an average age of onset of 25.4 years (*SD* = 1.2; Hudson et al., 2007). Diagnostic criteria of BED include: engaging in recurrent episodes of over-eating with loss of control (i.e., binge eating), and marked distress related to the binge eating (Fairburn and Wilson, 1993). Binge-eating disorder is associated with obesity and morbid obesity, which increases an individual's risk of mortality. As such, the specific rate of mortality for an individual with BED is not known, as the effect of obesity on mortality is difficult to quantify (Smink, Hoeken, and Hoek, 2012).

Eating Disorder Not Otherwise Specified

Individuals with an ED of clinical severity that did not meet specific diagnostic criteria for AN, BN, or BED, were previously classified as having EDNOS in the DSM-IV (in DSM-5 they may be categorized with other specified or unspecified feeding and ED). The symptoms of some individuals may resemble the full syndromes of another ED diagnostic category but fall short of meeting full clinical criteria for one or more diagnostic criteria. Other individuals may present with behavioural or cognitive features from two or more ED diagnoses combined and so may not meet strict criteria for either. Eating disorder not otherwise specified (EDNOS) was the

most prevalent diagnosis seen in outpatient treatment centres with up to 70% of patients falling within this category in the United States (Fairburn and Bohn, 2005; Mitchell et al., 2007).

All ED diagnoses have been found to be comorbid with almost all of the core DSM mood, anxiety, impulse-control, and substance use disorders (Hudson et al., 2007). Given the significant mental and physical burdens experienced by individuals diagnosed with an ED, it is no surprise that the number of treatment options for EDs have increased over the past few decades.

Bona Fide Psychotherapy vs. Non-Bona Fide Treatment

Wampold and colleagues (1997) identified “bona fide” psychotherapies as those that are intended to be therapeutic defined by the following three criteria: (1) delivered by trained therapists; (2) individualized and based on face-to-face meetings where the therapist and client develop a relationship; and (3) have specific psychologically valid components containing at least two of the following four criteria: (i) based on an established school of or approach to psychotherapy; (ii) makes reference to an established psychological change theory; (iii) uses a treatment manual, and; (iv) describes active ingredients of the psychotherapy. Historically, for EDs, CBT, a bona fide psychotherapy, has been the most researched and recommended psychotherapy (NICE, 2004; American Psychiatric Association, 2006). However, a number of non-bona fide treatment options, such as behavioural weight loss (Munsch et al., 2007; Grilo et al., 2011; Agras et al., 1994; Wilson et al., 2010), self-help (Bailer et al., 2004; Peterson et al., 1998; Peterson et al., 2009), and supportive therapy (Safer et al., 2010), have shown promising outcomes in several RCTs.

Non-bona fide treatments do not contain all the aforementioned criteria of a bona fide psychotherapy. For example, a treatment with no underlying theoretical rationale on which

treatment is based (e.g., cognitive behavioural theory, psychodynamic theory, interpersonal theory) or that does not involve face-to-face meetings with a therapist (e.g., self-help) would be considered a non-bona fide treatment. Non-bona fide treatments yield significantly worse outcomes compared to bona fide psychotherapy and pharmacotherapy in the treatment of depression (Spielmans, Berman, and Usitalo, 2011; Spielmans et al., 2007; Wampold, Minami, Baskin, and Callen Tierney, 2002). Additionally, in a meta-analysis that compared the effects of bona fide psychotherapy and non-bona fide treatment for EDs, Spielmans et al. (2013) found bona fide psychotherapy to have superior results compared to non-bona fide treatment when primary and secondary outcomes are combined. Due to methodological issues in the Spielmans et al. (2013) meta-analysis, which I outline and address in study one of this dissertation, I chose to separate primary and secondary outcomes when comparing bona fide psychotherapy and non-bona fide treatment for EDs.

It is important to evaluate and compare the efficacy of all bona fide psychotherapies. If some bona fide psychotherapies are clearly superior (relative efficacy), then that would justify their use as treatments and inclusion in treatment guidelines as first-line treatment options. However, if several or many bona fide psychotherapies yield similar outcomes, then this would argue for promoting more treatment options for individuals with an ED. Furthermore, evaluating whether certain non-bona fide treatments are effective for EDs is important as some of these treatments may be more accessible (i.e., self-help programs) and cost-effective (i.e., behavioural weight loss groups, support groups, etc.) than bona fide psychotherapy, which has to be delivered by a trained therapist. Having more evidence-based treatment options for individuals with an ED may help to improve the overall treatment outcomes for EDs.

Group Psychotherapy vs. Individual Psychotherapy

Group psychotherapy allows for more patients to be treated per clinician hour. If group psychotherapy is as effective as individual psychotherapy, it would likely be a more cost-effective treatment option. One of the largest and most recent meta-analyses that compared the efficacy of individual and group psychotherapy formats included 68 trials across a range of clinical samples (Burlingame et al., 2016). Results showed no differences in outcomes when identical individual and group psychotherapies were compared (i.e., individual CBT vs. group CBT) across homogenous patient samples of individuals with depression and individuals with anxiety, and across diverse patient samples of individuals with mixed diagnoses. Further, there were no differences between individual versus group psychotherapy in rates of treatment acceptance or dropout (Burlingame et al., 2016).

In the field of EDs, a meta-analysis of 10 RCTs comparing group psychotherapy to no treatment or to individual psychotherapy for individuals with BN indicated that group CBT was more effective than no treatment in remission outcomes at post-treatment (Polnay et al., 2014). There were insufficient data to determine whether there were differences between group psychotherapy and individual psychotherapy at post-treatment. Furthermore, for other non-CBT group treatments (i.e., behavioural therapy and nutritional counseling) the number of studies was too low to be able to conduct analyses. The authors stated that due to a limited amount of data they could only report low quality evidence for group CBT when compared to no treatment. The inability to draw clear conclusions based on statistical analyses (small sample sizes) may be due to the strict inclusion criteria used for the meta-analysis (i.e., including only BN samples, including only RCTs that compare group psychotherapy and no treatment or individual treatment, and limited search terms). With growing service demands in mental health centers,

and an increasing need for economical psychotherapeutic interventions, there has been a rise in the number of group psychotherapy trials for EDs, and findings from these studies have been positive.

It is likely because of this that the most recent American Psychiatric Association treatment guidelines and recommendations (Yager et al., 2014) include bona fide group psychotherapy, specifically CBT and IPT (Wilfley et al., 1993; Wilfley et al., 2002), as having strong research support. If individual and group formats are equally effective, as some research suggests (Burlingame et al., 2016), then more evidence-based treatment options will be available for clinicians and individuals with an ED. Study two of this dissertation is a meta-analysis of RCTs of bona fide group psychotherapy for EDs. I compare the effects of bona fide group psychotherapy to bona fide individual psychotherapy, and other active treatments. Additionally, I compare the effects of bona fide group CBT to bona fide group non-CBT psychotherapy.

Treatment Guidelines and Recommendations

The American Psychiatric Association has been developing standardized treatment guidelines to assist clinicians' in their decision making when treating patients, since 1991. Each guideline is developed by a group of psychiatrists and researchers who are selected based on their knowledge and experience in the area of interest (e.g., EDs). American Psychiatric Treatment guidelines for all disorders follow the same standardized format. Part A includes treatment recommendations, which contains an executive summary of recommendations, formulation and implementation of a treatment plan, and specific clinical features influencing the treatment plan. Part B includes background information and a review of available evidence, which contains disease definition, epidemiology, natural history, and a review and synthesis of available evidence. Finally, part C includes future research needs.

The American Psychiatric Association treatment guidelines are developed based on evidence from research and clinical consensus. The American Psychiatric Association uses a coding system that helps developers evaluate which research studies provide the most reliable and valid results. The American Psychiatric Association, American Psychological Association, and NICE propose that results from RCTs are the most influential, and so these groups use findings from RCTs to determine what treatments to recommend in their guidelines (American Psychiatric Association, 2006; American Psychological Association, 2006; NICE, 2004). The American Psychiatric Association Practice Guideline Development Process states that evidence is gathered through research studies and clinical consensus. If gaps exist within the scientific literature, evidence is derived via clinical consensus of experts in the area of interest. Eight types of research studies are hierarchically rated based on their validity and reliability. Randomized, double-blind clinical trials, which are not an option in psychotherapy research, are ranked first, followed by RCTs, clinical trials, cohort or longitudinal studies, control studies, reviews with secondary data analysis (i.e., meta-analysis), reviews (i.e., qualitative review), and opinion pieces or case studies (American Psychiatric Association, n.d.). Although single RCTs may provide some valuable information on a given treatment, the results are typically based on small samples and so estimates of effects will likely be biased (Field and Gillett, 2010). By integrating the findings from all available RCTs that examine a specific treatment one can: 1) estimate the mean and variance of underlying population effects; 2) compute the variability of treatment effects across all studies; and, 3) examine potential moderator variables. Meta-analysis has become increasingly popular in psychological research as a statistical technique that quantitatively synthesizes and summarizes the results of multiple studies examining the same research question. Because of this, results of study one and study two of this dissertation, both of

which are meta-analyses of psychotherapy for EDs, have the potential to influence later revisions of ED treatment guidelines and recommendations.

Testing Equivalence of Treatments

The notion advanced by some treatment guidelines like the American Psychiatric Association treatment guidelines (Yager et al., 2014) is that a limited number of bona fide psychotherapies, like CBT and IPT, yield superior results when compared to other bona fide psychotherapies or non-bona fide treatments, and have more sufficient research evidence. For example, Yager et al. (2014) recommend CBT as “the most effective and best-studied intervention”. The recommendations imply that certain treatments have specific effects and mechanisms of change that are different from and perhaps superior to other treatments. There has been debate among psychotherapy researchers and practitioners about what are mechanisms of change in psychotherapy. Wampold and Imel (2015) discuss three models to understand mechanisms of change and treatment effects: the medical model, common factors, and contextual model.

The medical model of psychotherapy consists of five components: (1) a classifiable illness or disorder; (2) a research-based psychological explanation for the illness or disease; (3) a theoretical conceptualization of the disease or illness that is sufficient and a mechanism for psychological change that can be identified; (4) the therapist derives a set of specific psychotherapeutic ingredients and administers them to the patient; and (5) these specific ingredients are responsible for the benefits of the psychotherapy (Wampold, Ahn, and Coleman, 2001). The medical model of psychotherapy predicts that certain psychotherapies will be superior to other psychotherapies in treating mental illnesses or disorders due to specificity of the treatment that target the mechanisms that result in the mental illness.

In contrast, the common factors model (Frank and Frank, 1993; Wampold and Imel, 2015) posits that therapeutic change is the result of several factors that are common across psychotherapies. Common factors considered necessary and sufficient for change include: (1) a bond between the therapist and client; (2) a healing and confiding therapy setting; (3) a psychologically and culturally derived explanation for the client's distress; (4) adaptive explanations and advice for overcoming distress that is accepted by the client; and (5) a series of techniques and practices engaged in by the therapist and client that lead to positive change (Wampold, 2015). Thus, the common factors model suggests that any psychotherapy that contains the aforementioned criteria, regardless of the specificity of interventions, will be effective.

The contextual model merges aspects from both the common factors model and the medical model (Wampold and Imel, 2015). The contextual model posits that there are three pathways through which psychotherapy yields benefits: (1) the personal relationship between therapist and client (a common factor); (2) the expectation that psychotherapy will enable clients to overcome or cope with their issues (a common factor); and (3) the specific therapeutic ingredients of the psychotherapy will encourage the client to engage in healthy behaviours (treatment specificity; Wampold and Imel, 2015). Overall, the contextual model suggests that psychotherapies that include the aforementioned criteria may be equally effective for a given mental illness.

Identifying whether specific psychotherapies for EDs are superior to others or whether they are equally effective has important implications. Testing the notion of treatment specificity for EDs will indicate whether some treatments and not others will be effective. If one type of psychotherapy emerges as superior to others, presumably, this would suggest that the particular

ingredients and purported mechanisms of that psychotherapy were responsible for the positive outcomes. Alternatively, if all bona fide psychotherapies are equally effective, then more treatment options would be available for individuals with an ED because clinicians trained in any bona fide psychotherapy with a symptom focus (a psychotherapy with a focus on decreasing ED symptoms either directly or indirectly) would be able to provide adequate treatment and clients might expect positive outcomes. In study one and study two of this dissertation I test whether bona fide psychotherapies are equivalent by comparing bona fide CBT, the most recommended and studied psychotherapy for EDs, to bona fide non-CBT psychotherapies.

Study Quality in Psychotherapy Trials

Another key aspect that future authors of treatment guidelines might consider when interpreting the results of meta-analyses is the quality of the RCTs that inform the guidelines. Study three of this dissertation is a comprehensive and systematic examination of the quality of all the RCTs that were included in study one and study two. Along with describing the quality of the RCTs in detail, I also compare the quality of the ED psychotherapy RCTs to psychotherapy RCTs for other mental disorders, and I assess whether study quality is related to the magnitude of effect sizes.

It is well known in other areas of research that the quality of a study can influence the magnitude of effects reported (Chambless and Hollon, 1998; Moher et al., 1995; Schulz, Chalmers, Hayes, and Altman, 1995). In general, a lower quality study is likely to contain greater risk of bias than a higher quality study, leading to an overestimation of treatment effects (Moher et al., 1998). Thus, the magnitude of effect sizes reported could be due in part to poor study quality rather than a true effect size estimate (Valentine, 2009). Because of this, it is important to take into consideration the methodological quality of psychotherapy trials when interpreting the

results. Not doing so can have a negative impact on practice, as findings from low quality studies will influence treatment guidelines and recommendations, and ultimately clinicians' practice (Armijo-Olivo, Fuentes, Ospina, Saltaji, and Hartling, 2013). Similarly, findings from RCTs influence future research directions. If an RCT has poor methodological quality that leads to overestimated effects, future research may be misguided.

Unfortunately, there is limited research in the field of psychotherapy on the impact of study quality on estimates of psychotherapy outcomes (Cuijpers, van Straten, Bohlmeijer, Hollon, and Andersson, 2010). The few studies that do examine the relationship between study quality and the magnitude of effect sizes in psychotherapy trials, address different aspects of study quality, and findings are inconsistent. Some studies report negative associations between study quality and treatment effects (Thoma et al., 2012; Cuijpers et al., 2010), whereas studies examining only a specific quality indictor (i.e., treatment integrity and intent-to-treat analyses) found no association (Gerber et al., 2011; Leichsenring et al., 2008; Perepletchikova, Treat, and Kazdin, 2007; Webb, DeRubeis, and Barber, 2010; Barth et al., 2013). Given that the findings of RCTs of psychotherapy for EDs are used to develop treatment guidelines, and that the quality of such studies can influence the estimate of the magnitude of effect sizes, it is imperative to evaluate the quality of the trials. Two measures that evaluate the quality of RCTs of psychotherapy are the Cochrane Risk of Bias Tool (Higgins, Altman, and Sterne, 2011), and the Randomized Controlled Trials-Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010).

Cochrane Risk of Bias Tool

In an effort to create a clear and well-defined scale that measures risk of bias in RCTs that examine the efficacy of a healthcare intervention (e.g., medical treatment trials,

pharmacology trials, etc.) the Cochrane Collaboration developed the Cochrane Risk of Bias Tool (Higgins, Altman, and Sterne, 2011). Bias is defined as a systematic error in the estimation and interpretation of an observed effect (Munder and Barth, 2018). Since its development, the Cochrane Risk of Bias Tool (Higgins et al., 2011) has become a widely accepted measure that is used to assess the internal validity of clinical trials. With this tool, studies are scored as having a "low risk of bias," "high risk of bias," or "unclear risk of bias" based on seven criteria, including: (1) random sequence generation; (2) allocation concealment; (3) performance bias; (4) detection bias; (5) incomplete outcome data; (6) selective reporting, and; (7) other bias (i.e., anything else, ideally pre-specified).

To assess the magnitude of bias in the group psychotherapy RCTs included in study two, I used only the Cochrane Risk of Bias criteria that are pertinent to psychotherapy trials (e.g., random sequence generation, allocation concealment, incomplete outcome data, and selective reporting). However, given that the Cochrane Risk of Bias tool was developed for medical and pharmacological contexts, it does not address some important aspects of psychotherapy research. As such, I also included additional study quality criteria that have been shown to affect outcomes in psychotherapy trials, such as: (1) researcher allegiance (Munder, Brutsch, Leonhart, Gerger, and Barth, 2013); (2) control of non-specific treatment factors (Ilardi and Craighead, 1994); (3) therapist effects (Chatoor and Kurpnick, 2001); (4) implementation quality (Chambless and Hollon, 1998); (5) analyses done with the intent-to-treat samples (Chakraborty and Gu, 2009); and, (6) addressing non-independence of the grouped data (Baldwin, Murray, and Shadish, 2005).

Given the Cochrane Risk of Bias Tool's limitations for assessing quality or bias issues in psychotherapy trials, in study one and study three I chose to use the RCT-Psychotherapy Quality

Rating Scale (PQRS; Kocsis et al., 2010) to assess the quality of all the psychotherapy RCTs. Since its development, the RCT-PQRS has been used to evaluate the quality of RCTs of psychotherapy for other mental disorders (Keefe, McCarthy, Dinger, and Barber, 2014; Thoma et al., 2012;), however, no such evaluation has been conducted for RCTs of psychotherapy for EDs.

Randomized Controlled Trials – Psychotherapy Quality Rating Scale

In 2004 the American Psychiatric Association Committee on Research appointed a subcommittee of senior psychotherapy and pharmacotherapy researchers from a variety of theoretical backgrounds to develop a rating scale to assess the methodological and reporting quality specifically for psychotherapy trials. Items were compared with the Consolidated Standards of Reporting Trials (CONSORT; Schulz et al., 2010) standards and other quality measures designed for treatment trials to ensure that all major quality criteria were included (Gerber et al., 2011). The Randomized Controlled Trials - Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010) was established as a new standard for the design and execution of psychotherapy RCTs. The RCT-PQRS is a 25-item standardized scale that is divided into 6 domains: (1) description of participants; (2) definition and delivery of treatment; (3) outcome measures; (4) data analysis; (5) treatment assignment; and (6) overall study quality. Unlike the dichotomous rating system of the Cochrane Risk of Bias Tool, items 1 through 24 on the RCT-PQRS are scored 0, 1, or 2, with no option for "not applicable" or "unclear" as scores are given for reporting quality as well as the methodology. A score of 0 is given for poor execution and description of the item. A score of 1 is given if the item is moderately described and executed, poorly described but well-executed, or well-described but poorly executed. Lastly, a score of 2 is given for well-described and executed items. Item 25 of the RCT-PQRS is an omnibus quality rating from 1 (exceptionally poor) to 7 (exceptionally good) of the overall quality of the study

(Gerber et al., 2011). In study three, I evaluate and describe the quality of RCTs of psychotherapy for EDs using the RCT-PQRS (Kocsis et al., 2010). Additionally, I examine whether study quality is related to the magnitude of treatment effects reported in study one.

The Current Dissertation

The overall goal of the current dissertation is to systematically and comprehensively evaluate the efficacy of psychotherapy for EDs in a number of ways with three different studies. It is important to note that the number of trials examining the efficacy of psychotherapy for AN is limited. Few trials in the current dissertation include individuals with AN, however, to avoid listing the specific disorders included in each analysis, I use the term ED throughout the dissertation. Additionally, the bona fide non-CBT psychotherapies that are compared directly to bona fide CBT include: IPT; exposure plus response prevention (ERP); short term focal therapy (STF); behavioural therapy (BT); hypnobehavioural therapy; emotional and social mind training (ESM), and; group psychodynamic interpersonal psychotherapy (GPIP). To avoid listing the specific non-CBT psychotherapies included in each analysis, I use the term "bona fide non-CBT psychotherapy" when describing results.

The first study is a meta-analysis of RCTs of bona fide individual and group psychotherapy for EDs. The goals of the first study are to: (1) estimate the effect of bona fide psychotherapy for adults with EDs compared to wait-list controls and to non-bona fide treatments (behavioral weight loss, self-help, supportive therapy, or non-directive therapy); and (2) test aspects of the equivalence of bona fide psychotherapy in this population by comparing the effect of bona fide CBT to bona fide non-CBT psychotherapies. In this meta-analysis I conduct three comparisons, including; (1) bona fide psychotherapy vs. wait-list controls; (2)

bona fide psychotherapy vs. non-bona fide treatment, and; (3) bona fide CBT vs. bona fide non-CBT.

The second study is a meta-analysis of RCTs of bona fide group psychotherapy for EDs. The goals of the second study are to: (1) estimate the effect of bona fide group psychotherapy for EDs compared to wait-list controls, to other active treatments, and when possible, to individual psychotherapy, and; (2) compare the effects of bona fide group CBT to bona fide non-CBT group psychotherapy. In this meta-analysis I conduct three comparisons, including; (1) bona fide group psychotherapy vs. wait-list controls; (2) bona fide group psychotherapy vs. other active treatments, and; (3) bona fide group CBT vs. bona fide non-CBT group psychotherapy.

The overall goal of the third study is to evaluate the quality of RCTs of bona fide psychotherapy for EDs using the RCT-PQRS. I examine whether study quality is related to the magnitude of treatment effects reported in the primary studies. Further, I examine whether the quality of RCTs of psychotherapy for EDs is comparable to the quality of RCTs of psychotherapy for other mental disorders.

Study One

Psychotherapy for Eating Disorders: A Meta-Analysis of Direct Comparisons*

Renee Grenon

Samantha Carlucci

University of Ottawa

Agostino Brugnera

University of Bergamo

Dominique Schwartze

Jena University Hospital

Nicole Hammond

University of Ottawa

Iryna Ivanova

Ottawa Hospital Research Institute

Nancy Mcquaid

The Ottawa Hospital

Genevieve Proulx

Giorgio A. Tasca

University of Ottawa and The Ottawa Hospital

*A version of this manuscript is published in *Psychotherapy Research*
DOI: 10.1080/10503307.2018.1489162

Abstract

Introduction: I conducted a meta-analysis of randomized controlled trials (RCTs) of bona fide individual and group psychotherapy for adults with an eating disorder (ED). **Method:** Thirty-five RCTs with 54 comparisons were included. The majority of RCTs included participants with bulimia nervosa and/or binge-eating disorder, whereas two RCTs included participants with anorexia nervosa, and three RCTs included participants with an ED not otherwise specified. **Results:** There was a clear advantage of bona fide psychotherapy over wait-list controls. Bona fide psychotherapy was superior to non-bona fide treatment; however, the majority of results were not stable. That is, when a single study was removed from analyses, the results became non-significant. There were no significant differences between bona fide cognitive-behavioural therapy (CBT) and bona fide non-CBT, with the exception of bona fide CBT resulting in greater reductions in ED psychopathology assessed by the Eating Disorder Examination, which primarily assesses maintenance factors according to the CBT model. **Discussion:** Generally, the results indicate that the bona fide psychotherapies included in the current meta-analysis are equally effective. Although the number of trials remains modest, I hope that as more research becomes available, treatment guidelines can be updated, and more evidence-based treatment options will be available for treating EDs.

Psychotherapy for Eating Disorders: A Meta-Analysis of Direct Comparisons

Individuals diagnosed with an eating disorder (ED) experience mental and physical impairments (National Institute for Health and Care Excellence [NICE], 2004). Historically, cognitive-behavioral therapy (CBT) has been the most researched and recommended treatment for bulimia nervosa (BN) and binge-eating disorder (BED; NICE, 2004; APA Work Group on Eating Disorders, 2006). American Psychiatric Association guidelines were updated in 2012 (Yager et al., 2014) to include both CBT and interpersonal psychotherapy (IPT: Wilfley et al., 1993; Wilfley et al., 2002) as first-line treatment options for BN and BED. By recommending some treatments and not others, the guidelines give the impression that certain therapies possess specific therapeutic ingredients that will result in superior efficacy for reducing ED symptoms.

Findings from several meta-analyses for other mental disorders (post-traumatic stress disorder, depression, generalized anxiety, and panic disorder) have not supported the theory of treatment specificity (Benish, Imel, and Wampold, 2008; Cuijpers, van Straten, Andersson, and van Oppen, 2008; Spielmans et al., 2007). These meta-analyses are consistent with a common factors model (Frank and Frank, 1993; Wampold and Imel, 2015) which suggests that psychotherapies that are fully intended to be therapeutic perform similarly.

There are several past meta-analyses that examine the efficacy of psychotherapy for EDs, however, many include trials with adolescents and non-RCTs (Hay, Caludino, Touyz, and Abd Elbaky, 2015; Hay, Bacaltchuk, Stefano, and Kashyap, 2009). In a recent meta-analysis, Spielmans et al. (2013) synthesized the results of 53 trials with 77 direct comparisons in order to identify whether there was support for treatment specificity among psychological treatments for BN and BED. Results indicated that: bona fide psychotherapies outperformed non-bona fide treatment; bona fide CBT marginally outperformed bona fide non-CBT for combined primary

outcomes (at post-treatment and only when BN and BED trials were combined), and full CBT treatments yielded similar effects as CBT with some active components removed. The authors point out that the finding showing CBT to be superior to non-CBT may be due to study confounds including different treatment dosages which favoured CBT. Spielmans et al. (2013) concluded that their meta-analysis generally supported the common factors model of psychotherapy. However, the 53 trials in the Spielmans meta-analysis included a variety of different study designs (i.e., RCTs, non-RCTs, dismantling studies, sequential designs). Randomized controlled trials are considered the “gold standard” method of evaluating treatment effects, and results from RCTs are often used to determine what treatments are recommended in guidelines (NICE, 2004; APA Work Group on Eating Disorders, 2006; Yager et al., 2014). By including many uncontrolled designs, the reliability of the results reported by Spielmans et al. (2013) may be limited. Researchers have shown that the biases that can result from uncontrolled designs lead to exaggerated intervention effects while hindering the reliability of results and compromising conclusions about treatments (Higgins and Green, 2011). The current meta-analysis includes only studies with RCT designs.

In addition to including non-RCTs, the Spielmans et al. (2013) meta-analysis combined psychotherapy trials with the following diverse characteristics: (a) samples of children, adolescents, and adults; (b) designs with treatment confounds (i.e., participants concurrently receiving other interventions); and (c) participants with sub-threshold or no ED diagnosis. Lastly, Spielmans et al. (2013) combined all primary outcomes together and secondary outcomes together in analyses. Although these inclusion criteria resulted in a larger pool of studies and comparisons, they may also have limited the reliability of the findings.

The current study aims to avoid potential confounds by: (a) including only RCT designs that directly compare a bona fide psychotherapy to a wait-list control condition or to non-bona fide treatment; (b) reviewing only studies with adults diagnosed with an ED; and (c) separately reporting primary (abstinence rates, frequency of binge eating and/or purging, and ED psychopathology) and secondary (depressive symptoms, self-concept, and interpersonal problems) outcomes. As CBT has been the most researched and recommended psychotherapy for BN and BED (NICE, 2004; APA Work Group on Eating Disorders, 2006; Yager et al., 2014), I also compare the effects of bona fide CBT to bona fide non-CBT. I expect that bona fide psychotherapy will be superior to wait-list controls and to non-bona fide treatments. I expect our findings to be most consistent with the common factors model (Frank and Frank, 1993; Wampold and Imel, 2015) such that bona fide CBT and bona fide non-CBT will yield similar effects.

Method

This meta-analysis builds upon a previous meta-analysis that investigated the effect of group psychotherapy for EDs (Grenon et al., 2017). However, the current meta-analysis differs by including both group and individual psychotherapy for EDs, and specifically compares the effects of bona fide psychotherapy and non-bona fide treatments. Effect sizes for group psychotherapy RCTs were taken from Grenon et al. (2017), and effect sizes for individual psychotherapy RCTs were newly extracted for this study.

Inclusion and Exclusion Criteria

The following inclusion and exclusion criteria were specified *a priori*. To be included, studies had to: (a) be published after 1979; (b) have only adult participants (> 17 years of age) diagnosed with an ED based on DSM-III, DSM-IV, or DSM 5 criteria (and revised versions); (c)

include at least one psychotherapy condition that was deemed a bona fide therapy (Wampold et al., 1997); (d) bona fide psychotherapy was compared to a wait-list control condition or a non-bona fide treatment (e.g. behavioral weight loss, self-help, supportive therapy, or non-directive therapy); and (e) use an RCT design.

Search Strategy

A systematic literature search of electronic databases (PsycINFO, MEDLINE, Cochrane Central Register of Controlled Trial) and a clinical trials registry (ClinicalTrials.gov) was conducted separately for group and individual psychotherapy RCTs.

Coding Procedures

For the purposes of classifying treatments as bona fide CBT, bona fide non-CBT, bona fide psychotherapy, or non-bona fide treatment, all studies were reviewed and coded by two psychology doctorate students. In the case of a disagreement or uncertainty about the type of treatment included in an RCT, a third rater (G.A.T.) reviewed the treatment descriptions and the three coders came to consensus.

CBT vs. Non-CBT

Interventions with both cognitive and behavioral components were classified as CBT. An intervention that contained only behavioral components was classified as behavioral therapy (BT).

Bona Fide Psychotherapy Versus Non-Bona Fide Treatment

To be classified as a bona fide psychotherapy, interventions had to include the following based on the Wampold et al. (1997) definition: (1) delivered by a therapist trained in the type of treatment and with at least a Master’s degree; (2) individualized and based on face-to-face meetings so that the therapist and client develop a relationship; and (3) have specific

psychologically valid components containing at least two of the following: (i) based on an established school or approach to psychotherapy; (ii) has an established psychological change theory; (iii) uses a treatment manual, and; (iv) describes active ingredients of the psychotherapy. Any other treatment that is intended to be effective, but is not psychotherapy, was coded as a non-bona fide treatment.

Outcome Measures

Data were extracted for the following outcome measures at post-treatment, short-term follow-up (≤ 6 months), and long-term follow-up (> 6 months). Primary outcomes included: abstinence rates; frequency of binge eating and/or purging, and ED related psychopathology. Secondary outcomes included: depressive symptoms, self-concept, and interpersonal problems.

Effect Sizes

Effect sizes for continuous outcomes were computed using Hedges' *g* (standardized mean difference). Hedges' *g* results in a more precise estimate of the effect size than Cohen's *d* (Higgins and Green, 2011), especially for studies with small sample sizes. Effect sizes for dichotomous outcomes were computed using relative risk (*RR*). The *RR* compares two groups on the probability of an outcome occurring. In the current meta-analysis, a $RR > 1.00$ indicates greater abstinence rates for bona fide psychotherapy. In the analyses comparing bona fide CBT and bona fide non-CBT, a $RR > 1.00$ indicates greater abstinence rates for CBT.

When a study reported both completer analyses and intent-to-treat analyses for the same outcome, I used the intent-to-treat data. When studies provided more than one measure for an outcome (e.g. two scales of ED related psychopathology) the effect sizes were averaged so that there was only one effect size per outcome per study (Cuijpers, 2016; Borenstein, Hedges, Higgins, and Rothstein, 2009). A correlation of 1.00 between outcomes was assumed, which

underestimates the precision of the summary effect. Effect size values of 0.20 were considered "small", 0.50 were considered "moderate", and 0.80 were considered "large" (Cohen, 1988). These effect size indices apply to between-group effect sizes, and not within-group (i.e., pre-post) effect sizes (Barber, Barrett, Gallop, Rynn, and Rickels, 2012).

Heterogeneity

The *Q*-statistic and Higgins' I^2 was used to assess for heterogeneity of effect sizes. The *Q*-statistic examines whether the variability of effect sizes is greater than would be expected by chance, and Higgins' I^2 represents the proportion of the overall variability that is beyond sampling error (Borenstein et al., 2009). When heterogeneity was significant, I examined whether treatment modality (i.e. group vs. individual psychotherapy) was a significant moderator. The Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2011) recommends conducting subgroup analyses when at least 10 studies are included. When heterogeneity was non-significant I did not conduct moderator analyses due to too few studies.

Meta-Analysis

A random-effects model was used to aggregate effects across studies. This model allows results to be generalized beyond the studies included in the meta-analysis and assumes the presence of between-study heterogeneity. In addition, a random effects model yields a more conservative effect size estimate (Cooper, Hedges, and Valentine, 2009). Three separate meta-analyses were conducted for between-group effect sizes of: (1) bona fide psychotherapy compared to a wait-list controls; (2) bona fide psychotherapy compared to non-bona fide treatment, and; (3) bona fide CBT compared to bona fide non-CBT. Meta-analyses were conducted for each outcome at post-treatment, short-term (≤ 6 months), and long-term (> 6

months) follow-up. One study removed sensitivity analyses were conducted to test the stability of the mean effect size with respect to the primary outcomes. This analysis gauges the impact of any one study on the combined effect by computing the overall effect after removing sequentially each study. Effect size calculations and analyses were computed in the Comprehensive Meta-Analysis program (CMA; Version 3; Borenstein, Hedges, Higgins, and Rothstein, 2005).

Publication Bias

For comparisons that included at least ten studies, I examined publication bias by observing the funnel plot of the primary outcome measure (Higgins and Green, 2011). When asymmetry of the funnel plot was observed and publication bias was significant as per Egger's test, I conducted the random-effects version of the Duval and Tweedie's trim and fill procedure (Duval and Tweedie, 2000). By this method, I adjusted effect sizes to correct for the number and location of missing studies and for their possible effect on the primary outcome.

Study Quality

The quality rating of all included RCTs can be found in Table 1. Quality was assessed using the omnibus rating scale (item 25) of the Randomized Controlled Trials-Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010). The omnibus quality rating ranges from one (exceptionally poor) to seven (exceptionally good) and takes into consideration the six quality domains: (1) description of participants, (2) definition and delivery of treatment, (3) full description and well qualified therapists, (4) outcome measures, (5) data analysis, and (6) treatment assignment. The correlation between the total RCT-PQRS score (items 1-24 summed) and the omnibus rating (item 25; Kocsis et al., 2010) is high, $r = 0.88$.

Results

For detailed information regarding the following: search terms and strategies; Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study flow chart; references list of included primary studies; exclusion criteria; excluded studies that underwent a full review and reasons for exclusion; primary and secondary outcome measures; references for outcome measures; detailed descriptions of non-bona fide treatments; interrater reliability, and; forest plots for abstinence rates at post-treatment, please see Dissertation Appendix A.

Study Selection

The literature search resulted in 3,464 hits (2,793 after removing duplicates). Of these, 2721 were excluded based on the title and/or abstract. The full texts of 72 studies were examined. Sixty studies were excluded because they did not meet inclusion criteria, resulting in 12 RCTs of bona fide individual psychotherapy. Twenty-three RCTs of bona fide group psychotherapy were included from a recent meta-analysis (Grenon et al., 2017). A total of 35 RCTs with 54 comparisons (see Table 1 for characteristics of included comparisons) were included in this meta-analysis.

Study Characteristics

The 35 RCTs included the following comparisons: 23 bona fide psychotherapy versus a wait-list control, 17 bona fide psychotherapy versus non-bona fide treatment, and 14 bona fide CBT versus bona fide non-CBT. The mean percentage of female participants was 99.27%. Ten RCTs compared bona fide psychotherapy to wait-list controls and reported post-treatment analyses only. Four RCTs with comparisons not including wait-list controls reported post-treatment analyses only. The average length of follow-up for the remaining 21 RCTs in months was 10.62, and only two trials included a follow-up over 12 months. Two trials included

participants with AN, 18 trials included participants with BN, 18 trials included participants with BED, and three trials included participants with an eating disorder not otherwise specified (EDNOS). Out of the 54 comparisons, 41 included CBT as the bona fide psychotherapy.

Thirty-four of 35 RCTs reported the number of participants that dropped out of treatment and/or wait-list. Mean dropout rates were comparable for bona fide psychotherapy (17.51%, SD = 11.94; range 0-53.13%; N = 45) and wait-list controls (13.20%, SD = 14.31; range 0-50%; N = 17). Non-bona fide treatment had a significantly greater mean dropout rate (29.12%, SD = 10.97; range 0-45.83%; N = 16) than the other two conditions ($F(2,75) = 7.63$, $p = .001$).

Comparison 1: Bona Fide Psychotherapy Versus Wait-List Control

Post-Treatment

Results of comparison 1 can be seen in Table 2. Bona fide psychotherapy for EDs showed statistically significantly higher abstinence rates from binge eating and/or purging (k = 14), had a greater effect in reducing the frequency of binge eating and/or purging frequency (k = 16), and was more effective in reducing ED psychopathology (k = 19) compared to wait-list control at post-treatment. The overall abstinence rates for bona fide psychotherapy and wait-list controls at post-treatment were 53.89% and 8.92%, respectively. Similar results were obtained for secondary outcomes. Statistically significant effects in favour of bona fide psychotherapy versus wait-list control were found for depressive symptoms (k = 16) and self-concept (k = 13). Results were not statistically significant for improving interpersonal problems (k = 6). Heterogeneity was not statistically significant for any analyses and so no moderators were tested. One study-removed analyses did not change the statistical significance or direction of any finding. No follow-up analyses were conducted for this comparison as wait-listed participants did not provide data following post-treatment.

Comparison 2: Bona Fide Psychotherapy Versus Non-Bona Fide Treatment

Post-Treatment

Results of comparison 2 can be seen in Table 3. Bona fide psychotherapy showed statistically significantly higher abstinence rates (51.13%) from binge eating and/or purging compared to non-bona fide treatment (40%; k = 10). However, no statistically significant difference was found between bona fide psychotherapy and non-bona fide treatment in reducing the frequency of binge eating and/or purging at post-treatment (k = 12). Heterogeneity was statistically significant for both abstinence and frequency of binge eating and/or purging. Treatment modality was not a statistically significant moderator of abstinence rates (between-level Q = 1.32, df = 1, p = .251) or frequency of binge eating and/or purging (between-level Q = 1.89, df = 1, p = .169). One likely explanation for the heterogeneity among studies is the different types of non-bona fide treatments (i.e. self-help, behavioral weight loss, non-directive/supportive therapy, nutritional counseling, self-monitoring) included in this analysis. With too few comparisons for any one treatment type, a moderator analysis was not possible. One study-removed analyses did not change the statistical significance or direction of either finding. When compared to non-bona fide treatment, bona fide psychotherapy resulted in a statistically significantly greater reduction in ED psychopathology (k = 13). Heterogeneity was not statistically significant. Two studies had a pronounced effect on the results of this analysis. Removal of Wilson, Wilfley, Agras, and Bryson (2010) or Safer, Robinson, and Jo (2010) resulted in no statistically significant difference between bona fide psychotherapy and non-bona fide treatment in reducing ED psychopathology.

No statistically significant difference was found between bona fide psychotherapy and non-bona fide treatment in reducing depressive symptoms (k = 12) or improving self-concept (k

= 4). Heterogeneity was not statistically significant and one study-removed analyses did not change the statistical significance or direction of these findings.

Follow-Up

At short-term follow-up (≤ 6 months), there was no statistically significant difference in abstinence rates between bona fide psychotherapy and non-bona fide treatment (k = 5). Heterogeneity was not statistically significant. Removal of Safer et al. (2010) lead to a statistically significant difference in abstinence rates in favour of bona fide psychotherapy. Bona fide psychotherapy showed a statistically significantly greater reduction in frequency of binge eating and/or purging compared to non-bona fide treatment (k = 3), however, removal of the Grilo, Masheb, Wilson, Gueorguieva, and White (2011) study resulted in no statistically significant difference. Bona fide psychotherapy yielded a statistically significantly greater reduction in depressive symptoms (k = 3) compared to non-bona fide treatment and no statistically significant difference was found between bona fide psychotherapy and non-bona fide treatment in reducing ED psychopathology (k = 3) or improving self-concept (k = 2). Heterogeneity was not statistically significant for any of these analyses. No studies reported interpersonal problems outcomes for this comparison.

At long-term follow-up (> 6 months), bona fide psychotherapy showed statistically significantly higher abstinence rates compared to non-bona fide treatment (k = 7), however, removal of either Safer et al. (2010) or Wilson et al. (2010) results in no statistically significant difference. There were no other statistically significant differences between bona fide psychotherapy and non-bona fide treatment at long-term (> 6 months) follow-up. Heterogeneity was not statistically significant for any of these analyses. No studies reported interpersonal problems as an outcome for this comparison at any time-point.

Comparison 3: Bona Fide CBT Versus Bona Fide Non-CBT

Post-Treatment

Results of comparison 3 can be seen in Table 4. There were no statistically significant differences between bona fide CBT and bona fide non-CBT in abstinence rates (58.92% and 48.15%, respectively; $k = 7$), frequency of binge eating and/or purging ($k = 5$), depressive symptoms ($k = 7$), self-concept ($k = 5$), or interpersonal problems ($k = 4$). Heterogeneity was not statistically significant for these analyses with the exception of frequency. However, with too few comparisons ($k = 5$), moderator analyses were not conducted. Bona fide CBT was superior to bona fide non-CBT in reducing ED psychopathology ($g = .31$, $p < .001$) at post-treatment ($k = 12$). The effect was small, heterogeneity was not statistically significant, and a one study removed analysis did not change these findings

Follow-up

At short-term (≤ 6 months) follow-up, there were no statistically significant differences between bona fide CBT and bona fide non-CBT in abstinence rates ($k = 3$), depressive symptoms ($k = 4$), self-concept ($k = 4$), or interpersonal problems ($k = 3$). Heterogeneity was not statistically significant for these analyses with the exception of self-concept. Moderator analyses were not conducted as there were too few comparisons ($k = 4$). Only one study (Wolf and Crowther, 1992) reported frequency of binge eating and/or purging at short-term follow-up and results were statistically significantly in favour of behavior therapy (BT) over CBT. Bona fide CBT was superior to bona fide non-CBT in reducing ED psychopathology ($g = .24$, $p = .013$; $k = 6$) at short-term (≤ 6 months) follow-up. The effect was small and heterogeneity was not statistically significant. A one study-removed analysis did not change these findings, with the exception of depressive symptoms. Removal of Tasca et al. (2006) lead to a statistically significant difference

between bona fide CBT and bona fide non-CBT (in favour of CBT) in improving depressive symptoms.

At long-term (> 6 months) follow-up, there were no statistically significant differences between bona fide CBT and bona fide non-CBT for any outcome. Heterogeneity was not statistically significant for these analyses with the exception of abstinence. To attempt to explain the variability among effect sizes for abstinence I computed effects for group psychotherapy and individual psychotherapy separately. Treatment modality was not a statistically significant moderator of abstinence rates (between-level $Q = 2.91$, $df = 1$, $p = .088$). A one study-removed analysis did not change the statistical significance or direction of any of these findings.

Publication Bias

Eleven analyses included ten or more trials. Publication bias was detected in two analyses: (1) bona fide psychotherapy vs. wait-list for depressive symptoms outcome (adjusted for one study missing); and, (2) bona fide psychotherapy vs. wait-list for the self-efficacy outcome (adjusted for four studies missing). After adjusting with the trim and fill procedure, the overall effect sizes for depressive symptoms and self-efficacy remained non-significant. The conclusions are therefore robust even if it is assumed that one and four studies are missing.

Study Quality

The average omnibus rating (item 25) was 4.51 ($SD = 1.22$; range = 2, 7) out of a possible range of one (exceptionally poor) to seven (exceptionally good). Twelve RCTs (34.3%) received an “average” rating (omnibus rating of 4) whereas only two RCTs (5.7%) received an “exceptionally good” rating (omnibus rating of 7). See Table 1 for all omnibus ratings.

Discussion

The primary goal of the current meta-analysis was to estimate the effect of bona fide psychotherapy for adults with EDs. Compared to wait-list control conditions, bona fide psychotherapy yielded superior results at post-treatment with small to large effect sizes (*g* range = .24 to .81). Participants who received bona fide psychotherapy were, on average 5.64 times more likely to be abstinent from binge eating and/or purging than participants in the wait-list control conditions. On average, 2.2 (numbers needed to treat; *NNT*) individuals would have to receive bona fide psychotherapy (instead of wait-list) for one additional individual to be abstinent at post-treatment. These findings are consistent with the general psychotherapy research in which bona fide psychotherapy is more effective than wait-list control conditions (Wampold et al., 1997).

Compared to non-bona fide treatment, participants who underwent bona fide psychotherapy were, on average 1.41 times more likely to be abstinent from binge eating and/or purging at post-treatment. However, many of the significant differences between bona fide psychotherapy and non-bona fide treatment were not reliable once single studies were removed from the analyses, and so results should be interpreted with caution. Significant heterogeneity was observed for abstinence and frequency of binge eating and/or purging at post-treatment. I suspect that this is a result of the differences in the focus of the non-bona fide treatments. For example, whereas all the bona fide psychotherapies are fairly similar with regard to effects on the outcomes included, the non-bona fide treatments vary in their focus (e.g., nutritional therapy, supportive therapy) and so might yield different effects on these outcomes. The majority of analyses of other primary outcomes showed similar outcomes for bona fide psychotherapy and non-bona fide treatment. These findings are in contrast to the Spielmans et al. (2013) meta-

analytic findings in which bona fide psychotherapy was superior to non-bona fide treatment for combined primary outcomes at both post-treatment and follow-up. A likely explanation for these contradictory findings is that unlike the Spielmans et al. review, the current meta-analysis included only RCTs. Also, I did not combine primary outcomes and secondary outcomes. The non-bona fide treatments included in this review may contain key common factors (e.g., therapeutic alliance, instillation of hope, therapist empathy, intention to be effective) that may account for the relative effectiveness of these interventions. Similarly, some of the non-bona fide treatments may contain some specific ingredients as well. I encourage researchers to continue investigating non-bona fide treatments so that clear conclusions can be drawn for these treatments. Currently, based on the available literature, I recommend that adults with an ED seek out bona fide psychotherapy to optimize their treatment outcomes, although I recognize that some non-bona fide interventions on average may be just as effective. Nine individuals would have to receive bona fide psychotherapy instead of non-bona fide treatment for one additional individual to be abstinent at post-treatment.

Our second goal was to investigate whether bona fide psychotherapies yielded equivalent effects by comparing bona fide CBT to bona fide non-CBT psychotherapy. Bona fide psychotherapy, regardless of treatment orientation (CBT vs. non-CBT), performed similarly for almost all outcomes at all time points. However, at post-treatment ($g = .31$) and short-term follow-up (≤ 6 months; $g = .24$), bona fide CBT showed small superior effects for improving ED psychopathology compared to bona fide non-CBT. One possible explanation for this finding is that the majority (75%) of the studies used the Eating Disorder Examination (EDE; Fairburn and Cooper, 1993) and the EDE-Questionnaire (EDEQ; Fairburn and Beglin, 1994) as a measure of ED psychopathology. These measures were designed to assess the cognitive and behavioural

maintenance factors of EDs (Fairburn and Cooper, 1993; Fairburn and Beglin, 1994). For example, items ask participants about their desire to lose weight, fear of losing control over eating, and whether they have deliberately limited the amount of food they consume. Thus, it is not surprising that the items on these measures are particularly sensitive to CBT interventions. By contrast bona fide CBT and bona fide non-CBT did not differ on depression, self-concept, and interpersonal problems which are mechanisms purported by the interpersonal model to underlie binge eating (Wilfley et al., 2002). Our findings are consistent with what was demonstrated in CBT trials for depression. For example, Shapiro et al. (1994) compared CBT and psychodynamic interpersonal psychotherapy for depression and found no differences between the two treatments on eight outcome measures except for the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, and Erbaugh, 1961), which resulted in a large effect in favour of CBT, a treatment also developed by Beck, Rush, Shaw, and Emery (1979). Similarly, in a meta-analysis, Oei and Free (1995) demonstrated that change in cognitions following psychotherapy were related to decreases in depressive symptoms as measured only by the BDI (Beck et al., 1961) but on no other depression scale.

Our findings are similar to those reported in the Spielmans et al. (2013) meta-analysis with one exception. Spielmans et al. found bona fide CBT to be superior to bona fide non-CBT on combined primary outcomes (including abstinence, frequency of binge eating and/or purging, and ED psychopathology) at post-treatment. The authors pointed out that this result may be due to study confounds which favoured CBT. Given the results of the current meta-analysis, there does not appear to be any clear advantage of bona fide CBT over other bona fide psychotherapies in improving abstinence rates, reducing the frequency of binge eating and/or purging. On average, 9.3 individuals (*NNT*) would have to receive bona fide CBT (instead of bona fide non-

CBT) for one additional individual to be abstinent at post-treatment. Although treatment guidelines recommend CBT and IPT as first-line treatment options (NICE, 2004; APA Work Group on Eating Disorders, 2006; Yager et al., 2014), the available evidence in the ED literature, where a bona fide non-CBT psychotherapy is directly compared to CBT, does not show any one psychotherapy to be superior. Although, CBT may be more effective in reducing CBT-defined ED psychopathology. The findings suggest that non-CBT therapies should be more extensively evaluated as first-line treatment options.

Identifying what specific factors are responsible for psychotherapeutic change for EDs is beyond the scope of our study. However, past meta-analyses that examined the effects of both common (i.e. alliance, goal consensus, expectations, etc.) and specific (i.e. differences between treatment types, adherence to protocol, specific ingredients, etc.) psychotherapy factors for other mental disorders (see Norcross, 2011; Baldwin and Imel, 2013; Wampold et al., 1997; Bell, Marcus, and Goodlad, 2013; Ahn and Wampold, 2001; Webb, DeRubeis, and Barber, 2010) indicated that effects produced by the common factors were much larger than those produced by specific factors. More research is needed in the ED field to replicate these findings regarding common factors; however, I suspect that similar results will be obtained among those with an ED. For example, with regard to alliance, a meta-analysis by Graves et al. (2017) indicated that early symptom improvement was related to subsequent therapeutic alliance level, and that previous alliance level was related to subsequent symptom improvement in individuals with an ED.

An implication of these findings is that therapists skilled in one of the efficacious bona fide psychotherapies included in the current meta-analysis may be able to effectively treat individuals with an ED. Recommending only one or two bona fide psychotherapies for

individuals with an ED, as is done by some treatment guidelines, may not be consistent with the cumulative evidence. Taken together, the RCTs included in the current meta-analysis (that directly compare CBT to another bona fide psychotherapy) highlight the lack of evidence supporting a hypothesis of superiority of one treatment approach over another. Current guidelines recommend the few psychotherapies that have shown statistically significant superiority mostly to no treatment. Division 12 of the American Psychological Association define some of these therapies as “strongly supported” (also referred to as research supported or empirically supported; APA Presidential Task Force on Evidence-Based Practice, 2006). However, in a meta-scientific review, Sakaluk and colleagues (2019) recently found that most of the RCTs cited in support of these treatments are not high quality. The notion that more research support is better (regardless of the quality of the research) should be reconsidered when future treatment guidelines are developed. Ideally, the quality of single research studies and the cumulative evidence from meta-analyses both should be considered before identifying any psychotherapy as having strong research support. Ultimately, including a wider range of bona fide psychotherapies as treatment options would give providers and individuals with an ED more treatment options. Some have argued that to enhance outcomes, common or contextual factors need to be taken into consideration when delivering treatment (Wampold and Imel, 2015). For example, the therapist can: a) adapt treatment to the patient’s culture and values; and, b) develop and maintain a therapeutic alliance (Wampold and Imel, 2015).

Limitations and Future Directions

Many of the analyses in this study included a small number of comparisons, and there was insufficient power to detect differences or conduct meaningful moderator analyses. The small number of comparisons may also have influenced the results of the sensitivity analyses.

For example, the removal of some comparisons that lead to a different conclusion could have been due to their large sample sizes compared to the other comparisons in the analysis. Thus, results should be interpreted with an appropriate degree of caution. As more trials become available, the number of comparisons needed for these analyses will increase, the power of the analyses will be improved, and the effects of important moderators (i.e. type of non-bona fide treatment, ED diagnosis) can be assessed. There is a need for more RCTs on bona fide psychotherapy for AN, however, high treatment refusal rates, low recruitment numbers, and an approximately 40% dropout rate make completing such trials difficult (Halmi, 2008). Additionally, over half (53%) of the bona fide vs. non-bona fide treatment comparisons include CBT as the bona fide psychotherapy. I encourage researchers to continue investigating all types of bona fide psychotherapies for EDs so that treatment guidelines can be updated accordingly, and more evidence-based treatment options can be available for individuals with an ED.

Although I evaluated the quality of all included RCTs, I did not directly examine the effect of overall quality or single quality items (i.e., researcher allegiance, therapist allegiance, etc.) on the magnitude of effect sizes. I recommend that in the future, the quality of RCTs of psychotherapy for EDs be investigated and the effect of quality on the magnitude of effect sizes be reported.

The current meta-analysis used a different coding procedure for CBT vs. non-CBT bona fide psychotherapies than Spielmans et al. (2013), which can lead to different results. For example, in the current study, behaviour therapy without a cognitive component and cognitive therapy without a behavioural component were not coded as CBT. The categorization of types of psychotherapies is something that is often overlooked in psychotherapy research but should be taken into consideration in the future.

Our data came from published trials (with the exception of one dissertation; Allen, 1997) and publication bias was identified in two analyses. To improve the results of future meta-analyses, I suggest all trials be registered, have a published protocol, and all results be reported in peer-reviewed journals following the CONSORT (CONsolidated Standards of Reporting Trials) standards. Finally, as is the case in ED research in general, males and other ethnic groups were underrepresented.

Conclusion

Overall, results of this meta-analysis indicate that bona fide psychotherapy is effective in the treatment of EDs. Although bona fide psychotherapy and non-bona fide treatments yielded similar results for many outcomes, bona fide psychotherapy outperformed non-bona fide treatments on several outcomes, though the latter results may not be stable. Until further research in conducted and more accurate conclusions can be drawn, I recommend individuals with an ED seek out a bona fide psychotherapy as a first-line treatment option. As more research on non-bona fide treatments becomes available, it will be possible to identify whether any of these treatment options (i.e. self-help, supportive therapy, behavioural weight loss) yield outcomes equivalent to or better than those of bona fide psychotherapy.

My results are most consistent with common factors model, in that bona fide psychotherapy, regardless of treatment orientation, yielded similar effects. Therapeutic alliance is the most researched common factor in psychotherapy research (Huibers and Cuijpers, 2015). However, more research is needed to identify mechanisms by which alliance and other common factors are responsible for therapeutic change in ED treatment trials. Future research should be designed to enable researchers to isolate and manipulate key common and specific factors while treating individuals with an ED in order to examine causal links to outcomes. My conclusion at

this time is in line with that of Wampold (2005) who posits that *how* a bona fide psychotherapy is conducted is more important than *what* bona fide psychotherapy is conducted.

References

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling. *Journal of Counseling Psychology, 48,* 251–257.

Allen, H. N. (1997). *Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder* (doctoral dissertation). University of North Carolina at Chapel Hill, North Carolina. Retrieved from ProQuest Dissertations and Theses.

American Psychiatric Association Work Group on Eating Disorders (2006). *Treatment of patients with eating disorders.* Retrieved from http://psychiatryonline.org/ content.aspx?bookid=28§ionid=1671334#138660

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285.

Bailer, U., de Zwaan, M., Leisch, F., Strnad, A., Lennkh-Wolfsberg, C., El-Giamal, N., ... Kasper, S. (2004). Guided self-help versus cognitive-behavioral group therapy in the treatment of bulimia nervosa. *International Journal of Eating Disorders*, *35,* 522-537.

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change (6th ed.,* pp. 258-297). New York: Wiley.

Barber, J. P., Barrett, M. S., Gallop, R., Rynn, M. A., & Rickels, K. (2012). Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *Journal of Clinical Psychiatry, 73*, 66-73.

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

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York, NY: Guilford Press.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*, 722-736.

Benish, S. G., Imel, Z. E., & Wampold, B. E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review, 28,* 746-758.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 3) [Computer Software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* United Kingdom: John Wiley & Sons, Inc.

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

Cooper, H., Hedges, L., & Valentine, J. C. (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). Russell Sage Foundation.

Cuijpers, P. (2016). *Meta-analyses in mental health research. A practical guide.* Retrieved from: https://www.researchgate.net/publication/301815425_Meta-analyses_in_mental_health_research_A_practical_guide

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76,* 909-922.

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.

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

Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination. 12th ed. In C. G. Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). Baltimore, MD: Johns Hopkins University Press.

Graves, T.A., Tabri, N., Thompson-Brenner, H., Franko, D.L., Eddy, K.T., Bourion-Bedes, S., … Thomas, J. J. (2017). A meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders. *International Journal of Eating Disorders, 50*, 323-340.

Grenon, R., Schwartze, D., Hammond, N., Ivanova, I., Mcquaid, N., Proulx, G., & Tasca, G. A. (2017). Group psychotherapy for eating disorders: A meta-analysis. *International Journal of Eating Disorders, 50,* 997-1013. doi: 10.1002/eat.22744

Grilo, C. M., Masheb, R. M., Wilson, G. T., Gueorguieva, R., & White, M. A. (2011). Cognitive–behavioral therapy, behavioral weight loss, and sequential treatment for obese patients with binge-eating disorder: A randomized controlled trial. *Journal of consulting and clinical psychology*, *79*, 675-685.

Halmi, K. A. (2008). The perplexities of conducting randomized, double-blind, placebo-controlled treatment trials in anorexia nervosa patients. *American Journal of Psychiatry, 165,* 1227-1228.

Hay, P. P. J., Bacaltchuk, J., Stefano, S., & Kashyap, P. (2009). Psychological treatments for bulimia nervosa and bingeing (review). *Cochrane Database of Systematic Reviews, 4,* 1-94. doi: 10.1002/14651858.CD000562.pub3.

Hay, P. P. J., Caludino, A. M., Touyz, S., & Abd Elbaky, G. (2015). Individual psychological therapy in the outpatient treatment of adults with anorexia nervosa (review). *Cochrane Database of Systematic Reviews, 7*, 1-97. doi: 10.1002/14651858.CD003909.pub2.

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration, 2011. Retrieved from www.cochrane-handbook.org.

Huibers, M. J. H., & Cuijpers, P. (2015). Common (nonspecific) factors in psychotherapy. In: R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (1st ed.). New York: John Wiley & Sons.

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., & Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry, 51,* 319-324.

Laessle, R. G., Zoettl, C., & Pirke, K. M. (1987). Meta analysis of treatment studies for bulimia. *International Journal of Eating Disorders, 6*, 647-653.

National Institute for Health and Care Excellence. (2004). *Eating disorders: Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders.* Retrieved from http://www.nice.org.uk/nicemedia/live/10932/29220/29220.pdf

Norcross, J. C. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness.* New York: Oxford University Press.

Oei, T. P. S., & Free, M. S. (1995). Do cognitive behaviour therapies validate cognitive models of mood disorders? A review of the empirical evidence. *International Journal of Psychology, 30,* 145-180.

Peterson, C. B., Mitchell, J. E., Crow, S. J., Crosby, R. D., & Wonderlich, S. A. (2009). The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. *American Journal of Psychiatry, 166,* 1347-1354.

Peterson, C.B., Mitchell, J.E., Engbloom, S., Nugent, S., Mussell, M.P., & Miller, J. P. (1998). Group cognitive-behavioral treatment of binge-eating disorder: A comparison of therapist-led versus self-help formats. *International Journal of Eating Disorders, 24*, 125–136.

Safer, D. L., Robinson, A. H., & Jo, B. (2010). Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. *Behavior Therapy, 41*, 106-120.

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522-534.

Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology, 128,* 500-509. doi: 10.1037/abn0000421

Spielmans, G. I., Benish, S. G., Marin, C., Bowman, W. M., Menster, M., & Wheeler, A. J. (2013). Specificity of psychological treatments for bulimia nervosa and binge eating disorder? A meta-analysis of direct comparisons. *Clinical Psychology Review, 33*, 460-469.

Spielmans, G. I., Pasek, L. F., & McFall, J. P. (2007). What are the active ingredients in cognitive and behavioral psychotherapy for anxious and depressed children? A meta-analytic review. *Clinical Psychology Review*, *27*, 642-654.

Tasca, G. A., Ritchie, K., Conrad, G., Balfour, L., Gayton, J., Lybanon, V., & Bissada, H. (2006). Attachment scales predict outcome in a randomized controlled trial of two group therapies for binge eating disorder: An aptitude by treatment interaction. *Psychotherapy Research, 16*, 106-121.

Vocks, S., Schulte, D., Busch, M., Grönemeyer, D., Herpertz, S., & Suchan, B. (2011). Changes in neuronal correlates of body image processing by means of cognitive-behavioural body image therapy for eating disorders: a randomized controlled fMRI study. *Psychological Medicine, 41*, 1651-1663.

Wampold, B. E. (2005). Establishing specificity in psychotherapy scientifically: Design and evidence issues. *Clinical Psychology: Science and Practice, 12,* 194-197.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York: Routledge.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially, 'all must have prizes'. *Psychological Bulletin, 122,* 203–215.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 70,* 299-310.

Wilfley, D. E., Agras, W. S., Telch, C. F., Rossiter, E. M., Schneider, J. A., Cole Golomb, … Raeburn, S. D. (1993). Group cognitive-behavioral therapy and group interpersonal

psychotherapy for the nonpurging bulimic individual: A controlled comparison. *Journal of Consulting and Clinical Psychology, 61*, 296-305.

Wilfley, D. E., Welch, R. R., Stein, R. I., Spurrell, E. B., Cohen, L. R., Saelens, B. E., … Matt, G. E. (2002). A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the tretament of overweight individuals with binge-eating disorder. *Archives of General Psychiatry, 59*, 713-721.

Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2010). Psychological treatments for binge eating disorder. *Archives of General Psychiatry, 67,* 94-101.

Wolf, E. M., & Crowther, J. H. (1992). An evaluation of behavioral and cognitive-behavioral group interventions for the treatment of bulimia nervosa in women. *International Journal of Eating Disorders, 11,* 3-15.

Yager, J., Devlin, M. J., Halmi, K. A., Herzog, D. B., Mitchell, J. E., Powers, P., & Zerbe, K. F. (2014). Guideline watch (August 2012): Practice guideline for the treatment of patients with eating disorders (3rd ed.) *Focus, 12*, 416–431.

Table 1

Characteristics of Included Studies and Comparisons

Study	*C*	*IG*	*Tx Type*	*CG*	N_{IG}	N_{CG}	*Subjects*		
							Diagnosis	*Age M (SD)*	*% Female*
Bona fide psychotherapy versus wait-list control									
Agras *et al*. 1989	US	CBT	Ind.	WL	39	19	BN	29.2 (8.6)	100
Agras *et al.* 1995	US	CBT	Grp.	WL	39	11	BED	47.6 (10.10)	86
Allen, 1997	US	CBT (AAT)	Grp.	WL	15	14	BED	21 (na)	100
Freeman *et al.* 1988	UK	CBT	Ind.	WL	32	20	BN	24.2 (5.6)	100
		BT	Ind.	WL	30	20	-	-	-
Gorin *et al.* 2003	US	CBT	Grp.	WL	32	31	BED	45.2 (10.03)	100
Griffiths *et al.* 1994	AU	CBT	Ind.	WL	23	28	BN	25.91 (5.73)	100
Laessle *et al.* 1987	DE	CBT	Grp.	WL	8	9	BN	23.39 (2.5)	100
Lee *et al.* 1986	US	CBT	Grp.	WL	15	15	BN	27.7 (5.3)	100
Leitenberg *et al.* 1988	US	CBT	Grp.	WL	12	12	BN	26 (6.04)	100
		CBT+ERP	Grp.	WL	11	12	-	-	-
Peterson *et al.* 1998	US	CBT	Grp.	WL	16	11	BED	42.4 (10.2)	100
Peterson *et al.* 2009	US	CBT	Grp.	WL	60	69	BED	47.10 (10.4)	90
Schlup *et al.* 2009	CH	CBT	Grp.	WL	18	18	BED	44.3 (10.3)	100
Tasca *et al.* 2006	CA	CBT	Grp.	WL	50	46	BED	42.75 (10.76)	91
		GPIP	Grp.	WL	50	46	-	-	-
Telch et al. 1990	US	CBT	Grp.	WL	23	21	BED	42.5 (8.4)	100

Study	*C*	*IG*	*Tx Type*	*CG*	N_{IG}	N_{CG}	*Subjects*		
							Diagnosis	*Age M (SD)*	*% Female*
Telch *et al.* 2001	US	DBT	Grp.	WL	22	22	BED	50 (9.1)	100
Vocks *et al.* 2011	DE	CBT (BIT)	Grp.	WL	32	30	AN/BN/EDNOS	28.22 (6.73)	100
Wilfley *et al.* 1993	US	CBT	Grp.	WL	18	20	BED	44.3 (8.3)	100
		IPT	Grp.	WL	18	20	-	-	-
Wolf *et al.* 1992	US	CBT	Grp.	WL	15	12	BN	26 (na)	100
		BT	Grp.	WL	15	12	-	-	-
Bona fide psychotherapy versus non-bona fide treatment									
Agras et al. 1994	US	CBT	Grp.	BWL	72	37	BED	45 (10)	100
Bailer *et al.* 2004	AT	CBT	Grp.	Self-help	41	40	BN	23.76 (4.5)	na
Esplen *et al.* 1998	CA	Guided Imagery	Ind.	ND	28	30	BN	26.6 (6.0)	96.65
Freeman *et al.* 1988		CBT	Ind.	SG	32	30	-	-	-
		BT	Ind.	SG	30	30	-	-	-
Grilo *et al.* 2011	US	CBT	Grp.	BWL	45	45	BED	44.8 (9.4)	67
Hsu *et al.* 2001	US	CT	Ind.	SG	26	24	BN	24.5 (6.4)	100
		CT	Ind.	NT	26	23	-	-	-
Kirkley *et al.* 1985	US	CBT	Grp.	ND Group	14	14	BN	28.3 (na)	100
McIntosh *et al.* 2005	NZ	CBT	Ind.	SCM	19	16	AN	na	100
	NZ	IPT	Ind.	SCM	21	16	AN	na	100
Munsch *et al.* 2007	CH	CBT	Grp.	BWL	44	36	BED	46.1 (11.65)	89
Peterson *et al.* 1998	US	CBT	Grp.	Self-help	16	15	BED	42.4 (10.2)	100

Study	*C*	*IG*	*Tx Type*	*CG*	N_{IG}	N_{CG}	*Subjects*		
							Diagnosis	*Age M (SD)*	*% Female*
Peterson *et al.* 2009	US	CBT	Grp.	Self-help	60	67	BED	47.10 (10.4)	90
Safer *et al.* 2010	US	DBT	Grp.	ST	50	51	BED	52.2 (10.6)	85
Wilson *et al.* 2010	US	IPT	Ind.	BWL	75	64	BED	na	100
		IPT	Ind.	Self-Help	75	66	-	-	-
Bona fide CBT versus bona fide non-CBT									
Agras *et al.* 2000	US	CBT	Ind.	IPT	110	110	BN	28.1 (7.2)	na
Cooper *et al.* 1995	UK	CBT	Ind.	ERP	13	14	BN	23.8 (na)	100
Fairburn *et al.* 1986	UK	CBT	Ind.	STF	12	12	BN	22.9 (4.4)	100
Fairburn *et al.* 1991	UK	CBT	Ind.	BT	25	25	BN	24.2 (na)	100
		CBT	Ind.	IPT	25	25	-	-	-
Fairburn *et al.* 2015	UK	CBT	Ind.	IPT	65	65	BN/BED/EDNOS	25.9 (7.7)	97.7
Freeman *et al.* 1988	UK	CBT	Ind.	BT	32	30	BN	24.2 (5.6)	100
Griffiths *et al.* 1994	AU	CBT	Ind.	Hypno.	23	27	BN	25.91 (5.73)	100
Lavender *et al.* 2012	UK	CBT	Grp.	ESM	37	37	EDNOS/BN	27.7 (7.45)	91
McIntosh *et al.* 2005	NZ	CBT	Ind.	IPT	19	21	AN	na	100
Tasca *et al.* 2006	CA	CBT	Grp.	GPIP	50	50	BED	42.75 (10.76)	91
Wilfley *et al.* 1993	US	CBT	Grp.	IPT	18	18	BED	44.3 (8.3)	100
Wilfley *et al.* 2002	US	CBT	Grp.	IPT	81	81	BED	45.25 (9.61)	82.72
Wolf *et al.* 1992	US	CBT	Grp.	BT	15	15	BN	26 (na)	100

Note: The reference list of included primary studies can be found in the Dissertation Appendix A. *Abbreviations:* C: country; US: United States; UK: United Kingdom; DE: Germany; CH: Switzerland; AU: Australia; CA: Canada; NZ: New Zealand; SE: Sweden; IG: intervention group; CBT: cognitive-behaviour therapy; AAT: appetite awareness training; GPIP: group psychodynamic interpersonal psychotherapy; DBT: dialectical behaviour therapy; BT: behaviour therapy; BIT: body image therapy; CT: cognitive

therapy; IPT: interpersonal psychotherapy; CG: comparison group; Ind.: individual; Grp: group; WL: wait-list; ERP: exposure plus response-prevention; ESM: emotional and social mind training; ND: non-directive; STF: short term focal therapy; SG: support group; ST: supportive therapy; Pharma.: pharmacotherapy; Hypno.: hypnobehavioural Therapy; NT: nutritional therapy; SCM: supportive clinical management; BWL: behavioural weight loss; N_{IG}: sample size of intervention group; N_{CG}: sample size of comparison group; Tx: treatment; *M*: mean; *SD*: standard deviation; na: authors did not report standard deviations for the mean age or percentage of females.

Table 2

Effects of Bona Fide Psychotherapy Compared to Wait-List

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	Q	*p(Q)*	I^2	*ES* range
Post-Tx	Abstinence	All	11(14)		5.64	3.92; 8.12	< .001	5.44	.964	0%	1.00, 16.58
	Frequency	All	13(16)	.81		.58; 1.04	< .001	24.54	.056	38.88%	.05, 2.90
	ED psychopathology	All	15(19)	.54		.39; .70	< .001	13.14	.783	0%	0, 1.10
	Depression	All	13(16)	.47		.28; .66	< .001	19.92	.175	24.68%	0, 1.44
	Self-concept	All	10(13)	.30		.13; .47	.001	8.03	.783	0%	0, .98
	Interpersonal problems	All	4(6)	.24		-.04; .51	.091	1.24	.942	0%	0, .49

Notes: Tx: Treatment; N(*k*): number of studies (number of comparisons); *g*: Hedge's *g*; *RR*: relative risk; *ES* range: effect size range of all studies included in the analysis.

Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures. Depression: valid self-report measures of depressive symptoms. Self-concept: valid self-report measures of self-esteem and/or self-efficacy. Interpersonal problems: Inventory of Interpersonal Problems.

Table 3

Effects of Bona Fide Psychotherapy Compared to Non-Bona Fide Treatment

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
Post-Tx	Abstinence	All	8(10)		1.41	1.10; 1.80	.007	19.65	.020	54.19%	.53, 2.89
		Group	5(5)		1.55	.94; 2.54	.084	10.50	.033	61.90%	.53, 2.89
		Individual	3(5)		1.14	.99; 1.32	.076	2.94	.568	0%	1.08, 1.99
	Frequency	All	10(12)	-.03		-.49; .43	.892	87.07	< .001	87.37%	-3.32, 1.55
		Group	6(6)	.29		-.05; .63	.093	12.51	.028	60.04%	-.50, .61
		Individual	4(6)	-.41		-1.36; .53	.393	66.55	< .001	92.49%	-3.32, 1.55
	ED psychopathology	All	11(13)	.17		.03; .32	.022	15.75	.203	23.83%	-.55, .64
		Group	7(7)	.13		-.08; .34	.210	8.60	.197	30.26%	-.55, .40
		Individual	4(6)	.22		-.01; .45	.061	6.91	.228	27.60%	-.12, .64
	Depression	All	10(12)	.12		-.05; .28	.181	14.92	.186	26.25%	-.42, .54
	Self-concept	All	4(4)	.01		-.24; .27	.912	3.83	.280	21.68%	-.27, .32
≤ 6 Months	Abstinence	All	5(5)		1.48	.91; 2.41	.117	7.14	.129	43.94%	.67, 3.46
	Frequency	All	3(3)	.31		.05; .56	.018	0.58	.749	0%	.17, .43

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
≤ 6 Months	ED Psychopathology	All	3(3)	.24		-.02; .49	.066	0.31	.855	0%	.17, .32
	Depression	All	3(3)	.31		.06; .57	.017	0.57	.754	0%	.28, .61
	Self-concept	All	2(2)	-.05		-.37; .28	.770	.001	.972	0%	-.05, -.04
> 6 Months	Abstinence	All	6(7)		1.18	1.00; 1.39	.045	4.90	.556	0%	.65, 1.39
	Frequency	All	5(5)	.13		-.17; .44	.385	7.01	.136	0%	-.28, .52
	ED Psychopathology	All	8(9)	.11		-.04; .25	.162	3.79	.804	0%	-.24, .28
	Depression	All	6(7)	.11		-.08; .29	.261	2.59	.763	0%	-.18, .30
	Self-concept	All	4(4)	-.02		-.25; .21	.847	0.49	.921	0%	-.23, .06

Notes: Tx: Treatment; N(*k*): number of studies (number of comparisons); *g*: Hedge's *g*; *RR*: relative risk; *ES* range: effect size range of all studies included in the analysis; ED: eating disorder. No data available for interpersonal problems at any time-point.

Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures. Depression: valid self-report measures of depressive symptoms. Self-concept: valid self-report measures of self-esteem and/or self-efficacy. Interpersonal problems: Inventory of Interpersonal Problems.

Table 4

Effects of Bona Fide CBT Compared to Bona Fide Non-CBT

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
Post-Tx	Abstinence	All	6(7)		1.16	.93; 1.45	.183	11.27	.080	46.74	0.63, 2.80
	Frequency	All	4(5)	-.10		-.92; .71	.807	25.19	< .001	84.12%	-2.04, .72
	ED Psychopathology	All	11(12)	.31		.16; .46	< .001	10.68	.470	0%	-.35, .82
		Group	4(4)	.19		-.06; .45	.138	1.88	.589	0%	-.35, .37
		Individual	7(8)	.37		.18; .56	< .001	7.62	.367	8.11%	-.14, .82
		vs. IPT	6(6)	.28		.09; .48	.004	5.84	.322	14.33%	-.35, .68
	Depression	All	7(7)	.08		-.11; .27	.417	6.52	.367	8.02%	-.48, .64
	Self-concept	All	5(5)	.08		-.15; .31	.506	5.47	.242	26.90%	-.43, .37
	Interpersonal problems	All	4(4)	.04		-.20; .27	.768	4.09	.252	26.69%	-.53, .19
≤ 6 Months	Abstinence	All	3(3)		1.10	.81; 1.49	.563	4.68	.096	57.24%	.90, 1.64
	Frequency	All	1(1)	-.96		-1.70; -.21	.012	NA	NA	NA	NA
	ED Psychopathology	All	6(6)	.24		.05; .43	.013	1.29	.936	0%	.47, .12
		Group	3(3)	.17		-.08; .42	.187	0.34	.845	0%	.12, .34

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
≤ 6 Months	ED Psychopathology	Individual	3(3)	.33		.05; .62	.023	0.23	.891	0%	.29, .47
	Depression	All	3(4)	.18		-.05; .42	.128	2.85	.415	0%	-.14, .55
	Self-concept	All	4(4)	-.45		-1.25; .35	.273	40.51	<.001	0	-2.0, .11
	Interpersonal Problems	All	3(3)	.14		-.08; .35	.209	0.30	.860	0%	.02, .17
> 6 Months	Abstinence	All	5(6)		1.23	.88; 1.73	.232	12.24	.032	59.15%	.82, 5.50
		Group	2(2)		1.01	.78; 1.30	.950	1.43	.233	29.85%	.91, 1.19
		Individual	3(4)		1.73	.85; 3.54	.133	7.90	.048	62.04%	.82, 5.50
	Frequency	All	1(1)	-.47		-1.26; .33	.254	NA	NA	NA	NA
	ED Psychopathology	All	5(6)	.16		-.02; .34	.079	2.63	.756	0%	.04, .47
	Depression	All	4(5)	.29		-.06; .63	.106	8.78	.067	54.41%	-.07, 1.11
	Self-concept	All	4(4)	.01		-.21; .22	.947	1.71	.634	0%	-.04, .50
	Interpersonal Problems	All	2(2)	.09		-.15; .32	.474	0.47	.494	0%	0, .17

Notes: CBT: cognitive-behaviour therapy; Tx: treatment; N(*k*): number of studies (number of comparisons); ED: eating disorder; *g*: Hedge's *g*; *RR*: relative risk; BED: binge-eating disorder; BN: bulimia nervosa; *ES* range: effect size range of all studies included in the analysis. Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures. Depression: valid self-report measures of depressive symptoms. Self-concept: valid self-report measures of self-esteem and/or self-efficacy. Interpersonal problems: Inventory of Interpersonal Problems.

Study Two

Group Psychotherapy for Eating Disorders: A Meta-Analysis*

Renee Grenon

University of Ottawa

Dominique Schwartze

Jena University Hospital

Nicole Hammond

University of Ottawa

Iryna Ivanova

Ottawa Hospital Research Institute

Nancy Mcquaid

The Ottawa Hospital

Genevieve Proulx

Giorgio A. Tasca

University of Ottawa and The Ottawa Hospital

*A version of this manuscript is published in the *International Journal of Eating Disorders*
DOI: 10.1002/eat.22744

Abstract

Introduction: In the current meta-analysis, I review the effect of group psychotherapy compared to both wait-list controls and other active treatments for adults with an eating disorder (ED). **Method:** Twenty-seven randomized controlled trials (RCTs) with a total of 1,853 participants were included. **Results**: Group psychotherapy is significantly more effective than wait-list controls at achieving abstinence rates of binge eating and/or purging (RR = 55.51, 95% CI: 3.73, 8.12), decreasing the frequency of binge eating and/or purging (g = 50.70, 95% CI: 0.51, 0.90), and reducing ED-related psychopathology (g = 50.49, 95% CI: 0.32, 0.66) after treatment. The effects of group psychotherapy and other active treatments (e.g., behavioural weight loss, self-help, individual psychotherapy) did not differ on any outcome at post-treatment or at follow-ups. Group cognitive-behaviour therapy (CBT) and other forms of group psychotherapy did not differ significantly on outcomes at any time point. **Discussion:** Additional research is needed to evaluate other group psychotherapy approaches, along with CBT, to provide more evidence-based treatment options for individuals with an ED. Group psychotherapy appears as effective as other common treatments for EDs.

Group Psychotherapy for Eating Disorders: A Meta-Analysis

Evaluating the efficacy of various treatment options for eating disorders (EDs) is important and necessary to inform both treatment guidelines and clinicians' practices. Results from randomized controlled trials (RCTs) are often used to determine what treatments are recommended in practice guidelines (American Psychiatric Association, 2006; American Psychological Association, 2006; NICE, 2004; Yager et al., 2014). The number of RCTs that investigate the efficacy of group psychotherapy for EDs has increased over the years, and results have been promising. One useful way to synthesize the findings from individual RCTs is by meta-analyses, which allow one to estimate reliable average effect sizes from the existing research. However, there is no meta-analysis that I know of that evaluates the overall effects of group psychotherapy for all EDs using only RCTs and direct comparisons. This study is a comprehensive meta-analysis that includes RCTs of direct comparisons of the effect of group psychotherapy compared to wait-list controls and other active treatments for adults with EDs.

Group Psychotherapy

Group psychotherapy has an evidence base indicating its efficacy for a variety of disorders (Burlingame, Strauss, and Joyce, 2013). Group psychotherapy theorists have long held that groups represent a social microcosm in which interpersonal factors that underlie psychological distress and symptoms can be more easily or effectively addressed due to the psychotherapeutic nature inherent to group dynamics (Yalom and Leszcz, 2005). Group psychotherapies that include group therapeutic factors like peer interpersonal feedback, social learning, emotional expression, and group cohesion (Yalom and Leszcz, 2005) may be well situated to target the specific interpersonal problems that maintain psychopathology secondary to an ED for some individuals. Many theories of the maintenance of ED symptoms point to the role

of interpersonal problems and affect dysregulation. The transtheoretical cognitive-behaviour therapy (CBT)-enhanced model (Fairburn, 2008) posits that for some individuals with EDs, interpersonal problems, mood intolerance, low self-esteem, and clinical perfectionism are maintenance factors. Similarly, the interpersonal model of binge eating posits that interpersonal problems lead to binge eating mediated by affect dysregulation or low mood (Wilfley, Frank, Welch, Spurrell, and Rounsaville, 1998). Both CBT-enhanced and interpersonal models for EDs have received empirical support when used to predict ED symptoms in transdiagnostic samples (Tasca et al., 2011; Wilfley et al., 1998). Thus, group therapy may be an efficacious and cost-effective modality to address some underlying maintenance factors in EDs.

Group Psychotherapy for Eating Disorders: Findings of Past Meta-Analyses

The majority of meta-analyses of psychological treatments for EDs combined individual and group psychotherapy trials by indirect comparisons, and they did not isolate the effects of group psychotherapy. Nor did they compare the effect of group psychotherapy to other common treatments used for EDs, such as behavioural weight loss and self-help programs. I found five past meta-analyses that compared the effects of group and individual psychotherapy for EDs, however, the majority of these are out of date, targeted only bulimia nervosa (BN), and reported conflicting results (Erford et al., 2013; Fettes and Peters, 1992; Hartmann, Herzog, and Drinkmann, 1992; Polnay et al., 2014; Thompson-Brenner, Glass, and Westen, 2003). Two past meta-analyses found no difference between group and individual psychotherapy in the treatment of BN (Erford et al., 2013; Hartmann et al., 1992), whereas another found individual psychotherapy to be superior to group psychotherapy (Thompson-Brenner et al., 2003). One explanation for these conflicting results is that they were based on indirect comparisons, meaning the group and individual psychotherapy conditions were conducted in separate trials. Indirect

comparisons cannot be interpreted as randomized comparisons. For example, some trials compare individual psychotherapy and wait-list controls, and other trials compare group psychotherapy and wait-list controls. Using these trials to compare individual and group psychotherapy represents an indirect comparison. The Cochrane Handbook for Systematic Reviews for Interventions (Higgins and Green, 2011) recommends that treatments be compared by direct comparisons (e.g., using only trials that directly compare group and individual psychotherapy). Using indirect comparisons may be unreliable due to study-level confounds among the trials (i.e., participants are randomized within trials but are not randomized to the trials [Higgins and Green, 2011]). The most recent meta-analysis by Polnay and colleagues (2014) focused on direct comparisons of CBT for BN. They reported that group CBT was more effective than no treatment for BN. Results of this meta-analysis are limited due to the small number of primary studies that were included (N = 10). The authors suggested further research be conducted to refine and improve the quality of their findings (Polnay et al., 2014). This study adds to the literature by conducting a meta-analysis of a larger number of direct comparisons that examined the effects of group psychotherapy and other common treatments for adults with EDs.

In previous years, The National Institute of Clinical Excellence (NICE, 2004) and the American Psychiatric Association Workgroup on Eating Disorders (American Psychiatric Association, 2006) recommended individual CBT as the first-line treatment option for EDs. Historically, individual CBT has been the most recommended and researched psychotherapy for EDs. Currently, the American Psychiatric Association Practice Guideline for the Treatment of Patients with Eating Disorders (Yager et al., 2014) and American Psychological Association Task Force on Evidence-Based Practice (American Psychological Association, 2006) both include group and individual CBT and group and individual interpersonal psychotherapy (IPT;

Wilfley et al., 1993, 2002) as the two psychotherapies that have shown strong research support for BN and binge-eating disorder (BED). Although other group psychotherapies have been tested (i.e., group psychodynamic interpersonal psychotherapy [GPIP; Tasca et al., 2006] and dialectical behaviour therapy [DBT; Telch, Agras, and Linehan, 2001]), group IPT currently is the most researched of these three (GPIP, DBT, and IPT) interventions. Other forms of psychotherapy need to be evaluated further, along with CBT and IPT, to provide more evidence-based treatment options for individuals with an ED.

The Current Study

I conducted a meta-analysis of RCTs examining the effects of group psychotherapy for adults with EDs. Identifying how group psychotherapy compares to no treatment (e.g., wait-list controls) and other active treatments, as well as comparing the effects of group CBT to other forms of group psychotherapy will help to inform clinicians' practice and future ED treatment guidelines. I included trials comparing group psychotherapy to wait-list control conditions, and to other active treatments (i.e., individual psychotherapy, behavioural weight loss, self-help, pharmacotherapy, supportive therapy, and nondirective treatment). As CBT is the most recommended and studied treatment for BN and BED (Yager et al., 2014), I also compared the effects of group CBT to other forms of group psychotherapy. Several RCTs have compared different types of group psychotherapy (e.g., group CBT vs. group IPT), but findings from these trials have yet to be synthesized. The goals of this meta-analysis are to: (1) estimate the effect of group psychotherapy for EDs compared to wait-list controls, other active treatments (e.g., behavioural weight loss, self-help, individual psychotherapy, pharmacotherapy, supportive therapy, or nondirective therapy), and when possible, individual psychotherapy alone; (2)

compare the effect of group CBT to other forms of group psychotherapy; and (3) examine the short-term (6 months), and long-term (>6 months) effects of group psychotherapy.

Method

Identification and Selection of Studies

Eligible studies were identified in three ways. First, a systematic literature search was conducted using a number of electronic databases (PsycINFO, MEDLINE, Cochrane Central Register of Controlled Trials, ClinicalTrials.gov) from January, 1980 to December, 2016. The search terms included those indicative of the population combined with intervention terms and study design terms. ClinicalTrials.gov and the Cochrane Registry of Controlled trials were screened to identify unpublished research. A detailed description of search terms can be found in Dissertation Appendix B. Second, the reference lists of previous meta-analyses and systematic reviews of psychological treatments for EDs were manually screened. Third, reference lists of all eligible primary studies were screened. Independent raters carried out the screening and selection of studies, and disagreements were resolved through consensus. No language restrictions were applied. To be included in the meta-analysis, studies had to meet the following criteria: (a) published after 1980; (b) participants were only adults (>17 years of age) diagnosed with an ED based on DSM-III, DSM-IV, or DSM 5 criteria (and revised versions); (c) one of the treatments in the study was a group format; and (d) one of the intervention conditions was deemed a bona fide psychotherapy by meeting at least two of the following criteria (Wampold et al., 1997): (a) delivered by trained therapists, (b) offered as a viable treatment (e.g., based on professional books or manuals), and (c) contained specific treatment components based on theories of change. Further, based on criteria outlined in other reviews, group psychotherapy (e) was defined as treatment in which at least three participants received therapy together at the same time (Huntley,

Araya, and Salisbury, 2012); (f) had to be performed by a professional leader; (g) contained at least five sessions; (h) was compared to a wait-list control condition or another active treatment; and (i) studies used an RCT design. The psychotherapy type was coded as CBT if it consisted primarily of both cognitive and behavioural techniques. If a psychotherapy contained only behavioural or only cognitive components, it was coded as behavioural therapy or cognitive therapy, respectively. The following types of studies were excluded: (a) dismantling studies, where group psychotherapy was compared with the same treatment with one component added or removed; (b) studies of in-patient group psychotherapy in which the effects of group psychotherapy alone were impossible to isolate; and (c) studies that did not provide information necessary to compute effect sizes even after I attempted to contact original authors.

Data Extraction

Descriptive information about each primary study and the participants were extracted, including; study author, year of publication, total sample size, mean age of sample, percentage of sample that were women, and ED diagnosis. Details about the conditions were also extracted, including; theoretical orientation, type of comparison group (wait-list control or active treatment), and the number of therapy sessions. To calculate effect sizes, I extracted data for the following outcome measures at post-treatment, short-term follow-up (6 months), and long-term follow-up (>6 months). Primary outcomes included: (a) abstinence rates: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Both observer and self-reported measures were included. Observer measures involved asking participants about binge-eating and/or purging behaviours. This was done using a calendar recall method, the Rating of Anorexia and Bulimia Interview (RABI; Clinton and Norring, 1999), or with a version of the Eating Disorder Examination (EDE;

Fairburn and Cooper, 1993). Self-report measures involved the same questions but in a written format (i.e., EDE Questionnaire [EDEQ]; Fairburn and Beglin, 1994); standardized food records, or a food diary). (b) Frequency of binge eating/purges: continuous outcome of the number of binge-eating and/or purging episodes in the last 7 or 28 days. This outcome was measured using the same or similar methods as abstinence rates, with the addition of the Eating Disorders Questionnaire (EDQ; Mitchell, Hatsukami, Eckert, and Pyle, 1985). (c) Valid observer or self-report measures of ED related psychopathology. Scales included the EDE (Fairburn and Cooper, 1993), EDEQ (Fairburn and Beglin, 1994), Eating Disorder Inventory (EDI; Garner, Olmstead, and Polivy, 1983), Three Factor Eating Questionnaire (TFEQ; Stunkard and Messick, 1985), Binge Eating Scale (BES; Gormally, Black, Dastin, and Rardin, 1982), Body Shape Questionnaire (BSQ; Cooper, Taylor, Cooper, and Fairburn, 1987), Eating Attitudes Test (EAT; Garner and Garfinkel, 1979), and RABI (Clinton and Norring, 1999). Secondary outcomes included: (a) continuous outcomes of valid depressive symptoms measures. Scales included the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, and Erbaugh, 1961), depression subscale of the Symptom Checklist-90 (SCL-90; Derogatis, 1977), Hamilton Depression Rating Scale (HDRS; Hamilton, 1960), Centre for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977), and the Inventory for Depressive Symptomology-Self-Report (IDS-SR; Rush et al., 1986); (b) continuous outcomes of overall or subscale scores of valid self-esteem, self-efficacy, and self-concept measures. Scales included the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1979), Lawson Social Self-Esteem Scale (LSE; Lawson, Marshall, and McGrath, 1979), Frankfurt Self-Concept Scale (FKSN; Deusinger, 1986), and General Self-Efficacy Scale (SWE; Jerusalem and Schwarzer, 1999); and, (c) continuous outcome of the

overall score of the Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureño, and Villaseñnor, 1988).

Study Quality

Studies included in the meta-analysis were coded for the following risk of bias criteria that were developed by the Cochrane Collaboration (Higgins and Green, 2011): (a) random sequence generation (high risk of bias if there was inadequate randomization of participants); (b) allocation concealment (high risk of bias if there was inadequate concealment of participant allocation prior to assignment); (c) blinding of outcome assessment (high risk of bias if there was knowledge of the allocated interventions by outcome assessors); (d) incomplete outcome data (high risk of bias if there was missing outcome data and only completer analyses performed). I also coded for three additional criteria (Chambless and Hollon, 1998; Munder, Brutsch, Leonhart, Gerger, and Barth, 2013): (e) therapist effects (high risk of bias if therapists of the intervention and comparison groups have different levels of experience, therapists have an affiliation to just one of the treatments, and/ or training was only provided for therapists of one treatment); (f) implementation quality (high risk of bias if no treatment manual was used, adherence was not measured, no therapist training or support); (g) researcher allegiance (high risk of bias if a study author developed or revised/adapted the treatment protocol, hypotheses favoured one of the treatments). When information regarding a specific quality category was not mentioned, the criterion was coded as “unclear,” and when a category was not applicable, the criterion was coded as “not applicable.” As some of the quality criteria were coded for each effect size, when a study had both high risk and low-risk outcomes, the criterion was coded as “mixed.”

Effect Size Calculation

Effect sizes for continuous outcomes were computed using Hedge's *g*, which is a variation of Cohen's *d*, that adjusts effect sizes using a correction. Hedge's *g* results in a more precise estimate of the effect size than Cohen's *d* (Higgins and Green, 2011), especially for studies with small sample sizes. To compute Hedge's *g*, the mean of the comparison condition is subtracted from the mean of the intervention condition, and divided by the pooled standard deviations of both conditions (Hedges and Olkin, 1985). Effect sizes for abstinence from binge eating and/ or purging were computed using relative risk (*RR*). The *RR* is an intuitive way to compare two groups on the probability of an outcome occurring. In this study, an $RR > 1.00$ indicates greater abstinence rates for group psychotherapy. In the analyses comparing group CBT and other group psychotherapies, an $RR > 1.00$ indicates greater abstinence rates for group CBT. An $RR = 3.00$ indicates the probability of abstinence from binge eating and/or purging for group psychotherapy is 3.00 times that of the comparison group. To calculate the relative risk, I divided the risk of event (i.e., abstinence from binge eating and/or purging) in the intervention condition by the risk of the same event in the control condition. Calculating *RR* requires at least one event (one person being abstinent from binge eating and/or purging) in the control condition. When *RR* was zero, I replaced the zero with a one in the cell. This resulted in a more conservative effect size estimate (Higgins and Green, 2011). When a study reported both completer-analyses and intent-to-treat analyses for the same outcome, I used the intent-to-treat data for analyses. When studies provided more than one measure for an outcome (e.g., two scales of ED-related psychopathology), the effect sizes were averaged so that there was only one effect size per outcome, per study. To compute an average effect size of multiple outcomes, a correlation of 1.00 between outcomes was assumed, which underestimates the precision of the

summary effect (Borenstein, Hedges, Higgins, and Rothstein, 2009; Cuijpers, 2016). When studies provided more than two conditions (i.e., group CBT, group IPT, wait-list control) and data for more than one comparison, I split the sample sizes in half to ensure that effect sizes were not inflated due to double counting participants. When means and standard deviations were not reported I calculated effect sizes using other data presented in the studies. Effect size values of 0.20 are considered "small effects," 0.50 are considered "moderate effects," and 0.80 are considered "large effects" (Cohen, 1988). Effect size indices apply to between-group effect sizes only (Barber, Barrett, Gallop, Rynn, and Rickels, 2012).

Meta-Analysis

A random-effects model was used to aggregate effects across studies. This model allowed results to be generalized beyond the studies included in the meta-analysis, assumes the presence of between-study heterogeneity, and yields a more conservative effect size estimate (Cooper, Hedges, and Valentine, 2009). Three separate meta-analyses were conducted that included an estimation of between-groups effect sizes of: (a) group psychotherapy compared to a wait-list control condition; (b) group psychotherapy compared to another active treatment condition (i.e., individual psychotherapy, pharmacotherapy, self-help, supportive/nondirective therapy, behavioural weight loss); and (c) group CBT compared to another group psychotherapy (i.e., group IPT, GPIP). Each meta-analysis was conducted for outcomes at post-treatment, short-term (6 months), and long-term (>6 months) follow-up. An effect size was considered an outlier when the 95% *CI* did not overlap with the 95% *CI* of the pooled effect size (Cuijpers et al., 2014). One study removed sensitivity analyses were conducted to test the stability of the mean effect size with respect to the primary outcomes.

Heterogeneity

To assess heterogeneity of effect sizes, the *Q*-statistic was computed and tested for statistical significance. Higgins' I^2 was used to determine the amount of heterogeneity that exists in effect sizes. The *Q*-statistic examines whether the variability is greater than would be expected by chance, and Higgins' I^2 represents the proportion of the overall variability that is beyond sampling error (the proportion of variability that can be considered "true differences") (Borenstein et al., 2009). I^2 values of 25% indicate "low," 50% indicate "moderate," and 75% indicate "high" levels of heterogeneity across studies (Higgins, Thompson, Deeks, and Altman, 2003). Low I^2 values indicate less "true" heterogeneity and more variability due to sampling error. I also reported the ranges of effect sizes for each analysis.

Moderators of Effect Size

When the heterogeneity across studies was moderate to high, I conducted moderator analyses to attempt to explain the dispersion of effect sizes. To examine the influence of ED diagnosis, I conducted a mixed-effect model in which subgroups were pooled with the random-effects model. To test for differences between subgroups I used a fixed-effects model because there are fewer studies in subgroup analyses, less variability in effect sizes among studies, and lower power. The between-level *Q*-statistic (i.e., the proportion of the overall variability that is explained by the moderator) was used to examine the effect of a categorical moderator variable on treatment outcomes. A significant between-level *Q*-statistic indicates that the categorical moderator variable explains a significant proportion of the variability in effect sizes across studies. If subgroups had less than two studies of comparisons, moderator analyses were not conducted.

To assess the reliability of our findings, I conducted meta-regression analyses to examine whether the power of a study moderated the magnitude of the effect size. I calculated the power of each individual study using G*Power software (Düsseldorf, Germany) (Faul, Erdfelder, Lang, and Buchner, 2007).

Publication Bias

For comparisons that included at least 10 studies, I examined publication bias by observing the funnel plot of the primary outcome measure (Higgins and Green, 2011). When asymmetry of the funnel plot was observed and publication bias was significant as per Egger's regression test of the intercept, I conducted the random-effects version of the Duval and Tweedie's trim and fill procedure (Duval and Tweedie, 2000). By this method, I adjusted effect sizes to correct for the number and location of missing studies, and for the effect that these studies might have on the primary outcome. All effect size calculations and analyses were computed in the Comprehensive Meta-Analysis program (CMA; Version 3 [Borenstein, Hedges, Higgins, and Rothstein, 2005]).

Results

Selection and Inclusion of Studies

The database literature search resulted in 501 hits (418 after removing duplicates). See Figure 1 for a flow chart summarizing the study selection process. Based on the title and/or abstract, 328 studies were excluded. The full texts of 90 studies were examined. Sixty-three studies were excluded because they did not meet inclusion criteria. Twenty-seven RCTs (see Dissertation Appendix B for a complete reference list of included studies) with 36 comparisons (see Table 1 for characteristics of included studies and comparisons) were included. All RCTs are published manuscripts, with the exception of one dissertation (Allen, 1997).

Characteristics of Included Studies

The 27 RCTs included 19 comparisons of group psychotherapy versus a wait-list control condition. The majority of the group psychotherapy conditions were CBT or included primarily CBT principles ($k = 15$). There were 12 comparisons of group psychotherapy and another active treatment. The group psychotherapy conditions in these comparisons were almost all CBT ($k = 9$). Other active treatment comparisons included: behavioural weight loss ($k = 3$), self-help ($k = 3$), supportive nondirective therapy ($k = 2$), individual CBT+IPT ($k = 2$), and individual CBT ($k = 1$). Finally, there were five comparisons of group CBT to another form of group psychotherapy (Emotional and Social Mind Training [ESM], GPIP, and IPT). Binge-eating disorder was the most studied ED diagnosis ($N = 19$), followed by BN ($N = 10$). One study treated participants with ED not otherwise specified (EDNOS; Nevonen & Broberg, 2005). The remaining two studies used a mix of participants with anorexia nervosa (AN), BN, and EDNOS (Vocks et al., 2011), or a mix of participants with BN and EDNOS (Lavender et al., 2012).

Of 27 RCTs, 26 reported the number of participants that dropped out of treatment and/or wait-list. The mean participant dropout rates were comparable for group psychotherapy (16.47%, $SD = 13.46$; range 0–42.11%; $N = 31$), other active treatments (24.49%, $SD = 10.11$; range 7.14–38.89%; $N = 12$), and wait-list control groups (16.63%, $SD = 15.77$; range 0–50%; $N = 15$). Differences in dropout rates between the three conditions were not significant ($F(2,55) = 1.66$, $p = .200$), however, with small sample sizes, power may not have been adequate to detect differences. The analysis did yield a medium effect size ($n^2 = 0.057$) indicating that 5.7% of the variance in dropout rates was due to condition (group psychotherapy, other active treatment, and wait-list control.

The quality ratings of the 27 RCTs and the treatments are shown in Table 1. I did not conduct analyses using the study quality scores, as there were a large number of unclear, mixed, and not applicable ratings. Sufficient information was not reported thus: I could not confidently rate criteria, some comparisons and outcomes yielded different quality ratings than others within the same RCT, and some criteria were not applicable to some RCTs and comparisons. Only 8.33% of comparisons in the current meta-analysis had at least 80% power to detect a moderate effect (d = .50), and so to assess the impact of this I examined power as a potential moderator of effect sizes.

Group Psychotherapy Versus Wait-list Control

Post-Treatment

Group psychotherapy showed statistically significantly higher abstinence rates from binge eating and/or purging compared to wait-list controls (RR = 5.51; 95% CI: 3.73, 8.12, p < .001, k = 13, RR range = 1.00, 16.58; see Figure 2). The probability of being abstinent from binge eating and/or purging for group psychotherapy participants was, on average 5.57 times that of wait-list controls. Heterogeneity was not statistically significant. Egger's test (p = .773) indicated no publication bias (*intercept*: -0.15; 95% CI: -1.24, 0.95). Group psychotherapy had an overall abstinence rate of 51.38% whereas wait-list control conditions had an overall abstinence rate of 6.51%.

Group psychotherapy had a statistically significantly greater effect in reducing binge eating and/or purging frequency compared to wait-list control conditions (g = 0.79; 95% CI: 0.51, 1.08, p < .001, k = 13), with moderate and statistically significant heterogeneity in the effect sizes (I^2 = 48.61%, g range = 0.05, 2.90). After removal of an outlier (Allen, 1997), the effect size reduced to g = 0.70 which was medium-sized (95% CI: 0.51, 0.90, p < .001, k = 12),

and heterogeneity was not statistically significant (I^2 = 3.92%, *g* range = 0.05, 1.25). Egger's test (*p* = .962) indicated no publication bias (*intercept*: -0.05; 95% *CI*: -2.09, 2.00).

Group psychotherapy was statistically significantly more efficacious than wait-list controls in reducing ED-related psychopathology, with a medium effect size (*g* = 0.49; 95% *CI*: 0.32, 0.66, $p < .001$, *k* = 17, *g* range = 0, 1.10). Heterogeneity was not statistically significant. Egger's test (*p* = .030) indicated publication bias (*intercept*: 1.29; 95% *CI*: 0.14, 2.43). After adjusting for publication bias, the effect size was reduced to *g* = 0.40 (95% *CI*: 0.24, 0.55; number of missed studies: 5). One study removed analyses did not change the direction or significance of any findings with respect to abstinence, frequency, or ED psychopathology.

Similar results were obtained for secondary outcomes. Statistically significant effects in favour of group psychotherapy versus control conditions were found for reducing depressive symptoms (*g* = 0.38, 95% *CI*: 0.22, 0.54, $p < .001$, *k* = 16, *g* range = 0, 0.81), and improving self-concept (*g* = 0.30, 95% *CI*: 0.13, 0.48, *p* = .001, *k* = 13, *g* range = 0, 1.44) with small effects (Table 2). Results were not statistically significant for improving interpersonal problems (*g* = 0.26, 95% *CI*: -0.03, 0.55, *p* = .081, *k* = 5, *g* range = 0, 0.49). Heterogeneity was not statistically significant for any of these analyses. No follow-up analyses were conducted for these comparisons as wait-listed participants did not provide data following post-treatment. Results of the meta-regression analyses, to assess the reliability of our findings, indicated that the frequency of binge eating and/or purging (*b* = .861, *SE* = .282, *Z* = 3.05, *p* = .002; *M* power = .519, *SD* = .299, range of *power* = .060, 1.00) and ED psychopathology (*b* = .932, *SE* = .372, *Z* = 2.50, *p* = 0.012; *M power* = .449, *SD* = .261, range of *power* = .050, .940) effect sizes were moderated by the power of the studies. Studies with lower power had smaller effect sizes. The relationship

between abstinence rates and power was not statistically significant (b = 1.154, SE = .691, Z = 1.67, p = .095; M $power$ = .697, SD = .350, range of $power$ = 0, 1.00).

Group Psychotherapy Versus Other Active Treatments

Post-treatment

No statistically significant differences in abstinence rates were found between group psychotherapy and other active treatments (RR = 1.20; 95% CI: 0.71, 2.02, p =.500, k = 8; see Figure 3). Heterogeneity was statistically significant (Q = 29.76, df = 7, p < .001, I^2 = 76%, RR range = 0.17, 2.88). To attempt to explain the variability among effect sizes, I computed effects for BED (RR = 1.22, 95% CI: 0.58, 2.57, p = .606, k = 4; Q = 23.17, df = 3, p < .001, I^2 = 87.05%, RR range = 0.51, 2.88), BN (RR = 1.01, 95% CI: 0.35, 2.94, p = .982, k = 3; Q = 3.90, df = 2, p = .142, I^2 = 48.70%, RR range = 0.17, 1.63), and EDNOS (RR = 2.83, 95% CI: 0.33, 24.67, k = 1) separately. Diagnosis was not a statistically significant moderator of abstinence rates (between-level Q = 1.15, df = 2, p = .562). One likely explanation for the heterogeneity between studies is the different types of comparison groups that are pooled in this analysis (i.e., self-help, individual CBT, behavioural weight-loss, individual CBT+IPT, supportive therapy). Unfortunately, with too few comparisons, a moderator analysis for type of other active treatment was not possible. A one study removed analysis did not change the significance or direction of the finding with respect to abstinence at post-treatment.

No statistically significant difference was found between group psychotherapy and other active treatments in reducing the frequency of binge eating and/or purging (g = 0.24, 95% CI: -0.08, .57, p = .146, k = 8). There was statistically significant heterogeneity among effect sizes (Q = 18.74, df = 7, p = .009, I^2 = 62.65%, g range = -0.50, .83). I computed effect sizes separately for BED (g = 0.49, 95% CI: 0.26, 0.72, p < .001, k = 4, g range = 0.28, 0.61) and BN (g = -0.04,

95% *CI*: -0.56, 0.48, $p = .875$, $k = 4$, g range = -0.50, 0.83). Heterogeneity was not statistically significant for both analyses. Diagnosis was a statistically significant moderator (between-level $Q = 10.87$, $df = 1$, $p = .001$). Compared to other active treatments, group psychotherapy had a greater effect on reducing binge eating for individuals diagnosed with BED than for those with BN. Two studies had a pronounced effect on the results of this analysis. Removal of Chen and colleagues (2003) and Bailer and colleagues (2004) studies led to a statistically significant difference between group psychotherapy and other active treatments (in favour of group psychotherapy) in reducing the frequency of binge eating and/or purging ($g = 0.33$, 95% *CI*: 0.01, 0.65, $p = .045$, $k = 8$) and ($g = 0.36$, 95% *CI*: 0.09, 0.63, $p = .008$, $k = 8$), respectively.

There was no statistically significant difference between group psychotherapy and other active treatments in reducing ED-related psychopathology ($g = 0.12$; 95% *CI*: -0.03, 0.26, $p = .107$, $k = 11$), depressive symptoms ($g = 0.12$; 95% *CI*: -0.05, 0.30, $p = .160$, $k = 11$), or improving self-concept ($g = 0.05$; 95% *CI*: -0.14, 0.24, $p = .616$, $k = 6$). Heterogeneity was not statistically significant in any of these analyses. A one study removed analysis indicated that removal of the Bailer and colleagues (2004) study lead to a statistically significant difference between group psychotherapy and other active treatments (in favour of group psychotherapy) in reducing ED-related psychopathology ($g = 0.17$; 95% *CI*: 0.21, 0.32, $p = .025$, $k = 11$).

Three studies directly compared group psychotherapy and individual psychotherapy (group CBT vs. individual CBT (Chen et al., 2003); and, group CBT+IPT vs. individual CBT+IPT (Nevonen and Broberg, 2005, 2006). With the available data, I was able to compare the effects of group psychotherapy and individual psychotherapy at posttreatment for abstinence, ED psychopathology, and depressive symptoms. No statistically significant differences were found between group psychotherapy and individual psychotherapy for abstinence rates (*RR* =

0.96; 95% *CI*: 0.26, 3.48, $p = .947$, $k = 3$), ED psychopathology ($g = -0.06$; 95% *CI*: -0.35, 0.23, $p = .672$, $k = 3$), or depressive symptoms ($g = -0.09$; 95% *CI*: -0.38, 0.20, $p = .541$, $k = 3$). Heterogeneity was not statistically significant in any of these analyses. Due to limited data, follow-up analyses could not be conducted.

Results of the meta-regression analyses indicated that abstinence rates ($b = 1.02$, $SE = .531$, $Z = 1.92$, $p = .055$; *M power* = .456, *SD* = .337, range of *power* = .091, .986), frequency of binge eating and/or purging ($b = .576$, $SE = .518$, $Z = 1.11$, $p = .267$; *M* power = .472, *SD* = .282, range of *power* = .118, .923), and ED psychopathology ($b = -.019$, $SE = .423$, $Z = -.04$, $p = .965$; *M power* = .176, *SD* = .177, range of *power* = .055, .517) effect sizes were not statistically significantly moderated by the power of the studies. The meta-regression for abstinence was approaching statistical significance.

Follow-Up

At short-term (6 months) and long-term (> 6 months) follow-up there were no statistically significant differences in the effects of group psychotherapy compared to other active treatments for any primary or secondary outcomes. Similar to the post-treatment results, at the short-term (6 months) follow-up, group psychotherapy significantly reduced binge eating frequency for individuals with BED ($g = 0.31$; 95% *CI*: 0.05, 0.56, $p = .018$, $k = 3$) when compared to other active treatments. Group psychotherapy and other active treatments showed similar effects in reducing binge eating and/or purging frequency for individuals with BN ($g = -0.33$; 95% *CI*: -0.84, 0.18, $p = .199$, $k = 1$). At short-term (6 months) follow-up, removal of the Chen and colleagues (2003) study lead to statistically significant differences between group psychotherapy and other active treatments (in favour of group psychotherapy) in reducing frequency of binges and/or purges ($g = 0.31$; 95% *CI*: 0.05, 0.56, $p = .018$, $k = 4$). At long- term

(> 6 months) follow-up, a one study removed analysis did not change the significance or direction of the finding with respect to any primary outcomes.

Group CBT versus Other Group Psychotherapies

Post-Treatment

Five studies directly compared group CBT with another form of group psychotherapy (ESM, GPIP, and IPT; Table 4). The following results should be interpreted with caution, as there were limited comparisons available for each of the outcomes. There was no statistically significant difference in the effect of group CBT compared to other group psychotherapies for abstinence rates (RR = 1.08; 95% CI: 0.95, 1.24, p = .435, k = 3), frequency of binge eating and/or purging (g = -0.30; 95% CI: -1.23, 0.64, p = .534, k = 1), or ED-related psychopathology (g = 0.19; 95% CI: -0.06, 0.45, p = .138, k = 4). Similarly, results for depressive symptoms (g = -0.07; 95% CI: -0.35, 0.21, p = .607, k = 3), self-concept (g = 0.13; 95% CI: -0.25, 0.51, p = .513, k = 3), and interpersonal problems (g = -0.04; 95% CI: -0.37, 0.29, p = .805, k = 3) were not statistically significant. All effect sizes were small. Heterogeneity was not statistically significant in each of these analyses.

Follow-Up

At short-term (< 6 months) follow-up there were no statistically significant differences in the effects of group CBT compared to other forms of group psychotherapy in abstinence rates (RR = 0.95; 95% CI: 0.77, 1.16, p = .615, k = 2), reducing ED psychopathology (g = 0.21; 95% CI: -0.04, 0.46, p = .095, k = 3), reducing depressive symptoms (g = 0.93; 95% CI: -0.27, 0.46 , p = .617, k = 2), or improving self-concept (g = -1.00; 95% CI: -2.92, 0.92, p = .308, k = 2). Again, these findings are based on few comparisons and definite conclusions cannot be drawn. One

study removed analyses did not change the significance or direction of results with respect to any primary outcome at post-treatment or follow-up.

Discussion

When compared to wait-list control conditions, group psychotherapy yielded superior results at post-treatment. Participants who underwent group psychotherapy were, on average 5.51 times more likely abstain from binge eating and/or purging than wait-list controls. Group psychotherapy had a medium to large effect at post-treatment on decreasing the frequency of binge eating and/or purging (g = 0.70) compared to wait-list controls. That is, the average participants who underwent group psychotherapy was estimated to have a greater decrease in binge eating and/or purging at post-treatment than did 76% of participants in the wait-control condition. Additionally, group psychotherapy exhibited small to medium effects at post-treatment on decreasing ED-related psychopathology (g = 0.49), improving depressive symptoms (g = 0.38), and improving self-concept (g = 0.30) compared to wait-list controls. These findings were expected as most studies that compare a bona fide therapy to a wait-list control condition demonstrate that the psychological treatment is more effective (Wampold et al., 1997). These findings are also consistent with those of Polnay and colleagues who found that when compared to no treatment, group CBT yielded superior results at posttreatment (Polnay et al., 2014). The effect sizes in our study are similar in range to what has been found for psychotherapy versus wait-list control conditions for other mental disorders (Lambert, 2013). The only outcome for which group psychotherapy was not statistically significantly superior at post-treatment was improvement of interpersonal problems (g = 0.26). Only five comparisons out of 19 included interpersonal problems as an outcome and as a result the analysis was low in statistical power.

Group psychotherapy had similar effects on outcomes when compared to other active treatments at post-treatment, short-term follow-up (≤ 6 months), and long-term follow-up (> 6 months). It is important to note that the effect sizes for abstinence rates varied significantly across the studies in the analysis. I speculated that the type of comparison treatment was a likely moderator of these effects, however, due to limited data, I was unable to conduct this moderator analysis. I examined each comparison's individual effect size, along with other study characteristics (how abstinence was measured, type of group psychotherapy, number of sessions, etc.) to try and identify what may be contributing to the heterogeneity. No one variable stood out as a possible explanation for the heterogeneity, and too few comparisons preclude any conclusions.

I found only three RCTs that directly compared the effects of group and individual psychotherapy (Chen et al., 2003; Nevonen and Broberg, 2005, 2006). At post-treatment, group and individual psychotherapy had similar effects for abstinence from binge eating and/ or purging, improving ED-related psychopathology, and decreasing depressive symptoms. These results should be interpreted with caution. I hope that in the future, more RCTs will directly compare the effects of group psychotherapy and other active treatments, including individual psychotherapy, and researchers will be able to reliably compare the different treatment modalities and conduct the appropriate moderator analyses. Such studies should include a cost-effectiveness component.

There was significant heterogeneity among effect sizes for frequency of binge eating and/or purging when comparing group psychotherapy to other active interventions at post-treatment. Eating disorder diagnosis was a significant moderator, such that group psychotherapy had a statistically significantly greater effect on reducing binge eating frequency for individuals

diagnosed with BED but not for those with BN. One explanation for this may be the type of active treatments used as comparisons. Behavioural weight loss and self-help were the comparisons for BED, whereas individual CBT, self-help, nondirective group therapy, and pharmacotherapy were the comparisons for BN. The active treatments compared to group psychotherapy in the BN studies may be more effective in reducing symptoms than those used in the BED studies. The range in effect sizes for BN was wide (g = -.50, .83), and heterogeneity was approaching significance. However, with only one comparison per treatment type, no subgroup analyses could be conducted. A second explanation for this finding may be that interpersonal problems (i.e., social isolation, weight related stigma, social skill deficits) could be a more salient issue for the maintenance of BED symptoms than for BN. If this were true, individuals with BED would benefit from group psychotherapeutic factors such as interpersonal learning, cohesion, and peer feedback (Yalom and Leszcz, 2005) more than individuals with BN. This is speculative but might be a possible explanation for the differential outcomes.

Finally, five studies directly compared group CBT to another form of bona fide non-CBT group psychotherapy. I found that effect sizes at posttreatment, short-term follow-up (≤ 6 months), and long-term follow-up (> 6 months) were comparable across group treatment types. Although treatment guidelines recommend CBT and IPT as first-line treatment options (American Psychiatric Association, 2006; NICE, 2004; Yager et al., 2014), the current meta-analysis suggests that there are no differences in outcomes between group CBT and some bona fide non-CBT group psychotherapies for EDs. These results are consistent with findings in the general psychotherapy literature that show few or no differences between bona fide psychotherapies for disorders like depression (see Wampold and Imel, 2015 for a review). I speculate that factors like group support, group cohesion, safe emotional expression,

interpersonal learning and feedback, and the provision of a clearly articulated rationale for symptoms and treatments may be also contributing to the efficacy of bona fide group psychotherapies (Yalom and Leszcz, 2005). Group process research in EDs is a limited but emerging area of inquiry that may shed light on the effective elements of group treatment (Tasca, Compare, Zarbo, and Brugnera, 2016). Other group psychotherapy approaches, along with CBT and IPT, should be investigated further to allow for more evidence-based treatment options for EDs. Having a wider array of treatment options, especially for those who do not respond to CBT or IPT, may help improve the overall outcomes for individuals with an ED.

Benefits of Group Psychotherapy for Eating Disorders

The findings indicate that group psychotherapy may be a suitable treatment option for individuals with EDs. Group psychotherapy may be more cost-effective than individual psychotherapy (McCrone et al., 2005; Otto, Pollack, and Maki, 2008; Roberge, Marchand, Reinharz, and Savard, 2008), and treatment acceptance and dropout rates of group psychotherapy are equivalent to individual psychotherapy (Burlingame et al., 2016). Research evaluating and comparing the costs of group psychotherapy compared to other forms of treatment for mental disorders is limited, but shows some promise. In two separate studies, researchers have found that group psychotherapy for panic disorder was more cost-effective than individual psychotherapy or pharmacotherapy (Otto et al., 2008; Roberge et al., 2008). In a study of children who were sexually abused, McCrone and colleagues found that group and individual therapy were equally effective, but the cost of individual psychotherapy was significantly more than for group psychotherapy (McCrone et al., 2005). Despite this limited research, practitioners have noted the cost-effectiveness of group psychotherapy for years (American Group Psychotherapy Association, 2010).

The rates of treatment acceptance and dropout among studies that directly compared group and individual psychotherapy for various mental illnesses have been examined in a recent meta-analysis (Burlingame et al., 2016). Findings indicated no significant difference in rates of treatment acceptance between group (M = 84.83%, SD = 26.40) and individual (M = 88.76%, SD = 23.18) psychotherapy. Treatment dropout rates were also equivalent between group (M = 17.28%, SD = 14.30) and individual (M = 14.96%, SD = 14.29) psychotherapy (Burlingame et al., 2016). In the three RCTs in the current meta-analysis that compared group and individual psychotherapy for EDs (Chen et al., 2003; Nevonen and Broberg, 2005, 2006), similar dropout rates were reported for both modalities. Group psychotherapy for EDs may be particularly apt from a clinical standpoint since group psychotherapy processes address the additional maintenance factors (interpersonal problems, mood intolerance, low self-esteem, and clinical perfectionism) (Fairburn, 2008) that may underlie ED symptoms for some individuals. These maintenance factors can be effectively addressed through group psychotherapy that includes group therapeutic factors like group cohesion, peer interpersonal feedback, safe emotional expression, and social skill building (Yalom and Leszcz, 2005).

Limitations and Future Directions

With the limited number of comparisons available, many analyses, especially moderator analyses, were not possible or lacked sufficient power. Only one study (Chen et al., 2003) directly compared the same kind of group and individual psychotherapy, meaning no analyses could be done to compare the effects of the same treatment delivered in different formats. Only 5 studies out of 27 included interpersonal problems as an outcome. Given that interpersonal problems may be a key factor in developing and maintaining ED symptoms for some individuals (Ansell, Grilo, and White, 2012; Fairburn, 2008; Ivanova, Tasca, Proulx, and Bissada, 2015), I

suggest that future researchers include this as an outcome variable. Group psychotherapy investigators should also measure group processes including working alliance, group climate, group cohesion, and group therapeutic factors. Specific measures to assess these group processes can be found in the CORE Battery-Revised by the American Group Psychotherapy Association (Burlingame et al., 2006). The number of studies with moderators of interest (e.g., number of sessions, type of active treatment comparison) also was too small to conduct statistical analyses. As more studies become available, the number of comparisons needed for these analyses will increase, the power of the analyses will be improved, and the effects of important moderators can be assessed.

Initially, one of our goals was to examine whether study quality had an effect on outcome. When examining the effects of psychotherapy for depression, Cuijpers, van Straten, Bohlmeijer, Hollon, and Andersson (2010) found that higher-quality studies were statistically significantly associated with lower effect sizes. However, due to the large number of quality criteria that I rated as unclear, not applicable, or mixed in the primary studies, I was unable to evaluate the effects of study quality. Future studies examining group psychotherapy for EDs should report greater details of study quality (e.g., how participants were randomized to treatment groups, whether the assignment was concealed), treatment quality (e.g., use of treatment manuals, training and qualifications of therapists), and data analysis quality (e.g., whether the effects of hierarchically nested data for accounted for) in more detail.

After one-study-removed analyses, the meta-analytic findings were generally reliable and stable, with the exception of two results. Removal of Chen and colleagues (2003) and/or Bailer and colleagues (2004) studies resulted in group psychotherapy showing significantly greater effects in reducing the frequency of binge eating and/or purging at post-treatment compared to

other active treatments. Similarly, removal of Bailer and colleagues study resulted in a significantly greater reduction of ED psychopathology at post-treatment for group psychotherapy compared to other active treatments.

Due to small sample sizes, only 3 out of 36 of comparisons had at least 80% power to detect a moderate effect (d = .50). Underpowered comparisons lead to less reliable effect size estimates, and so results should be interpreted with caution. There is debate in the literature about whether to include or exclude studies with small sample sizes from meta-analyses. Kraemer, Gardner, Brooks, and Yesavage (1998) argue that the only small trials that are published are those with significant treatment effects and this leads to overestimated effect sizes in a meta-analysis (Kraemer et al., 1998). Other researchers support the inclusion of all trials in meta-analyses, regardless of sample size (Cuijpers, 2016; Sackett and Cook, 1993; Schulz and Grimes, 2005). In a study examining the bias in meta-analytic estimates of the absolute efficacy of psychotherapy, Staines and Cleland (2007) concluded that assigning less weight to studies with small sample sizes greatly reduces the degree of overestimation (Staines and Cleland, 2007). To assess the reliability of our meta-analytic findings, I examined whether the power of a study moderated the magnitude of the effect size for primary outcomes for group psychotherapy versus waitlist comparisons. Results show that studies with lower power had smaller effect sizes. This indicates that effect sizes may in fact be underestimated. Given that most psychotherapy trials have small sample sizes I felt it would result in a more accurate estimate of the current state of treatment effects to include all available RCTs. I recommend that future researchers conduct studies with the minimum sample size necessary to detect effects with at least 80% power.

There were no outpatient trials of group psychotherapy for individuals diagnosed with AN. Inpatient care for AN typically includes a variety of treatments making the effects of group

psychotherapy alone impossible to isolate. There is a great need for more RCTs on treatments for AN. However, high treatment refusal rates, low recruitment numbers, and a high dropout rate of approximately 40% represent a severe challenge in completing such trials (Halmi, 2008).

Our data came from published studies (with the exception of one dissertation), and publication bias was identified in one analysis. Publishing results of all trials, regardless of outcomes, would improve the results of future meta-analyses by lowering the effects of publication bias. Ideally, all psychotherapy trials should be adhering to the same safeguards put in place in the biomedical field to ward off untrustworthy findings (Coyne, 2016). That is, psychotherapy trials should be registered, the protocol published before the study begins, and results of the trial should be reported in peer-reviewed journals and following CONSORT (CONsolidated Standards of Reporting Trials [Schulz, Altman, and Moher, 2010]) standards. Lastly, agreements between coders during study selection were not recorded and so inter-rater reliability was not reported for the search process.

Conclusion

Overall, results of this meta-analysis indicate that group psychotherapy is effective in the treatment of EDs. Group psychotherapy yielded similar effects to other treatments such as individual psychotherapy, self-help, behavioural weight-loss, and pharmacotherapy. When compared to other active treatments, group psychotherapy had a significantly greater effect on reducing binge eating for individuals diagnosed with BED, whereas group psychotherapy and other active treatments yielded similar effects on reducing binge eating and/or purging for individuals diagnosed with BN.

Group and individual CBT and IPT are the current first-line treatments recommended for individuals with an ED (American Psychological Association, 2006; Yager et al., 2014). Our

results showed that as a whole group CBT and other bona fide group treatments are equally effective for BN and BED. The number of studies of group interventions of EDs has grown steadily over the years, but the overall numbers of studies remain modest. I encourage researchers to continue investigating these treatments and improve and report study quality indicators so that more definite conclusions can be drawn, and treatment guidelines updated accordingly.

References

Agras, W. S., Telch, C. F., Arnow, B., Eldredge, K., Wilfley, D. E., Raeburn, S. D., … Marnell, M. (1994). Weight loss, cognitive-behavioral, and desipramine treatments in binge eating disorder: An additive design. *Behavior Therapy, 25*, 225–238.

Agras, W. S., Telch, C. F., Arnow, B., Eldredge, K., Detzer, M. J., Henderson, J., & Marnell, M. (1995). Does interpersonal therapy help patients with binge eating disorder who fail to respond to cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology, 63,* 356–360.

Allen, H. N. (1997). Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder (doctoral dissertation). Chapel Hill, NC: University of North Carolina.

American Group Psychotherapy Association. (2010, April 26). Practitioners tout cost-effectiveness of group psychotherapy [Mental Health Weekly]. Retrieved from: https://www.agpa.org/docs/default-source/media/practitioners-tout-cost-effectiveness-of-group-therapy.pdf?sfvrsn=478385a9_0

American Psychiatric Association. (2006). Practice guideline for the treatment of patients with eating disorders, 3rd edition. *American Journal of Psychiatry, 163*, 1–54.

American Psychological Association. (2006). Task force on evidence-based practice. Evidence-based practice in psychology. *American Psychologist, 61,* 271–285.

Ansell, E., Grilo, C. M., & White, M. A. (2012). Examining the interpersonal model of binge eating and loss of control over eating in women. *International Journal of Eating Disorders, 1,* 43–50.

Bailer, U., de Zwaan, M., Leisch, F., Strnad, A., Lennkh-Wolfsberg, C., El- Giamal, N., . . . Kasper, S. (2004). Guided self-help versus cognitive behavioral group therapy in the treatment of bulimia nervosa. *International Journal of Eating Disorders, 35*, 522–537.

Barber, J. P., Barrett, M. S., Gallop, R., Rynn, M. A., & Rickels, K. (2012). Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: A randomized, placebo-controlled trial. *Journal of Clinical Psychiatry, 73*, 66–73.

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

Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. (2005). Comprehensive Meta-Analysis, version 3 [Computer Software]. Englewood, NJ: Biostat,

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* Chichester, UK: Wiley.

Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., et al. (2016). Outcome differences between individual and group formats when identical and non-identical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy, 53,* 446–461.

Burlingame, G., Strauss, B., & Joyce, A. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Ed.), *Bergin and Garfields handbook of psychotherapy and behavior change* (6th ed., pp. 640–689) New York: Wiley.

Burlingame, G. M., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K. R., Ogrodniczuk, J., & Taylor, S. M. (2006). *CORE battery-revised: An assessment tool kit for promoting optimal group selection, process and outcome.* New York, NY: American Group Psychotherapy Association.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Chen, E., Touyz, S. W., Beumont, P. J., Fairburn, C. G., Griffiths, R., Butow, P., … Basten, C. (2003). Comparison of group and individual cognitive-behavioral therapy for patients with bulimia nervosa. *International Journal of Eating Disorders, 33,* 241–254.

Clinton, D., & Norring, C. (1999). The Rating of Anorexia and Bulimia (RAB) interview: Development and preliminary validation. *European Eating Disorder Review, 7,* 362–371.

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

Cooper, H., Hedges, L., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.) Russell Sage Foundation, New York, NY.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the body shape questionnaire. *International Journal of Eating Disorder, 6*, 485–494.

Coyne, J. (2016). Stalking a Cheshire cat: Figuring out what happened in a psychotherapy intervention trial (Internet). (cited January 13, 2016). Retrieved from http://blogs.plos.org/mindthebrain/2016/01/13/stalking-a-cheshire- cat-figuring-out-whathappened-in-a-psychotherapy-intervention-trial/

Cuijpers, P. (2016). Meta-analyses in mental health research. A practical guide. Retrieved from: https://indd.adobe.com/view/5fc8f9a0-bf1e-49d3-bf5f-a40bfe5409e0

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: A meta-analysis. *Clinical Psychology Review, 34,* 130–140.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychology Medicine, 40,* 211–223.

Derogatis, L. R. (1977). *SCL-90, administration, scoring and procedures manual for the revised version.* Baltimore: John Hopkins School of Medicine.

Deusinger, I. M. (1986). *Die Frankfurter Selbstkonzeptskalen (FSKN).* Handanweisung. Gottingen: Hogrefe.

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.

Erford, B. T., Richards, T., Peacock, E., Voith, K., McGair, H., Muller, B., . . . Chang, C. Y. (2013). Counseling and guided self-help outcomes for clients with bulimia nervosa: A meta-analysis of clinical trials from 1980 to 2010. *Journal of Counseling & Development, 91,* 152–172.

Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders.* New York, NY: 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.

Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination. 12th ed. In C. G.

Fairburn, C.G., & Wilson, G. T. (1993). *Binge eating: Nature assessment, and treatment* (pp. 317–360). New York, NY: Guilford Press.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioural Research Methods, 39,* 175–191.

Fettes, P. A., & Peters, J. M. A. (1992). Meta-analysis of group treatments for bulimia nervosa. *International Journal of Eating Disorders, 11,* 97–110.

Garner, D. M., & Garfinkel, P. E. (1979). The eating attitudes test: An index of symptoms of anorexia nervosa. *Psychological Medicine, 9,* 273–279.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2,* 15–34.

Gorin, A. A., Le Grange, D., & Stone, A. A. (2003). Effectiveness of spouse involvement in cognitive behavioral therapy for binge eating disorder. *International Journal of Eating Disorders, 33,* 421–433.

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

Grilo, C. M., Masheb, R. M., Wilson, G. T., Gueorguieva, R., & White, M. A. (2011). Cognitive-behavioral therapy, behavioral weight loss, and sequential treatment for obese patients with binge-eating disorder: a randomized controlled trial. *Journal of Consulting and Clinical Psychology, 79,* 675–685.

Halmi, K. A. (2008). The perplexities of conducting randomized, double blind, placebo-controlled treatment trials in anorexia nervosa patients. *American Journal of Psychiatry, 165,* 1227–1228.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological Neurosurgery Psychiatry, 23,* 56–62.

Hartmann, A., Herzog, T., & Drinkmann, A. (1992). Psychotherapy of bulimia nervosa: What is effective? A meta-analysis. *Journal of Psychosomatic Research, 36,* 159–167.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0.* John Wiley & Sons: The Cochrane Collaboration.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of interpersonal problems: Psychometirc properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56,* 885–892.

Huntley, A. L., Araya, R., & Salisbury, C. (2012). Group psychological therapies for depression in the community: Systematic review and meta-analysis. *British Journal of Psychiatry, 200,* 184–190.

Ivanova, I. V., Tasca, G. A., Proulx, G., & Bissada, H. (2015). Does the interpersonal model apply across eating disorder diagnostic groups? A structural equation modeling approach. *Comprehensive Psychiatry, 63,* 80–87.

Jacobi, C., Dahme, B., & Dittmann, R. (2002). Cognitive-behavioural, fluoxetine and combined treatment for bulimia nervosa: Short- and long-term results. *European Eating Disorders Review, 10,* 179–198.

Jerusalem, M., & Schwarzer, R. (1999). Skala zur allgemeinen selbstwirksamkeit- serwartung (SWE). In R. Schwarzer & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schulermerkmalen. Dokumentation der psychometrischen Ver- fahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbst- wirksame Schulen.* Berlin, Germany: Freie Universita't Berlin.

Kirkley, B. G., Schneider, J. A., Agras, W., & Bachman, J. A. (1985). Comparison of two group treatments for bulimia. *Journal of Consulting and Clinical Psychology, 53*, 43–48.

Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods, 3,* 23–31.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York, NY: Wiley.

Laessle, R. G., Waadt, S. S., & Pirke, K. M. (1987). A structured behaviorally oriented group treatment for bulimia nervosa. *Psychotherapy and Psychosomatics, 48,* 141–145.

Lavender, A., Startup, H., Naumann, U., Samarawickrema, N., DeJong, H., Kenyon, M., . . . van Schmidt, U. (2012). Emotional and social mind training: A randomised controlled trial of a new group-based treatment for bulimia nervosa. *PLoS One, 7*, 1–9.

Lawson, J. S., Marshall, W. L., & McGrath, P. (1979). The social self-esteem inventory. *Educational Psychology Measurement, 39,* 803–811.

Lee, N. F., & Rush, A. J. (1986). Cognitive-behavioural group therapy for bulimia. *International Journal of Eating Disorders, 5,* 599–615.

Leitenberg, H., Rosen, J. C., Gross, J., Nudelman, S., & Vara, L. S. (1988). Exposure plus response-prevention treatment of bulimia nervosa. *Journal of Consulting and Clinical Psychology, 56,* 535–541.

McCrone, P. R., Weeramanthri, T., Knapp, M. R. J., Rushton, A., Trowell, J., Miles, G., et al. (2005). Cost-effectiveness for individual versus group psychotherapy for sexually abused girls. *Child and Adolescent Mental Health, 10,* 26–31.

Mitchell, J. E., Hatsukami, D., Eckert, E. D., & Pyle, R. L. (1985). Eating disorders questionnaire. *Psychopharmacological Bulletin, 21,* 1025–1043.

Munder, T., Brutsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychological Review, 33,* 501–511.

Munsch, S., Biedert, E., Meyer, A., Michael, T., Schlup, B., Tuch, A., & Margraf, J. (2007). A randomized comparison of cognitive behavioral therapy and behavioral weight loss treatment for overweight individuals with binge eating disorder. *International Journal of Eating Disorders, 40,* 102–113.

Nevonen, L., & Broberg, A. G. (2005). A comparison of sequenced individual and group psychotherapy for eating disorder not otherwise specified. *European Eating Disorder Review, 13,* 29–37.

Nevonen, L., & Broberg, A. G. (2006). A comparison of sequenced individual and group psychotherapy for patients with bulimia nervosa. *International Journal of Eating Disorders, 39,* 117–127.

NICE. (2004). *Eating disorders—core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders.* London: NICE.

Otto, M. W., Pollack, M. H., & Maki, K. M. (2008). Empirically supported treatments for panic disorder: Costs, benefits, and stepped care. *Journal of Consulting and Clinical Psychology, 68,* 556–563.

Peterson, C. B., Mitchell, J. E., Crow, S. J., Crosby, R. D., & Wonderlich, S. A. (2009). The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. *The American Journal of Psychiatry, 166,* 1347–1354.

Peterson, C. B., Mitchell, J. E., Engbloom, S., Nugent, S., Mussell, M. P., & Miller, J. P. (1998). Group cognitive-behavioral treatment of binge-eating disorder: A comparison of therapist-led versus self-help formats. *International Journal of Eating Disorders, 24,* 125–136.

Polnay, A., James, V. A., Hodges, L., Murray, G. D., Munro, C., & Lawrie, S. M. (2014). Group therapy for people with bulimia nervosa: Systematic review and meta-analysis. *Psychological Medicine, 44,* 2241–2245.

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

Roberge, P., Marchand, A., Reinharz, D., & Savard, P. (2008). Cognitive behavioral treatment for panic disorder with agoraphobia: A randomized, controlled trial and cost-effectiveness analysis. *Behaviour Modification, 32*, 333–351.

Rosenberg, M. (1979). *Conceiving the self.* New York, NY: Basic Books.

Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Research, 18,* 65–87.

Sackett, D. L., & Cook, D. J. (1993). Can we learn anything from small trials? *Annals of the New York Academy of Sciences, 703,* 25–32.

Safer, D. L., Robinson, A. H., & Jo, B. (2010). Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. *Behavior Therapy, 41*, 106–120.

Schlup, B., Munsch, S., Meyer, A. H., Margraf, J., & Wilhelm, F. H. (2009). The efficacy of a short version of a cognitive-behavioral treatment followed by booster sessions for binge eating disorder. *Behaviour Research and Therapy, 47,* 628–635.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340,* c332.

Schulz, K. F., & Grimes, D. A. (2005). Sample size calculations in randomized trials: Mandatory and mystical. *Lancet, 365,* 1348–1353.

Staines, G. L., & Cleland, C. M. (2007). Bias in meta-analytic estimates of the absolute efficacy of psychotherapy. *Review of General Psychology, 11*, 329-347.

Stunkard, A. J., & Messick, S. (1985). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Tasca, G. A., Compare, A., Zarbo, C., & Brugnera, A. (2016). Therapeutic alliance and binge-eating outcomes in a group therapy context. *Journal of Counseling Psychology, 63*, 443–451.

Tasca, G. A., Presniak, M. D., Demidenko, N., Balfour, L., Krysanski, V., Trinneer, A., & Bissada, H. (2011). Testing a maintenance model for eating disorders in a sample seeking treatment at a tertiary care center: A structural equation modeling approach. *Comprehensive Psychiatry, 52,* 678–687.

Tasca, G. A., Ritchie, K., Conrad, G., Balfour, L., Gayton, J., Lybanon, V., & Bissada, H. (2006). Attachment scales predict outcome in a randomized controlled trial of two group therapies for binge eating disorder: An aptitude by treatment interaction. *Psychotherapy Research, 16,* 106–121.

Telch, C. F., Agras, W. S., Rossiter, E. M., Wilfley, D., & Kenardy, J. (1990). Group cognitive-behavioral treatment for the non-purging bulimic: An initial evaluation*.* *Journal of Consulting and Clinical Psychology, 58,* 629–635.

Telch, C. F., Agras, W. S., & Linehan, M. M. (2001). Dialectical behavior therapy for binge eating disorder. Journal of Consulting and Clinical Psychology, 69, 1061–1065.

Thompson-Brenner, H., Glass, S., & Westen, D. (2003). A multidimensional meta-analysis of psychotherapy for bulimia nervosa. *Clinical Psychology-Science and Practice, 10,* 269–287.

Vocks, S., Schulte, D., Busch, M., Grönemeyer, D., Herpertz, S., & Suchan, B. (2011). Changes in neuronal correlates of body image processing by means of cognitive-behavioural body image therapy for eating disorders: A randomized controlled fMRI study. *Psychology Medicine, 41,* 1651–1663.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially, ‘all must have prizes. *Psychological Bulletin, 122*, 203–215.

Wilfley, D. E., Agras, W. S., Telch, C. F., Rossiter, E. M., Schneider, J. A., Cole Golomb, A., . . . Raeburn, S. D. (1993). Group cognitive behavioral therapy and group interpersonal psychotherapy for the non-purging bulimic individual: A Controlled Comparison. *Journal of Consulting Clinical Psychology, 61,* 296–305.

Wilfley, D. E., Frank, M. A., Welch, R., Spurrell, E. B., & Rounsaville, B. J. (1998). Adaptive interpersonal psychotherapy to a group format (IPTG) for binge eating disorder: Toward a model for adaptive empirically supported treatments. *Psychotherapy Research, 8,* 379–391.

Wilfley, D. E., Welch, R. R., Stein, R. I., Spurrell, E. B., Cohen, L. R., Saelens, B. E., . . . Matt, G. E. (2002). A randomized comparison of group cognitive behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. *Archives of General Psychiatry, 59,* 713–721.

Wolf, E. M., & Crowther, J. H. (1992). An evaluation of behavioral and cognitive-behavioral group interventions for the treatment of bulimia nervosa in women. *International Journal of Eating Disorders, 11,* 3–15.

Yager, J., Devlin, M. J., Halmi, K. A., Herzog, D. B., Mitchell, J. E., Powers, P., & Zerbe, K. F. (2014). Guideline watch (August 2012): Practice guideline for the treatment of patients with eating disorders (3rd ed.) *Focus*, *12,* 416–431.

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy*. New York, NY: Basic Books

Table 1

Characteristics of Included Studies and Comparisons

Study	C	IG	*N* sessions	CG	N_{IG}	N_{CG}	Study Quality[a]	Tx Quality[b]	*Subjects* Diagnosis	Age *M (SD)*	% Female
Group psychotherapy vs. wait-list											
Agras, 1995	US	CBT	12	WL	39	11	u u na -	na u na	BED	47.6 (10.10)	86
Allen, 1997	US	CBT (AAT)	8	WL	15	14	+ u na -	na u na	BED	21 (na)	100
Gorin, 2003	US	CBT	12	WL	32	31	+ u u +	na + na	BED	45.2 (10.03)	100
Laessle, 1987	DE	CBT	24	WL	8	9	u u na +	na u na	BN	23.39 (2.5)	100
Lee, 1986	US	CBT	12	WL	15	15	u u u -	na u na	BN	27.7 (5.3)	100
Leitenberg, 1988	US	1) CBT 2) CBT+ERP	24	WL	1) 12 2) 15	17	u u na - u u na -	na + na na + na	BN	26 (6.04)	100
Peterson, 1998	US	CBT	14	WL	16	11	u u u +	na u na	BED	42.4 (10.2)	100
Peterson, 2009	US	CBT	15	WL	60	69	u + + ?	na + na	BED	47.10 (10.4)	90
Schlup, 2009	CH	CBT	8	WL	18	18	u u na +	na + na	BED	44.3 (10.3)	100
Tasca, 2006	CA	1) CBT 2) GPIP	16	WL	1) 50 2) 50	46	u u + - u u + -	na + na na + na	BED	42.75 (10.76)	91
Telch,, 1990	US	CBT	10	WL	23	21	+ u na ?	na u na	BED	42.5 (8.4)	100
Telch, 2001	US	DBT	20	WL	22	22	u u - -	na u na	BED	50 (9.1)	100
Wilfley, 1993	US	1) CBT 2) IPT	16	WL	1) 18 2) 18	20	u u u + u u u +	na + na na + na	BED	44.3 (8.3)	100
Wolf, 1992	US	1) CBT 2) BT	10	WL	1) 15 2) 15	12	u u u - u u u -	na u na na u na	BN	26 (na)	100
Vocks, 2011	DE	CBT (BIT)	10	WL	32	30	+ u na -	na u na	AN, BN, EDNOS	28.22 (6.73)	100

Study	C	IG	*N* sessions	CG	N_{IG}	N_{CG}	Study Quality[a]	Tx Quality[b]	Subjects		
									Diagnosis	Age *M (SD)*	% Female
Group psychotherapy vs. other active treatment											
Agras, 1994	US	CBT	12	BWL	72	37	u u + -	+ u -	BED	45 (10)	100
Bailer, 2004	AT	CBT	18	Self-help	41	40	u u na ?	- - u	BN	23.76 (4.5)	---
Chen, 2003	AU	CBT	19	Ind. CBT	35	36	+ u u +	+ + -	BN	25.8 (7.24)	100
Grilo, 2011	US	CBT	16	BWL	45	45	+ u u +	+ + u	BED	44.8 (9.4)	67
Jacobi, 2002	DE	CBT	20	Pharma.	19	16	u u u +	na + -	BN	26 (5.8)	100
Kirkley, 1985	US	CBT	16	ND Group	14	14	- u na -	+ u u	BN	28.3 (na)	100
Munsch, 2007	CH	CBT	16	BWL	44	36	u u - ?	- + +	BED	46.1 (11.65)	89
Nevonen, 2005	SE	CBT+IPT	23	Ind. CBT+IPT	18	17	u u ? +	u u +	EDNOS	20.5 (na)	100
Nevonen, 2006	SE	CBT+IPT	23	Ind. CBT+IPT	44	42	u u ? +	u ++	BN	20.5 (na)	100
Peterson, 1998	US	CBT	14	Self-Help	16	15	u u u -	na u +	BED	42.4 (10.2)	100
Peterson, 2009	US	CBT	15	Self-Help	60	67	u + + ?	na + +	BED	47.10 (10.4)	90
Safer, 2010	US	DBT	20	ST	50	51	u u u +	+ + na	BED	52.2 (10.6)	85
Group CBT vs. other group psychotherapy											
Lavender, 2012	UK	CBT	12	ESM	37	37	+ u + +	+ + -	EDNOS, BN	27.7 (7.45)	91
Tasca, 2006	CA	CBT	16	GPIP	50	50	u u + -	+ + -	BED	42.75 (10.76)	91
Wilfley, 1993	US	CBT	16	IPT	18	18	u u u +	- + -	BED	44.3 (8.3)	100
Wilfley, 2002	US	CBT	20	IPT	81	81	u u - ?	+ + -	BED	44.3 (8.3)	82,72
Wolf, 1992	US	CBT	10	BT	15	15	u u u +	+ u u	BN	26 (na)	100

[a)] (+): study quality criteria has been adequately met (low-risk); (-): study quality criteria has not been met (high-risk); na: criteria not applicable; u: unclear whether criteria was met; ?: mixed (some outcomes were high-risk and some low-risk). The order of study quality criteria in this column is as follows: selection bias (biased allocation to interventions due to inadequate generation of a randomised sequence); selection bias (biased allocation to interventions due to inadequate concealment of allocations prior to assignment); performance bias (blinding of participants and personnel); attrition bias (incomplete outcome data).

[b] (+): treatment quality criteria has been adequately met (low-risk); (-): treatment quality criteria has not been met (high-risk); na: criteria not applicable; u: unclear whether criteria was met; ?: mixed (some outcomes were high-risk and some low-risk). The order of treatment quality criteria in this column is as follows: therapist effects; implementation quality; researcher allegiance.

Abbreviations: *C*: country; US: United States; DE: Germany; CH: Switzerland; CA: Canada; AT: Austria; AU: Australia; SE: Sweden; UK: United Kingdom; IG: intervention group; CG: comparison group; N_{IG}: sample size of intervention group; N_{CG}: sample size of comparison group; WL: wait-list; BWL: behavioural weight loss; Ind.: individual; ND: non-directive; Pharma.: pharmacotherapy; ST: supportive therapy; CBT: cognitive-behaviour therapy; BT: behaviour therapy; CT: cognitive therapy; BWL: behavioural weight loss therapy; GPIP: group psychodynamic interpersonal psychotherapy; ESM: emotional and social mind Training, ERP: exposure plus response-prevention; AAT: appetite awareness training; BIT: body image therapy; Tx: treatment; *M*: mean; *SD*: standard deviation; na: authors did not report standard deviations for the mean age.

Table 2

Effects of Group Psychotherapy for Eating Disorders Compared to Wait-List

Time	Outcome	Disorder	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p(Q)*	I^2	*ES* range
Post-treatment	Abstinence	Combined	10(13)		5.51	3.73; 8.12	< .001	5.08	.957	0%	1.00, 16.58
		BED	9(11)		5.82	3.98; 8.52	< .001	3.58	.964	0%	2.58, 16.58
		BN	1(2)		2.53	.50; 12.71	.259	.75	.388	0%	1.00, 4.36
	Frequency	Combined[a]	11(13)	.79		.51; 1.08	< .001	23.35	.025	48.61%	.05, 2.90
		BED[a]	7(7)	.92		.46; 1.37	< .001	22.21	.001	72.99%	.05, 2.90
		BN	4(6)	.66		.28; 1.03	.001	.77	.979	0%	.34, .94
	Frequency	Combined[b]	10(12)	.70		.51; .90	< .001	11.45	.407	3.92%	.05, 1.25
		BED[b]	6(6)	.73		.38; 1.07	< .001	10.59	.060	52.79%	.05, 1.25
	ED psychopathology	Combined	13(17)	.49		.32; .66	< .001	9.90	.872	0%	0, 1.10
		BED	9(11)	.46		.27; .64	< .001	6.70	.754	0%	0, 1.10
		BN	3(5)	.73		.29; 1.18	.001	1.74	.783	0%	.19, 1.00
		Mixed	1(1)	.33		-.35; 1.02	.338	-	-	-	-

Notes: [a]: outlier included in analysis; [b]: outlier excluded from analysis; ED: eating disorder; N(*k*): number of studies (number of comparisons); *g*: Hedge's *g*; *RR*: relative risk; BED: binge-eating disorder; BN: bulimia nervosa; *ES* range: effect size range of all studies included in the analysis.

Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures.

Table 3

Effects of Psychotherapy for Eating Disorders Compared to Other Active Treatments

Time	Outcome	Tx Type	N(*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
Post-treatment	Abstinence	Combined	8(8)		1.20	.71; 2.02	.500	29.76	< .001	76.00%	.17, 2.88
		BED	4(4)		1.22	.58; 2.57	.606	23.17	< .001	87.05	.51, 2.88
		BN	3(3)		1.01	.35; 2.94	.982	3.90	.142	48.70%	.17, 1.63
		EDNOS	1(1)		2.83	.33; 24.67	.346	-	-	-	-
	Frequency	Combined	8(8)	.24		-.08; .57	.146	18.74	.009	62.65%	-.50, .83
		BED	4(4)	.49		.26; .72	< .001	1.04	.791	0%	.28, .61
		BN	4(4)	-.04		-.56; .48	.875	6.83	.077	56.09%	-.50, .83
	ED Psychopathology	Combined	11(11)	.12		-.03; .26	.107	9.40	.500	0%	-.55, .40
		BED	6(6)	.23		.05; .40	.014	1.27	.938	0%	.09, .40
		BN	4(4)	-.10		-.39; .20	.534	4.08	.253	26.54%	-.55, .21
		EDNOS	1(1)	.07		-.58; .73	.828	-	-	-	-

Notes: N(*k*): number of studies (number of comparisons); *g*: Hedge's *g*; *RR:* relative risk; *ES* range: effect size range of all studies included in the analysis.
Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures.

Table 4

Effects of Group Cognitive-Behavioural Psychotherapy for Eating Disorders Compared to Other Group Treatments

Time	Outcome	Disorder	N (*k*)	*g*	*RR*	95% *CI*	*p*	*Q*	*p*(*Q*)	I^2	*ES* range
Post-treatment	Abstinence	BED	3(3)		1.08	.95; 1.24	.255	1.67	.435	0%	.63, 1.14
	Frequency	BN	1(1)	-.30		-1.23; .64	.534	-	-	-	-
	ED Psychopathology	Combined	4(4)	.19		-.06; .45	.138	1.88	.598	0%	-.35, .37
		BED	3(3)	.18		-.09; .45	.185	1.74	.419	0%	-.35, .37
		BN	1(1)	.37		-.57; 1.30	.444	-	-	-	-
≤ 6 Months	Abstinence	BED	2(2)		.95	.77; 1.16	.615	.45	.505	0%	.90, 1.04
	Frequency	BN	1(1)	-.96		-1.70; -.21	.012	-	-	-	-
	ED Psychopathology	Combined	3(3)	.21		-.04; .46	.095	.14	.933	0%	.19, .34
		BED	2(2)	.19		-.07; .46	.152	0	.955	0%	.19, .21
		BN	1(1)	.34		-.37; 1.04	.352	-	-	-	-
> 6 Months	Abstinence	BED	2(2)		1.01	.78; 1.30	.950	1.43	.233	29.85	.91, 1.19
	ED Psychopathology	BED	1(1)	.05		-.28; .37	.782	-	-	-	-

Notes: N(*k*): number of studies (number of comparisons); ED: eating disorder; *g*: Hedge's *g*; *RR:* relative risk; BED: binge-eating disorder; BN: bulimia nervosa; *ES* range: effect size range of all studies included in the analysis.

Abstinence: dichotomous outcome of the number of participants in each condition with cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Frequency: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. ED psychopathology: continuous outcome of subscale scores of valid eating disorder psychopathology measures.

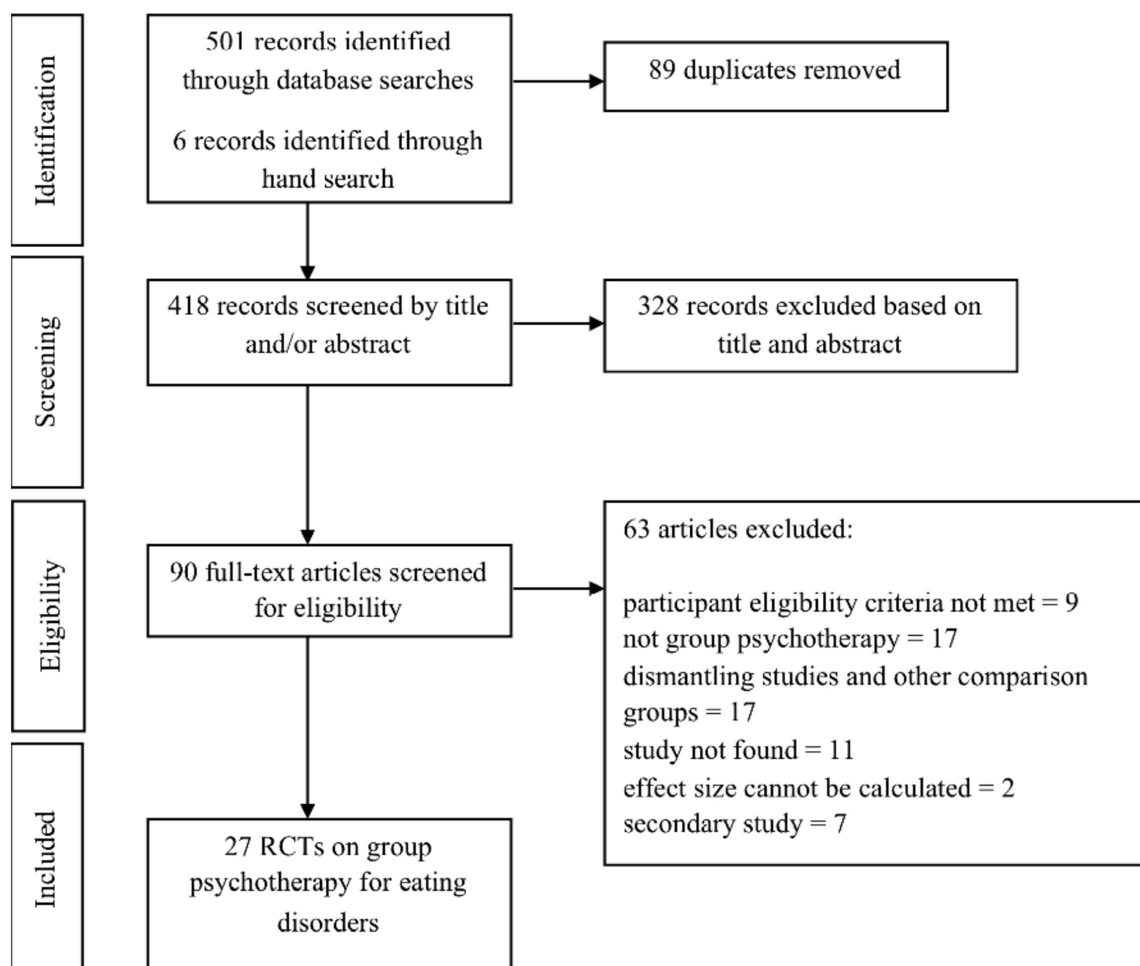


Figure 1. Flow chart of the study selection process of group psychotherapy for the treatment of eating disorders. *Note.* RCT: randomized controlled trial.

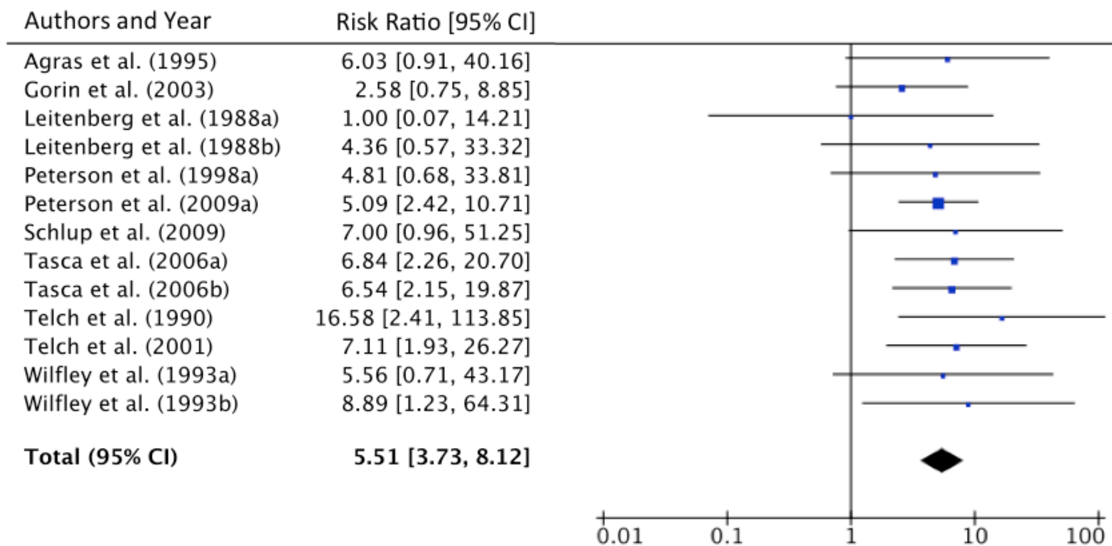


Figure 2. Forest plot of between-group effect sizes of abstinence from bingeing and/or purging of group psychotherapy as compared to wait-list controls at post-treatment. *Notes.* The squares represent the effect sizes for each study, the size of the square the relative weighting of the study in the analysis. Error bars indicate a 95% confidence interval. CI: confidence interval.

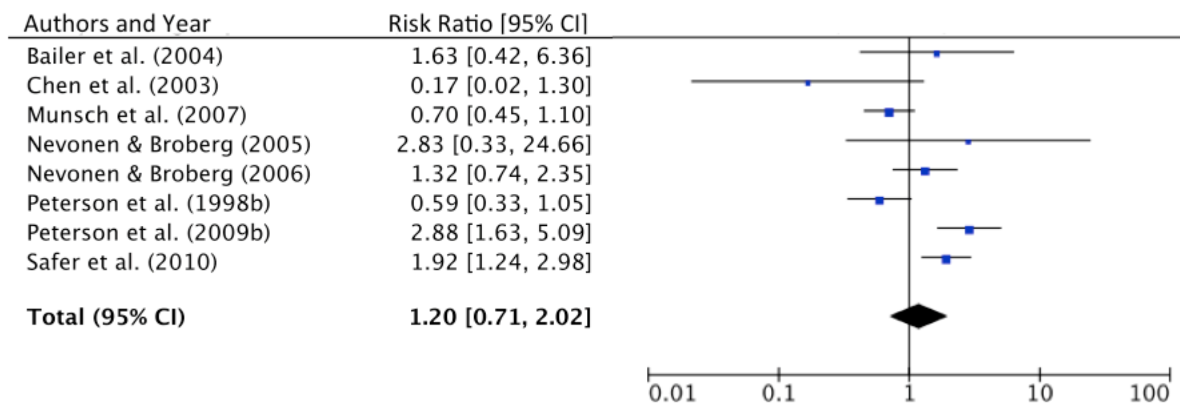


Figure 3. Forest plot of between-group effect sizes of abstinence from bingeing and/or purging of group psychotherapy as compared to other active treatments at post-treatment. *Note.* The squares represent the effect sizes for each study, the size of the square the relative weighting of the study in the analysis. Error bars indicate a 95% confidence interval. CI: confidence interval.

Study Three

The Quality of Randomized Controlled Trials of Psychotherapy for Eating Disorders*

Renee Grenon
Alena McKenna
Hilary Maxwell
Samantha Carlucci
University of Ottawa
Agostino Brugnera
University of Bergamo
Dominique Schwartze
Jena University Hospital
Nicole Hammond
Iryna Ivanova
Ottawa Hospital Research Institute
Nancy Mcquaid
The Ottawa Hospital
Genevieve Proulx
University of Ottawa and The Ottawa Hospital
Giorgio A. Tasca
University of Ottawa and Ottawa Hospital Research Institute

*A version of this manuscript is published in the *International Journal of Methods in Psychiatric Research*
DOI: 10.1002/mpr.1734

Abstract

Introduction: I investigated the quality of randomized controlled trials (RCTs) that included direct comparisons (treatments have been compared head to head in a primary study) of psychotherapy for adults with an eating disorder (ED). **Method:** Thirty-five direct comparison RCTs of psychotherapy for adults diagnosed with an ED were rated using the Randomized Controlled Trials Psychotherapy Quality Rating Scale (RCT-PQRS). **Results:** The mean total RCT-PQRS score (M = 28.26; SD = 7.04) was in line with those that were reported for RCTs of psychotherapy for depression and anxiety disorders. Several standards of quality were unfulfilled by over half of the RCTs of treatment for EDs, including: therapist supervision while treatment was being provided (62.9% unfulfilled); outcome assessment performed by raters blind to treatment group/condition (54% unfulfilled), and; adequate sample size (66% unfulfilled). More recent RCTs were of higher quality, and higher quality was moderately associated with lower effect sizes. **Discussion:** To improve the quality of RCTs of psychotherapy of EDs, I recommend that researchers address the quality criteria listed in the RCT-PQRS. Psychotherapy trials should be registered, have a published protocol, and be reported following the Consolidated Standards of Reporting Trials (CONSORT) guidelines. Authors should consider the quality of the research when using that research to inform ED treatment guidelines.

The Quality of Randomized Controlled Trials of Psychotherapy for Eating Disorders

Eating disorders (EDs) are described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), as a disturbance in eating or eating-related behaviour that significantly impairs an individual's functioning (American Psychiatric Association, 2013). Researchers commonly use randomized controlled trials (RCTs) to evaluate the effects of interventions for EDs (National Collaborating Centre for Mental Health [NICE], 2004; American Psychiatric Association [APA] Workgroup on Eating Disorders, 2006; Yager et al., 2014). Treatment guidelines often rely on the results of RCTs to inform clinicians' interventions for EDs. Despite the influence that RCTs have in the development of treatment guidelines, researchers and clinicians often do not take into account the methodological quality of RCTs when interpreting treatment effects (Moher et al., 1998). This is important because research indicates that study quality may be negatively related to the magnitude of psychotherapy effect sizes for other mental disorders (Cuijpers et al., 2010).

Researchers argue that randomized controlled trial (RCT) designs yield valid results of treatment efficacy because they control for unmeasured confounds (Higgins and Green, 2011). Despite this, RCTs of psychotherapy are still susceptible to methodological flaws that threaten internal validity such as not analyzing the intent-to-treat sample, using outcome measures that are not validated, and having less than adequate sample sizes that reduce statistical power. These issues, along with others, can lead to misleading findings because the magnitude of the treatment effects can be affected by poor study quality (Jüni, Altman, and Egger, 2001).

Researchers have discussed how to objectively define and rate study quality. Specific measures and guidelines such as the Cochrane Risk of Bias tool (Higgins et al., 2011) and the Consolidated Standards of Reporting Trials (CONSORT; Schulz et al., 2010) have been

developed in an attempt to increase the quality of RCTs and their reporting. The development of checklists and reporting guidelines may have had an impact on improving the quality of RCTs over time. For example, Watson et al. (2017) found that the quality of 96 RCTs of prevention programs for EDs was significantly and positively associated with publication year. However, the Cochrane tool and the CONSORT guidelines do not address quality issues specific to psychotherapy trials, such as length of follow-up, the training or supervision of therapists, adherence to treatment manuals, or consideration of therapist and/or site effects.

In 2004, a subcommittee of the American Psychiatric Association Committee on Research developed a rating scale to assess the methodological and reporting quality of psychotherapy RCTs. This subcommittee established the Randomized Controlled Trials of Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010) as a new standard for the design and execution of psychotherapy RCTs. Since it's development, researchers have used the RCT-PQRS used to evaluate the quality of RCTs of psychotherapy for depression (Thoma et al., 2012) and anxiety (Keefe, McCarthy, Dinger, and Barber, 2014). No such evaluation has been conducted for RCTs of psychotherapy for EDs.

Study Quality and Effect Size

Meta-analysis has emerged an important method for aggregating effect sizes and to evaluate the efficacy of treatments (Egger, Smith, and Phillips, 1997). Past meta-analyses found that the quality of RCTs is related to the magnitude of reported effect sizes. For example, Cuijpers and colleagues (2010) conducted a meta-analysis examining the efficacy of psychotherapy for depression while taking study quality into account. Results indicated that higher quality studies had smaller effect sizes (when compared to lower quality studies), and as a

result, the overall effect sizes for psychotherapy for depression may be overestimated (Cuijpers et al., 2010).

Two recent meta-analyses (Grenon, Schwartze et al., 2017; Grenon, Carlucci et al., 2018) of RCTs demonstrated that psychotherapy for EDs is effective. Despite this, it is possible that the magnitude of the effects of psychotherapies for EDs may be biased by the methodological quality of the included RCTs (Moher et al., 1998; Schultz, Altman, and Moher, 2010; Higgins et al., 2011). Because results from RCTs drive treatment guidelines for EDs (NICE, 2004; Yager et al., 2014), I believe that it is essential to examine whether the quality of an RCT is related to the magnitude of its effect sizes.

Study Quality in Different Contexts and Groups

Several factors and contexts might be associated with study quality to affect the precision with which one can interpret effect sizes. First, psychotherapy for EDs may be delivered in a group and individual format, and both modalities are now included in some treatment guidelines and recommendations for EDs (Yager et al., 2014). Previous research demonstrated that the study quality of group and individual RCTs for both depression (Cuijpers et al., 2010) and generalized anxiety (Cuijpers, Sijbrandij, Koole, Huibers, Berking and Andersson, 2014) are equivalent, however, these studies did not control for year of publication. Recall that publication year has been found to be positively associated with higher study quality (Watson et al., 2017; Kocsis et al., 2010), and that higher study quality has been found to be related to lower effect sizes (Cuijpers et al., 2010).

Second, treatment guidelines and recommendations are specific to ED diagnoses such that the treatment recommendations for BN may be different than those for BED, for example. A recent meta-analysis suggested that the average effect size of group treatment versus other active

treatments for BED is higher than for BN (Grenon et al., 2017). Bulimia nervosa was introduced as an ED diagnosis in the DSM-III (American Psychiatric Association, 1980), whereas BED was introduced as research criteria for EDs in the DSM-IV (American Psychiatric Association, 1994). As a result, the average year of publication for BN trials (1992) is much earlier than the average year of BED trials (2002; see Table 1). Treatment for anorexia nervosa (AN) typically includes a variety of interventions making the effects of group or individual psychotherapy alone impossible to isolate. Further, there are few RCTs of psychotherapy for AN.

Third, some might argue that RCTs published in journals with higher impact factors (IFs) might be of better quality than RCTs published in journals with lower IFs. The IF was established in the 1960s by Eugene Garfield as a way to assess the importance or prominence of a journal by a calculation that reflects the average number of citations of studies recently published (Garfield, 2006). Some authors express skepticism about how the IF is calculated and whether it actually represents the quality of studies published in a journal (Misteli, 2013; Seglen, 1997; Baum, 2011). The IFs of journals have steadily increased over time independent of their initial IF (Althouse, West, Bergstrom, and Bergstrom, 2008). Some have attributed the inflation of IFs to some important changes in how research is cited. With the growth of research in many fields, there has been an increase in the number of references in each manuscript. Easy access to literature via the internet has led to an increased number of citations, which has increased the value of citations for individual authors. These changes have also increased the tendency of authors to cite work found within the journal they hope to be published in (Althouse, West, Bergstrom, and Bergstrom, 2008). Alternatively, one could argue that the increase of a journal's IF could be due to the increasing quality of studies published over time in that journal which drives more citations.

The Current Study

The overall goal of the current study is to comprehensively and systematically evaluate the quality of RCTs of psychotherapy for EDs. Additionally, I explore whether the quality of RCTs of psychotherapy for EDs is comparable to the quality of RCTs for other mental illnesses (Keefe et al., 2014; Thoma et al., 2012) while controlling for publication year. Having a better understanding of the quality of RCTs in the ED field is necessary in order to indicate how confident clinicians, researchers, and writers of treatment guidelines can be in the findings.

Due to the more recent development of quality checklists and journal editors requiring that authors adhere to publishing standards, I hypothesize that the quality of RCTs of psychotherapy for EDs has significantly improved over time. Based on the findings from Cuijpers et al. (2010), I hypothesize that study quality will be significantly and negatively related to effect size, with higher quality trials reporting smaller effect sizes than lower quality trials. In order to help interpret the effects of psychotherapy more precisely, I explore whether there is a significant difference in study quality between group and individual psychotherapy RCTs, and between BED and BN trials, while controlling for publication year. I hypothesize that publication year will be positively and significantly related to journal IF. Finally, with a mediation model, I explore whether the improvement in study quality over the years explains the positive relationship between publication year and journal IF. To my knowledge, this is the first study to assess these associations.

Method

Search Strategy

A systematic literature search of electronic databases (PsycINFO, MEDLINE, Cochrane Central Register of Controlled Trials, ClinicalTrials.gov) was conducted separately for group and

individual psychotherapy RCTs for adults with an ED. Full methodology is provided in two previous meta-analyses (Grenon, Schwartze et al., 2017; Grenon, Carlucci et al., 2018).

Inclusion and Exclusion Criteria

The following inclusion and exclusion criteria were specified *a priori*. To be included, studies had to: (a) be published after 1980; (b) have only adult participants (> 17 years of age) diagnosed with an ED based on DSM-III, DSM-IV, or DSM 5 criteria (and revised versions); (c) include at least one psychotherapy condition that was deemed a bona fide therapy (see Wampold *et al.* 1997 for a complete definition of bona fide psychotherapy); (d) compare a bona fide psychotherapy to another bona fide treatment, a wait-list control condition, or a non-bona fide treatment; and (e) use an RCT design. The following types of studies were excluded: (a) dismantling studies; (b) studies of in-patient psychotherapy; (c) studies that did not provide information necessary to compute effect sizes; (d) studies with treatment confounds; and (e) studies that include participants with subthreshold diagnoses.

Study Quality

Quality was assessed using the Randomized Controlled Trials of Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010), a 25-item standardized scale that is divided into 6 domains: (1) description of participants; (2) definition and delivery of treatment; (3) outcome measures; (4) data analysis; (5) treatment assignment; and (6) and overall study quality. The scale yields a 24-item total score and a one-item omnibus rating scale.

Items 1 through 24 were scored as 0, 1, or 2. A score of 0 was given for poor execution and description of the item. A score of 1 was given if the item was moderately described and executed, poorly described but well-executed, or well-described but poorly executed. A score of 2 was given for well-described and executed items. The total quality score is the sum of items 1

through 24, and ranges from 0 to 48. Gerber et al. (2011) proposed that a score of 24 or higher indicated an "average" quality study. Item 25 of the RCT-PQRS is an omnibus quality rating from 1 (exceptionally poor) to 7 (exceptionally good) of the overall quality of the study (Gerber et al., 2011). See Dissertation Appendix C for a full copy of the RCT-PQRS. Previous research reported that the scale demonstrated good internal consistency, with a Cronbach's alpha of 0.87, and a correlation of 0.88 between the total quality score (items 1-24 summed) and the omnibus rating (item 25; Kocsis et al., 2010). The scale also demonstrated good external validity, with correlations of 0.51 and 0.47 between year of publication and total quality score and year of publication and the omnibus rating, respectively (Kocsis et al., 2010). Criterion validity was demonstrated by RCT-PQRS ratings of the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Elkin, Parloff, Hadley, and Autry, 1989) and the Treatment of Adolescents with Depression Study (March et al., 2004). These two trials are widely regarded as high quality and received scores of 40 and 38, respectively, on the RCT-PQRS (Kocsis et al., 2010).

The first author (R.G.) established reliability of her ratings with those of the developers of the RCT-PQRS by coding 20 RCTs selected at random from the Gerber et al. (2011) study. Interrater reliability (calculated as intra-class correlation coefficient [ICC]; Shrout and Fleiss, 1979) was excellent for both the total score ($ICC = .94$) and the omnibus rating ($ICC = .89$). The first author (R.G.) then rated all RCT's in the current study. The third author (H.M.) rated 20 randomly selected studies (57%) to assess reliability specific to the current study's ratings. Inter rater reliability was excellent for both the total score ($ICC = .98$) and the omnibus rating ($ICC = .93$). In the current study, Cronbach's alpha was .82 for the total score, and the correlation between total quality score and the omnibus rating was $r = .97$. Given the high positive

correlation between the total quality score and the omnibus score, I chose to only report analyses using the total quality score. The RCT-PQRS scores from other studies to which I compared our ratings were also reliably rated as reported in those studies (Thoma et al., 2012; Keefe, McCarthy, Dinger, and Barber, 2014).

Effect Sizes

Between group (psychotherapy vs. wait-list) effect sizes for continuous primary outcomes at post-treatment were extracted from a previously published meta-analysis (Grenon, Carlucci et al., 2018). The effect sizes for frequency of binge eating and/or purging and ED psychopathology were computed using Hedge's *g* (standardized mean difference). All effect size calculations and analyses were computed in the Comprehensive Meta-Analysis program (CMA; Version 3; Borenstein *et al.* 2005). For full details on effect size calculations see Grenon, Carlucci et al. (2018).

Impact Factors

The IFs of journals that published the RCTs included in the current study were obtained through the Journal Citation Reports database (Journal Citation Reports Social Sciences Edition, 2016). Two IFs were obtained for each RCT: one IF for the journal in the year the study was published and one IF for the journal two years after the study was published. The correlation between the two IF values was $r(29) = .98$ ($p < .001$). Thus, I chose to report only analyses using the IF for the journal two years after the study was published because it represents the impact of the journal for two years during which the study in question was included in the calculation. The Journal Citation Reports database includes IFs dating from 1997 to present. Impact Factors of studies that were published before 1997 were obtained via the Social Science Citation Index (SSCI) Journal Citation Reports (Garfield, 1985; Garfield, 1986; Garfield, 1987; Garfield, 1988;

Garfield, 1990; Garfield, 1991; Garfield, 1992; Garfield, 1993; Garfield, 1994; Garfield, 1995; Garfield, 1996).

One RCT (Allen, 1997) is a dissertation and was not included in any IF analysis. The IF for two years after the date of publication could not be calculated for the Fairburn et al. (2015) study as the 2017 IF for the journal was not available. Thus, Fairburn et al. (2015) was excluded from IF analyses. I was also unable to obtain the IF two years after publication for the Telch et al. (1990) study, however, I did find the 1991 IF for the journal, and substituted this value in the analysis.

Data Analysis

To examine whether publication year and study quality were positively and significantly correlated, I conducted a regression between publication year and the total RCT-PQRS score. I conducted two separate Pearson correlations to examine whether study quality was associated with the magnitude of effect sizes for two separate primary outcomes obtained from Grenon, Carlucci et al. (2018): (1) frequency of binge eating and/or purging; and (2) ED psychopathology. To compare study quality between group and individual psychotherapy RCTs I conducted an ANCOVA with psychotherapy modality (group or individual) as the independent variable, total RCT-PQRS score as the dependent variable, and publication year as a covariate. To compare study quality between BED RCTs and BN RCTs, I conducted an ANCOVA with diagnosis as the independent variable, total RCT-PQRS score as the dependent variable, and publication year as a covariate. To assess whether publication year predicted the magnitude of the IF, and whether study quality mediated this relationship, I tested a mediation model with year of publication as the predictor variable, IF as the outcome variable, and total RCT-PQRS score as the potential mediator. I ran this model using the PROCESS (Hayes, 2012) macro in SPSS.

Finally, I compared the study quality of the 35 RCTs of psychotherapy for EDs to 120 RCTs of psychotherapy for depression (Thoma et al., 2012) and 14 RCTs of psychotherapy for anxiety (Keefe, McCarthy, Dinger, and Barber, 2014). An ANCOVA was conducted with type of RCT (ED, anxiety, or depression) as the independent variable, total RCT-PQRS score as the dependent variable, and publication year as a covariate. All analyses were conducted using IBM SPSS Statistics (IBM Corp, 2016).

Results

Study Quality

A total of 35 RCTs of direct comparisons of psychological treatments for eating disorders that were published between 1985 and 2015 (see Dissertation Appendix C for a complete reference list of included studies) were included in this study. Characteristics of the 35 RCTs can be seen in Table 1.

The average omnibus rating of RCTs for EDs (item 25) was 4.51 (*SD* = 1.22; range = 2 to 7). Twenty-eight RCTs (80%) received an omnibus rating of at least 4 (average quality; Gerber et al., 2011), but only two RCTs (5.7%) received an “exceptionally good” rating (omnibus rating of 7). The average total RCT-PQRS score (items one to 24 summed) was 28.26 (*SD* = 7.04; range = 15 to 42). Twenty-seven RCTs (77%) received a total RCT-PQRS score of at least 24 indicating “average” quality, and only four (11%) received a total score of at least 40 indicating “exceptionally good” quality. The percentage of studies that scored 0 (item unfulfilled), 1 (item partially fulfilled), or 2 (item fulfilled) for each item (1 to 24) can be seen in Figure 1. The following criteria were fulfilled by over half the included trials: (1) study conclusions justified by results; (2) balance of allegiance to types of treatment by practitioners; (3) comparison group from same population and time frame as experimental group; (4) appropriate statistical tests; (5)

intent-to-treat analyses; (6) outcomes specified a priori; (7) treatments sufficiently described or referenced to allow for replication; (8) description of number of subjects screened, included, and excluded; and, (9) description of inclusion/exclusion criteria. The following criteria were unfulfilled by over half the included trials: (1) appropriate consideration of therapist and site effects; (2) adequate sample size; (3) discussion of safety and adverse events; (4) outcome assessment by raters blind to treatment group and with established reliability; (5) therapist supervision while treatment being provided; and, (6) description of relevant participant comorbidities.

Comparing Quality of RCTs for EDs, Anxiety, and Depression

There were no statistically significant differences ($F_{(2,166)} = .665, p = .516$, *partial eta*^{2} = .008) in total RCT-PQRS scores between the 35 RCTs of psychotherapy for EDs (*M* = 28.26, *SD* = 7.04), 120 RCTs of cognitive behavioural therapy (CBT) for depression (*M* = 25.59, *SD* = 8.86; Thoma et al., 2012), or 14 RCTs of psychodynamic psychotherapy for anxiety disorders (*M* = 30.00, *SD* = 10.10; Keefe, McCarthy, Dinger, and Barber, 2014), while controlling for publication year. Publication year was a significant control variable with a moderate effect ($F_{(2,166)} = 58.83, p < .001, n_p^2 = .26$).

Publication Year and Study Quality

As hypothesized, the total RCT-PQRS score was positively and statistically significantly related to publication year ($r(33) = .50, p < .001$; see path a in Figure 2). The effect was large with more recent RCTs having higher study quality than older RCTs.

Effect Size and Study Quality

There was no statistically significant relationship between study quality and the effect sizes for frequency of binge eating and/or purging ($r(11) = -.44, p = .131$). Similarly, there was

no statistically significant relationship between study quality and the effect sizes for ED psychopathology ($r(13) = -.38$, $p = .161$). However, both correlation coefficients were moderate in effect size, suggesting that these analyses were underpowered.

Quality of Group vs. Individual Psychotherapy RCTs

There was no statistically significant difference in total RCT-PQRS scores ($F_{(1,33)} = .105$, $p = .748$, $partial\ eta^2 = .003$) between group psychotherapy RCTs ($M = 28.39$, $SD = 6.81$) and individual psychotherapy RCTs ($M = 28.00$, $SD = 7.76$), while controlling for publication year. The effect of publication year as a control variable was significant with a large effect ($F_{(1,33)} = 26.51$, $p < .001$, $n_p^2 = .45$).

Quality of BN vs. BED RCTs

There was no statistically significant difference in total RCT-PQRS scores ($F_{(1,29)} = .499$, $p = .486$, $partial\ eta^2 = .018$) between RCTs for BN ($M = 24.20$, $SD = 4.80$) and RCTs for BED ($M = 31.13$, $SD = 7.30$), while controlling for publication year. The effect of publication year as a factor was significant with a large effect ($F_{(1,29)} = 12.73$, $p = .001$, $n_p^2 = .31$).

Publication Year and Impact Factor Mediated by Study Quality

As mentioned above, there was a direct significant and positive relationship between publication year and study quality (path a; Figure 2). The direct relationship between study quality and IF was not statistically significant and the effect was small ($r(31) = .21$, $p = .081$; path b; Figure 2). There was no significant direct relationship between publication year and IF ($r(31) = .06$, $p = .503$). Finally, there was a statistically significant indirect effect of publication year on IF through study quality (*path c* = 0.11, 95% CI: 0.01, 0.29; Figure 2). Total RCT-PQRS score accounted for 64% (0.637) of the total effect, indicating a large effect.

Discussion

The overall goal of the current study was to evaluate the quality of RCTs of psychotherapy for adults diagnosed with an ED. Although mean RCT-PQRS quality scores were approximately what Kocsis et al. (2010) deemed “average”, only two of the RCTs of psychotherapy for EDs were rated as “exceptional” in quality, and so there is considerable room for improvement. Many of the RCTs did not fulfill specific standards commonly expected from treatment trials. Nevertheless, the mean total RCT-PQRS (Kocsis et al., 2010) score of the RCTs of psychotherapy for EDs was similar to that obtained from 120 RCTs of CBT for depression (Thoma et al., 2012) and 14 RCTs of psychodynamic psychotherapy for anxiety (Keefe, McCargthy, Dinger, and Barber, 2014). The results of the current study suggest that psychotherapy trials for EDs are of comparable quality and rigour as those found in other prominent areas of psychotherapy research.

I evaluated several factors that might be related to quality and therefore effect size. More recent RCTs of psychotherapy for EDs were more likely to have higher RCT-PQRS scores. Similar findings were reported from meta analyses examining the quality RCTs of ED prevention (Watson et al., 2017) and of psychodynamic psychotherapy for a variety of mental illnesses (Kocsis et al., 2010). A likely explanation for this positive relationship is greater awareness of methodological issues and the implementation of quality checklists that have been developed over the years. Journal editors are increasingly requiring that authors follow reporting guidelines such as CONSORT, and this has likely resulted in higher quality RCTs and/or better reporting of RCTs (Schulz et al., 2010).

Because the results of RCTs of psychotherapy for EDs have widespread impact on treatment recommendations and guidelines (American Psychiatric Association, 2006; American

Psychological Association, 2006; NICE, 2004; Yager et al., 2014), I examined whether the quality of an RCT was related to the magnitude of the effect size of primary outcomes at post-treatment. I found non-statistically significant correlations between study quality and effect sizes for primary outcome (frequency of binges and/or purges and ED psychopathology). However, these analyses were likely low in statistical power, and the correlations were negative and moderate in size, thus suggesting that lower quality may be related to higher effect sizes. This interpretation of our finding is consistent with reports by Cuijpers et al. (2010) for RCTs of psychotherapy for depression. This negative association between quality and effect size suggests that lower methodological quality may result in an overestimation of treatment effects.

I also evaluated contextual factors that may be related to study quality and, as a result, the magnitude of effects. First, I compared the quality of group and individual psychotherapy RCTs for EDs. Our results demonstrated that the quality of RCTs of group psychotherapy and individual psychotherapy are comparable. These findings are in line with those of Cuijpers et al. (2010) who reported no significant difference in quality between group and individual psychotherapy RCTs for depression. These findings support the use of both individual and group psychotherapy studies and meta analyses in informing treatment guidelines for eating disorders (Yager et al., 2014).

Second, given that treatment guidelines are written separately for each ED diagnosis, I assessed and described study quality by ED diagnosis. A comparison of the means in which publication year was not controlled indicated that trials for BED had higher mean study quality scores than trials for BN ($t(29) = 3.17$, $p = .003$). This is not surprising because the average publication year of BED trials is 2002 while the average publication year of BN trials is 1992. The quality of RCTs of psychotherapy for BED and BN were not significantly different after

publication year was controlled. Such a finding offers a more nuanced interpretation of the results of a previous meta-analysis that reported the effects of group psychotherapy versus another active treatment modality was significantly larger for BED than for BN (Grenon, Schwartze et al., 2017). Recall that RCTs of BED are more recent, and therefore of higher quality leading to more conservative estimates of treatment effects. Hence, the difference in group treatment effect for BED versus BN is likely underestimated and larger than previously reported. There was only one outpatient trial of psychotherapy for individuals diagnosed with AN (McIntosh et al., 2005) included in the current meta-analysis making it impossible to conduct analyses comparing AN, BN, and BED trials. Although there is a great need for more RCTs on psychotherapy for AN, high treatment refusal rates, low recruitment numbers, and high dropout rates make completing such trials challenging (Halmi, 2008).

Finally, I assessed whether there was an indirect effect of publication year on IF, through study quality. There was a significant indirect effect suggesting that there is an association between publication year and impact factor, through study quality. Critics have suggested that IFs can be inflated by journals that opt to publish studies that are likely to be highly cited, such as meta-analyses and systematic reviews (Misteli, 2013). Some authors attribute IF inflation to an increase in the number of references in each manuscript over the years (Althouse et al., 2008). Our findings suggest that while these may be factors causing IF inflation over time, it is also possible that this increase in IFs is partly due to improved study quality in more recently published RCTs. Readers should interpret this finding with some caution, and I encourage researchers to replicate these results in RCTs of psychotherapy for other disorders.

Limitations and Future Directions

There are three notable limitations of the current study, which are true for all studies that asses the quality of RCTs. First, the methodological quality of a study cannot easily be distinguished from reporting quality. Ideally, authors should be able to fully report study methodology, however, this is not the case given that many journal impose strict page limits. One way to address this would be for editors to encourage authors to report full methodological procedures via online supplementary materials. This would require additional guidelines for reporting complete details of such procedures. For example, journals could require authors of intervention RCTs to complete a checklist (with quality items from the RCT-PQRS; Kocsis et al., 2010) outlining how authors addressed each item. Without a thorough reporting of the methodology, it is difficult for reviewers, readers, future resesarchers, and developers of treatment guidelines, no matter how knowledgable, to to take into account study quality when evaluating aggregate effect sizes. The RCT-PQRS (Kocsis et al., 2010) represents an advancement in the field in that the scoring of items takes into consideration both the implementation and reporting of the quality criteria.

Second, the RCT-PQRS (Kocsis et al., 2010) has limited psychometric evaluation due to its recent development. To our knowledge, only a few studies have applied the RCT-PQRS and described their findings in detail. Although the omnibus rating and total quality score have high interrater reliability, many of the individual item scores may not be reliable. This may limit the number of analyses that researchers can conduct, and prevents one from examining quality criteria from a single item. For example, being able to reliably rate single items such as “balance of allegiance to type of treatment” within the research team (item 23) would allow one to assess the impact of multi-allegiance teams on effect sizes. One way to improve single-item interrater

reliability would be to develop a detailed user manual that outlines what to look for in the RCTs and how to score each item. As well, there are no generalizable norms indicating what scores are indicative of a high quality study. I hope that as more psychotherapy researchers use the RCT-PQRS, there will be more opportunities to compare findings, and establish norms for high, moderate, and low quality RCTs.

Finally, the inclusion/exclusion crtieria that I employed for research evaluated in this study have the potential to affect the results I reported. First, non-RCTs were excluded, which likely lead to higher quality scores than had they been included. Second, RCTs that had participants with subthreshold ED diagnoses were excluded from our study. It is unknown how the inclusion of studies with these participants might affect quality scores. Third, I excluded studies in which participants received concurrent treatments including hospitalization and medications. Because of this, I excluded an RCT of the treatment of AN by Zipfel and colleagues (2014), for example. I urge researchers to address questions of the effects of loosening or tightening inclusion and exclusion criteria so future authors of treatment guidelines for EDs can determine whether findings from some studies should be considered when developing treatment recommendations.

Conclusions and Recommendations

The quality of psychotherapy RCTs for EDs has been continuously improving since 1985, likely due to the development and use of several quality checklists and reporting guidelines such as CONSORT (Schulz et al., 2010). Despite this increase in quality, there remains room for improvement on many important quality criteria. Figure 1 presents the most common weaknesses found in psychotherapy trials for EDs. Six criteria were not fulfilled in over half of the included RCTs. Based on the findings presented in Figure 1, I recommend that ED

treatment researchers: report relevant comorbidities of all participants; have and report appropriate therapist supervision while treatment is being provided; include blind outcome assessment raters to treatment groups and report the reliability among raters; use adequate sample sizes and report power calculations; consider and statistically model therapist and/or site effects in analyses; and, discuss and report safety and adverse events. Additionally, I recommend researchers: register all psychotherapy RCTs with organizations like ClinicalTrials.gov; publish the RCT protocol, which outlines primary and secondary outcomes, hypotheses, and analyses prior to the start of the trial; use quality rating scales and checklists to help design studies (i.e., RCT-PQRS; Kocsis et al., 2010); and, follow the CONSORT guidelines for reporting (Schulz et al., 2010).

Given the correlations I found between study quality and effect size, and between year of publication and quality ratings in RCTs for EDs, I urge authors of meta-analyses to evaluate and report the quality of studies, or control for quality in their analyses, so that more accurate conclusions can be drawn regarding the magnitude of treatment effects. In addition, I suggest that authors of ED treatment guidelines consider the treatment effects reported in RCTs and meta-analyses for EDs within the context of the quality of the RCTs. This would require a more detailed, comprehensive, and nuanced evaluation of the existing research to inform treatment guidelines.

References

Allen, H. N. (1997). Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder (doctoral dissertation). University of North Carolina at Chapel Hill.

Althouse, B. M., West, J. D., Bergstrom, T., & Bergstrom, C. T. (2008). Differences in impact factor across fields and over time. *Journal of the Association for Information Science and Technology*, *60*, 27-34. doi: 10.1002/asi.20936

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.)*. Washington, DC: American Psychiatric Publishing.

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

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association Work Group on Eating Disorders (2006). Treatment of patients with eating disorders. Retrieved from http://psychiatryonline.org/content.aspx?bookid=28§ionid=1671334#138660

American Psychological Association (2006). Task force on evidence-based practice. Evidence-based practice in psychology. *American Psychologist, 61*, 271-285.

Baum, J. (2011). Free-riding on power laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies. *Organization*, *18*, 449-466. doi:10.1177/1350508411403531

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis, Version 3 [Computer Software]. Englewood, NJ: Biostat.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine, 40*, 211-223. doi: 10.1017/S0033291709006114

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: A meta-analysis. *Clinical Psychology Review, 34,* 130-140.

Egger, M., Smith, D. G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

Elkin, I., Shea, T, Watkins, J. T, Imber S. D., Sotsky, S. M., Collins, J. F, Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fies- ter, S. J., & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971- 982.

Fairburn, C. G., Bailey-Straebler, S., Basden, S., Doll, H. A., Jones, R., Murphy, R., O'Connor, M. E., Cooper, Z. (2015). A transdiagnostic comparison of enhanced cognitive behavior therapy (CBTE) and interpersonal psychotherapy in the treatment of eating disorders. *Behavior Research and Therapy, 70*, 64-71.

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1985 Annual, vol. 7 (Philadelphia: Institute for Scientific Information, 1986).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1986 Annual, vol. 7 (Philadelphia: Institute for Scientific Information, 1987).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1987 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1988).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1988 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1989).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1990 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1991).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1991 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1992).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1992 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1993).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1993 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1994).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1994 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1995).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1995 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1996).

Garfield, E. *SSCI Journal Citation Reports;* a bibliometric analysis of social science journals in the ISI Data Base. *Social Sciences Citation Index®* 1996 Annual, vol. 6 (Philadelphia: Institute for Scientific Information, 1997).

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association, 295*, 90-93.

Gerber, A. J., Kocsis, J. H., Milrod, B. L., Roose, S. P., Barber, J. P., Thase, M. E., & Leon, A. C. (2011). A quality-based review of randomized controlled trials of psychodynamic psychotherapy. *American Journal of Psychiatry*, *168*, 19-28. doi:10.1176/appi.ajp.2010.08060843

Grenon, R., Schwartze, D., Hammond, N., Ivanova, I., Mcquiad, N., Proulx, G., & Tasca, G. A. (2017). Group psychotherapy for eating disorders: A meta-analysis. *International Journal of Eating Disorders, 50,* 997-1013. doi: 10.1002/eat.22744

Grenon, R., Carlucci, S., Brugnera, A., Schwartze, D., Hammond, N., Ivanova, I., Mcquaid, N., Proulx, G., & Tasca, G. A. (2018). Psychotherapy for eating disorders: A meta-analysis of direct comparisons. *Psychotherapy Research, 29,* 1-13. doi: 10.1080/10503307.2018.1489162

Halmi, K. A. (2008). The perplexities of conducting randomized, double-blind, placebo-controlled treatment trials in anorexia nervosa patients. *American Journal of Psychiatry, 165,* 1227-8.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. Retrieved from http://www.afhayes.com/public/process2012.pdf

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *The British Medical Journal, 343*, d5928. doi: http://doi.org/10.1136/bmj.d5928

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011] Retrieved from: http://www.chochrane-handbook.org.

IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 24.0 [computer software]. Armonk, NY: IBM Corp.

Journal Citation Reports Social Sciences Edition. (2016). Retrieved from Thomson Reuters website: https://jcr.incites.thomsonreuters.com/

Jüni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Medical Journal*, *323*, 42–46.

Keefe, J. R., McCarthy, K. S., Dinger, U., Zilcha-Mano, S., Barber, J. P. (2014). A meta-analytic review of psychodynamic therapies for anxiety disorders. *Clinical Psychology Review, 34*, 309-323.

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., & Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry* , *51*, 319-324. doi:10.1016/j.comppsych.2009.07.001

Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the of Association for Information Science and Technology, 63*, 2140–2145. doi:10.1002/asi.22731

March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., Burns, B., Domino, M., McNulty, S., Vitiello, B., & Severe, J. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *Journal of the American Medical Association*, *292*, 807-820.

Misteli, T. (2013). Eliminating the impact of the Impact Factor. *The Journal of Cell Biology*, *201*, 651-652. doi:10.1083/jcb.201304162

Moher, D., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P., & Klassen, T. P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet*, *352*, 609-613.

National Institute for Health and Care Excellence. (2004). Eating disorders: Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. Retrieved from http://www.nice.org.uk/nicemedia/live/10932/29220/29220.pdf

Schulz, K. F., Altman, D. G., & Moher, D. for the CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, *152*, 726-732.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*, 498-502. doi:10.1136/bmj.314.7079.497

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Telch, C. F., Agras, W. S., Rossiter, E. M., Wilfley, D., Kenardy, J. (1990). Group cognitive-behavioral treatment for the nonpurging bulimic: An initial evaluation. *Journal of Consulting and Clinical Psychology, 58,* 629-635.

Thoma, N. C., McKay, D., Gerber, A. J., Milrod, B. L., Edwards, A. R., & Kocsis, J. H. (2012). A Quality-Based Review of Randomized Controlled Trials of Cognitive-Behavioral Therapy for Depression: An Assessment and Metaregression. *American Journal of Psychiatry*, *169*, 22-30. doi:10.1176/appi.ajp.2011.11030433

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, 'all must have prizes'. *Psychological Bulletin, 122,* 203-215.

Watson, H. J., Goodman, E. L., McLagan, N. B., Joyce, T., French, E., Willan, V., & Egan, S. J. (2017). Quality of randomized controlled trials in eating disorder prevention. *International Journal of Eating Disorders, 50*, 459-470. doi: 10.1002/eat.22712.

Yager, J., Devlin, M. J., Halmi, K. A., Herzog, D. B., Mitchell, J. E., Powers, P., & Zerbe, K. F. (2014). Guideline watch (August 2012): Practice guideline for the treatment of patients with eating disorders (3rd ed.) *Focus*, 12, 416–431.

Zipfel, S., Wild, B., Groß, G., Friederich, H. C., Teufel, M., Schellberg, D., ... & Burgmer, M. (2014). Focal psychodynamic therapy, cognitive behaviour therapy, and optimised treatment as usual in outpatients with anorexia nervosa (ANTOP study): randomised controlled trial. *The Lancet*, *383*, 127-137.

Table 1

Characteristics of Included Randomized Controlled Trials

Study	*C*	*IG*	*CG*	N_{IG}	N_{CG}	*Total Quality Score*[a]	*Omnibus Rating*[b]	*IF*[a]	*IF*[b]	*Diagnosis*
Agras *et al*. 1989	US	iCBT	WL	39	19	30	5	na	2.941	BN
Agras *et al*. 1994	US	gCBT	BWL	72	37	26	4	1.434	2.494	BED
Agras *et al.* 1995	US	gCBT	WL	39	11	23	4	3.225	3.057	BED
Agras *et al.* 2000	US	iCBT	iIPT	110	110	35	6	11.778	11.622	BN
Allen, 1997	US	gCBT (AAT)	WL	15	14	17	3	na	na	BED
Bailer *et al.* 2004	AT	gCBT	Self-help	41	40	25	4	1.867	1.839	BN
Cooper *et al.* 1995	UK	iCBT	iERP	13	14	24	4	2.290	2.078	BN
Esplen *et al.* 1998	CA	iGuided Imagery	ND	28	30	21	3	3.124	3.412	BN
Fairburn *et al.* 1986	UK	iCBT	iSTF	12	12	30	5	1.467	1.727	BN
Fairburn *et al.* 1991	UK	iCBT	iBT	25	25	25	4	7.918	9.505	BN
		iCBT	iIPT	25	25					
		iBT	iIPT	25	25					
Fairburn *et al.* 2015	UK	iCBT	iIPT	65	65	40	6	3.798	na	BN, BED, EDNOS
Freeman *et al.* 1988	UK	iCBT	WL	32	20	15	2	3.136	3.758	BN
		iBT	WL	30	20					

Study	*C*	*IG*	*CG*	N_{IG}	N_{CG}	*Total Quality Score*[a]	*Omnibus Rating*[b]	*IF*[a]	*IF*[b]	*Diagnosis*
Freeman *et al.* 1988 *continued*	UK	iCBT	iBT	32	30	15	2	3.136	3.758	BN
		iCBT	SG	32	30					
		iBT	SG	30	30					
Gorin *et al.* 2003	US	gCBT	WL	32	31	31	5	1.540	1.915	BED
Griffiths *et al.* 1994	AU	iCBT	WL	23	28	21	3	na	0.263	BN
		iCBT	Hypno.	23	27					
Grilo *et al.* 2011	US	gCBT	BWL	45	45	42	7	4.848	5.228	BED
Hsu *et al.* 2001	US	iCT	SG	26	24	24	4	3.119	3.133	BN
		iCT	NT	26	23					
Kirkley *et al.* 1985	US	gCBT	ND	14	14	22	3	1.967	2.173	BN
Laessle *et al.* 1987	DE	gCBT	WL	8	9	20	3	0.282	0.315	BN
Lavender *et al.* 2012	UK	gCBT	ESM	37	37	33	5	3.730	3.234	BN/EDNOS
Lee *et al.* 1986	US	gCBT	WL	15	15	24	4	1.392	1.199	BN
Leitenberg *et al.* 1988	US	gCBT	WL	12	17	25	4	2.007	2.633	BN
		gCBT+ERP	WL	15	17					
McIntosh *et al.* 2005	NZ	iCBT	iIPT	19	21	31	5	8.286	9.127	AN
		iCBT	SCM	19	16					
		iIPT	SCM	21	16					
Munsch *et al.* 2007	CH	gCBT	BWL	44	36	33	5	2.269	2.797	BED
Peterson *et al.* 1998	US	gCBT	WL	16	11	25	4	1.139	1.336	BED
Peterson *et al.* 2009	US	gCBT	WL	60	69	36	6	12.522	12.539	BED

Study	*C*	*IG*	*CG*	N_{IG}	N_{CG}	*Total Quality Score*[a]	*Omnibus Rating*[b]	*IF*[a]	*IF*[b]	*Diagnosis*
Safer *et al.* 2010	US	gDBT	ST	50	51	35	6	2.408	2.911	BED
Schlup *et al.* 2009	CH	gCBT	WL	18	18	34	5	2.995	3.295	BED
Tasca *et al.* 2006	CA	gCBT	WL	50	46	42	7	0.930	1.579	BED
		GPIP	WL	50	46					
		gCBT	GPIP	50	50					
Telch *et al.* 1990	US	gCBT	WL	23	21	30	5	2.633	2.941	BED
Telch *et al.* 2001	US	gDBT	WL	22	22	24	4	3.566	3.252	BED
Vocks *et al.* 2011	DE	gCBT (BIT)	WL	32	30	24	4	6.159	5.428	AN, BN, EDNOS
Wilfley *et al.* 1993	US	gCBT	WL	18	20	25	4	3.066	5.428	BED
		gIPT	WL	18	20					
Wilfley *et al.* 2002	US	gCBT	gIPT	81	81	35	6	11.622	11.207	BED
Wilson *et al.* 2010	US	iIPT	BWL	75	64	40	6	10.782	13.772	BED
		iIPT	Self-Help	75	66					
Wolf *et al.* 1992	US	gCBT	WL	15	12	22	3	1.167	1.536	BN
		gBT	WL	15	12					

Abbreviations: *C*: Country; US: United States; AT: Austria; UK: United Kingdom; AU: Australia; CA: Canada; NZ: New Zealand; CH: Switzerland; DE: Germany; SE: Sweden; IG: Intervention group; i: individual; g: group; CBT: cognitive-behaviour therapy; BT: behaviour therapy; CT: cognitive therapy; IPT: interpersonal psychotherapy; CG: comparison group; WL: wait-list; ERP: exposure plus response-prevention; ND: non-directive; STF: short term focal therapy; SG: support group; Hypno.: hypnobehavioural therapy; NT: nutritional therapy; SCM: supportive clinical management; BWL: behavioural weight loss; GPIP: group psychodynamic interpersonal psychotherapy; ST: supportive therapy; ESM: emotional and social mind training; BIT: body image therapy; N_{IG}: sample size of intervention group; N_{CG}: sample size of comparison group; Tx: treatment; BED: binge-eating disorder; BN: bulimia nervosa; AN: anorexia nervosa; EDNOS: eating disorder not otherwise specified; Total quality score[a]: Total quality score from the Randomized Controlled Trial-Psychotherapy Quality Rating Scale (RCT-PQRS); Omnibus rating[b]: Omnibus rating of overall quality from the RCT-PQRS; 1: exceptionally poor; 2: very poor; 3: moderately poor; 4: average; 5: moderately good; 6: very good; 7: exceptionally

good. *IF*[a]: impact factor for the year the study was published; *IF*[b]*:* impact factor two years after the study was published; na: impact factor not available.

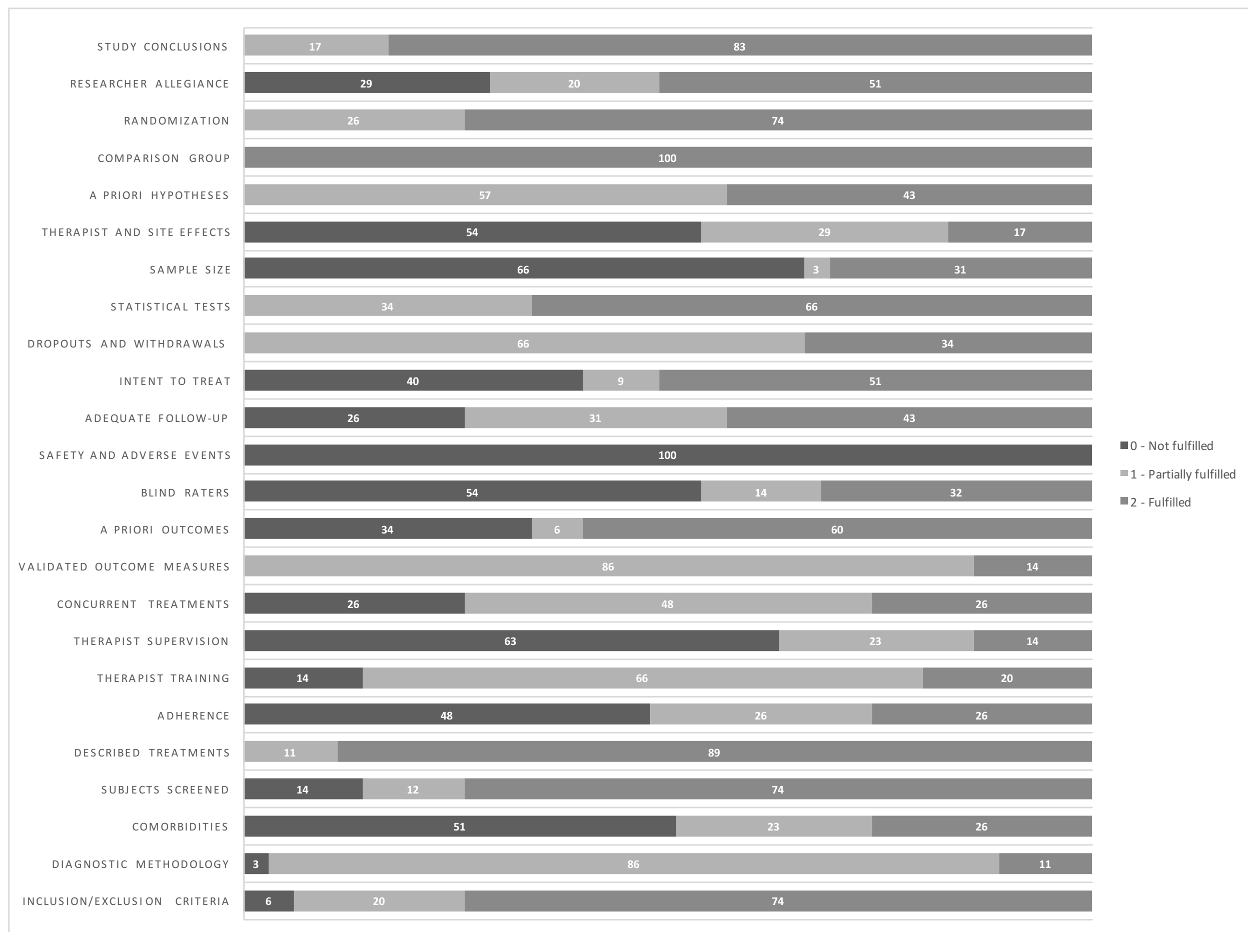


Figure 1. Percent of psychotherapy for eating disorders randomized controlled trials (RCTs) by item score (N = 35) on the RCT – Psychotherapy Quality Rating Scale (PQRS). Items are rated from 0 to 2. Descriptions of items: *Inclusion/exclusion criteria* = diagnostic method and criteria for inclusion and exclusion. *Diagnostic methodology* = documentation or demonstration of reliability of diagnostic methodology. *Comorbidities* = Description of relevant comorbidities. *Subjects screened* = Description of numbers of subjects screened, included, and excluded. *Described treatments* = Treatment(s) (including control/comparison groups) are sufficiently described or referenced to allow for replication. *Adherence* = Method to demonstrate that treatment being studied is treatment being delivered (only satisfied by supervision if transcripts or tapes are explicitly reviewed). *Therapist training* = Therapist training and level of experience in the treatment(s) under investigation. *Therapist supervision* = Therapist supervision while treatment is being provided. *Concurrent treatments* = Description of concurrent treatments (eg, medication) allowed and administered during course of study (if patients on medication are included, a rating of 2 requires full reporting of what medications were used; if patients on medications are excluded, this alone is sufficient for a rating of 2). *Validated outcome measures* = Validated outcome measure(s) (either established or newly standardized). *A priori outcomes* = Primary outcome measure(s) specified in advance (although does not need to be stated explicitly for a rating of 2). *Blind raters* = Outcome assessment by raters blinded to treatment group and with established reliability. *Safety and adverse events* = Discussion of safety and adverse events during study treatment(s). *Adequate follow-up* = Assessment of long-term post-termination

outcome (should not be penalized for failure to follow comparison group if this is a wait- list or non-treatment group that is subsequently referred for active treatment). *Intent to treat* = Intent-to-treat method for data analysis involving primary outcome measure. *Dropouts and withdrawals* = Description of dropouts and withdrawals. *Statistical tests* = Appropriate statistical tests (eg, use of Bonferroni correction, longitudinal data analysis, adjustment only for a priori identified confounders). *Sample size* = Adequate justification and sample size. *Therapist and site effects* = Appropriate consideration of therapist and site effects. *A priori hypotheses* = A priori relevant hypotheses that justify comparison group(s). *Comparison group/s* = Comparison group(s) from same population and time frame as experimental group. *Randomization* = Full and appropriate method of randomization performed after screening and baseline assessment. *Researcher allegiance* = Balance of allegiance to types of treatment by practitioners. *Study conclusions* = Conclusions of study justified by sample, measures, and data analysis, as presented (note: useful to look at conclusions as stated in study abstract).

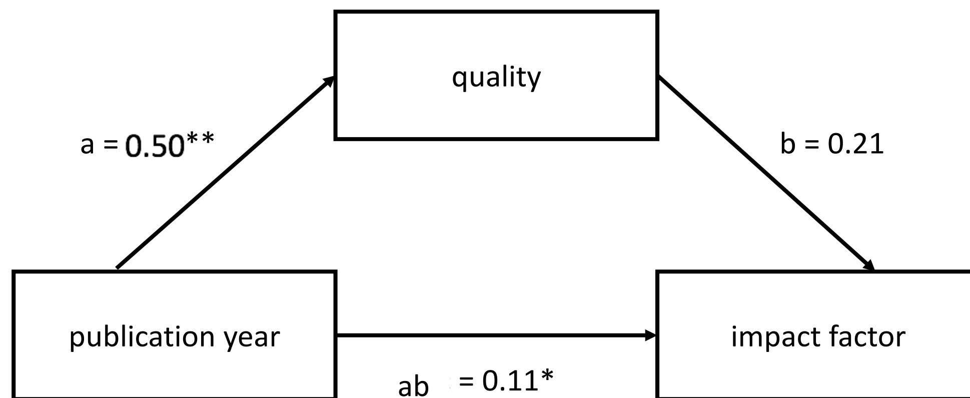


Figure 2. Standardized regression coefficients for the relationship between publication year and impact factor (IF) as mediated by study quality. The standardized regression coefficient between publication year and IF, after controlling for study quality, is represented by the c path. Note: ** $p < .001$, * $p < .05$

General Discussion

Treatment guidelines for EDs published by the American Psychiatric Association (Yager et al., 2014) recommend individual and group CBT and individual IPT as the best studied first-line treatment options for BN and BED. This may suggest that CBT and IPT possess specific therapeutic ingredients that will result in superior efficacy compared to other interventions for reducing ED symptoms. However, there is limited understanding about how different psychotherapies for EDs compare, and whether individual and group psychotherapy modalities are equally effective. Additionally, given that the findings of RCTs of psychotherapy for EDs are used to develop treatment guidelines, which can influence clinician's practice, and that the quality of such studies can influence the magnitude of treatment outcomes, it is imperative to evaluate the quality of RCTs in order to better interpret the findings. Thus, the overall goals of this dissertation were to systematically and comprehensively evaluate the efficacy of psychotherapy for EDs with two meta-analyses of RCTs of direct comparisons, and to evaluate and describe the quality of said RCTs. Taken together, results from the three manuscripts included in this dissertation will give readers a better understanding of the efficacy of psychotherapy for EDs, so ultimately, clinicians and researchers can be better informed when interpreting the findings of studies and recommendations in treatment guidelines. Further, developers of treatment guidelines for EDs and clinicians' who treat individuals with EDs can consider the findings of these meta analyses when updating treatment guidelines and developing treatment plans for patients.

Review of the Findings

Study One

The first study is a meta-analysis of RCTs of bona fide psychotherapy for EDs. The first study: (1) estimated the effect of bona fide psychotherapy for adults with EDs compared to wait-list controls and to non-bona fide treatments (behavioural weight loss, self-help, supportive therapy, or non-directive therapy); and (2) compared the effect of bona fide CBT to bona fide non-CBT psychotherapies. In this meta-analysis, I conducted three comparisons.

First, I compared bona fide psychotherapy vs. wait-list controls. At post-treatment, individuals who underwent bona fide psychotherapy yielded significantly better outcomes than those in wait-list control groups. These findings are consistent with the psychotherapy literature, in general. Most studies that compare bona fide psychotherapy to wait-list controls demonstrate that bona fide psychotherapy is significantly more effective (Wampold et al., 1997).

Second, I compared bona fide psychotherapy vs. non-bona fide treatment. At post-treatment, bona fide psychotherapy yielded greater abstinence rates and reductions in ED psychopathology compared to non-bona fide treatments. There were no significant differences between bona fide psychotherapy and non-bona fide treatment in reducing the frequency of binge eating and/or purging, reducing depressive symptoms, or improving self-concept. At short-term follow-up, there were no differences between bona fide psychotherapy and non-bona fide treatment, with the exception of reducing the frequency of binge eating and/or purging and reducing depressive symptoms. At long-term follow-up, bona fide psychotherapy showed statistically significantly higher abstinence rates compared to non-bona fide treatment. Many of the significant differences between bona fide psychotherapy and non-bona fide treatment,

however, were not stable. Until further research is available on the efficacy of specific non-bona fide treatments, bona fide psychotherapy should remain the first-line treatment recommendation.

Finally, I compared bona fide CBT vs. bona fide non-CBT psychotherapy. There were no significant differences between bona fide CBT and bona fide non-CBT on any outcome at any time-point, with the exception of reducing ED psychopathology at post-treatment, and short-term follow-up. This finding may be due to the particular nature of the measurements used in these studies. As such, based on the current literature, there does not appear to be any clear advantage of bona fide CBT over other bona fide psychotherapies for EDs.

Study Two

The second study of this dissertation is a meta-analysis of RCTs of bona fide group psychotherapy. The goals of the second study were to: (1) estimate the effect of bona fide group psychotherapy for EDs compared to wait-list controls, other active treatments, and when possible, individual psychotherapy alone, and; (2) compare the effects of bona fide group CBT to bona fide non-CBT group psychotherapy. In this meta-analysis, I conducted three comparisons.

First, I compared bona fide group psychotherapy vs. wait-list controls. At post-treatment, individuals who underwent bona fide group psychotherapy yielded significantly better outcomes than those in wait-list control groups for all outcomes, with the exception of interpersonal problems which was approaching significance (better interpersonal problems outcomes for participants in group psychotherapy compared to wait-list controls). Again, this finding was expected as this outcome is commonly found in the psychotherapy literature.

Second, I compared bona fide group psychotherapy vs. other active treatments. There were no differences in outcomes between bona fide group psychotherapy and other active treatments at any time point with two exceptions. Compared to other active treatments, group

psychotherapy yielded greater reductions in the frequency of binge eating and/or purging at post-treatment and short-term follow-up for individuals with BED. Group psychotherapy and other active treatments showed similar effects in reducing the frequency of binge eating and/or purging for individuals with BN. Effect sizes varied significantly across studies included in these analyses, and I suspect this was due to the range of other active treatments included in the analyses. However, with too few comparisons, I could not conduct moderator analyses. Only three studies directly compared the effects of group and individual psychotherapy for EDs, and findings showed similar effects for abstinence rates, improving ED psychopathology, and decreasing depressive symptoms.

Finally, I compared bona fide group CBT vs. bona fide non-CBT group psychotherapy. There were no significant differences between bona fide group CBT and bona fide non-CBT group psychotherapy on any outcome at any time-point. Findings should be interpreted with caution as results were based on a limited number of comparisons.

Study Three

The goal of the third study of this dissertation was to evaluate and describe the quality of RCTs of bona fide psychotherapy for EDs using the RCT-PQRS. As expected, more recent RCTs had significantly higher study quality scores than older RCTs. The quality of RCTs of psychotherapy for EDs was similar to the quality of RCTs of psychotherapy for other mental illnesses. Study quality scores were not significantly related to the magnitude of effect sizes for primary outcomes, however, correlation coefficients were moderate in size, suggesting that analyses were likely underpowered. Study quality scores for group psychotherapy RCTs and individual psychotherapy RCTs were similar. Finally, results of a mediation analysis indicated

that higher study quality partially explained the positive relationship between publication year and IF.

Implications for Treatment Guidelines for Eating Disorder

Findings from study one and study two of this dissertation demonstrate that bona fide psychotherapy is effective in the treatment of EDs. When bona fide psychotherapy was compared to a wait-list control condition, bona fide psychotherapy resulted in significantly better outcomes, with moderate to large effect sizes. This finding is clinically important as it emphasizes the benefit of undergoing bona fide psychotherapy for an ED versus not seeking treatment. In other words, this comparison highlights the positive impact of bona fide psychotherapy relative to the natural course of an ED. Additionally, the bona fide psychotherapy versus wait-list control results support the inclusion of bona fide psychotherapy as a first-line treatment option in ED treatment guidelines.

Findings from study one and study two also suggest that the bona fide psychotherapies included in the meta-analyses are effective in the treatment of EDs, regardless of treatment orientation. This is in contrast to current treatment guidelines for EDs (e.g., Yager et al., 2014), which recommend only individual and group CBT and individual IPT as first-line treatments for BN and BED. However, the quality of many of the studies cited in treatment guidelines may not have been not considered when the guidelines were developed. In a recent meta-scientific review, only 19% of psychological treatments cited in the research supported treatments list by Division 12 of the American Psychological Association were based on high quality studies (Sakaluk et al., 2019). In the third study of this dissertation, we rated only two studies (5.7%) as "exceptionally good" in quality using the RCT-PQRS. That is the vast majority of RCTs for the

treatment of EDs were missing some key components that might ensure replicability of the findings.

Treatments regarded as having “strong support” by some treatment guidelines are those with more than one RCT that shows statistically significant improvements often when compared to no treatment. Another potential issue with this approach is that individual studies tend not to replicate easily, especially if sample sizes are small (Maxwell, Lau, and Howard, 2015). This likely occurs because effect sizes of multiple studies distribute along a continuum, and a single effect size represents only one point on that continuum. Using results from meta-analyses of direct comparisons and considering the quality of included RCTs may be valuable in guiding future treatment recommendations. Further, in order for patients with an ED to access a wider range of treatments, researchers would have to test a broader sampling of bona-fide psychotherapies other than those already evaluated. For example, individuals who do not respond to CBT or IPT may respond positively to other bona fide psychotherapies tested for specific EDs (e.g., GPIP for BED, focal psychodynamic psychotherapy for AN, and others).

In study one I compared the effects of bona fide psychotherapy and non-bona fide treatments. Results indicated that some non-bona fide therapies may be as effective as some bona fide psychotherapy in the treatment of EDs, however, results were not stable. It is important that more trials be conducted in which bona fide psychotherapy and non-bona fide treatment are directly compared so clear conclusions can be drawn about the efficacy of certain non-bona fide treatments. If certain non-bona fide treatments are found to be as effective as bona fide psychotherapy, then it would important that this be reflected in treatment guidelines as some of these treatments may be more accessible (i.e., self-help programs) and cost-effective (i.e.,

behavioural weight loss groups, support groups, etc.) than bona fide psychotherapy, which has to be delivered by a trained therapist.

The American Psychiatric Association, American Psychological Association, and the National Institute for Health and Care Excellence (NICE) consider the results from RCTs to be the most influential in developing their treatment guidelines (American Psychiatric Association Work Group on Eating Disorders, 2006; American Psychological Association, 2006; NICE, 2004). However, given that single psychotherapy RCTs are typically based on small samples, the estimates of effects may be biased (Field and Gillett, 2010). A meta-analysis integrates the findings from all available RCTs and quantitatively synthesizes and summarizes the results of multiple studies examining the same research. A well conducted meta-analysis has the potential to provide reliable estimates of effects that are superior to any single RCT. Findings from meta-analyses should have more of an influence on the development of treatment guidelines than single RCTs. Creating a sub-committee responsible for reviewing meta-analyses to aide in the development of treatment guidelines would be a worthwhile undertaking for the aforementioned organizations.

Implications for Group Psychotherapy for Eating Disorders

The results of this dissertation provide additional evidence that group psychotherapy is effective in the treatment of ED symptoms and improving depressive symptoms and self-concept. Both the CBT-enhanced model (Fairburn, 2008) and interpersonal model of binge eating (Wilfley et al., 1993) suggest that interpersonal problems play a key role in the maintenance of ED symptoms. Given that many group psychotherapies include therapeutic factors that specifically address interpersonal functioning, such therapies are able to target the interpersonal problems that maintain psychopathology secondary to an ED and ED symptoms

(Ivanova et al., 2015). Assessing interpersonal problems in patients with an ED who want to undergo psychotherapy can provide therapists with an understanding of whether patients may be better suited for individual or group psychotherapy. For example, patients with an ED who score high on an interpersonal problems measure may benefit more from a group psychotherapy that includes therapeutic factors like peer interpersonal feedback, social learning, and emotional expression.

There is an increasing need for economical psychotherapeutic interventions as service demands in mental health centers grow. Compared to individual psychotherapy, group psychotherapy may be more cost-effective than individual psychotherapy (McCrone et al., 2005; Otto, Pollack, and Maki, 2008; Roberge, Marchand, Reinharz, and Savard, 2008). While the research evaluating and comparing the costs of group psychotherapy compared to other forms of treatment for mental disorders is limited, practitioners have noted the cost-effectiveness of group psychotherapy for years (American Group Psychotherapy Association, 2010). Group psychotherapy allows for more patients to be treated per clinician hour, likely resulting in more cost-effective treatment compared to individual psychotherapy.

Implications for the Equivalence of Treatments

The results of the current dissertation, in which few outcome differences between treatment modalities were noted, provide some support for the notion that therapeutic change is the result of several factors that are common across psychotherapies. Thus, the bona fide psychotherapies included in the current studies, regardless of treatment orientation, and that are intended to be therapeutic, may be equally effective. Our findings also suggest that some non-bona fide treatments may be as effective as some bona fide psychotherapies. Non-bona fide

treatments may capitalize on common factors (e.g., therapeutic alliance, instillation of hope, empathy, intention to be effective, patient expectations) or specific ingredients.

Treatment equivalence could also be consistent with the medical model of psychotherapy, which suggests that it is the specificity (specific ingredients) of the treatment that targets the mechanisms that result in the ED. For example, it may be that the specific ingredients in each of the bona fide psychotherapies are equally effective in minimizing ED symptoms because they target a specific though different maintaining factor. However, given that psychotherapy research has constistantly shown the importance of common factors in treatment outcomes (Wampold & Imel, 2015), it is unlikely that it is solely the various specific ingredients that are contributing to positive outcomes. It is more likely that both common factors and specific ingredients are working together to lead to the positive thrapuetic changes reported in study one and study two. This is in line with the contextual model which is a combination of both the common factors and medical model of psychotherapy. The contextual model suggests that psychotherapies that include key common factors and specific ingredients may be equally effective for a given mental disorder.

While treatment guidelines recommend CBT and IPT as the best studied first-line psychotherapy options, other psychotherapy approaches, along with CBT and IPT, should be investigated further to allow for more diversity in evidence-based treatment options for patients with EDs. In addition to this, non-bona fide treatments that may be effective, partly due to common factors or specific ingredients, should be further investigated and if indicated, included in ED treatment guidelines. Having a wider array of treatment options, especially for those who do not respond to CBT or IPT, may help improve the overall outcomes for individuals with an ED.

Implications for Considering the Quality of Trials of Psychotherapy for Eating Disorders

In study two (which I conducted prior to study one) I used the Cochrane Risk of Bias tool to rate the quality of RCTs of group psychotherapy for EDs. Since the Cochrane Risk of Bias tool was developed for medical and pharmacological contexts, it did not address some important quality aspects of psychotherapy research and several items were not applicable to psychotherapy trials making the results difficult to interpret. Given these limitations for assessing quality or bias issues in psychotherapy trials, in study one and three I chose to use the RCT-PQRS (Kocsis et al., 2010) to assess the quality of all the psychotherapy RCTs. The relationship between the magnitude of the effect size of primary outcomes at post-treatment and study quality were not statistically significant, however, these analyses were likely low in statistical power, and the correlations were negative and moderate in size. This interpretation of our finding is consistent with reports by Cuijpers et al. (2010) for RCTs of psychotherapy for depression, and suggests that lower methodological quality may result in an overestimation of treatment effects. Further, older RCTs were of lower quality and their effect sizes were larger.

All psychotherapy trials should be evaluated for quality before using the results to aide in the development of treatment guidelines. Further, I recommend that when describing the findings from influential RCTs, the authors of the guidelines should also report the quality of such RCTs. In particular, older RCTs that continue to influence current treatment guidelines should be examined carefully and considered cautiously, and their influence on guidelines may need to be moderated in light of these findings.

Future Directions

Based on the findings of the three studies included in this dissertation, I have several recommendations for future researchers who will be evaluating the efficacy of various treatments

for EDs. First, I encourage researchers to continue investigating the efficacy of non-bona fide treatments (i.e., behavioural weight loss, self-help, etc.) for EDs, so that clear conclusions can be drawn for each of these treatments. Based on the available literature, it appears that some non-bona fide interventions may be as effective as bona fide psychotherapy. Similarly, I recommend that future researchers continue evaluating the efficacy of diverse bona fide psychotherapies (i.e., GPIP, DBT, IPT, etc.), including CBT, for EDs. Findings from both study one and study two suggest that the included bona fide psychotherapies, regardless of theoretical orientation, may be effective in the treatment of EDs, but there is a lack of diversity in the available and tested treatments for EDs. Future researchers should investigate what specific and common factors are responsible for the psychotherapeutic change in individuals with an ED. Furthermore, it has been estimated that 25% of all individuals in the community with an ED are male but only 1% of ED research is focused on males with an ED (Sweeting, Walker, MacLean, Patterson, Räisänen, and Hunt, 2015). For example, in study one of this dissertation, only 3 out of eighteen studies that compared bona fide psychotherapy and a wait-list control condition included males. I recommend that future researchers evaluate the efficacy of current treatment options for males. The low base rates of males with some EDs makes recruiting only men into a treatment study difficult if not impossible. However, researchers could make a concerted effort to include males in their trials, and report or discuss their outcomes separately. Doing so would help guide future research and potentially lead to evidence-based treatment options specifically for males.

Second, I urge researchers to conduct RCTs that directly compare the effects of bona fide group psychotherapy and bona fide individual psychotherapy. Only three such RCTs currently exist in the literature making clear conclusions of direct comparisons of group and individual psychotherapy difficult. This has implications for the utility of group therapy which may target

specific maintenance factors for EDs. I recommend that such studies also include a cost-effectiveness component in order to assess another component of the potential utility of providing therapies in groups that may be of particular interest to health care systems.

Finally, all trials should be registered and the protocol published before the study begins, with results reported in peer-reviewed literature following CONSORT (CONsolidated Standards of Reporting Trials) standards. I recommend that journal editors require authors of RCTs to report full methodological procedures via online supplementary materials. Furthermore, I urge authors of meta-analyses to evaluate and report the quality of included studies, or to control for quality in their analyses, so that more accurate conclusions can be drawn regarding the magnitude of treatment effects. In particular, RCT-PQRS is particularly suited to assessing the quality of psychotherapy trials. In addition, I suggest that developers of ED treatment guidelines consider the treatment effects reported in RCTs and meta-analyses for EDs within the context of the quality of the RCTs.

Conclusions

This doctoral dissertation adds to the ED literature by: (1) systematically and comprehensively evaluating the efficacy of psychotherapy with two meta-analyses of RCTs of direct comparisons; (2) evaluating and describing the quality of the included RCTs, and; (3) assessing whether study quality was related to the magnitude of reported effects. Overall, the findings suggest that the included bona fide psychotherapies, regardless of theoretical orientation or modality (group or individual) are effective in the treatment of EDs, thus providing support for the equivalence of psychotherapies, and the need for more research to increase diversity of available treatments. While bona fide psychotherapy and non-bona fide treatments yielded

similar results for many outcomes, findings were not stable, and bona fide psychotherapy out-performed non-bona fide treatments on several outcomes.

Similarly, group psychotherapy for EDs yielded similar effects to other treatments, such as individual psychotherapy, self-help, behavioural weight-loss, and pharmacotherapy. Group and individual CBT are the current first-line treatments recommended for individuals with an ED, however our results suggest that as a whole other group therapies (ESM, GPIP, IPT and BT) may be as effective as group CBT in the treatment of BN and BED.

The quality of psychotherapy RCTs for EDs has been continuously improving since 1985, likely due to the development and use of several quality checklists and reporting guidelines. Despite this increase in quality, this dissertation highlighted that there remains room for improvement on many important quality criteria. Future research directions should include: identifying key common factors while treating individuals with an ED in order to examine causal links to outcomes; continuing to investigate diverse forms of group and individual psychotherapy so that more definite conclusions can be drawn and treatment guidelines can be updated accordingly; and evaluating and reporting the quality of studies and controlling for quality in meta-analyses. Such research will allow for more accurate conclusions to be drawn regarding the magnitude of treatment effects, and as a result, guidelines may provide more useful direction to clinicians who will be able to optimize outcomes for patients with an ED.

References

Agras, W. S., Telch, C. F., Arnow, B., Eldredge, K., Wilfley, D. E., Raeburn, S. D., … Marnell, M. (1994). Weight loss, cognitive-behavioral, and desipramine treatments in binge eating disorder: An additive design. *Behavior Therapy, 25,* 225–238.

American Group Psychotherapy Association (2010). Practitioners tout cost-effectiveness of group psychotherapy. Retrieved from http://www.agpa.org/docs/default-source/media/practitioners-tout-cost-effectiveness-of-group-therapy.pdf?sfvrsn=0

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders (4th ed. revised)*. Washington, DC: American Psychiatric Publishing.

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

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.)*. Washington, DC: American Psychiatric Publishing.

American Psychiatric Association (n.d.). *Practice Guideline Development Process.* Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association Work Group on Eating Disorders (2006). Treatment of patients with eating disorders. Retrieved from http://psychiatryonline.org/content.aspx?bookid=28§ionid=1671334#138660

American Psychological Association (2006). Task force on evidence-based practice. Evidence-based practice in psychology. *American Psychologist, 61*, 271-285.

Armijo-Olivo, S., Fuentes, J., Ospina, M., Saltaji, H., & Hartling, L. (2013). Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. *BMC Medical Research Methodology*, *13*, 116.

Bailer, U., de Zwaan, M., Leisch, F., Strnad, A., Lennkh-Wolfsberg, C., El- Giamal, N., . . . Kasper, S. (2004). Guided self-help versus cognitive behavioral group therapy in the treatment of bulimia nervosa. *International Journal of Eating Disorders, 35*, 522–537.

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology*, *73*, 924.

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., … Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, *10*, e1001454. doi:10.1371/journal.pmed.1001454

Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., et al. (2016). Outcome differences between individual and group formats when identical and non-identical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy, 53,* 446–461.

Chakraborty, H., & Gu, H. (2009). *A mixed model approach for intent-to-treat analysis in longitudinal clinical trials with missing values*. (RTI Press Publication No. MR-0009-0903). Research Triangle Park, NC: RTI Press. doi: 10.3768/rtipress.2009.mr.0009.0903

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Chatoor, I., & Kurpnick, J. (2001). The role of non-specific factors in treatment outcome of psychotherapy studies. *European Child & Adolescent Psychiatry*, *10*, S19.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine, 40*, 211-223. doi: 10.1017/S0033291709006114

Fairburn, C.G., & Bohn, K. (2005). Eating disorder NOS (EDNOS): an example of the troublesome "not otherwise specified" (NOS) category in DSM-IV. *Behaviour Research Therapy, 43*, 691-701. doi: 10.1016/j.brat.2004.06.011

Fairburn, C.G., & Harrison, P. J. (2003). Eating disorders. *The Lancet, 361*, 407-416.

Fairburn, C.G., & Wilson, G. T. (1993). *Binge eating: Nature assessment, and treatment* (pp. 317–360). New York, NY: Guilford Press.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63,* 665-694. doi: org/10.1348/000711010X502733

Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). Baltimore, MD: Johns Hopkins University Press.

Gerber, A. J., Kocsis, J. H., Milrod, B. L., Roose, S. P., Barber, J. P., Thase, M. E., & Leon, A. C. (2011). A quality-based review of randomized controlled trials of psychodynamic psychotherapy. *American Journal of Psychiatry* , *168*, 19-28. doi:10.1176/appi.ajp.2010.08060843

Grilo, C. M., Masheb, R. M., Wilson, G. T., Gueorguieva, R., & White, M. A. (2011). Cognitive–behavioral therapy, behavioral weight loss, and sequential treatment for obese

patients with binge-eating disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *79*, 675-685.

Higgins, J.P.T., Altman, D.G., & Sterne, J.A.C. (2011). Assessing risk of bias in included studies. In J.P.T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (Vol. Version 5.1.0). Retrieved from http://www.cochrane-handbook.org:

Hudson, J. I., Hiripi, E., Pope, H. G., Jr., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry, 61*, 348-358. doi: 10.1016/j.biopsych.2006.03.040

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, *1*, 138-155.

Ivaova, I. V., Tasca, G. A., Hammond, N., Balfour, L., Ritchie, K., Koszycki, D., & Bissada, H. (2015). Negative Affect Mediates the Relationship Between Interpersonal Problems and Binge-Eating Disorder Symptoms and Psychopathology in a Clinical Sample: A Test of the Interpersonal Model. *European Eating Disorders Review, 23,* 133-138. doi: 10.1002/erv.2344

Keefe, J. R., McCarthy, K. S., Dinger, U., Zilcha-Mano, S., Barber, J. P. (2014). A meta-analytic review of psychodynamic therapies for anxiety disorders. *Clinical Psychology Review, 34*, 309-323.

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., & Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry* , *51*, 319-324. doi:10.1016/j.comppsych.2009.07.001

Leichsenring, F., & Rabung, S. (2008). Effectiveness of Long-Term Psychodynamic Psychotherapy: A Meta-analysis. *Journal of the American Medical Association, 300*, 1551-1565.

McCrone, P. R., Weeramanthri, T., Knapp, M. R. J., Rushton, A., Trowell, J., Miles, G., & Kolvin, I. (2005). Cost-effectiveness for individual versus group psychotherapy for sexually abused girls. *Child and Adolescent Mental Health*, *10*, 26-31. doi: 10.1111/j.1475-3588.2005.00113.x

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist, 70*, 487–498. doi: 10.1037/a0039400

Mitchell, J. E., Crosby, R. D., Wonderlich, S. A., Hill, L., le Grange, D., Powers, P., & Eddy, K. (2007). Latent profile analysis of a cohort of patients with eating disorders not otherwise specified. *International Journal of Eating Disorders, 40 Suppl*, S95-98. doi: 10.1002/eat.20459

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, *16*, 62–73. doi: 10.1016/0197-2456(94)00031-W

Moher, D., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P., & Klassen, T. P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet*, *352*, 609-613.

Munder, T., & Barth, J. (2018). Cochrane’s risk of bias tool in the context of psychotherapy outcome research. *Psychotherapy Research*, *28*, 347-355.

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: an overview of reviews. *Clinical Psychology Review*, *33*, 501-511.

Munsch, S., Biedert, E., Meyer, A., Michael, T., Schlup, B., Tuch, A., & Margraf, J. (2007). A randomized comparison of cognitive behavioral therapy and behavioral weight loss treatment for overweight individuals with binge eating disorder. *International Journal of Eating Disorders, 40,* 102–113.

National Institute for Health and Care Excellence (NICE). (2004). Eating disorders: Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. Retrieved from http://www.nice.org.uk/nicemedia/live/10932/29220/29220.pdf

Otto, M. W., Pollack, M. H., & Maki, K. M. (2008). Empirically supported treatments for panic disorder: costs, benefits, and stepped care. *Journal of Consulting and Clinical Psychology, 68*, 556-63.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*, 829.

Peterson, C. B., Mitchell, J. E., Crow, S. J., Crosby, R. D., & Wonderlich, S. A. (2009). The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. *American Journal of Psychiatry, 166,* 1347-1354.

Peterson, C.B., Mitchell, J.E., Engbloom, S., Nugent, S., Mussell, M.P., & Miller, J. P. (1998). Group cognitive-behavioral treatment of binge-eating disorder: A comparison of

therapist-led versus self-help formats. *International Journal of Eating Disorders, 24*, 125–136.

Polnay, A., James, V. A., Hodges, L., Murray, G. D., Munro, C., & Lawrie, S. M. (2014). Group therapy for people with bulimia nervosa: Systematic review and meta-analysis. *Psychological Medicine, 44*, 2241–2245.

Roberge, P., Marchand, A., Reinharz, D., & Savard, P. (2008). Cognitive-behavioral treatment for panic disorder with agoraphobia: a randomized, controlled trial and cost-effectiveness analysis. *Behavior Modification, 32,* 333-351. doi: 10.1177/0145445507309025.

Safer, D. L., Robinson, A. H., & Jo, B. (2010). Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. *Behavior Therapy, 41*, 106-120.

Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology, 128,* 500-509. doi: 10.1037/abn0000421

Schulz, K. F., Altman, D. G., & Moher, D. for the CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, *152*, 726-732.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, *273*, 408-412.

Smink, F. R., Van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: incidence, prevalence and mortality rates. *Current Psychiatry Reports*, *14*, 406-414.

Spielmans, G.I., Benish, S.G., Marin, C., Bowman, W.M., Menster, M., & Wheeler, A.J. (2013). Specificity of psychological treatments for bulimia nervosa and binge eating disorder? A meta-analysis of direct comparisons. *Clinical Psychology Review, 33*, 460-469.

Spielmans, G. I., Berman, M. I., & Usitalo, A. N. (2011). Psychotherapy versus second-generation antidepressants in the treatment of depression: a meta-analysis. *The Journal of Nervous and Mental Disease*, *199*, 142-149.

Spielmans, G. I., Pasek, L. F., & McFall, J. P. (2007). What are the active ingredients in cognitive and behavioral psychotherapy for anxious and depressed children? A meta-analytic review. *Clinical Psychology Review*, *27*, 642-654.

Sweeting, H., Walker, L., MacLean, A., Patterson, C., Räisänen, U., & Hunt, K. (2015). Prevalence of eating disorders in males: a review of rates reported in academic research and UK mass media. *International Journal of Mens Health, 14,* 1-27. doi: 10.3149/jmh.1402.86.

Thoma, N. C., McKay, D., Gerber, A. J., Milrod, B. L., Edwards, A. R., & Kocsis, J. H. (2012). A Quality-Based Review of Randomized Controlled Trials of Cognitive-Behavioral Therapy for Depression: An Assessment and Metaregression. *American Journal of Psychiatry*, *169*, 22-30. doi:10.1176/appi.ajp.2011.11030433

Valentine, J.C. (2009). Judging the quality of primary research. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 129-146). New York: Russell Sage Foundation.

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry, 14,* 270-277.

Wampold, B. E., Ahn, H. N., & Coleman, H. L. (2001). Medical model as metaphor: Old

habits die hard. *Journal of Counseling Psychology, 48*, 268-273.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York: Routledge.

Wampold, B. E., Minami, T., Baskin, T. W., & Tierney, S. C. (2002). A meta-(re) analysis of the effects of cognitive therapy versus ‘other therapies’ for depression. *Journal of Affective Disorders*, *68*, 159-165.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially, ‘all must have prizes’. *Psychological Bulletin, 122,* 203–215.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*, 200-211.

Wilfley, D. E., Agras, W. S., Telch, C. F., Rossiter, E. M., Schneider, J. A., Cole Golomb, … Raeburn, S. D. (1993). Group cognitive-behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A controlled comparison. *Journal of Consulting and Clinical Psychology, 61*, 296-305.

Wilfley, D. E., Welch, R. R., Stein, R. I., Spurrell, E. B., Cohen, L. R., Saelens, B. E., … Matt, G. E. (2002). A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the tretament of overweight individuals with binge-eating disorder. *Archives of General Psychiatry, 59*, 713-721.

Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2010). Psychological treatments for binge eating disorder. *Archives of General Psychiatry, 67,* 94-101.

Yager, J., Devlin, M. J., Halmi, K. A., Herzog, D. B., Mitchell, J. E., Powers, P., & Zerbe, K. F. (2014). Guideline watch (August 2012): Practice guideline for the treatment of patients with eating disorders (3rd ed.) *Focus*, *12,* 416–431.

APPENDIX A

Study One

Supplementary Materials: For Online Publication Only

	Page
A) Search Terms	166
B) Search Strategies	169
C) PRISMA Study Flow Chart	170
D) Reference List of Included Primary Studies	171
E) Exclusion Criteria	174
F) List of Excluded Studies that Underwent a Full Review	174
G) Primary and Secondary Outcomes	177
H) References for Outcome Measures	178
I) Detailed Descriptions of Non-Bona Treatments	180
J) Interrater Reliability	185
K) Forest Plot: Abstinence Rates for Bona Fide Psychotherapy vs. Wait-List	186
L) Forest Plot: Abstinence Rates for Bona Fide Psychotherapy vs. Non-Bona Fide Treatments	187
M) Forest Plot: Abstinence Rates for Bona Fide Cognitive Behavioural Therapy vs. Bona Fide Non-Cognitive Behavioural Therapy	188

MEDLINE Search Terms

Diagnosis	(“eat* disord*”[Title/Abstract] OR anorexi*[Title/Abstract] OR bulimi*[Title/Abstract] OR “binge eat*”[Title/Abstract] OR EDNOS[Title/Abstract] OR “purg* disord*”[Title/Abstract])
AND	
Intervention	(intervention[Title/Abstract] OR therap*[Title/Abstract] OR psychotherap*[Title/Abstract] OR psychoanaly*[Title/Abstract] OR cognitive behav* therap*[Title/Abstract] OR CBT[Title/Abstract] OR counseling*[Title/Abstract] OR psychodyn*[Title/Abstract] OR behav* therap*[Title/Abstract] OR treatment[Title/Abstract])
AND	
Design	(random* assign*[Title/Abstract] OR random* allocat*[Title/Abstract] OR trial*[Title/Abstract] OR control group[Title/Abstract] OR experimental design[Title/Abstract] OR treatment effectiveness[Title/Abstract] OR treatment efficacy[Title/Abstract] OR treatment outcomes[Title/Abstract])

PsycINFO Search Terms

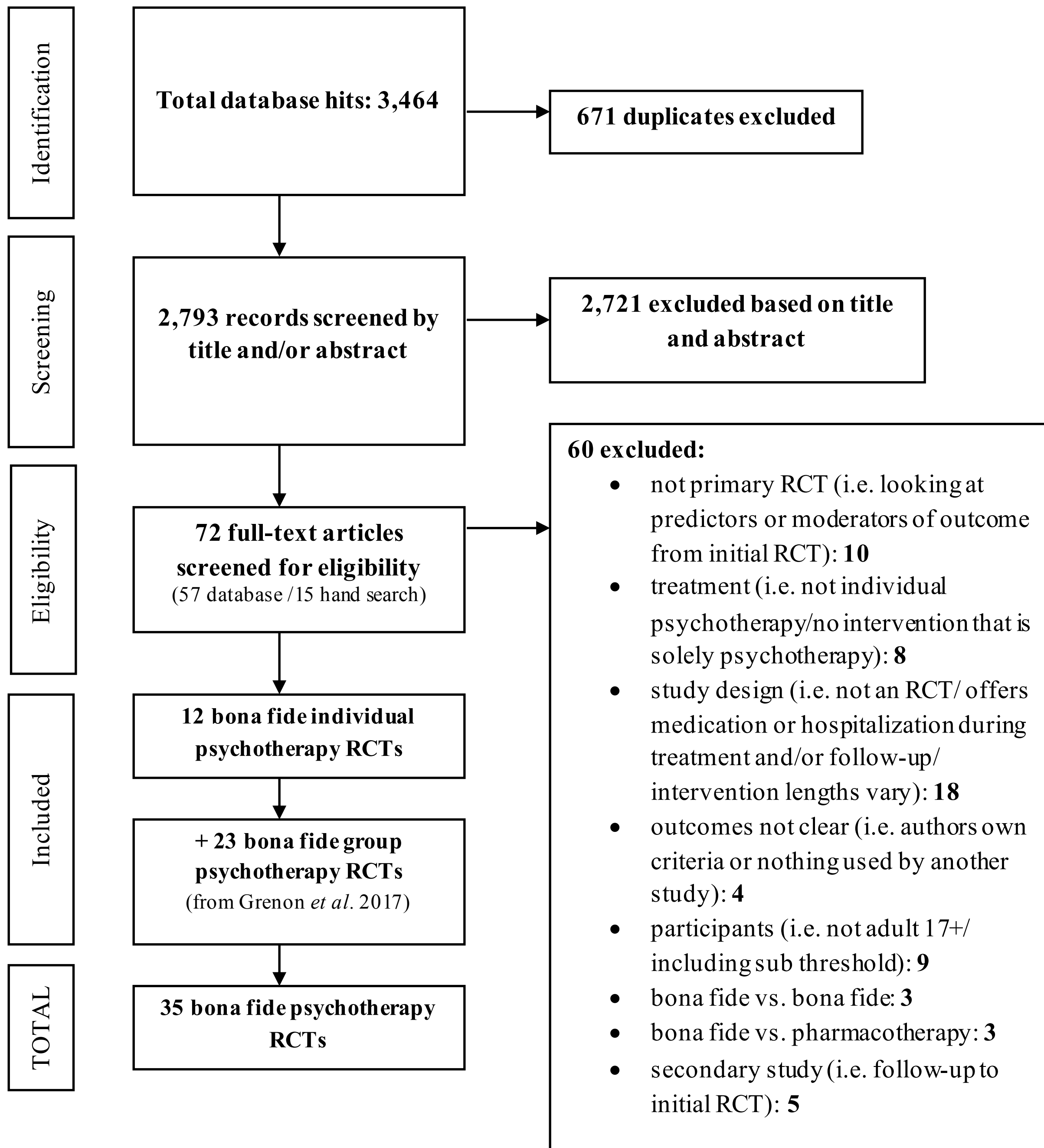
Diagnosis	TI: (eat* disord*) OR anorexi* OR bulimi* OR (binge eat*) OR EDNOS* OR (purg* disord*) OR AB: (eat* disord*) OR anorexi* OR bulimi* OR (binge eat*) OR EDNOS* OR (purg* disord*)
AND	
Intervention	TI: (intervention OR therap* OR psychotherap* OR psychoanaly* OR (cognitive behav* therap*) OR CBT OR counseling OR psychodyn* OR (behav* therap*) OR treatment) OR AB: intervention OR therap* OR psychotherap* OR psychoanaly* OR (cognitive behav* therap*) OR CBT OR counseling OR psychodyn* OR (behav* therap*) OR treatment
AND	
Design	TI: (random* assign*) OR (random* allocat*) OR trial* OR (control group) OR (experimental design) OR (treatment effectiveness) OR (treatment efficacy) OR (treatment outcomes) OR AB: (random* assign*) OR (random* allocat*) OR trial* OR (control group) OR (experimental design) OR (treatment effectiveness) OR (treatment efficacy) OR (treatment outcomes)

CENTRAL Search Terms

Diagnosis	(eat* disord* or anorexi* or bulimi* or binge eat* or EDNOS or purg* disord*):ti,ab,kw
AND	
Intervention	(intervention or therap* or psychotherap* or psychoanaly* or cognitive behav* therap* or psychodyn* or CBT or behavioral or counselling or treatment):ti,ab,kw
Design	
	Only includes RCTs.

Search Strategies

Search terms and strategies were identical for both individual and group psychotherapy RCTs, with the exception of "group" being used in the intervention terms. Electronic databases were screened from January 1st, 1980 to August 21st, 2017. Search terms included those indicative of the population (eating disorder OR anorexia OR bulimia OR binge-eating disorder OR purging disorder) combined with intervention terms (psychotherapy OR therapy, etc.) and study design terms (randomized trials, etc.). Reference lists of previous meta-analyses and systematic reviews of psychological treatments for EDs were manually screened, and reference lists of all eligible primary studies were screened. Independent raters carried out the screening and selection of studies, and disagreements were resolved through consensus. No language restrictions were applied.



PRISMA Study Flow Chart: Flow chart of the study selection process of individual psychotherapy for the treatment of eating disorders. The 23 group psychotherapy RCTs were included from a previous meta-analysis (Grenon et al., 2017). *Note.* RCT: randomized controlled trial

References of Included Primary Studies

Agras WS, Schneider JA, Arnow, B, Raeburn, SD, Telch, CF. Cognitive-Behavioral and Response-Prevention Treatments for Bulimia Nervosa. J Consult Clin Psych 1989;57:215-221.

Agras WS, Telch CF, Arnow B, Eldredge K, Detzer MJ, Henderson J, Marnell M. Does interpersonal therapy help patients with binge eating disorder who fail to respond to cognitive-behavioral therapy? J Consult Clin Psychol 1995;63:356-360.

Agras WS, Telch CF, Arnow B, Eldredge K, Wilfley DE, Raeburn SD, Henderson J, Marnell M. Weight loss, cognitive- behavioral, and desipramine treatments in binge eating disorder: An additive design. Behav Ther 1994;25:225-238.

Agras WS, Walsh T, Fairburn CG, Wilson GT, Kraemer HC. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. Arch Gen Psychiatry 2000;57:459-466.

Allen HN. Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder (doctoral dissertation). University of North Carolina at Chapel Hill 1997.

Bailer U, de Zwaan M, Leisch F, Strnad A, Lennkh-Wolfsberg C, El-Giamal N, Hornik K, Kasper S. Guided self-help versus cognitive-behavioral group therapy in the treatment of bulimia nervosa. Int J Eat Disord 2004;35:522-537.

Cooper PJ, Steere J. A comparison of two psychological treatments for bulimia nervosa: Implications for models of maintenance. Behav Res Ther 1995;33:875-885.

Esplen MJ, Garfinkel PE, Olmstead M, Gallop RM, Kennedy S. A randomized controlled trial of guided imagery in bulimia nervosa. Psychological Medicine 1998;28:1347-1357.

Fairburn CG, Bailey-Straebler S, Basden S, Doll HA, Jones R, Murphy R, O’Connor ME, Cooper Z. A transdiagnostic comparison of enhanced cognitive behavior therapy (CBT-E) and interpersonal psychotherapy in the treatment of eating disorders. Behav Res Ther 2015;70:64-71.

Fairburn CG, Jones R, Peveler RC, Carr SJ, Solomon RA, O’Connor ME, Burton J, Hope RA. Three psychological treatments for bulimia nervosa. Arch Gen Psychiatry 1991;48:463-469.

Fairburn CG, Kirk J, O’Connor M. A comparison of two psychological treatments for bulimia nervosa. Behav Res Ther 1986;24:629-643.

Freeman CPL, Barry F, Dunkeld-Turnbull J, Henderson A. Controlled trial of psychotherapy for bulimia nervosa. British Medical Journal 1988; 296: 521-525.

Gorin AA, Le Grange D, Stone AA. Effectiveness of spouse involvement in cognitive behavioral therapy for binge eating disorder. Int J Eat Disord 2003;33:421-433.

Griffiths RA, Hadzi-Pavlovic D, Channon-Little L. A controlled evaluation of hypnobehavioural treatment for bulimia nervosa: Immediate pre-post treatment effects. Eur Eat Disord Rev 1994;2:202-220.

Grilo CM, Masheb RM, Wilson GT, Gueorguieva R, White MA. Cognitive-behavioral therapy, behavioral weight loss, and sequential treatment for obese patients with binge-eating disorder: a randomized controlled trial. J Consult Clin Psychol 2011;79:675-685.

Hsu LKG, Rand W, Sullivan S, Liu DW, Mulliken B, McDonagh B, Kaye WH. Cognitive therapy, nutritional therapy and their combination in the treatment of bulimia nervosa. Psychological Medicine 2001;31:871-879.

Kirkley BG, Schneider JA, Agras W, Bachman JA. (1985). Comparison of two group treatments for bulimia. J Consult Clin Psychol 1985;53:43-48.

Laessle RG, Waadt SS, Pirke KM. (1987). A structured behaviorally oriented group treatment for bulimia nervosa. Psychother Psychosom 1987;48:141-145.

Lavender A, Startup H, Naumann U, Samarawickrema N, DeJong H, Kenyon M, van den Eynde F, Schmidt U. Emotional and social mind training: A randomised controlled trial of a new group-based treatment for bulimia nervosa. PLoS ONE 2012;7:1-9.

Lee NF, Rush AJ. Cognitive-behavioural group therapy for bulimia. Int J Eat Disord 1997;5:599–615.

Leitenberg H, Rosen JC, Gross J, Nudelman S, Vara LS. Exposure plus response-prevention treatment of bulimia nervosa. J Consult Clin Psychol 1988;56:535-541.

McIntosh VVW, Jordan J, Carter FA, Luty SE, McKenzie JM, Bulik CM, Frampton CMA, Joyce PR. Three psychotherapies for anorexia nervosa: A randomized, controlled trial. Am J Psychiatry 2005;162:741-747.

Munsch S, Biedert E, Meyer A, Michael T, Schlup B, Tuch A, Margraf J. A randomized comparison of cognitive behavioral therapy and behavioral weight loss treatment for overweight individuals with binge eating disorder. Int J Eat Disord 2007;40:102-113.

Peterson CB, Mitchell JE, Crow SJ, Crosby RD, Wonderlich SA. The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. Am J Psychiatry 2009;166:1347-1354.

Peterson CB, Mitchell JE, Engbloom S, Nugent S, Mussell MP, Miller JP. Group cognitive-behavioral treatment of binge-eating disorder: A comparison of therapist-led versus self-help formats. Int J Eat Disord 1998;24:125–136.

Safer DL, Robinson AH, Jo B. Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. Behav Ther 2010;41:106-120.

Schlup B, Munsch S, Meyer AH, Margraf J, Wilhelm FH. The efficacy of a short version of a cognitive-behavioral treatment followed by booster sessions for binge eating disorder. Behav Res Ther 2009;47:628-635.

Tasca GA, Ritchie K, Conrad G, Balfour L, Gayton J, Lybanon V, Bissada H. Attachment scales predict outcome in a randomized controlled trial of two group therapies for binge eating disorder: An aptitude by treatment interaction. Psychother Res 2006;16:106-121.

Telch CF, Agras WS, Linehan MM. Dialectical behavior therapy for binge eating disorder. J Consult Clini Psychol 2001:69;1061-1065.

Telch CF, Agras WS, Rossiter EM, Wilfley D, Kenardy J. Group cognitive-behavioral treatment for the nonpurging bulimic: An initial evaluation. J Consult Clin Psychol 1990;58:629-635.

Vocks S, Schulte D, Busch M, Grönemeyer D, Herpertz S, Suchan B. Changes in neuronal correlates of body image processing by means of cognitive-behavioural body image therapy for eating disorders: a randomized controlled fMRI study. Psychol Med 2011;41:1651-1663.

Wilfley DE, Agras WS, Telch CF, Rossiter EM, Schneider JA, Cole Golomb A, Sifford L, Raeburn SD. Group cognitive-behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A Controlled Comparison. J Consult Clin Psychol 1993;61:296-305.

Wilfley DE, Welch RR, Stein RI, Spurrell EB, Cohen LR, Saelens BE, Dounchis JZ, Frank MA, Wiseman CV, Matt GE. A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the tretament fo overweight individuals with binge-eating disorder. Arch Gen Psychiatry 2002;59:713-721.

Wilson GT, Wilfley DE, Agras WS, Bryson SW. Psychological Treatments for Binge Eating Disorder. Arch Gen Psychiatry 2010;67:94-101.

Wolf EM, Crowther JH. An evaluation of behavioral and cognitive-behavioral group interventions for the treatment of bulimia nervosa in women. Int J Eat Disord 1992;11:3-15.

Exclusion Criteria:

The following types of studies were excluded: (a) dismantling studies, where a full treatment was compared with one or more components of that same treatment; (b) studies of in-patient psychotherapy where a variety of treatments were provided and the effects of psychotherapy alone were impossible to isolate; (c) studies that did not provide information necessary to compute effect sizes even after we attempted to contact original authors; and (d) studies with treatment confounds, where medication, hospitalization, or another active treatments (e.g. nutritional counseling) was provided to participants outside of the trial and these participants were not excluded from analyses.

Excluded studies that underwent a full review

AUTHORS	YEAR	REASON FOR EXCLUSION
		NOT PRIMARY RCT (i.e., protocol, follow-up study)
Accurso et al.	2015	
Accurso et al.	2016	
Agras et al.	1994	
Carter et al.	2003	
Carter et al.	2011	
Daniel et al.	2016	
Elbaky et al.	2014	
Esplen et al.	1998	
Fairburn et al.	1993	
Fairburn et al.	1993	
Fairburn et al.	1995	
Fischer et al.	2014	
Griffiths et al.	1996	
Jones et al.	1993	
Schmidt et al.	2016	
		STUDY DESIGN (i.e., non-RCT, medication/hospitalization during treatment or follow-up/ intervention lengths not equal, dismantling)
Bachar et al.	1999	
Balestrieri et al.	2015	
Barga et al.	2004	
Brambilla et al.	2014	
Bulik et al.	1998	
Byrne et al.	2017	
Channon et al.	1989	
Ciano et al.	2002	
Davis et al.	1997	
Davis et al.	1999	
Dicker	2003	
Fairburn et al.	2009	
Garner et al.	1993	
Ghaderi et al.	2006	

Hilbert et al.	2004	
Jager et al.	1997	
Le Grange et al.	2002	
Leitenberg et al.	1998	
Loeb et al.	2000	
Olmstead et al.	1991	
Painot et al.	2001	
Pendleton et al.	2002	
Richter-Reno et al.	1992	
Russel et al.	1987	
Schmidt et al.	1989	
Schmidt et al.	2012	
Schmidt et al.	2013	
Schmidt et al.	2015	
Schutzmann et al.	2009	
Tantillo et al.	2003	
Thiels et al.	1998	
Touyz et al.	2013	
Treasure et al.	1999	
Walsh et al.	1997	
Weinstein et al.	1994	
Wilson et al.	1986	
Wilson et al.	1991	
		TREATMENT (i.e., not individual psychotherapy, no intervention that is solely psychotherapy)
Alfonsson et al.	2015	
Dingemans et al.	2007	
Grilo et al.	2005	
Grilo et al.	2012	
Katzman et al.	2010	
Laessle et al.	1991	
Ordman et al.	1985	
Shag et al.	2015	
Stice et al.	2015	
		PARTICIPANTS (i.e., not adult, subthreshold)
Cassin et al.	2008	
Freeman et al.	1985	
Goodrick et al.	1998	
Hall et al.	1987	
Nauta et al.	2000	
Raynaud et al.	1999	
Safer et al.	2001	
Schmidt et al.	2007	
Serfaty et al.	1999	
Stein et al.	2013	
Thackwray et al.	1993	

Wild et al.	2009	
Wonderlich et al.	2014	
Zipfel et al.	2014	
		OUTCOMES/DATA (i.e., outcomes not clear, unable to use data provided)
Dare et al.	2001	
Gowers et al.	1994	
Leitenberg et al.	1994	
Sundgot-Borgen et al.	2002	
Treasure et al.	1995	

Primary Outcomes

(1) Abstinence rates: dichotomous outcome of the number of participants in each condition with 100% cessation of binge eating and/or purging in the last seven, 14, 21, or 28 days. Both observer and self-reported measures were included. Observer measures involved asking participants about bingeing and/or purging behaviours. This was done using a calendar recall method, the Rating of Anorexia and Bulimia Interview (RABI; Clinton & Norring, 1999), or with a version of the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993). Self-report measures involved the same questions but in a written format (i.e. EDE-Questionnaire [EDEQ], Fairburn & Beglin, 1994); standardized food records, or a food diary). The number of participants randomized was used as the denominator regardless of whether completer data or intent-to-treat data were reported in order to avoid overestimation of effects.
(2) Frequency of binge eating/purging: continuous outcome of the number of binge-eating and/or purging episodes in the last seven or 28 days. This outcome was measured using the same or similar methods as abstinence rates, with the addition of the Eating Disorders Questionnaire (EDQ; Mitchell, Hatsukami, Eckert, & Pyle, 1985).
(3) ED related psychopathology: continuous outcome of subscale scores of valid ED psychopathology measures. Scales included the EDE, EDEQ, Eating Disorder Inventory (EDI; Garner, Olmstead, & Polivy, 1983), Three Factor Eating Questionnaire (TFEQ; Stunkard & Messick, 1985), Binge Eating Scale (BES; Gormally, Black, Dastin, & Rardin, 1982), Body Shape Questionnaire (BSQ; Cooper, Taylor, Cooper, & Fairburn, 1987), Eating Attitudes Test (EAT; Garner & Garfinkel, 1979), and RABI.

Secondary Outcomes

(1) Depressive symptoms: measures included continuous outcomes of overall or subscale scores of valid depressive symptoms measures. Scales included the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), depression subscale of the Symptom Checklist-90 (SCL-90; Derogatis, 1977), Hamilton Depression Rating Scale (HDRS; Hamilton, 1960), Centre for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977), and the Inventory for Depressive Symptomology-Self-Report (IDS-SR; Rush et al., 1987).
(2) Self-concept: measures included continuous outcomes of overall or subscale scores of valid self-esteem, self-efficacy, and self-concept measures. Scales included the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1979), Lawson Social Self-Esteem Scale (LSE; Lawson, Marshall, & McGrath, 1979), Frankfurt Self-Concept Scale (FKSN; Deusinger, 1986), and General Self-Efficacy Scale (SWE; Jerusalem & Schwarzer, 1999)
(3) Interpersonal problems: continuous outcome of the overall score of the Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureño, & Villaseñnor, 1988).

References for Outcome Measures

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

Clinton, D., & Norring, C. (1999). The Rating of Anorexia and Bulimia (RAB) interview: Development and preliminary validation. *European Eating Disorder Review, 7,* 362–371.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the body shape questionnaire. *International Journal of Eating Disorder,* 6, 485–494.

Dare, C., Eisler, I., Russell, G., Treasure, J., & Dodge, L. (2001). Psychological therapies for adults with anorexia nervosa. *The British Journal of Psychiatry, 178,* 216-221.

Derogatis, L. R. (1977). *SCL-90, administration, scoring and procedures manual for the revised version*. Baltimore: John Hopkins School of Medicine.

Deusinger, I. M. (1986). *Die Frankfurter Selbstkonzeptskalen* (FSKN). Handanweisung. Gottingen: Hogrefe.

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

Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination (12th edition). In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: nature assessment, and treatment* (pp. 317-360). New York, NY: Guilford Press.

Garner, D. M., & Garfinkel, P. E. (1979). The eating attitudes test: An index of symptoms of anorexia nervosa. *Psychological Medicine, 9,* 273–279.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological Neurosurgery Psychiatry, 23*, 56–62.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2,* 15–34.

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

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñnor, V. S. (1988). Inventory of interpersonal problems: Psychometirc properties and clinical applications. *Journal of Consulting Clinical Psychology, 56,* 885–892.

Jerusalem, M., & Schwarzer, R. (1999). Skala zur allgemeinen selbstwirksamkeit- serwartung (SWE). In R. Schwarzer & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schulermerkmalen. Dokumentation der psychometrischen Ver- fahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbst- wirksame Schulen*. Berlin, Germany: Freie Universita€t Berlin.

Lawson, J. S., Marshall, W. L., & McGrath, P. (1979). The social selfesteem inventory. *Educational Psychology Measurement, 39,* 803–811.

Mitchell, J. E., Hatsukami, D., Eckert, E. D., & Pyle, R. L. (1985). Eating disorders questionnaire. *Psychopharmacological Bulletin, 21*, 1025–1043.

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

Rosenberg, M. (1979). *Conceiving the self*. New York, NY: Basic Books.

Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Research, 18,* 65–87.

Stunkard, A. J., & Messick, S. (1985). *Society and the adolescent selfimage*. Princeton, NJ: Princeton University Press

Non-Bona Fide Treatments: Description
Each description below was taken directly from the RCT.

<u>Behavioural Weight Loss (BWL)</u>

Agras et al., 1994:
Therapy was based on the LEARN Program for Weight Control (Brownell, 1985), which was modified for use with this particular population and extended to a 30-week program and material specifically dealing with methods to overcome binge eating (or loss of control over eating) was removed from the manual to minimize overlap with cognitive-behavioral therapy. The focus of therapy was on making gradual lifestyle changes in diet, exercise, and eating habits. A diet aiming at a weight loss of about one pound per week was suggested, and participants were encouraged to experiment in order to find the level of caloric intake that allowed them to meet this goal. The importance of reducing fat intake and increasing the intake of complex carbohydrates was stressed. Consumption of three meals each day and eating heart-healthy food. Relapse prevention techniques were introduced toward the end of the program. Participants self-monitored different aspects of their behavior over the 30-session program, including food intake, exercise, and thoughts concerning food. At each session, the self-monitoring of participants was reviewed, problems were aired, and the group was challenged to find solutions to particular problems. Therapy did not focus on the problem of binge eating, rather the focus was on weight loss, and weight was monitored at each session throughout treatment.

- **Intervention focused on weight-loss and lacked any psychologically valid component.**

Grilo et al., 2011:
Based on a manual from Brownell, 2000 (16 group session (60 min.) over a 24 week period). LEARN is an acronym for lifestyle, exercise, attitudes, relationships and nutrition. It focus on making gradual lifestyle changes with goals of moderate caloric restriction and increased physical activity to produce gradual weight losses.

- **Intervention focused on weight-loss and lacked any psychologically valid component.**

Munsch et al., 2007:
Based on the manual "Weight Loss with Xenical" (Margraf, 2000). Instructing patients to normalize fat intake and to achieve balanced nutrition.

- **Intervention focused on weight-loss and lacked any psychologically valid component.**

Wilson et al., 2010:
TheNational Institutes of Diabetes and Digestive and Kidney Diseases's Diabetes Prevention Program's manual was adapted for this study. The program includes both moderate caloric restriction and exercise. The treatment initially focused on dietary change toward a weight loss goal of 7% of one's starting weight. Participants were first asked to reduce fat intake to 25% of calories from fat. If satisfactory progress in terms of weight loss was not being made, a calorie goal was set based on initial weight. Self-monitoring of exercise, fat intake, and (if necessary) caloric intake is an essential aspect of the program. The exercise goal was 2.5 hours of moderate exercise each week. The core curriculum consists of 16 individual weekly sessions each lasting 50 minutes and followed by 4 sessions at 2-week intervals aimed at continuing weight loss and

enhancing maintenance of such losses based on the National Institutes of Diabetes and Digestive and Kidney Diseases manual for contacts after the initial core 16 sessions.

- **Intervention focused on weight-loss and lacked any psychologically valid component.**

Non-Directive

Kirkley et al., 1985:
Group discussed their food choices, eating frequency, eating rate, binge patterns, vomiting rituals, the role of stress in their bulimia, and ideas about forbidden foods but were not instructed how to alter these behaviors. Emphasis was placed on self-discovery, understanding one's bulimia, and self-disclosure.

- **No use of a manual.**
- **Self-monitoring only.**

Esplen et al., 1998:
Subjects assigned to the control group attended 1 h weekly sessions conducted by the investigator for a period of 6 weeks. Subjects were asked to keep personal journals, recording eating patterns which were used as the central component of each visit. A treatment manual was used to maintain a focus on the content of the journals during each session. The therapist} investigator assessed each day and commented on observed patterns (e.g. times of day, foods eaten, associated mood level and eating behaviors). Every effort was made to refrain from providing specific guidelines to alter eating attitudes or behavior.

- **Intervention lacked any psychologically valid component.**

Nutritional Therapy

Hsu et al., 2001:
Conducted according to a written manual, was administered by three experienced registered dieticians (who have had previous experience in counselling eating disorder patients). During the 3" # year study period, three dieticians treated the patients randomized to NT and CNT. Briefly, NT was aimed at helping the patient to understand the principles of good nutrition, her nutritional needs, and the nutritional relationship between over-restrictive eating and binge eating. In addition, the program aimed at helping the patient to establish and maintain a pattern of regular eating through meal planning, including buying and preparing healthy food.

- **Intervention focused on nutrition and lacked any psychologically valid component.**

Self-Help

Peterson et al., 1998:
In the structured self-help (SH) condition (n=15), subjects viewed the same psychoeducational videotape for the first half and led their own discussion and review of homework for the second half, based on specific discussion topics provided in writing for each session. A group member

was assigned to facilitate the discussion in each SH group on a rotating basis. Staff involvement in the SH condition was limited to turning on the videotape and collecting forms at the beginning of each session.

- **Not delivered by trained professional.**
- **No mention of a manual.**
- **Lacked any psychologically valid component.**
- **Not individualized or based on face-to-face meetings where the therapist and client develop a relationship.**

Peterson et al., 2009:
In the self-help groups, participants watched psychoeducational videotape during the first half of each session and conducted their own homework review and discussion during the second half. Participants in the self-help groups were given comprehensive instructions with detailed guidelines and time allotments for each discussion session. In addition, group members were assigned the role of lead facilitator on a rotating basis. Participants assigned to the waiting list condition received therapist-led treatment at the end of the 20-week waiting period.

- **Not delivered by trained professional.**
- **No mention of a manual.**
- **Lacked any psychologically valid component.**
- **Not individualized or based on face-to-face meetings where the therapist and client develop a relationship.**

Bailer et al., 2004:
A self-care manual for sufferers of BN that contains cognitive-behavioral educational and treatment strategies (Schmidt & Treasure, 1993) was translated into German (Schmidt & Treasure, 1996). Guided self-help was facilitated by three female first-year and second-year residents in psychiatry, none of whom had any experience with eating-disordered patients or formal psychotherapy training. Each therapist and each resident was trained to implement the forms of treatment described below. In the self-help condition, the patients received the self-help manual and were told that their progress using the manual would be reviewed. They were told that the manual contained all the skills needed to overcome their bulimia and were asked to work through the book at their own pace, to put what they had learned into practice, and to complete the exercises within the book. Appointments with residents in psychiatry were used to help and encourage the use of the self-help book and to tackle obstacles such as poor motivation, depression, and acute crisis. Patients were offered 18 weekly visits lasting no longer than 20min.

- **Not delivered by trained professional.**
- **Not individualized or based on face-to-face meetings where the therapist and client develop a relationship.**

Wilson et al., 2010:
This manualized treatment is based on Fairburn's book Overcoming Binge Eating and is performed under the guidance of a therapist. The book provides education about binge eating and a step-by-step self-help program. This intervention is derived from manual-based CBT. The primary focus is developing a regular pattern of moderate eating using self-monitoring, self-control strategies, and problem solving. Relapse prevention is emphasized to promote

maintenance of behavioral change. The principal role of the therapist is to explain the rationale for the use of the self-help manual, generate a reasonable expectancy for a successful outcome, and to motivate the patient to focus on using the manual. There were 10 treatment sessions, each lasting approximately 25 minutes, except for the first session, which was 60 minutes long. The first 4 sessions were weekly, the next 2 occurred at 2-week intervals, and the last 4 occurred at 4-week intervals. The therapists were first- or second-year graduate students with no experience in CBTgsh or treating BED, 4 at Rutgers University and 4 at Washington University. Dr. Fairburn conducted initial training in CBTgsh in a 3-hour workshop. The therapists did not receive regularly scheduled supervision. As with the other 2 treatments, quarterly meetings across sites were held throughout the study.

- **Not delivered by trained professional.**
- **Not individualized or based on face-to-face meetings where the therapist and client develop a relationship.**

Support Group/Supportive Therapy

Safer et al., 2010:
Active Comparison Group Treatment (ACGT) was developed with the goal of creating a comparison therapy whose rationale and procedures would be credible enough to generate therapeutic factors in common with DBT-BED (i.e., therapeutic alliance, treatment expectations, therapeutic optimism) while lacking the specific elements of DBT-BED and other BED treatments. Interested readers are referred to Safer and Hugo (2006) for a detailed discussion of ACGT's design. The ACGT manual was modeled after Markowitz and Sacks' (2002) manual of supportive therapy for chronic depression and subsequently modified to address binge eating for the current study. The manual instructs therapists to follow a Rogerian approach (Rogers, 1951). Self-esteem and self-efficacy are bolstered by highlighting patients' strengths (i.e., bolstering self-esteem to enhance the ability to stop binge eating). The therapy encourages patients to find answers within themselves instead of providing patients with specific techniques or skills. ACGT's ingredients (e.g., bolstering self-esteem) were intended to be indistinguishable from those evoked by the common factors of therapeutic alliance and development of therapeutic optimism.

- **Intervention focused on improving patient’s self-esteem and lacked specific psychological components in DBT and other BED treatments.**
- **Intervention is described as a placebo control.**

Hsu et al., 2001:
These support groups, based on self-help principles and therefore not led by a therapist, have been a part of our eating disorders treatment program for several years and have been quite popular. Each group session lasted about 90 min. Support groups were `open' and on average each group had six to eight patients (all females with anorexia or bulimia nervosa). For practical reasons, each group had, at any one time, one to two study patients, while the rest were non-study clinic patients. The study patients attended 14 sessions. All group sessions were audiotaped, but there was no written manual or format. A large variety of issues were brought up and discussed during these sessions, and sometimes experiential and psychodrama techniques were used.

- **Not delivered by trained professional.**
- **No mention of a manual.**

Freeman et al., 1988:
The group therapy sessions were supportive and educational in orientation. The emphasis was on mutual support and on providing information on bulimia nervosa. Sessions were semi-structured in that there was a specific topic for each week, which the therapists talked about for around 15 minutes. Group members then discussed the topic, with only non-directive intervention from the therapists. Behavioral tasks whenever possible related to the topic which had been discussed- were agreed on each week for each group member and discussed in the subsequent session.

- **Intervention focused on education and lacked any psychologically valid component.**

Supportive Clinical Management

McIntosh et al., 2005:
Nonspecific supportive clinical management was developed for the present study, and its aim was to mimic outpatient treatment that could be offered to individuals with anorexia nervosa in usual clinical practice. It combined features of clinical management and supportive psychotherapy. Clinical management includes education, care, and support and fostering a therapeutic relationship that promotes adherence to treatment. Supportive psychotherapy aims to assist the patient through use of praise, re assurance, and advice. The abnormal nutritional status and dietary patterns typical of anorexia nervosa were central to nonspecific supportive clinical management, which emphasized the resumption of normal eating and the restoration of weight (32) and provided information on weight maintenance strategies, energy requirements, and relearning to eat normally. Information was provided verbally and as written handouts. Other therapy content was dictated by the patient, with the therapist constrained to avoid specific strategies or foci of interpersonal psychotherapy or cognitive behavior therapy.

- **No mention of a manual.**
- **Supportive therapy focused on praise and advice and lacked specific strategies of interpersonal or cognitive behavioral therapy.**

Inter-rater reliability for study quality coding

The first author (R.G.) established reliability with the developers of the RCT-PQRS by coding 20 RCTs selected at random from Gerber et al. (2011). Interrater reliability (intra-class correlation coefficient [ICC]; Shrout & Fleiss, 1979) was excellent for the omnibus rating (ICC = .893). The first author (R.G.) rated all included RCTs. A second rater rated 20 RCTs (57%) selected at random. Interrater reliability was excellent for the omnibus rating (ICC = .928).

Gerber, A. J., Kocsis, J. H., Milrod, B. L., Roose, S. P., Barber, J. P., Thase, M. E., & Leon, A. C. (2011). A quality-based review of randomized controlled trials of psychodynamic psychotherapy. *American Journal of Psychiatry*, *168*, 19-28.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

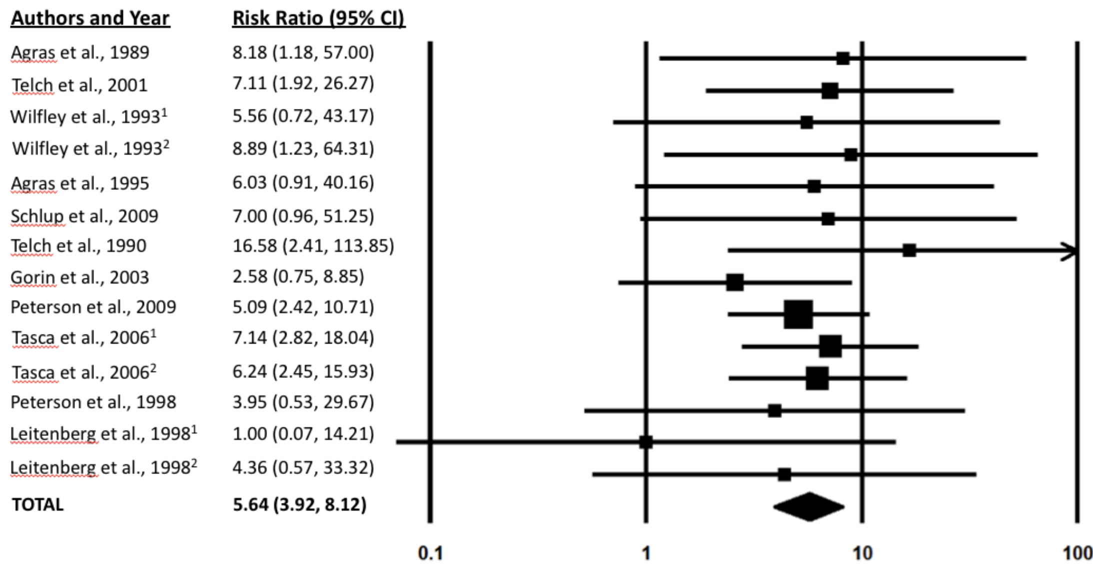
Inter-rater reliability for group psychotherapy effect size coding

Interrater reliability was assessed by comparing the ratings of Rater 1 (first author: R.G.) with those of Rater 2 (fourth author: D.S.). Both are PhD students in psychology with a background in eating disorder and psychotherapy research. Following training, Rater 1 independently coded all 27 group psychotherapy RCTs (four of these RCTs were excluded from the current meta-analysis). Rater 2 independently coded 6 of 27 (22.22%) group psychotherapy RCTs. Reliability analyses were conducted on the effect sizes extracted from the group psychotherapy RCTs and are presented below. Three groups of effect sizes were examined: (a) differences between group psychotherapy and wait-list control groups, (b) differences between group psychotherapy and other active treatments, and (c) differences between group CBT and group non-CBT. Intraclass correlation coefficients (absolute agreement) ranged from .996 to 1.00. In summary, interrater reliability was excellent. Once the independent ratings were complete, a consensus rating was reached for any disagreements between raters.

Inter-rater reliability for individual psychotherapy effect size coding

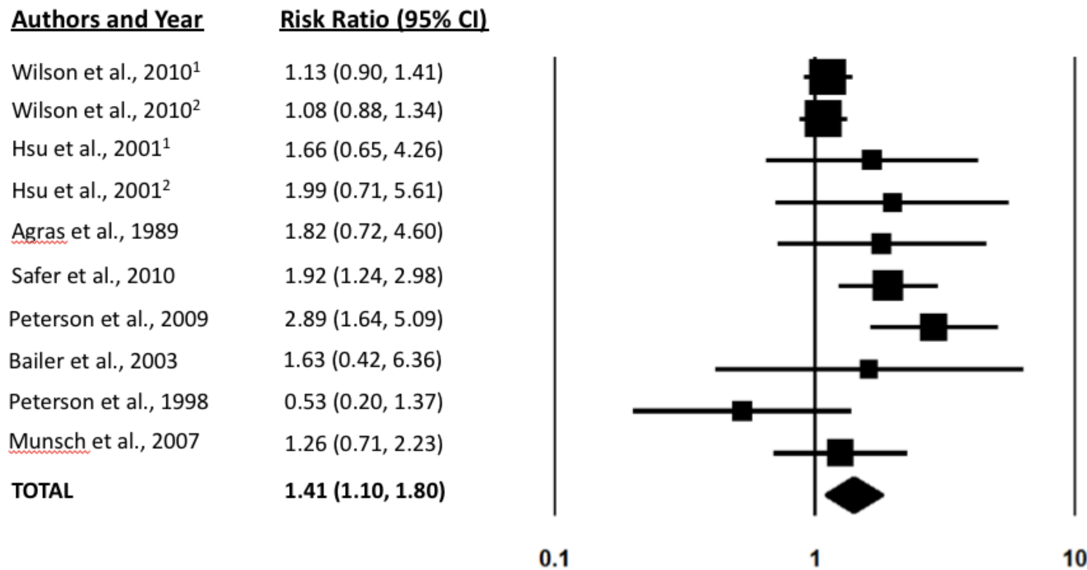
Interrater reliability was assessed by comparing the ratings of Rater 1 (first author: R.G.) with those of Rater 2 (second author: S.C.) and Rater 3 (third author: A.B.). All three raters are PhD students in psychology with a background in eating disorder and psychotherapy research. Following training, Rater 1 independently coded all 12 individual psychotherapy RCTs. Raters 2 and 3 independently coded 8 of 12 (66.67%) group psychotherapy RCTs. Reliability analyses were conducted on the effect sizes extracted from the group psychotherapy RCTs and are presented below. Three groups of effect sizes were examined: (a) differences between bona fide psychotherapy and wait-list control groups, (b) differences between bona fide psychotherapy and non-bona fide treatment, and (c) differences between bona fide CBT and bona fide non-CBT. Intraclass correlation coefficients (absolute agreement) ranged from .974 to 1.00. In summary, interrater reliability was excellent. Once the independent ratings were complete, a consensus rating was reached for any disagreements between raters.

Forest Plot for Abstinence: Bona Fide Psychotherapy vs. Wait-List



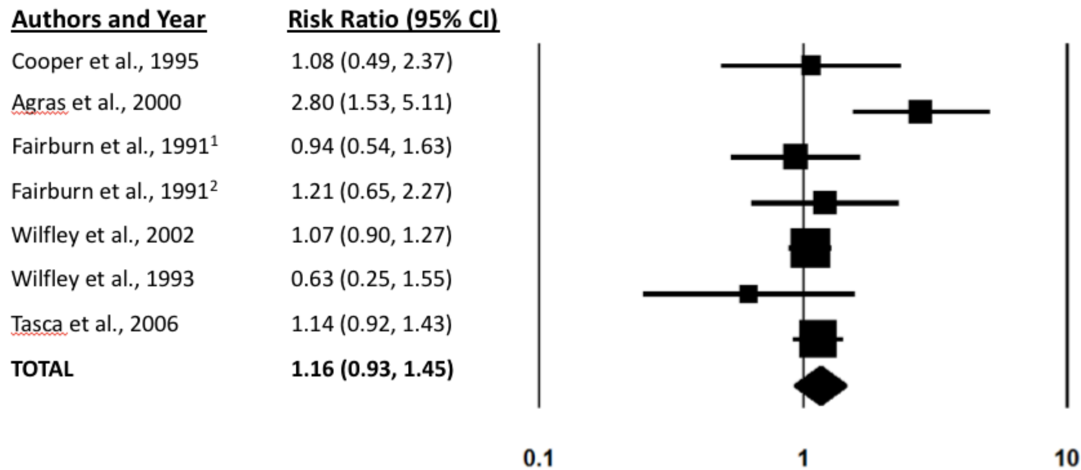
Forest plot of between-group effect sizes of abstinence from binge eating and/or purging of bona fide psychotherapy as compared to wait-list control at post-treatment. The squares represent the effect sizes for each study, the size of the square is the relative weighting of the study in the analysis. Error bars indicate a 95% confidence interval. CI: confidence interval; Wilfley et al., 1993[1]: CBT vs. wait-list; Wilfley et al., 1993[2]: IPT vs. wait-list; Tasca et al., 2006[1]: CBT vs. wait-list; Tasca et al., 2006[2]: group psychodynamic interpersonal psychotherapy vs. wait-list; Leitenberg et al., 1998[1]: CBT vs. wait-list; Leitenberg et al., 1998[2]: CBT+emotion-response-prevention vs. wait-list. Note: CBT: cognitive-behavioural therapy; IPT: interpersonal psychotherapy.

Forest Plot for Abstinence: Bona Fide Psychotherapy vs. Non-Bona Fide Treatment



Forest plot of between-group effect sizes of abstinence from binge eating and/or purging of bona fide psychotherapy as compared to non-bona fide treatment at post-treatment. The squares represent the effect sizes for each study, the size of the square the relative weighting of the study in the analysis. Error bars indicate a 95% confidence interval. CI: confidence interval; Wilson et al., 2010[1]: IPT vs. BWL; Wilson et al., 2010[2]: IPT vs. self-help; Hsu et al., 2001[1]: CT vs. support group; Hsu et al., 2001[2]: CT vs. nutritional therapy. Note: IPT: interpersonal psychotherapy; BWL: behavioural weight-loss; CT: cognitive therapy.

Forest Plot for Abstinence: Bona Fide CBT vs. Bona Fide Non-CBT



Forest plot of between-group effect sizes of abstinence from binge eating and/or purging of bona fide CBT as compared to bona fide non-CBT at post-treatment. The squares represent the effect sizes for each study, the size of the square the relative weighting of the study in the analysis. Error bars indicate a 95% confidence interval. CI: confidence interval; Fairburn et al., 1991[1]: CBT vs. behaviour therapy; Fairburn et al., 1991[2]: CBT vs. interpersonal psychotherapy. Note: CBT: cognitive-behavioural therapy.

APPENDIX B

Study Two

Supplementary Materials: For Online Publication Only

Search terms

MEDLINE

Diagnosis	(eat* disord*[Title/Abstract] OR anorexi*[Title/Abstract] OR bulimi*[Title/Abstract] OR binge eat*[Title/Abstract] OR EDNOS[Title/Abstract] OR purg* disord*[Title/Abstract]).
AND	
Intervention	(group CBT[Title/Abstract] OR group intervention[Title/Abstract] OR group treatment[Title/Abstract] OR group psychotherap*[Title/Abstract] OR group therap*[Title/Abstract] OR group counseling*[Title/Abstract] OR group format[Title/Abstract] OR group program*[Title/Abstract] OR group psychoanaly*[Title/Abstract] OR group-focused[Title/Abstract] OR group-centered[Title/Abstract] OR group-delivered[Title/Abstract] OR CBGT[Title/Abstract] OR group intervention[Title/Abstract] OR group treatment[Title/Abstract] OR group cognitive behavioral[Title/Abstract] OR group behavioral intervention[Title/Abstract] OR group psychodynamic[Title/Abstract])
AND	
Design	(random*[Title/Abstract] OR assign*[Title/Abstract] OR allocat*[Title/Abstract] OR trial*[Title/Abstract] OR control group[Title/Abstract] OR experimental design[Title/Abstract] OR placebo[Title/Abstract] OR treatment effectiveness[Title/Abstract] OR treatment efficacy[Title/Abstract] OR treatment outcomes[Title/Abstract])

COCHRANE CENTRAL REGISTER OF CONTROLLED TRIALS

Diagnosis	(eat* disord* or anorexi* or bulimi* or binge eat* or EDNOS or purg* disord*):ti,ab,kw
AND	
Intervention	(group next (treatment or intervention or setting or session) or group near/3 (therap* or psychotherap* or psychoanaly* or cognitive behav* therap* or CBT or training or format or exposure or program or counseling or approach or support*) or "group based" or "group focused" or "group centered" or "group delivered" or CBGT or "group vs individual" or "group versus individual"):ti,ab,kw
Design	
	Only includes RCTs.

PsycINFO

Diagnosis	TI(eat* disord* OR anorexi* OR bulimi* OR binge eat* OR EDNOS* OR purg* disord*) OR AB(eat* disord* OR anorexi* OR bulimi* OR binge eat* OR EDNOS* OR purg* disord*)
AND	
Intervention	TI("group treatment" OR "group intervention" OR "group setting" OR "group therap*" OR "group psychotherap*" OR "group psychoanaly*" OR "group cognitive behav* therap*" OR "group CBT" OR "group counselling" OR "group focused" OR "group centred" OR "group delivered" OR "CBGT") OR AB(group treatment OR group intervention OR group setting OR group therap* OR group psychotherap* OR group psychoanaly* OR group cognitive behav* therap* OR group CBT OR group counseling OR group focused OR group centred OR group delivered OR CBGT)
AND	
Design	TI(random* assign* OR random* allocat* OR trial* OR control group OR experimental design OR treatment effectiveness OR treatment efficacy OR treatment outcomes) OR AB(random* OR assign* OR allocat* OR trial* OR control group OR experimental design OR placebo OR treatment effectiveness OR treatment efficacy OR treatment outcomes)

References of included primary studies

Agras WS, Telch CF, Arnow B, Eldredge K, Detzer MJ, Henderson J, Marnell M. Does interpersonal therapy help patients with binge eating disorder who fail to respond to cognitive-behavioral therapy? J Consult Clin Psychol 1995;63:356-360.

Agras WS, Telch CF, Arnow B, Eldredge K, Wilfley DE, Raeburn SD, Henderson J, Marnell M. Weight loss, cognitive- behavioral, and desipramine treatments in binge eating disorder: An additive design. Behav Ther 1994;25:225-238.

Allen HN. Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder (doctoral dissertation). University of North Carolina at Chapel Hill 1997.

Bailer U, de Zwaan M, Leisch F, Strnad A, Lennkh-Wolfsberg C, El-Giamal N, Hornik K, Kasper S. Guided self-help versus cognitive-behavioral group therapy in the treatment of bulimia nervosa. Int J Eat Disord 2004;35:522-537.

Chen E, Touyz SW, Beumont PJ, Fairburn CG, Griffiths R, Butow P, Russell J, Schotte DE, Gertler R, Basten C. Comparison of group and individual cognitive-behavioral therapy for patients with bulimia nervosa. Int J Eat Disord 2003;33:241-254.

Gorin AA, Le Grange D, Stone AA. Effectiveness of spouse involvement in cognitive behavioral therapy for binge eating disorder. Int J Eat Disord 2003;33:421-433.

Grilo CM, Masheb RM, Wilson GT, Gueorguieva R, White MA. Cognitive-behavioral therapy, behavioral weight loss, and sequential treatment for obese patients with binge-eating disorder: a randomized controlled trial. J Consult Clin Psychol 2011;79:675-685.

Jacobi C, Dahme B, Dittmann R. Cognitive-behavioural, fluoxetine and combined treatment for bulimia nervosa: Short- and long-term results. Eur Eat Disord Rev 2002;10:179-198.

Kirkley BG, Schneider JA, Agras W, Bachman JA. (1985). Comparison of two group treatments for bulimia. J Consult Clin Psychol 1985;53:43-48.

Laessle RG, Waadt SS, Pirke KM. (1987). A structured behaviorally oriented group treatment for bulimia nervosa. Psychother Psychosom 1987;48:141-145.

Lavender A, Startup H, Naumann U, Samarawickrema N, DeJong H, Kenyon M, van den Eynde F, Schmidt U. Emotional and social mind training: A randomised controlled trial of a new group-based treatment for bulimia nervosa. PLoS ONE 2012;7:1-9.

Lee NF, Rush AJ. Cognitive-behavioural group therapy for bulimia. Int J Eat Disord 1997;5:599–615.

Leitenberg H, Rosen JC, Gross J, Nudelman S, Vara LS. Exposure plus response-prevention treatment of bulimia nervosa. J Consult Clin Psychol 1988;56:535-541.

Munsch S, Biedert E, Meyer A, Michael T, Schlup B, Tuch A, Margraf J. A randomized comparison of cognitive behavioral therapy and behavioral weight loss treatment for overweight individuals with binge eating disorder. Int J Eat Disord 2007;40:102-113.

Nevonen L, Broberg AG. A comparison of sequenced individual and group psychotherapy for eating disorder not otherwise specified. Eur Eat Disord Rev 2005;13:29-37.

Nevonen L, Broberg AG. A comparison of sequenced individual and group psychotherapy for patients with bulimia nervosa. Int J Eat Disord 2006;39:117-127.

Peterson CB, Mitchell JE, Crow SJ, Crosby RD, Wonderlich SA. The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. Am J Psychiatry 2009;166:1347-1354.

Peterson CB, Mitchell JE, Engbloom S, Nugent S, Mussell MP, Miller JP. Group cognitive-behavioral treatment of binge-eating disorder: A comparison of therapist-led versus self-help formats. Int J Eat Disord 1998;24:125–136.

Safer DL, Robinson AH, Jo B. Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. Behav Ther 2010;41:106-120.

Schlup B, Munsch S, Meyer AH, Margraf J, Wilhelm FH. The efficacy of a short version of a cognitive-behavioral treatment followed by booster sessions for binge eating disorder. Behav Res Ther 2009;47:628-635.

Tasca GA, Ritchie K, Conrad G, Balfour L, Gayton J, Lybanon V, Bissada H. Attachment scales predict outcome in a randomized controlled trial of two group therapies for binge eating disorder: An aptitude by treatment interaction. Psychother Res 2006;16:106-121.

Telch CF, Agras WS, Linehan MM. Dialectical behavior therapy for binge eating disorder. J Consult Clini Psychol 2001:69;1061-1065.

Telch CF, Agras WS, Rossiter EM, Wilfley D, Kenardy J. Group cognitive-behavioral treatment for the nonpurging bulimic: An initial evaluation. J Consult Clin Psychol 1990;58:629-635.

Vocks S, Schulte D, Busch M, Grönemeyer D, Herpertz S, Suchan B. Changes in neuronal correlates of body image processing by means of cognitive-behavioural body image therapy for eating disorders: a randomized controlled fMRI study. Psychol Med 2011;41:1651-1663.

Wilfley DE, Agras WS, Telch CF, Rossiter EM, Schneider JA, Cole Golomb A, Sifford L, Raeburn SD. Group cognitive-behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A Controlled Comparison. J Consult Clin Psychol 1993;61:296-305.

Wilfley DE, Welch RR, Stein RI, Spurrell EB, Cohen LR, Saelens BE, Dounchis JZ, Frank MA, Wiseman CV, Matt GE. A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the tretament fo overweight individuals with binge-eating disorder. Arch Gen Psychiatry 2002;59:713-721.

Wolf EM, Crowther JH. An evaluation of behavioral and cognitive-behavioral group interventions for the treatment of bulimia nervosa in women. Int J Eat Disord 1992;11:3-15.

APPENDIX C

Study Three

Supplementary Material

References of included primary studies:

Agras WS, Schneider JA, Arnow, B, Raeburn, SD, Telch, CF. Cognitive-Behavioral and Response-Prevention Treatments for Bulimia Nervosa. J Consult Clin Psych 1989;57:215-221.

Agras WS, Telch CF, Arnow B, Eldredge K, Detzer MJ, Henderson J, Marnell M. Does interpersonal therapy help patients with binge eating disorder who fail to respond to cognitive-behavioral therapy? J Consult Clin Psychol 1995;63:356-360.

Agras WS, Telch CF, Arnow B, Eldredge K, Wilfley DE, Raeburn SD, Henderson J, Marnell M. Weight loss, cognitive- behavioral, and desipramine treatments in binge eating disorder: An additive design. Behav Ther 1994;25:225-238.

Agras WS, Walsh T, Fairburn CG, Wilson GT, Kraemer HC. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. Arch Gen Psychiatry 2000;57:459-466.

Allen HN. Appetite awareness training: A cognitive-behavioral intervention for binge eating disorder (doctoral dissertation). University of North Carolina at Chapel Hill 1997.

Bailer U, de Zwaan M, Leisch F, Strnad A, Lennkh-Wolfsberg C, El-Giamal N, Hornik K, Kasper S. Guided self-help versus cognitive-behavioral group therapy in the treatment of bulimia nervosa. Int J Eat Disord 2004;35:522-537.

Cooper PJ, Steere J. A comparison of two psychological treatments for bulimia nervosa: Implications for models of maintenance. Behav Res Ther 1995;33:875-885.

Esplen MJ, Garfinkel PE, Olmstead M, Gallop RM, Kennedy S. A randomized controlled trial of guided imagery in bulimia nervosa. Psychological Medicine 1998;28:1347-1357.

Fairburn CG, Bailey-Straebler S, Basden S, Doll HA, Jones R, Murphy R, O'Connor ME, Cooper Z. A transdiagnostic comparison of enhanced cognitive behavior therapy (CBTE) and interpersonal psychotherapy in the treatment of eating disorders. Behav Res Ther 2015;70:64-71.

Fairburn CG, Jones R, Peveler RC, Carr SJ, Solomon RA, O'Connor ME, Burton J, Hope RA. Three psychological treatments for bulimia nervosa. Arch Gen Psychiatry 1991;48:463-469.

Fairburn CG, Kirk J, O'Connor M. A comparison of two psychological treatments for bulimia nervosa. Behav Res Ther 1986;24:629-643.

Freeman CPL, Barry F, Dunkeld-Turnbull J, Henderson A. Controlled trial of psychotherapy for bulimia nervosa. British Medical Journal 1988; 296: 521-525.

Gorin AA, Le Grange D, Stone AA. Effectiveness of spouse involvement in cognitive behavioral therapy for binge eating disorder. Int J Eat Disord 2003;33:421-433.

Griffiths RA, Hadzi-Pavlovic D, Channon-Little L. A controlled evaluation of hypnobehavioural treatment for bulimia nervosa: Immediate pre-post treatment effects. Eur Eat Disord Rev 1994;2:202-220.

Grilo CM, Masheb RM, Wilson GT, Gueorguieva R, White MA. Cognitive-behavioral therapy, behavioral weight loss, and sequential treatment for obese patients with bingeeating disorder: a randomized controlled trial. J Consult Clin Psychol 2011;79:675-685.

Hsu LKG, Rand W, Sullivan S, Liu DW, Mulliken B, McDonagh B, Kaye WH. Cognitive therapy, nutritional therapy and their combination in the treatment of bulimia nervosa. Psychological Medicine 2001;31:871-879.

Kirkley BG, Schneider JA, Agras W, Bachman JA. (1985). Comparison of two group treatments for bulimia. J Consult Clin Psychol 1985;53:43-48.

Laessle RG, Waadt SS, Pirke KM. (1987). A structured behaviorally oriented group treatment for bulimia nervosa. Psychother Psychosom 1987;48:141-145.

Lavender A, Startup H, Naumann U, Samarawickrema N, DeJong H, Kenyon M, van den Eynde F, Schmidt U. Emotional and social mind training: A randomised controlled trial of a new group-based treatment for bulimia nervosa. PLoS ONE 2012;7:1-9.

Lee NF, Rush AJ. Cognitive-behavioural group therapy for bulimia. Int J Eat Disord 1997;5:599–615.

Leitenberg H, Rosen JC, Gross J, Nudelman S, Vara LS. Exposure plus responseprevention treatment of bulimia nervosa. J Consult Clin Psychol 1988;56:535-541.

McIntosh VVW, Jordan J, Carter FA, Luty SE, McKenzie JM, Bulik CM, Frampton CMA, Joyce PR. Three psychotherapies for anorexia nervosa: A randomized, controlled trial. Am J Psychiatry 2005;162:741-747.

Munsch S, Biedert E, Meyer A, Michael T, Schlup B, Tuch A, Margraf J. A randomized comparison of cognitive behavioral therapy and behavioral weight loss treatment for overweight individuals with binge eating disorder. Int J Eat Disord 2007;40:102-113.

Peterson CB, Mitchell JE, Crow SJ, Crosby RD, Wonderlich SA. The efficacy of selfhelp

group treatment and therapist-led group treatment for binge eating disorder. Am J Psychiatry 2009;166:1347-1354.

Peterson CB, Mitchell JE, Engbloom S, Nugent S, Mussell MP, Miller JP. Group cognitive-behavioral treatment of binge-eating disorder: A comparison of therapist-led versus self-help formats. Int J Eat Disord 1998;24:125–136.

Safer DL, Robinson AH, Jo B. Outcome from a randomized controlled trial of group therapy for binge eating disorder: Comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. Behav Ther 2010;41:106-120.

Schlup B, Munsch S, Meyer AH, Margraf J, Wilhelm FH. The efficacy of a short version of a cognitive-behavioral treatment followed by booster sessions for binge eating disorder. Behav Res Ther 2009;47:628-635.

Tasca GA, Ritchie K, Conrad G, Balfour L, Gayton J, Lybanon V, Bissada H. Attachment scales predict outcome in a randomized controlled trial of two group therapies for binge eating disorder: An aptitude by treatment interaction. Psychother Res 2006;16:106-121.

Telch CF, Agras WS, Linehan MM. Dialectical behavior therapy for binge eating disorder. J Consult Clini Psychol 2001:69;1061-1065.

Telch CF, Agras WS, Rossiter EM, Wilfley D, Kenardy J. Group cognitive-behavioral treatment for the nonpurging bulimic: An initial evaluation. J Consult Clin Psychol 1990;58:629-635.

Vocks S, Schulte D, Busch M, Grönemeyer D, Herpertz S, Suchan B. Changes in neuronal correlates of body image processing by means of cognitive-behavioural body image therapy for eating disorders: a randomized controlled fMRI study. Psychol Med 2011;41:1651-1663.

Wilfley DE, Agras WS, Telch CF, Rossiter EM, Schneider JA, Cole Golomb A, Sifford L, Raeburn SD. Group cognitive-behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A Controlled Comparison. J Consult Clin Psychol 1993;61:296-305.

Wilfley DE, Welch RR, Stein RI, Spurrell EB, Cohen LR, Saelens BE, Dounchis JZ, Frank MA, Wiseman CV, Matt GE. A randomized comparison of group cognitivebehavioral therapy and group interpersonal psychotherapy for the tretament fo overweight individuals with binge-eating disorder. Arch Gen Psychiatry 2002;59:713-721.

Wilson GT, Wilfley DE, Agras WS, Bryson SW. Psychological Treatments for Binge Eating Disorder. Arch Gen Psychiatry 2010;67:94-101.

Wolf EM, Crowther JH. An evaluation of behavioral and cognitive-behavioral group interventions for the treatment of bulimia nervosa in women. Int J Eat Disord 1992;11:3-15.

The Randomized Controlled Trial of Psychotherapy Quality Rating Scale (RCT-PQRS)

Please rate all items on the basis of the designated paper(s) describing the study. Items #4, #5, #9 and #16 specifically rate the description of certain elements of the study. All other items are designed to capture both description and quality of the study's elements. For these items, when non-standard elements are described, adequate justification of this measure or method is required to score a 2.

Description of subjects

Item #1 Diagnostic method and criteria for inclusion and exclusion
0 = poor description and inappropriate method/criteria
1 = full description or appropriate method/criteria
2 = full description and appropriate method/criteria

Item #2 Documentation or demonstration of reliability of diagnostic methodology
0 = poor or no reliability documentation
1 = brief reliability documentation (documentation in the literature is sufficient, even if it is not explicitly cited)
2 = full reliability documentation (documentation of within-study reliability necessary)

Item #3 Description of relevant comorbidities
0 = poor or no description of relevant comorbidities
1 = brief description of relevant comorbidities
2 = full description of relevant comorbidities

Item #4 Description of numbers of subjects screened, included, and excluded
0 = poor or no description of numbers screened, included, and excluded
1 = brief description of numbers screened, included, and excluded
2 = full description of numbers screened, included, and excluded

Definition and delivery of treatment

Item #5 Treatment(s) (including control/comparison groups) are sufficiently described or referenced to allow for replication
0 = poor or no treatment description or references
1 = brief treatment description or references (also if full description of one group and poor description of another)
2 = full treatment description or references (manual not required)

Item #6 Method to demonstrate that treatment being studied is treatment being delivered (only satisfied by supervision if transcripts or tapes are explicitly reviewed)
0 = poor or no adherence reporting
1 = brief adherence reporting with standardized measure or
full adherence reporting with non-standardized measure (e.g., non-independent rater)

2 = full adherence reporting with standardized measure (must be quantitative and completed by an independent rater)

Item #7 Therapist training and level of experience in the treatment(s) under investigation
0 = poor description and under-qualified therapists
1 = full description or well-qualified therapists
2 = full description and well-qualified therapists

Item #8 Therapist supervision while treatment is being provided
0 = poor description and inadequate therapist supervision
1 = full description or adequate therapist supervision
2 = full description and adequate therapist supervision

Item #9 Description of concurrent treatments (e.g., medication) allowed and administered during course of study (if patients on medication are included, a rating of 2 requires full reporting of what medications were used; if patients on medications are excluded, this alone is sufficient for a rating of 2).
0 = poor or no description of concurrent treatments
1 = brief description of concurrent treatments
2 = full description of concurrent treatments

Outcome measures

Item #10 Validated outcome measure(s) (either established or newly standardized)
0 = poor or no validation of outcome measure(s)
1 = brief validation of outcome measure(s) (shown or cited)
2 = full validation of outcome measure(s) (shown or cited)

Item #11 Primary outcome measure(s) specified in advance (though does not need to be stated explicitly for a rating of 2)
0 = poor or no specification of primary outcome measure(s) in advance
1 = brief specification of primary outcome measure(s) in advance
2 = full specification of primary outcome measure(s) in advance

Item #12 Outcome assessment by raters blinded to treatment group and with established reliability
0 = poor or no blinding of raters to treatment group (e.g., rating by therapist, non-blind independent rater, or patient self-report) and reliability not reported
1 = blinding of independent raters to treatment group or established reliability
2 = blinding of independent raters to treatment group and established reliability

Item #13 Discussion of safety and adverse events during study treatment(s)
0 = poor or no discussion of safety and adverse events
1 = brief discussion of safety and adverse events
2 = full discussion of safety and adverse events

Item #14 Assessment of long-term post-termination outcome (should not be penalized for failure to follow comparison group if this is a wait-list or non-treatment group that is subsequently referred for active treatment)
0 = poor or no post-termination assessment of outcome
1 = medium-term assessment of post-termination outcome (2 to 12 months post-termination)
2 = long-term assessment of post-termination outcome (greater or equal to 12 months post termination)

Data analysis

Item #15 Intent-to-treat method for data analysis involving primary outcome measure
0 = no description or no intent-to-treat analysis with primary outcome measure
1 = partial intent-to-treat analysis with primary outcome measure
2 = full intent-to-treat analysis with primary outcome measure

Item #16 Description of dropouts and withdrawals
0 = poor or no description of dropouts and withdrawals
1 = brief description of dropouts and withdrawals
2 = full description of dropouts and withdrawals (must be explicitly stated and include reasons for dropouts and withdrawals)

Item #17 Appropriate statistical tests (e.g., use of Bonferroni correction, longitudinal data analysis, adjustment only for a priori identified confounders)
0 = inappropriate statistics, extensive data dredging, or no information about appropriateness of statistics
1 = moderately appropriate, though unsophisticated, statistics and/or moderate data dredging
2 = fully appropriate statistics and minimal data dredging in primary findings

Item #18 Adequate sample size
0 = inadequate justification and inadequate sample size
1 = adequate justification or adequate sample size
2 = adequate justification and adequate sample size

Item #19 Appropriate consideration of therapist and site effects
0 = therapist and site effects not discussed or considered
1 = therapist and site effects discussed or considered statistically
2 = therapist and site effects discussed and considered statistically

Treatment assignment

Item #20 A priori relevant hypotheses that justify comparison group(s)
0 = poor or no justification of comparison group(s)
1 = brief or incomplete justification of comparison group(s)
2 = full justification of comparison group(s)

Item #21 Comparison group(s) from same population and time-frame as experimental group

0 = comparison group(s) from significantly different population and/or time-frame
1 = comparison group(s) from moderately different population and/or time frame
2 = comparison group(s) from same population and time-frame

Item #22 Randomized assignment to treatment groups
0 = poor (e.g., pseudo-randomization, sequential assignment) or no randomization
1 = adequate but poorly defined randomization procedure
2 = full and appropriate method of randomization performed after screening and baseline assessment

Overall quality of study

Item #23 Balance of allegiance to types of treatment by practitioners
0 = no information or poor balance of allegiance to treatments by study therapists (e.g., therapy in experimental and control groups both administered by therapists with strong allegiance to therapy being tested in the experimental group)
1 = some balance of allegiance to treatments by study therapists
2 = full balance of allegiance to treatments (e.g., therapies administered by therapists with allegiance to respective techniques)

Item #24 Conclusions of study justified by sample, measures, and data analysis, as presented (note: useful to look at conclusions as stated in study abstract)
0 = poor or no justification of conclusions from results as presented or insufficient information to evaluate (e.g., sample or treatment insufficiently documented, data analysis does not support conclusions, or numbers of withdrawals or dropouts makes findings unsupportable)
1 = some conclusions of study justified or partial information presented to evaluate
2 = all conclusions of study justified and complete information presented to evaluate

Omnibus Rating

Please provide an overall rating of the quality of the study taking into account the adequacy of description, the quality of study design, data analysis, and justification of conclusions
1 = exceptionally poor
2 = very poor
3 = moderately poor
4 = average
5 = moderately good
6 = very good
7 = exceptionally good

Quality Score (sum items 1-24) __________

Omnibus Score __________

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., & Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry*, *51*, 319-324. doi:10.1016/j.comppsych.2009.07.001

Excluded studies

AUTHORS	YEAR	REASON FOR EXCLUSION
		NOT PRIMARY RCT (i.e., protocol, follow-up study)
Accurso et al.	2015	
Accurso et al.	2016	
Agras et al.	1994	
Carter et al.	2003	
Carter et al.	2011	
Daniel et al.	2016	
Elbaky et al.	2014	
Esplen et al.	1998	
Fairburn et al.	1993	
Fairburn et al.	1993	
Fairburn et al.	1995	
Fischer et al.	2014	
Griffiths et al.	1996	
Jones et al.	1993	
Schmidt et al.	2016	
		STUDY DESIGN (i.e., non-RCT, medication/hospitalization during treatment or follow-up/ intervention lengths not equal, dismantling)
Bachar et al.	1999	
Balestrieri et al.	2015	
Barga et al.	2004	
Brambilla et al.	2014	
Bulik et al.	1998	
Byrne et al.	2017	
Channon et al.	1989	
Ciano et al.	2002	
Davis et al.	1997	
Davis et al.	1999	
Dicker	2003	
Fairburn et al.	2009	
Garner et al.	1993	
Ghaderi et al.	2006	
Hilbert et al.	2004	
Jager et al.	1997	
Le Grange et al.	2002	
Leitenberg et al.	1998	
Loeb et al.	2000	
Olmstead et al.	1991	
Painot et al.	2001	

Pendleton et al.	2002	
Richter-Reno et al.	1992	
Russel et al.	1987	
Schmidt et al.	1989	
Schmidt et al.	2012	
Schmidt et al.	2013	
Schmidt et al.	2015	
Schutzmann et al.	2009	
Tantillo et al.	2003	
Thiels et al.	1998	
Touyz et al.	2013	
Treasure et al.	1999	
Walsh et al.	1997	
Weinstein et al.	1994	
Wilson et al.	1986	
Wilson et al.	1991	
		TREATMENT (i.e., not individual psychotherapy, no intervention that is solely psychotherapy)
Alfonsson et al.	2015	
Dingemans et al.	2007	
Grilo et al.	2005	
Grilo et al.	2012	
Katzman et al.	2010	
Laessle et al.	1991	
Ordman et al.	1985	
Shag et al.	2015	
Stice et al.	2015	
		PARTICIPANTS (i.e., not adult, subthreshold)
Cassin et al.	2008	
Freeman et al.	1985	
Goodrick et al.	1998	
Hall et al.	1987	
Nauta et al.	2000	
Raynaud et al.	1999	
Safer et al.	2001	
Schmidt et al.	2007	
Serfaty et al.	1999	
Stein et al.	2013	
Thackwray et al.	1993	
Wild et al.	2009	
Wonderlich et al.	2014	
Zipfel et al.	2014	
		OUTCOMES/DATA

		(i.e., outcomes not clear, unable to use data provided)
Dare et al.	2001	
Gowers et al.	1994	
Leitenberg et al.	1994	
Sundgot-Borgen et al.	2002	
Treasure et al.	1995	