How Many Therapists? Practical Guidance on Investigating Therapist Effects in Randomized Controlled Trials for Eating Disorders

Doug Thompson, PhD[1,*], **Fary Cachelin, PhD**[2], **Ruth H. Striegel-Moore, PhD**[3], **Bruce Barton, PhD**[4], **Munyi Shea, PhD**[5], and **G. Terence Wilson, PhD**[6]

[1]Thompson Research Consulting LLC, Chicago, Illinois

[2]Department of Psychology, University of North Carolina, Charlotte, North Carolina

[3]Department of Psychology, Wesleyan University, Middletown, Connecticut

[4]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts

[5]Department of Psychology, California State University, Los Angeles, California

[6]Graduate School of Applied and Professional Psychology, Rutgers, the State University of New Jersey, New Jersey

Abstract

Objective—An important question in implementation/dissemination research is whether the efficacy of a given treatment varies in part based on the therapist delivering the treatment. This study sought to provide practical guidance to researchers in the field of eating disorders for building measurement of therapist effects into the design of a typical, relatively small randomized controlled trial (RCT).

Method—Using assumptions based on past trials of eating disorder treatments, Monte Carlo simulations were used to examine 12 different scenarios based on crossing the number of therapists (between two and five) and the estimated therapist effect size (small, medium, and large). Patient sample size and study design were held constant.

Results—There was reasonable power (70%) to detect the therapist effect with three or four therapists and a large effect size.

Discussion—Several practical implications for testing therapist effects in RCT are discussed.

Keywords

therapist effects; treatment outcome; randomized clinical trial

Introduction

In clinical psychology, a therapist effect is in evidence when treatment outcomes differ among therapists administering the same treatment. Past work has shown mixed results regarding therapist effects. The majority of studies examining the issue have found minimal

*Correspondence to: Doug Thompson, Thompson Research Consulting LLC, 151 N. Michigan #3308, Chicago, IL 60601. Doug.thompson@ThompsonResearchConsulting.com.

or no therapist effects[1–4] but some studies have reported significant therapist effects.[5–7] Evidence suggests that therapist effects are less likely when treatments are highly standardized (e.g., manual-based treatments) or therapists are highly skilled, experienced, and well trained.[3,4]

Most randomized controlled trials (RCTs) in clinical psychology have not been designed to have sufficient power to detect therapist effects. Also, the majority of RCTs have ignored therapist effects in the analysis.[4] These may be among the reasons why therapist effects have seldom been reported. Lack of attention to therapist effects is unfortunate, because therapist effects can have important implications. If treatment outcomes differ significantly among therapists, this points to barriers in generalizing the treatment to other settings—extensive training, clinical aptitude, or experience might be required to administer the treatment effectively. On the other hand, if there is no intertherapist variability in outcomes, this may point to a treatment that can be readily and effectively administered by a variety of therapists. Thus, therapist effects can have profound implications for the generalizability of trial results.

Although large trials offer the opportunity for a more definitive study of therapist effects, smaller trials are typical in the eating disorders field and may remain so in a limited funding climate. For example, one compilation of published, RCTs on binge eating disorder (BED)/ bulimia nervosa (BN) included 12 separate trials, with sample sizes ranging from 50 to 204 and an average sample size of 110.[8] For this reason, the present article focuses on approaches to studying therapist effects in small trials (200 or fewer participants, including both treatment and control/comparison groups).

The typical small sample size of RCTs in the eating disorders field limits the analyses and inferences that the results can support. A key consideration is whether therapist effects will be analyzed as fixed or random. In a fixed effects analysis, differences among the therapists in the trial are estimated, without attempting to generalize the difference among therapists in the trial to a larger population of therapists. In a random effects analysis, on the other hand, the therapists in the trial are regarded as a sample from a larger population and variation among therapists in the trial (sample) is used to estimate variation among therapists in the population. Major differences between the fixed and random effects approaches are summarized in Table 1.

For a definitive study of therapist effects, a random effects analysis is ideal, because it enables generalization to a broader population of therapists and it enables modeling of potentially important features of the data (heteroscedasticity and intra-therapist correlation). However, a random effects analysis may not be feasible in small RCTs. Brown and Prescott[9] recommended that there be at least five clusters for a random effects analysis (in this article, therapists are the clusters). Although this is only a rule of thumb, it has face validity—one would hesitate to generalize to a larger population of therapists after having observed fewer than five or so. Typically, one would want to observe more than five therapists before feeling comfortable about generalizing to a larger population. Another problem is that selection of therapists for an RCT often violates the assumption of representative (ideally random) sampling from a larger population required for a random effects analysis. Therapists in an RCT may be carefully selected from a limited group of therapists, for example, graduate students at a single university. Therapists may be selected based on their interest in the study, their aptitude, and their perceived ability to administer the treatment effectively. Given such selection criteria, it may not even be clear what larger population of therapists is appropriate for inference. Finally, based on the small sample size of patients in a typical RCT on eating disorders, the data may not be able to support the complex modeling of variance structures involved in a random effects analysis (heteroscedasticity,

intratherapist correlation). The variance estimates could easily be influenced by outliers and leverage points; thus, complex modeling of variance structures could be misleading rather than informative. For these reasons, it is arguable that a fixed effects approach may be the only feasible option in many small RCTs, despite the limitations of this approach.

The methodological literature on therapist effects has focused on random effects analyses,[4,10,11] while relatively little attention has been given to the fixed effects approach. An exception was Siemer and Joormann,[12] who contrasted random versus fixed therapist effects and recommended the latter due to greater power to detect therapist effects. In response, Serlin et al.[13] argued that this power is bought at the price of severe limitations to generalizability of the study findings. We agree with Serlin et al. that a random effects approach, when feasible, is needed for a conclusive study of therapist effects; power does not justify the choice of a fixed effects approach. However, the results of a random effects analysis for analyzing therapist effects are likely to be weakly supported and perhaps misleading in a typical, small RCT in the eating disorders field; therefore, this article focuses on the fixed effects approach, while acknowledging the limitations.

The main objective of this article is to provide practical guidance to researchers who are designing and analyzing RCTs on treatments for eating disorders (BN/BED), highlighting upsides and downsides of various approaches to investigating therapist effects. Attention is given to practical as well as design and analysis considerations. Among the key questions addressed are, if the investigator intends to estimate therapist effects in a typical, small RCT, what is the optimal number of therapists to employ? What are the implications of using different numbers of therapists?

To quantify the effect of varying the number of therapists, we estimate the power to detect therapist effects under several scenarios in a typical, small trial of BED/BN treatment. The assumed design involves a single treatment, administered by multiple therapists, compared with no-treatment control; this is a frequently used design in RCTs in eating disorders.[14–17] Monte Carlo simulations were used to estimate power to detect therapist effects, varying the number of therapists and the size of the therapist effect, while holding constant the study design and patient sample size.

Method

Power to estimate therapist effects under some specific scenarios is illustrated in the context of a RCT of guided self-help (GSH) compared with a no-treatment control, for the treatment of BED and BN. The primary outcome was the count of episodes of binge eating during the past month. This outcome is assumed to be measured at baseline and post-treatment (i.e., after 12 weeks of treatment). The primary hypothesis is that treatment with GSH (compared with no GSH) will result in a greater reduction of binge eating at post-treatment.

According to the study design, eligible participants are randomized to either GSH ($n = 50$) or no-treatment control ($n = 50$). Prior analyses estimated that a sample size of 50 per group would provide sufficient power to test the primary hypothesis regarding the treatment effect, assuming a 33% reduction in binge eating in the GSH group and a 10% reduction in the control group. The simulations assumed a population with specific characteristics (based on past literature on RCTs for BN/BED), as opposed to resampling from existing data. The assumptions used in the therapist-effect power calculations were based on comparable past studies,[15,18–22] specifically: (1) generalized linear models including a fixed effect for therapist is the analytic technique (normal distribution, identity link function); (2) the outcome (dependent) variable is frequency of binge eating in the last month, at post-treatment, represented as being on a continuous scale; (3) the only covariate is the frequency

of binge eating in the last month, at baseline; (4) the significance of the therapist effect is tested—the null hypothesis is: Δ therapist #1 = … = Δ therapist #n (n ranges from 2 to 5), where Δ is the difference in frequency of binge eating between baseline and control at post-treatment, for patients of a given therapist (in other words, this tests whether each therapists' patients have the same degree of remission in binge eating between baseline and post-treatment, compared to the control patients); (5) the alpha level for the tests will be 0.05; (6) equal numbers of participants are randomly assigned to the GSH and control groups; (7) in both groups, the raw mean (Standard Deviation (SD)) of frequency of binge eating at baseline is ~9.9 (14.4) per month; (8) the outcome is positively skewed and log-transformed before analysis; (9) there is a 10% reduction in binge eating in the control group (assuming some spontaneous recovery) and a 33% reduction in the GSH group (33% was the smallest reduction found in comparable past studies, and is used to be conservative); and (10) there is 30% attrition between baseline and post-treatment.

Statistical tests were done separately for the treatment main effect (H$_0$: degree of remission in binge eating in the treatment condition averaged across therapists = degree of remission in binge eating in the control condition) and for the therapist effect (H$_0$: degree of remission in binge eating in patients of therapist #1 = degree of remission in binge eating in patients of therapist #2 = … = degree of remission in binge eating in patients of therapist #n). Note that the assumed RCT design does not enable estimation of a treatment-by-therapist interaction in the usual sense (i.e., the differential therapist effect in treatment vs. control) because the therapists do not administer the control condition. An RCT design with multiple treatments, where each therapist administers each treatment/condition, would enable estimation of a treatment-by-therapist interaction, but such a design is beyond the scope of this article.

Factors varied in the simulations were (1) the number of therapists (between 2 and 5) and (2) the estimated therapist effect size (small, medium, and large). Patient sample size and study design were held constant. Within a simulation, all therapists were assigned the same number of patients, except when the number of patients was not evenly divisible by the number of therapists, in which case the remaining patients were randomly allocated to therapists. Following Grissom and Kim,[23] the therapist effect size was defined as (max. therapist mean) – (min. therapist mean)/common SD; a "large" effect was defined as a difference of four binge eating episodes (BE), "medium" as three BE and "small" as two BE. The simulations examined 12 different scenarios, defined by crossing number of therapists with therapist effect size (Table 2). Table 2 describes the average tendencies in each scenario. Because the Monte Carlo simulations incorporate a realistic degree of sampling variability, means may vary widely among samples drawn from the population described in a given scenario.

Monte Carlo simulations were used to estimate power to detect the therapist effect.[24,25] Briefly, using this method, assumptions are made about the true characteristics of the population. Then repeated samples are drawn with replacement from the specified population. There is a realistic element of randomness that induces variation between the samples. This simulates what would actually be observed if one repeatedly kept doing studies of random samples of participants drawn from the population. In each sample, the statistical test of interest is performed (in this study, this is the omnibus test of the therapist effect, where the null hypothesis is that none of the therapists differ in treatment outcomes). Power is the proportion of samples that have a significant test result.

Simulations were conducted using SAS 9.2 (SAS Institute, Cary, NC). To estimate power, we drew 500 random samples for each scenario and summarized the results of the test of therapist effect across samples. Details of the methodology, including SAS code, are available.[26]

Results

In all scenarios, power to detect the treatment main effect (i.e., the test of treatment averaged across therapists, compared with control) was greater than 80%. Power to detect therapist effects in each scenario is shown in Table 3. Only with the largest effect size, with two therapists, did power to detect the therapist effect exceed the conventional level of 80%. However, there was reasonable power (70%) to detect the therapist effect with three or four therapists and a large effect size, or two therapists and a medium effect size.

Discussion

The main objective of this article was to provide practical guidance to researchers on how best to approach therapist effects in the context of a typical, small RCT for eating disorders. Of the two main approaches for analyzing therapist effects (fixed vs. random), the fixed effects approach may be most feasible in small trials, although it has limitations in terms of generalizability and modeling of certain aspects of the data. Power using the fixed effects approach equates to power to detect differences among the specific therapists in the trial. This is not the same as power to estimate therapist effects in the population of therapists, regarding the therapists in the trial as a sample; this would require a random effects analysis. A properly designed random effects study would be the gold standard to estimate therapist effects, but it would require a large sample of therapists (at least five, preferably more) and therapists would need to be purposively sampled to enable inference to some specified population.

In the simulations, power to detect therapist effects increased as the size of the therapist effect increased and as the number of therapists decreased. The former finding is to be expected, although this study makes a contribution by quantifying the power in specific scenarios. The latter finding also makes sense—the smaller the number of therapists, the greater the number of patients within each (because the total patient sample size was held constant); thus, the more information available to make conclusions about the effectiveness of the therapist.

The simulation results imply that if therapist effects are analyzed as fixed, it is preferable to minimize the number of therapists, to maximize the information about the effectiveness of each therapist. Does this mean that the optimal number of therapists is two? We believe that the answer is no, based on three considerations. First, there are practical reasons to involve more than two therapists. Therapists may drop out during the course of a trial. It is useful to have a pool of trained backup therapists available if needed. If there are only two therapists involved in a trial, and one therapist stops participating, the trial timeframe is likely to be extended (because all treatment falls to a single therapist) and the trial itself might be jeopardized. For this reason alone, it is worthwhile to consider involving at least three or four therapists.

Second, there is little value in detecting small therapist effects. For nearly every treatment, there will probably be at least a very small therapist effect—it is unrealistic to think that therapists will be exactly equal in treatment effectiveness or that patients, even randomly assigned, will have the exact same response to therapy. However, there is no reason to set up a trial and analysis to detect therapist effects that are so small as to be practically unimportant. The important consideration is ability to detect large, practically important therapist effects. The simulation study presented above provided example definitions of small, medium, and large therapist effects. The results suggest that in a typical small RCT on eating disorders, three or four therapists can be employed with reasonable power (70%) to detect large therapist effects.

Third, although one may not be able to conclusively generalize results to a larger population of therapists based on a small trial with a nonrandom sample of therapists and a fixed effects analysis, using a greater number of therapists provides a better sense of intertherapist variation in administration of the treatment, which may be helpful for future research, including the design of larger trials designed specifically to estimate therapist effects. Variation among three or four therapists may provide a better sense of the therapist effect (at least qualitatively) than does variation among two therapists.

In this article, the therapist effect is defined as the intertherapist difference in outcomes for a single treatment—specifically, the difference between therapists in their patients' BE reduction between baseline and post-treatment, holding constant the type of treatment administered. In RCTs where each therapist administers multiple types of treatment, the therapist effect may also be defined in terms of the therapist-by-treatment interaction, that is, the difference in outcomes for a therapist administering treatment A versus treatment B. These are different ways of defining therapist effects. Both definitions are informative and intuitive.

How should therapists be selected for the typical, small RCT for eating disorders? The ideal may be to select three or four skilled therapists and provide them with extensive, standardized training. Then the results will indicate whether the treatment can be effectively and consistently administered by a group of skilled, well-trained therapists. In other words, the results will indicate whether the treatment can be effective under relatively favorable conditions. If there is a treatment effect but no therapist effect, this suggests that the treatment can be effectively and consistently administered under relatively favorable conditions. On the other hand, if there is a therapist effect (i.e., the therapists differ in their treatment outcomes), this suggests that the treatment does not consistently succeed under such conditions. If each therapist has a relatively large sample of patients, then having a sample of three or four skilled, well-trained therapists provides more certainty about the nature of therapist differences than does a sample of two therapists. If there is a therapist effect with two therapists, it is uncertain whether the treatment was ineffective or whether one of the therapists was ineffective. However, if there are three or four therapists and the majority show favorable treatment outcomes, this more strongly suggests that the treatment can be effective and the outlying therapist failed to administer the treatment effectively.

One might also view an initial, relatively small trial as a pilot study with respect to therapist effects. Leon et al.[27] recommend using pilot studies to test the feasibility of various aspects of a trial. Therapist effects in an initial trial might be viewed as a pilot indicating the feasibility of therapist selection and training methods. If a treatment effect is present but there is no therapist effect, this suggests that the therapist selection and training methods were adequate. On the other hand, if therapists differ in treatment outcomes, this points to inadequate therapist selection and/or training methods that need to be improved before conducting a larger trial.

The forgoing discussion assumes that patients are randomly assigned to therapists. Patients vary in the severity of their condition, so it is possible that a spurious therapist effect could result from assignment of sicker patients to one of the therapists, by chance. This is a possibility in any trial, but especially in small trials.

Although the objective of this article was to provide practical advice for designing and analyzing RCTs on eating disorders, most of the points discussed are applicable to psychotherapy RCTs in other fields as well. However, the power analysis results, here indicating that three or four therapists will be sufficient to detect relatively large therapist effects, will not necessarily be the same for RCTs for other disorders. Many RCTs for BN/

BED have used similar designs (pre/post, treatment vs. control) and measures (e.g., the Eating Disorders Examination), so it is possible to describe a typical RCT in BN/BED and to base power calculations on the results of such trials. In RCTs on other disorders, there may be less consistency in measurement methodologies and the primary outcome measures may have a different distribution; therefore, a different number of therapists might be recommended.

Importantly, the simulation results reported here assumed a particular RCT design, namely a single treatment administered by multiple therapists, compared with no-treatment control. This design has been frequently used in the eating disorders field (e.g., Refs. 14–17), but other designs have also been used. For example, some RCTs have compared multiple treatments, either administered by the same therapists (treatments crossed with therapists, e.g., Refs. 28–30) or with different therapists administering different treatments (treatments nested within therapists, e.g., Refs. 31–33). The recommended number of therapists will likely vary based on the RCT design among other factors. These other designs are beyond the scope of this article. Practical recommendations for approaching therapist effects using these designs in eating disorders research would be a worthwhile topic for future work.

The simulation results reported here are also applicable to RCT designs where a single treatment administered by multiple therapists is compared with another treatment that does not involve therapists, for example, medication-only (e.g., Ref. 34) or pure self-help (e.g., Ref. 35). In such designs, the assumed treatment main effect may be reduced, but the therapist effect (i.e., the difference between therapists in the therapist-administered treatment condition) will be approximately the same as reported here.

In summary, this study points to several practical suggestions for investigators conducting typical, small RCTs (200 total patients) for eating disorders (BN/BED): (1) use a fixed effects approach to analyze therapist effects; (2) regard the therapist effects analysis as conclusive with respect to the specific therapists in the trial, but as inconclusive with respect to the population of therapists; and (3) employing about three or four therapists is likely to be practical, while providing sufficient power to detect large therapist effects and insight into the nature of therapist differences (if there are any).

Acknowledgments

References

1. Loeb KL, Wilson GT, Labouvie E, Pratt EM, Hayaki J, Walsh BT, et al. Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: A study of process and outcome. J Consult Clin Psychol. 2005; 73:1097–1107. [PubMed: 16392983]

2. Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. J Consult Clin Psychol. 2010; 78:200–211. [PubMed: 20350031]

3. Wilson GT, Wilfley DE, Agras WS, Bryson S. Allegiance bias and therapist effects: Results of a randomized controlled trial of binge eating disorder. Clin Psychol: Sci Pract. 2011; 18:119–125.

4. Crits-Cristoph P, Mintz J. Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. J Consult Clin Psychol. 1991; 59:20–26. [PubMed: 2002139]

5. Jannoun L, Munby M, Catalan J, Gelder M. A home-based treatment program for agoraphobia: Replication and controlled evaluation. Behav Ther. 1980; 11:294–305.

6. Lewis M, Morley S, van der Windt DA, Hay E, Jellema P, Dziedzic K, et al. Measuring practitioner/therapist effects in randomised trials of low back pain and neck pain interventions in primary care settings. Eur J Pain. 2010; 14:1033–1039. [PubMed: 20444631]

7. Mathews AM, Johnston DW, Lancashire M, Munby M, Shaw PM, Gelder MG. Imaginal flooding and exposure to real phobic situations: Treatment outcome with agoraphobic patients. Br J Psychiatry. 1976; 129:361–371. [PubMed: 10036]

8. Franko DL, Thompson-Brenner H, Thompson DR, Boisseau CL, Davis A, Forbush KT, et al. Racial/ethnic differences in adult participants in randomized clinical trials of binge eating disorder. J Consult Clin Psychol. (in press).

9. Brown, H.; Prescott, R. Applied mixed models in medicine. New York: Wiley; 1999.

10. Walters SJ. Therapist effects in randomized controlled trials: What to do about them. J Clin Nurs. 2009; 19:1102–1112. [PubMed: 20492055]

11. Walwyn R, Roberts C. Therapist variations within randomized trials of psychotherapy: Implications for precision, internal and external validity. Stat Methods Med Res. 2010; 19:291–315. [PubMed: 19608603]

12. Siemer M, Joormann J. Power and measures of effect size in analysis of variance with fixed versus random nested factors. Psychol Methods. 2003; 8:497–517. [PubMed: 14664685]

13. Serlin RC, Wampold BE, Levin JR. Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joormann. Psychol Methods. 2003; 8:524–534. [PubMed: 14664687]

14. 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. [PubMed: 9697011]

15. Carter JC, Fairburn CG. Cognitive-behavioral self-help for bin-geeating disorder: A controlled effectiveness study. J Consult Clin Psychol. 1998; 66:616–623. [PubMed: 9735577]

16. Palmer RL, Birchall H, McGrain L, Sullivan V. Self-help for bulimic disorders: A randomized controlled trial comparing minimal guidance with face-to-face or telephone guidance. Br J Psychiatry. 2002; 181:230–235. [PubMed: 12204928]

17. Banasiak SJ, Paxton SJ, Hay P. Guided self-help for bulimia nervosa in primary care: A randomized controlled trial. Psychol Med. 2005; 35:1283–1294. [PubMed: 16168151]

18. Cachelin FM, Striegel-Moore RH, Regan PC. Factors associated with treatment seeking in a community sample of European American and Mexican American women with eating disorders. Eur Eat Disord Rev. 2004; 14:1–8.

19. Cachelin FM, Striegel-Moore RH. Help seeking and barriers to treatment in a community sample of Mexican American and European American women with eating disorders. Int J Eat Disord. 2006; 39:1544–1561.

20. Carter JC, Olmsted MP, Kaplan AS, McCabe RE, Mills JS, Aime A. Self-help for bulimia nervosa: A randomized controlled trial. Am J Psychiatry. 2006; 160:973–978. [PubMed: 12727703]

21. Ghaderi A, Scott B. Pure and guided self-help for full and sub-threshold bulimia nervosa and binge eating disorder. Br J Clin Psychol. 2003; 42:257–269. [PubMed: 14565892]

22. Pritchard BJ, Bergin JL, Wade TD. A case series evaluation of guided self-help for bulimia nervosa using a cognitive manual. Int J Eat Disord. 2004; 36:144–156. [PubMed: 15282684]

23. Grissom, RJ.; Kim, JJ. Effect sizes for research: A broad practical approach. Mahwah, NJ: Earlbaum; 2005.

24. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. Struct Equation Model. 2002; 4:599–620.

25. Williams MS, Ebel ED, Wagner BA. Monte Carlo approaches for determining power and sample size in low-prevalence applications. Prev Vet Med. 2007; 82:151–158. [PubMed: 17590459]

26. Thompson, D.; Cachelin, F.; Striegel-Moore, R.; Wilson, T. Power to detect therapist effects in randomized, controlled trials of interventions in clinical psychology. Proceedings of the Midwest SAS User Group; Oct 2010; Milwaukee, WI. (Available at: http://www.mwsug.org/proceedings/2010/stats/MWSUG-2010-76.pdf)

27. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res. 2011; 45:626–629. [PubMed: 21035130]

28. Agras WS, Walsh BT, 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. [PubMed: 10807486]

29. Walsh BT, Wilson GT, Loeb KL, Devlin MJ, Pike KM, Roose SP, et al. Medication and psychotherapy in the treatment of bulimia nervosa. Am J Psychiatry. 1997; 154:523–531. [PubMed: 9090340]

30. Wilfley DE, Welch RR, Stein RI, Spurrell EB, Cohen LR, Saelens BE, et al. A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. Arch Gen Psychiatry. 2002; 59:713–721. [PubMed: 12150647]

31. Wilson GT, Wilfley DE, Agras WS, Bryson SW. Psychological treatments for binge eating disorder. Arch Gen Psychiatry. 2010; 67:94–101. [PubMed: 20048227]

32. Durand MA, King M. Specialist treatment versus self-help for bulimia nervosa: A randomised controlled trial in general practice. Br J Gen Pract. 2003; 53:371–377. [PubMed: 12830564]

33. Bailer U, de Zwaan M, Leisch F, Strnad A, Lennkh-Wolfsberg C, El Giamal N, et al. Guided self-help versus cognitive-behavioral group therapy in the treatment of bulimia nervosa. Int J Eat Disord. 2004; 35:522–537. [PubMed: 15101068]

34. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs. medications in the treatment of moderate to severe depression. Arch Gen Psychiatry. 2005; 62:409–416. [PubMed: 15809408]

35. Treasure J, Schmidt U, Troop N, Tiller J, Todd G, Keilen M, et al. First step in managing bulimia nervosa: Controlled trial of therapeutic manual. Br Med J. 1994; 308:686–689. [PubMed: 8142791]

TABLE 1

Major differences between fixed versus random effects approaches to analyzing therapist effects

Dimension Compared	Approach to Analyzing Therapist Effects	
	Fixed Therapist Effects	Random Therapist Effects
Scope of inference (i.e., the results may be generalized to…)	The specific therapists in the trial	The population of therapists
Assumptions about the sample of therapists in the trial	None	The therapists are a representative sample from a specified population
Intratherapist correlation (i.e., correlation of patient outcomes within therapists)	Ignored (assumed to be zero)	Estimated in the analysis (may be different from zero)
Residual variance	Equal across patients (homoscedasticity)	May be unequal across patients (heteroscedasticity)
Recommended minimum sample size of therapists	At least two	At least five

TABLE 2

Average episodes of binge eating (BE) per month by visit (baseline vs. post-treatment), condition (control vs. treatment), therapist (#), number of therapists, and therapist effect size

Therapist Effect Size	Number of Therapists	BE at Baseline (Control and All Therapists)	BE at Post-Treatment							Maximum Difference Among Therapists
			Control	Treatment (Therapist #)						
				#1	#2	#3	#4	#5		
Largest	Five	9.9	8.9	4.7	5.6	6.6	7.6	8.6		4.0
	Four	9.9	8.9	4.7	5.6	7.6	8.6	n/a		4.0
	Three	9.9	8.9	4.7	6.6	8.6	n/a	n/a		4.0
	Two	9.9	8.9	4.7	8.6	n/a	n/a	n/a		4.0
Medium	Five	9.9	8.9	5.1	5.9	6.6	7.3	8.1		3.0
	Four	9.9	8.9	5.1	5.9	7.3	8.1	n/a		3.0
	Three	9.9	8.9	5.1	6.6	8.1	n/a	n/a		3.0
	Two	9.9	8.9	5.1	8.1	n/a	n/a	n/a		3.0
Smallest	Five	9.9	8.9	5.6	6.1	6.6	7.1	7.6		2.0
	Four	9.9	8.9	5.6	6.1	7.1	7.6	n/a		2.0
	Three	9.9	8.9	5.6	6.6	7.6	n/a	n/a		2.0
	Two	9.9	8.9	5.6	7.6	n/a	n/a	n/a		2.0

TABLE 3

Estimated power to detect therapist effects (%) by therapist effect size (large, medium, small) and number of therapists (2–5)

Therapist Effect Size	Number of Therapists			
	2	3	4	5
Large (%)	92	77	70	58
Medium (%)	75	46	45	37
Small (%)	44	28	25	23

Notes: Percentages in table represent the number of simulated trials (of 500 simulations) for which a significant therapist effect was found, which is an estimate of the power to detect the therapist effect size.