

Can Psychotherapists Function as Their Own Controls? Meta-Analysis of the Crossed Therapist Design in Comparative Psychotherapy Trials

Fredrik Falkenström, PhD; John C. Markowitz, MD; Hanske Jonker, MSc;
Björn Philips, PhD; and Rolf Holmqvist, PhD

ABSTRACT

Objective: Clinical trials sometimes have the same therapists deliver more than 1 psychotherapy, ostensibly to control for therapist effects. This “crossed therapist” design makes controlling for therapist allegiance imperative, as therapists may prefer one treatment they deliver to the other(s). Research has established a strong relationship between principal investigators’ allegiances and treatment outcome. Study therapists’ allegiances probably also influence outcome, yet this moderating factor on outcome has never been studied.

Data Sources: English language abstracts in PsycINFO and MEDLINE from January 1985 to December 2011 were searched for keywords *psychotherapy* and *randomized trial*.

Study Selection: The search yielded 990 abstracts that were searched manually. Trials using the same therapists in more than 1 condition were included.

Data Extraction: Thirty-nine studies fulfilled inclusion criteria. Meta-regression analyses assessed the influence of researchers’ allegiance on treatment outcome, testing the hypothesis that studies poorly controlling for therapist allegiance would show stronger influence of researcher allegiance on outcome. A single-item measure assessed researchers’ reported attempts to control for therapist allegiance.

Results: Only 1 of 39 studies (3%) measured therapist treatment allegiance. Another 5 (13%) mentioned controlling for, without formally assessing, therapist allegiance. Most publications (67%) did not even mention therapist allegiance. In studies not controlling for therapist allegiance, researcher allegiance strongly influenced outcome, whereas studies reporting control for therapist allegiance showed no differential researcher allegiance. Researchers with cognitive-behavioral therapy allegiance described controlling for therapist allegiance less frequently than other researchers.

Conclusions: The crossed therapist design is subject to bias due to differential psychotherapist allegiance. Worrisome results suggest that researchers strongly allied to a treatment may ignore therapist allegiance, potentially skewing outcomes. All clinical trials, and especially crossed therapist designs, should measure psychotherapist allegiance to evaluate this possible bias.

J Clin Psychiatry 2013;74(5):482–491

© Copyright 2012 Physicians Postgraduate Press, Inc.

Submitted: April 12, 2012; accepted August 10, 2012.

Online ahead of print: October 30, 2012 (doi:10.4088/JCP.12r07848).

Corresponding author: Fredrik Falkenström, PhD, Lustigkullevägen 17, SE-616 33 Åby, Sweden (Fredrik.Falkenstrom@liu.se).

One of the sacrosanct assumptions of a client is that their therapist believes in the treatment being delivered.

Bruce E. Wampold¹(p159)

Some psychotherapy researchers have designed clinical trials to “cross” therapists with treatments: that is, the same therapists deliver more than 1 type of psychotherapy. The rationale is that using the same therapist across therapies controls for therapist effects on treatment outcome. The design has attracted surprisingly little critical attention in the scientific literature. This article reviews studies that have used this design and discusses important research implications.

Researchers agree that individual therapists may generate different outcomes with patients, some having consistently greater efficacy than others,^{2,3} although the size of this therapist effect has stirred debate.^{4,5} In randomized clinical trials (RCTs) in which therapists are trained to treat a specific disorder according to a manual, between-therapist variation is most likely smaller than in general practice.⁶ Nonetheless, reviews of psychotherapy research recommend controlling for therapist effects by including the therapist factor as a random term in the statistical analysis of any clinical trial, regardless of whether the therapist effect reaches conventional significance levels. The reasoning is that even small therapist effects increase the type 1 error rate of the between-treatment comparison if the therapist factor is omitted from the statistical model.⁷ Crits-Christoph and Mintz⁶ recommend using an α level of .2 to .3 (rather than the usual .05) as the criterion for deciding to exclude the therapist factor in the main outcome analysis of a clinical trial.

The RCT is considered the optimal design to ensure internal validity of clinical trials, reducing the influence of variables other than the one(s) experimentally manipulated.^{8,9} Medical and psychotherapy outcome research deems this design the “gold standard.” Psychotherapy RCTs most commonly compare different teams of (ideally) similarly experienced, equally expert therapists providing the different study treatments—a practical approach, as there might not be so many therapists who are able to deliver more than 1 therapy correctly and competently. (As far as we know this latter issue has never been studied, but our personal experiences have demonstrated that it is difficult to learn more than 1 psychotherapy well.) Elkin¹⁰ enumerated the difficulties of this modal design, which “nests” therapists within treatments: ie, different teams of therapists provide the different treatments. This design potentially confounds treatment efficacy with therapist efficacy. Elkin¹⁰ described the difficulties of ensuring similar characteristics, experience, competence, training, and supervision of the competing teams of therapists.

- Therapist belief in treatment is likely to be a strong nonspecific effect of psychotherapy, yet this factor has almost never been studied.
- Researcher allegiance may influence study findings, in part, through the selecting of biased therapists in “crossed therapist” study designs.

The alternative design “crosses” psychotherapists over treatments: the same (crossed) therapists provide both the experimental and comparison treatment(s).⁶ Some advocates of this design have argued that it controls for therapist effects^{11–13}: the therapist provides his or her own control. Although this design might seem to disentangle therapists from treatments, the need persists to ensure that therapists have comparable training, experience, and competence in both experimental and comparison treatments. As several authors have noted, the crossed design raises additional problems: contamination between treatments and differential therapist allegiance to the treatments.^{6,7,10,12}

Contamination involves therapists using techniques from one treatment in the comparison condition that blur distinctions between treatments and should skew results toward finding no treatment differences. Still more crucial, therapeutic allegiance reflects therapist belief that one psychotherapy works better than another. We have uncovered no empirical research studying the connection between *therapist* allegiance and outcome. Several reviews,^{14–16} however, show that *researcher* (principal investigator) allegiance to a treatment correlates strongly with study outcome. This is a correlational finding, but if the relationship is indeed causal, one way that researcher allegiance could affect outcome is through researchers selecting study therapists who share the researcher’s allegiance. Researcher’s allegiance could affect outcome in other ways: for example, by selecting ineffective comparison treatments or choosing not to publish studies that run counter to their allegiance.¹⁴ Researcher allegiance and therapist allegiance are thus different but most likely interdependent concepts.

Study therapists who value one treatment above another would most likely communicate their bias to their patients, affecting both technical and nonspecific aspects of treatment: both treatment delivery and the conviction with which they deliver each therapy. Therapists may have greater natural affinity for and skill in one treatment than another. Therefore, even absent empirical data, it is plausible to hypothesize that differential therapist allegiance in a crossed therapist design may artificially enlarge between-treatment differences. Differential allegiance, competence, or both might also contribute to smaller between-treatment differences if therapists deliver an ineffective treatment more enthusiastically and competently than an inherently more potent treatment. The well-established relationship between research allegiance and good outcome suggests this is rarer,

however. Contamination and differential allegiance could still affect a study in which distinct teams of therapists provide the different treatments. Contamination of techniques from other therapeutic schools could occur if study therapists have previous training in other therapies, and differential allegiance could occur if one therapist team feels greater enthusiasm for its treatment model than another.

Therapist allegiance may thus constitute a moderating variable, comparable to variables such as therapist experience and training, which all outcome studies should consider. Nonetheless, the demanding arrangement of switching back and forth between competing models renders the crossed design particularly vulnerable to criticism. Indeed, having the same therapists deliver multiple modalities may obscure potential therapist bias. By using the crossed design to ostensibly control for therapist factors, researchers may ignore important allegiance differences that may receive greater scrutiny in studies comparing nested competing therapist teams, even though, paradoxically, the latter probably have less variability in therapeutic allegiance.

As the literature has never addressed this issue, we review studies that used the crossed therapist design to study how researchers have addressed therapist allegiance. A preliminary qualitative review (available from first author [F.F.] upon request) revealed that therapist allegiance was seldom, if ever, measured. We therefore chose as our primary research question whether researchers reported having *controlled for* therapist allegiance, and whether not controlling for therapist allegiance affected treatment outcome. We hypothesized that a stronger relationship would exist between researcher allegiance and study outcome when therapist allegiance was poorly controlled. Because methodological quality of RCTs has evolved,¹⁷ we also explored whether control for therapist allegiance increased over time.

METHOD

Selection of Studies

We searched PsycINFO and MEDLINE (EBSCO) for abstracts containing the terms *psychotherapy* and *randomized trial*. The search, limited to English language articles published in peer-reviewed journals between January 1985 and December 2011, yielded 990 hits (652 in PsycINFO, 338 in MEDLINE) that the first author (F.F.) scrutinized. He searched all abstracts and, when necessary, article method sections to determine whether the RCT used the same psychotherapists in more than 1 treatment condition. Requests sent to 2 psychotherapy research mailing lists sought leads on crossed studies. One was the e-mail list of the Society for Psychotherapy Research, an international multidisciplinary organization devoted to psychotherapy research of diverse theoretical orientations. The second was the Psychodynamic Research Listserv, open to people actively associated with psychodynamic research in an academic or clinical practice setting. Tips from these 2 mailing lists yielded 2 studies.^{18,19} Incorporating our own and colleagues’ knowledge of crossed therapist studies added another 3 studies.^{20–22}

The search excluded trials of component and dismantling studies within the same general therapy orientation, as differential competence and allegiance would most likely be less problematic in such studies. We included, however, component studies comparing models that are recognized as treatments in their own right: eg, comparisons between cognitive and behavioral treatments, because, although the 2 models are often combined as cognitive-behavioral therapy (CBT), both therapists and researchers may have strong allegiances in the continuing debate over the relative contributions of cognitive and behavioral techniques to CBT outcome. We excluded studies of the same therapy model delivered in different formats (eg, individual versus group format) because allegiance most likely differs more between competing theoretical models than between different delivery formats. Finally, we excluded studies published before 1985 because secular improvement in RCT methodological quality¹⁷ suggested that older studies would lack comparability to more modern ones.

This search yielded 39 trials^{11–13,18–54} published between 1986 and 2011, which we included in the study.

Measures

Control for therapist allegiance. The first author (F.F.) conducted a qualitative review of the articles. This review provided the basis for developing a measure for rating control for therapist allegiance in RCTs, the Falkenström Allegiance Control for Therapists measure (FACT). The FACT comprises a single item measuring researcher control for therapist allegiance on a scale from 0 to 3:

0. The article does not mention the issue of potentially differing therapist allegiances. If the authors mention the crossed therapist design at all, they do so only as a blanket disclaimer of therapist effects.
1. The issue of potentially different allegiances is mentioned but not assessed or controlled for in any way.
2. The authors report making some attempt to control for therapist allegiance, such as trying to recruit therapists who informally deny therapy preferences. However, they make no attempt to measure therapist allegiance or to statistically control for this variable.
3. The authors measure therapist allegiance and, if necessary, apply relevant statistical control for this variable.

Researcher allegiance. We measured researcher allegiance by using the methodology most prevalent in published studies.^{14,55} The method entails using manualized scoring criteria to score researcher allegiance according to the way published articles are written. Three raters (F.F., H.J., and B.P.) read published articles, scrutinizing especially the Introduction and methods section. The scale for researcher allegiance was adapted from Gaffan et al.⁵⁵ They developed the following criteria, which should be viewed as indicators rather than absolute rules:

0. No allegiance: (a) no evidence is presented that the treatment is effective; (b) the authors simply mention that the treatment will be included, without elaboration.
1. Weak allegiance: (a) authors describe previous research indicating effectiveness of the treatment, but the results are described as mixed or give no indication that the treatment should be effective for the population tested in the study; (b) the authors state a hypothesis with mixed predictions: eg, the treatment is expected to benefit some patients but not others.
2. Moderate allegiance: (a) reference is given to previous research indicating efficacy of the treatment relative to no treatment; (b) evidence is presented from the literature that the treatment is effective for this population of patients; (c) statements showing that the author(s) believes the treatment is effective or widely approved; and (d) a short but clear rationale for the treatment's procedure.
3. Strong allegiance: (a) reference to published research showing superiority of the treatment to some other treatment; (b) a specific hypothesis is presented as to why the treatment should be superior to the other treatment(s) included in the study; (c) a detailed description (at least 10 lines) explains the treatment's procedures and aims; and (d) the treatment was developed or first introduced by one of the authors (in our scoring this immediately gave an allegiance score of 3).

Following Gaffan et al.,⁵⁵ we converted the ratings to relative allegiance scores by subtracting the lowest score from the higher ones within each study. For example, if treatment A and treatment B had researcher allegiance scores of 3 (high allegiance), both would receive relative allegiance scores of 0. If a third treatment C in the same study had an absolute allegiance score of 1, the relative allegiance scores would be 2 for treatments A and B and 0 for C.

Treatment dose and therapist characteristics. We rated whether researchers had ensured that all treatments were delivered in equal dosage and whether therapists in all conditions had equal training, experience, and supervision. These were single-item variables with simple yes/no/not reported alternatives. Training was defined not only by the training the researchers provided for the study but also by all training therapists had had in the respective treatment modalities.

Effect sizes. The effect sizes were between-groups standardized mean change scores, Δ ,⁵⁶ calculated by first taking the mean pretreatment-posttreatment difference divided by the pretreatment standard deviation separately for treatment and control groups. The variance was then calculated as follows:

$$V_d = \left[\frac{1}{n} + \frac{d^2}{2n} \right] \times 2(1-r)$$

where n is the group size and r the correlation between pre-measurements and postmeasurements. Because the latter was

almost never reported, we assumed a correlation between pretreatment and posttreatment measurements of 0.60 for all measurements in all studies. Both d and V_d were then corrected for sample size bias:

$$g = J \times d$$

$$V_g = J^2 \times V_d$$

$$J = 1 - \frac{3}{4(df) - 1}$$

Finally, the between-groups standardized mean change scores were calculated as follows:

$$\Delta = g_T - g_C$$

$$V_{\Delta} = V_{g_T} + V_{g_C}$$

Nine studies reported some measures as proportions rather than continuous measures (15 of 94 measures). In these cases, Cohen h was used.⁵⁷ Four studies provided no information for calculating effect sizes for between-group contrasts that were statistically nonsignificant (14 of 27 measures). This was unfortunate, since we then had to assume zero differences between groups in these cases. The effect sizes were most likely not actually 0; yet, omitting these nonsignificant measures would have further distorted our data.

Two studies had to be excluded from the meta-regression analyses: one³⁴ because the effect size used was time-to-event (hazard ratio) from a survival analysis, which was not deemed comparable to Becker Δ in the other studies, and the other⁴⁸ because patients were crossed over between therapies at midtreatment, confounding interpretation of effect sizes.

Procedure

The number of comparisons for each study was the number of treatment groups minus 1. A study comparing 2 treatments had 1 comparison; a study comparing 3 treatments had 2 comparisons, etc, yielding 1 between-group effect size per comparison. Effect sizes were scored by using the group scoring lowest on researcher allegiance as the comparison group. When allegiance did not differ between groups, experimental/comparison group status was arbitrary. The relative allegiance score for the comparison group, being redundant, was not used.

Reliability

The first author (F.F.) rated all variables for all studies. Two other raters (H.J. and B.P.) rated all variables for 33 and 6 studies, respectively. Interrater agreement was calculated by using intraclass correlation (2-way random effects model) for continuous variables, Spearman ρ for ordinal variables, and Cohen κ for nominal variables.

Statistical Analyses

A problem with meta-analysis is that some data are statistically nonindependent. In our analysis, most studies used more than 1 outcome measure, with the result that outcome measures used for the same patients were dependent on each other. Several studies^{21,22,28,33,44,45,47,54} compared more than 2 therapies, creating additional nonindependence in data.

Fortunately, recent advances in statistical methodology provide solutions to this problem. Multilevel meta-analysis⁵⁸ enables explicit modeling of the covariances on both within- and between-study levels. However, this method ordinarily requires knowledge of the within-study intercorrelations (ie, correlations between different measures used), information seldom, if ever, available to a meta-analysis. To address this problem, Hedges et al⁵⁹ created a theorem allowing computation of regression estimators in meta-regression that are robust for violations of the nonindependence assumption. This theorem computes fixed regression estimators that take random effects (between-studies variation) into account. Being most concerned about within-study dependence of effect sizes, we chose the correlated effects model. The value of ρ , the mean intercorrelation of effect sizes, was first set to .5, deemed a reasonable guess of the values of ρ . We then ran sensitivity tests assigning different values to ρ between .1 and .9. Demonstrating the robustness of the approach, these different values of ρ had little effect on the estimated regression coefficients.

Computations used the Stata 11 macro `robumeta` to compute robust standard errors in meta-regression.⁶⁰ Despite conducting multiple statistical significance tests, we chose not to control for family-wise type I error. This first study on this subject was by nature exploratory, and the multiple tests mostly tested different research questions. The use of corrections for family-wise type I error has been questioned.⁶¹

RESULTS

Table 1 lists all 39 included studies.^{11,13,18-54}

Reliability of Ratings

Reliability estimates were good to excellent for the main predictor variables. For FACT, Spearman ρ was .91. For the meta-regression analyses, we collapsed the FACT into a dichotomous scale in which 0 meant no control for therapist allegiance (categories 0 and 1 from the full scale) and 1 meant some control for therapist allegiance (categories 2 and 3). The rationale for this decision was that mere mention that the same therapists do both treatments (category 1) probably does not influence the effect of researcher allegiance on outcome. Only 1 study⁴⁹ met category 3, raising concern that this single trial would excessively influence the statistical analyses. The collapsed scale had reliability somewhat lower than the original, but still acceptable ($\kappa = 0.72$). Relative researcher allegiance had an Intraclass Correlation of 0.82 for the ratings of a single rater, which was considered good reliability. The ratings of the first rater (F.F.) were used in the final analyses.

For treatment dose, therapist training, and supervision, reliability was adequate to good ($\kappa = 0.82, 0.64, 0.77$, respectively). For therapist experience, reliability was low ($\kappa = 0.53$), most likely because the initial scoring criteria were unclear. Because all disagreements arose between rater 1 and 2, rater 3 rated all disagreements, and we used these ratings as the final score for these studies.

Table 1. Studies Using the Crossed Therapist Design

Study	Treatments	Condition	N	Researcher Allegiance ^a	FACT Rating
Agras et al ²³	CBT vs IPT	Bulimia nervosa	220	CBT:1 IPT:0	0
Barkham et al ¹⁸	CBT vs psychodynamic-interpersonal	Depression	36	CBT: 0 Psychodynamic-interpersonal: 0	2
Barkham et al ²⁴	CBT vs psychodynamic interpersonal	Subsyndromal depression	116	CBT: 0 Psychodynamic interpersonal: 0	2
Beck et al ²⁵	CBT vs client-centered supportive	Panic disorder	33	CBT: 3 Client-centered supportive: 0	1
Bjornsson et al ²⁶	Cognitive-behavioral group therapy vs group psychotherapy	Social anxiety disorder	45	Cognitive-behavioral group therapy: 2 Group psychotherapy: 0	0
Borkovec et al ²⁷	CBT vs nondirective	Generalized anxiety disorder	30	CBT: 3 Nondirective: 0	0
Borkovec and Mathews ²⁸	Cognitive therapy vs coping desensitization vs nondirective	Generalized anxiety disorder and panic disorder without agoraphobia	30	Cognitive therapy: 3 Coping desensitization: 2 Nondirective: 0	0
Castelnuovo et al ²⁹	Brief strategic therapy vs CBT	Binge eating disorder	60	Brief strategic therapy: 1 CBT: 0	0
Clark et al ²⁰	Cognitive therapy vs exposure + applied relaxation	Social phobia	62	Cognitive therapy: 2 Exposure + applied relaxation: 0	0
Cottraux et al ¹¹	CBT vs Rogerian supportive therapy	Chronic PTSD	60	CBT: 2 Rogerian supportive therapy: 0	1
Cottraux et al ³⁰	CBT vs Rogerian supportive therapy	Borderline personality disorder	65	CBT: 2 Rogerian supportive therapy: 0	1
Dunn et al ³¹	Self-management therapy vs psychoeducational group therapy	Chronic PTSD + depression	101	Self-management therapy: 3 Psychoeducational group therapy: 0	0
Fairburn et al ³²	CBT vs short-term focal psychotherapy	Bulimia nervosa	24	CBT: 2 Short-term focal psychotherapy: 0	0
Fairburn et al ³³	CBT vs IPT vs behavior therapy	Bulimia nervosa	75	CBT: 2 IPT: 0 Behavior therapy: 0	1
Frank et al ³⁴	Interpersonal and social rhythm therapy vs intensive clinical management	Bipolar I disorder	175	Interpersonal and social rhythm therapy: 3 Intensive clinical management: 0	0
Goldman et al ³⁵	Process-experiential psychotherapy vs client-centered psychotherapy	Depression	38	Process-experiential psychotherapy: 1 Client-centered psychotherapy: 0	1
Gosselin et al ³⁶	CBT vs nonspecific supportive therapy	Generalized anxiety disorder + benzodiazepine dependency	61	CBT: 3 Nonspecific supportive therapy: 0	0
le Grange et al ³⁷	Family-based therapy vs supportive therapy	Anorexia nervosa	80	Family-based therapy: 3 Supportive therapy: 0	0
Greenberg and Watson ³⁸	Process-experiential psychotherapy vs client-centered psychotherapy	Depression	34	Process experiential psychotherapy: 1 Client-centered psychotherapy: 0	1
Guthrie et al ³⁹	Psychodynamic therapy vs supportive listening	Irritable bowel syndrome	101	Psychodynamic therapy: 3 Supportive listening: 0	1
Herbert et al ⁴⁰	CBT vs educational/supportive therapy	Social anxiety disorder	73	CBT: 2 Educational/supportive therapy: 0	0
Hopko et al ⁴¹	Behavioral activation vs problem-solving therapy	Depressed breast cancer patients	80	Behavioral activation: 1 Problem-solving therapy: 0	0
Hudson et al ⁴²	CBT vs group support and attention	Children and adolescents with anxiety disorders	112	CBT: 3 Group support and attention: 0	0
Lipsitz et al ⁴³	IPT vs supportive therapy	Social anxiety disorder	70	IPT: 2 Supportive therapy: 0	0
Luty et al ¹⁹	CBT vs IPT	Major depression	177	CBT: 0 IPT: 0	2
Marchand et al ⁴⁴	Cognitive therapy vs exposure, cognitive therapy + exposure vs supportive therapy	Panic disorder with agoraphobia	137	Cognitive therapy: 2 Exposure: 2 Cognitive therapy + exposure: 3 Supportive therapy: 0	0
Marks et al ⁴⁵	Exposure vs cognitive restructuring vs exposure combined with cognitive restructuring vs relaxation	PTSD	87	Exposure: 3 Cognitive restructuring : 2 Exposure + cognitive restructuring : 3 Relaxation: 0	0
Masheb et al ⁴⁶	CBT vs supportive therapy	Vulvodynia	50	CBT: 3 Supportive therapy: 0	0

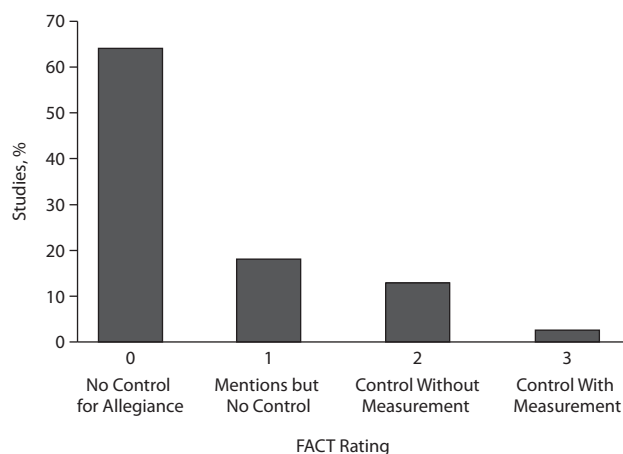
(continued)

Table 1 (continued). Studies Using the Crossed Therapist Design

Study	Treatments	Condition	N	Researcher Allegiance ^a	FACT Rating
McIntosh et al ²¹	CBT vs IPT vs nonspecific supportive clinical management	Anorexia nervosa	56	CBT: 1 IPT: 1 Nonspecific supportive clinical management: 0	0
Miklowitz et al ⁴⁷	Family-focused therapy vs interpersonal and social rhythm therapy vs CBT vs collaborative care	Bipolar depression	293	Family-focused therapy: 3 Interpersonal and social rhythm therapy: 3 CBT: 3 Collaborative care: 0	0
Shapiro and Firth ⁴⁸	Prescriptive vs exploratory therapy	Mixed depression and anxiety (mostly depression)	40	Prescriptive therapy: 0 Exploratory therapy: 0	2
Shapiro et al ¹³	CBT vs psychodynamic-interpersonal	Depression	117	CBT: 0 Psychodynamic-interpersonal: 0	2
Snyder and Wills ⁴⁹	Behavioral vs insight-oriented marital therapy	Marital distress	79	Behavioral marital therapy: 0 Insight-oriented marital therapy: 0	3
Strauman et al ⁵⁰	Self-system therapy vs cognitive therapy	Depression	45	Self-system therapy: 1 Cognitive therapy: 0	1
Tarrier et al ⁵¹	Imaginal exposure vs cognitive therapy	Chronic PTSD	72	Imaginal exposure: 0 Cognitive therapy: 0	0
Taylor et al ²²	CBT (prolonged exposure) vs EMDR vs relaxation training	PTSD	60	CBT: 2 EMDR: 1 Relaxation training: 0	0
Walsh et al ⁵²	CBT vs supportive therapy	Bulimia nervosa	120	CBT: 3 Supportive therapy: 0	0
Wilfley et al ⁵³	Group CBT vs group IPT	Binge eating disorder	162	Group CBT: 0 IPT: 0	0
Zettle and Rains ⁵⁴	Group comprehensive distancing vs group cognitive therapy vs group partial cognitive therapy	Depression	31	Group comprehensive distancing: 3 Group cognitive therapy: 2 Group partial cognitive therapy: 0	0

^aRelative researcher allegiance score for each treatment condition included in study: 0 = no allegiance, 1 = weak allegiance, 2 = moderate allegiance, 3 = strong allegiance.

Abbreviations: CBT = cognitive-behavioral therapy, EMDR = eye movement desensitization and reprocessing, FACT = Falkenström Allegiance Control for Therapists, IPT = interpersonal psychotherapy, PTSD = posttraumatic stress disorder.

Figure 1. Falkenström Allegiance Control for Therapists (FACT) Ratings of Control for Therapist Allegiance

Descriptive Statistics

No studies overlapped with the earlier meta-analysis by Luborsky et al,¹⁴ and only 3 overlapped with Gaffan et al.⁵⁵ The median publication year was 2002. Thirty-one studies (79%) included cognitive therapy, behavioral therapy, or CBT as at least 1 of the treatments. The 39 studies yielded 48 treatment comparisons. Among these 48 comparisons, 10 provided no way of ascertaining which treatment was

intended to be experimental and which was control treatment. Among the remaining 38 comparisons, the most common experimental treatment was CBT (18 comparisons). Including variants of CBT, such as behavior therapy, cognitive therapy, or self-management therapy, increased that figure (29 comparisons). The most common comparison treatments were supportive (14), client-centered (5), and relaxation (5) therapies.

Therapist Allegiance

Figure 1 shows the FACT ratings for all 39 studies. Only 1 study⁴⁹ (3%) earned the maximal FACT rating, ie, addressing measurement of therapist allegiance and its statistical control if needed. Another 5 studies (13%) reported having attempted to recruit therapists with balanced allegiances as a way of controlling for therapist allegiance, although none had actually measured therapist allegiance. The remaining studies did not mention therapist allegiance at all (26 studies; 67%) or mentioned it only as a study limitation or a variable that could be dismissed as unimportant (7 studies; 18%).

Comparing studies with researcher allegiance favoring CBT to those lacking such allegiance on the (collapsed) FACT scale indicated that CBT-allegiant researchers controlled for therapist allegiance less often than researchers without CBT allegiance (Fisher exact test $P = .002$).

Researcher Allegiance Predicting Outcome

We first ran a meta-regression analysis with researcher allegiance as the sole predictor. This test revealed a statistically significant effect of researcher allegiance on outcome ($b=0.10$, $SE=0.02$; $t=3.87$; $P<.001$; 95% CI, 0.05– 0.15). Each point increase in relative allegiance increased the between-groups effect size by 0.10.

Control for Therapist Allegiance Predicting Outcome

We next tested the direct effect of control for therapist allegiance (collapsed FACT) on outcome. This effect was not significant ($b= -0.07$, nonsignificant). Yet, only 5 studies reported controlling for allegiance, limiting statistical power.

Control for Therapist Allegiance Moderates Relationship Between Researcher Allegiance and Outcome

Our primary hypothesis was that researcher allegiance more strongly affects outcome in studies that do not control for therapist allegiance. We added an interaction term (researcher allegiance \times control for therapist allegiance) to the statistical model that contained main effects of researcher allegiance and control for therapist allegiance on outcome. This model proved impossible to run, however, as none of the studies that had attempted control for therapist allegiance had been scored as having any differential researcher allegiance. The interaction term was therefore almost completely collinear with the original control for therapist allegiance variable (ie, researcher allegiance \times control for therapist allegiance = control for therapist allegiance).

Exploring this further revealed a large difference in researcher allegiance between the studies that had not attempted to control for therapist allegiance (32) and those reporting having attempted some control (6). The former group had a mean relative allegiance score of 2.1 ($SD=0.97$), whereas the latter group had a mean of 0 (with no variation). This difference was highly statistically significant ($t_{32}=12.2$, $P<.001$).

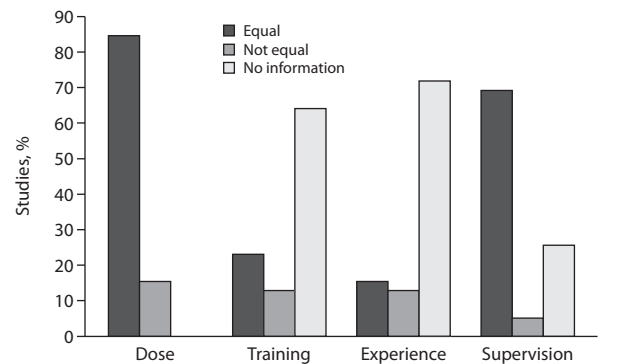
Publication Year

Testing whether recent studies controlled for therapist allegiance more often than older studies unexpectedly showed a significant relationship in the opposite direction from the expected one: more recent studies controlled *less* for therapist allegiance ($\rho = -.33$; $P=.04$). Using the dichotomized therapist allegiance control variable yielded an almost identical result. Further exploration revealed that 5 of the 6 studies^{13,18,24,48,49} that controlled for therapist allegiance were published between 1987 and 1996 (the sixth was published in 2007).

Reporting of Dose and Therapist Characteristics

Figure 2 shows reporting of equal therapy dose, therapist training, experience, and supervision. Most studies reported having ensured equal dose (85%) and amount of supervision (69%) in all treatments, but most studies inadequately

Figure 2. Proportion of Studies Reporting Having Ensured Equal Dose, Training, Therapist Experience, and Supervision in the Different Study Treatments



reported therapist training and experience. About two-thirds of the studies did not report how much training and experience the therapists had in each of the treatments they delivered (64% and 72%, respectively). Six studies reported higher treatment dose, 5 reported more therapist experience, and 5 reported more training in the treatment scoring higher on researcher allegiance.

DISCUSSION

Psychotherapy is a complex enterprise, involving many factors: patient, therapist, treatment factors, and complex interactions between all of these. The RCT design is ideally suited to isolating the treatment factor: when patients are randomized to treatments, confounding factors should be equally distributed between groups. However, because it is usually not feasible (and probably not desirable) to randomize therapists to treatments, the problem of confounding therapist and treatment effects remains.

Some researchers have advocated^{11–13} solving this problem by using the same psychotherapists to conduct more than 1 study treatment. Despite the salience of this design issue, no research has previously reviewed trials with the crossed therapist design. Our main finding was that the great preponderance of the 39 such studies we collected failed to report the key issue of therapist treatment allegiance. Despite the glaring need in a crossed design to ascertain therapist allegiance to the respective treatments, less than half (36% [14]) of the trials reviewed even mentioned therapist or researcher allegiance. Still fewer (13% [5]) explicitly reported attempting to control for and only 1 (3%) actually measured therapist allegiance. About two-thirds of the studies surveyed did not report therapists' previous training or experience in the treatments they provided. Most studies included involved CBT as a treatment (87% [34]), and researchers with CBT allegiance more often used this design without controlling for therapist allegiance. Thoma et al⁶² found that the average CBT trial of depression scored at the lower range of adequate methodological quality and that CBT trials did not, on average, show better quality than psychodynamic trials.

Although most researchers did not even mention therapist allegiance as a possible bias, some acknowledged that their study therapists believed the comparison treatment to be ineffective and that this may have biased results:

The choice of the same therapists for both treatments created a twofold problem. On the positive side, the effect of the same therapist was constant across the two therapeutic conditions. On the negative side, all were CBT-oriented therapists and this may have biased the treatment in favor of CBT as the therapists were requested to use methods they judged as noneffective.^{11(p108)}

Our analyses, corroborating earlier research^{14,55} while surveying largely different outcome trials, showed that researcher allegiance clearly influenced treatment outcome. Our main research question was whether this effect diminished in studies that controlled for therapist allegiance. Interestingly, we found no differential researcher allegiance in any of the studies that had reportedly attempted some control for therapist allegiance. This sharply contrasted with the studies not attempting such control, for which mean researcher allegiance score was 2.1 on a scale ranging from 0 to 3. The stronger the researchers' allegiances to 1 psychotherapy, the more likely they appeared to ignore therapist allegiance. If this indeed indicates how research using the crossed therapist design is being conducted, it is remarkable and worrisome that researchers strongly allied to 1 treatment consciously or unconsciously overlook the potential bias of therapist allegiance.

Some researchers may argue that treatment integrity checks provide an objective assessment of whether therapists deliver treatment as it is meant to be delivered; hence, integrity checks should reveal therapist allegiance effects as problems in the delivery of treatment. Although our review did not focus on integrity, we did note that almost 70% (27) of the included studies reported formal adherence analyses, and only 1 of these reported an adherence problem. Treatment integrity should ideally include ratings of therapist competence. Assessing competence is difficult, however, and a recent meta-analysis showed no significant relationship of competence ratings to outcome.⁶³ We doubt that integrity analyses based on fairly coarse measures will detect more subtle effects of therapists who technically adhere to the treatment protocol but do not deliver treatment with the same enthusiasm as therapists who really believe in their model.

Our analyses suggested that controlling for therapist allegiance has fallen out of fashion: 5 of the 6 studies that controlled for therapist allegiance were published between 1987 and 1996 and only 1 since that time. Our impression is that the psychotherapy research community has seldom discussed the issue. Therapist allegiance remains a crucial, unstudied factor in psychotherapy research. Strength of belief in a therapy may affect the therapist's comfort and authenticity in conducting treatment, the therapy's plausibility for the patient, and, thus, the strength of the therapeutic alliance. This belief matters when a therapist conducts a single therapy, and it matters still more when the therapist delivers more than 1 modality.

If therapist allegiance influences therapeutic outcome, then the crossed therapist design may prove a double-cross: while ostensibly controlling for therapist factors between therapists, it may merely obscure them within the same therapist. Although grant reviewers frequently suggest controlling for therapist effects by crossing therapists, the research supporting this injunction is essentially nonexistent. Does choosing a crossed therapist design solve the problem of confounding therapist effects with treatment effects? The answer depends on what kinds of therapist characteristics affect outcome. If general personality characteristics (warmth, empathy, interpersonal skills, etc) that could reasonably be considered independent of treatment type matter most for patient outcome, the crossed design might control for those effects. If therapist effects rather reflect competence in or enthusiasm for a particular treatment model, however, the alternative design of "amicably competitive" balanced therapist teams⁶⁴ is a better choice. In this design, equally experienced, competent, and adherent teams of psychotherapists compete in friendly rivalry, demonstrating in effect the optimal effects of a given intervention. Rather than these therapists attempting multiple competencies, their goal is expertise in their assigned modality, with variables such as experience and competence counterbalanced across groups.

One could argue that researcher allegiance will out, regardless of design: the alternative design structure of nested "amicably competitive" expert therapist teams⁶⁴ could be equally compromised if the principal investigator chooses less competent or experienced therapists for comparator conditions. We recommend that all clinical trials (crossed or nested) measure and report therapist allegiance, at least until more is known about the effect of therapist allegiance on the process and outcome of psychotherapy. To our knowledge, no instrument for measuring therapist allegiance has yet been published. In our own ongoing studies, we have developed simple self-report instruments for this purpose.

Limitations

This is a new field of research. We sought but may not have located all recent studies using the crossed therapist design. How common this approach is relative to nested competitive therapy teams delivering single modalities is unclear; the latter seems more prevalent. The relatively small number of studies included limits the conclusions that can be drawn, especially about the effect of controlling for therapist allegiance, as this occurred in only 6 of the studies. A key limitation (and key study finding) is the absence of therapist allegiance ratings in the psychotherapy trials under study.

Recommendations for future research. Researchers should rigorously report therapists' backgrounds when using the crossed therapist design, clarifying the training and experience study therapists have in conducting each of the different treatments they use in the trial and not just trial-related training but also previous training and experience. Researchers should report whether they have tried to statistically control for therapist effects (including allegiance) in their main outcome analyses; and, if not, why not? (That

is, what significance criterion did they use in preliminary analyses to rule out therapist effects?) A crossed therapist design, if employed, should control for therapist \times treatment interaction effects, not just therapist main effects. (Some researchers infer an absence of therapist effects in their study when they find no differences in outcome between individual therapists.²⁰ Yet finding no therapist differences in outcome does not rule out the possibility that all therapists deliver 1 treatment better than the other due to either greater competency or allegiance to that treatment.)

All clinical trials should routinely measure therapist allegiance and relate it to outcome. The FACT could be added to existing RCT quality measures¹⁷ in reviews and meta-analyses of RCTs.

Given the potential problems of the crossed therapists design, why use it? We recommend that, in standard efficacy trials, about which the predominant question is differential efficacy between 2 treatments, the crossed design be considered a weaker alternative to the amicably competitive therapist teams approach. The crossed therapists design could have utility for studying interactions between therapist characteristics and treatment method effects on outcome. Another indication for this design might be practical, in which geography or other factors constrain patient access to multiple therapists practicing different treatments. We recommend that researchers avoid crossed therapists designs unless the research questions clearly focus on issues that this design facilitates, or unless practicalities such as geographic isolation preclude alternative designs. If used, crossed therapist designs must measure therapist allegiance.

Therapist allegiance warrants consideration equal to researcher allegiance in designing and interpreting the results of psychotherapy trials. Therapists in every clinical trial could rate on a simple, anchored Likert-type scale their belief as well as their self-perceived skill in all the study's therapies, their judgment as to how well each therapy fits their clinical perspectives, and their prediction of the prospective study outcome. Such evaluations would help to control for potential study biases and might yield interesting findings about psychotherapy outcome trials. Combining such data with psychotherapy process and outcome variables might elucidate the suitability of particular therapists for particular therapies.

Author affiliations: Department of Behavioral Sciences, Linköping University, Linköping (Drs Falkenström, Philips, and Holmqvist); Centre for Clinical Research Sörmland, Uppsala University, Uppsala (Dr Falkenström); Department of Clinical Neuroscience, Karolinska Institutet, and Center for Dependency Disorders, Stockholm County Council, Stockholm (Dr Philips), Sweden; New York State Psychiatric Institute and Columbia University College of Physicians & Surgeons, New York, New York (Dr Markowitz and Ms Jonker); and Faculty of Social Sciences, University of Utrecht, Utrecht, Holland (Ms Jonker).

Financial disclosures: Dr Falkenström receives salary from the Sörmland County Council. Dr Markowitz receives salary support from New York State Psychiatric Institute and grant support from the National Institute of Mental Health (grant MH079078); receives small book royalties from American Psychiatric Press, Basic Books, and Oxford University Press; and receives an editorial stipend from Elsevier Press. Dr Philips receives salary from Linköping University and Stockholm County Council, and small book royalties from Liber. He has grant support from the Swedish Council for Working Life and Social Research (grant 2007-0457) and

from Karolinska Institutet. Dr Holmqvist and Ms Jonker report no financial conflicts of interest.

Funding/support: This study received no direct funding.

Acknowledgments: The authors are indebted to A.C. Del Re, PhD, of Stanford University School of Medicine and VA Palo Alto, California, for instructions on effect size calculations. Dr Del Re reports no conflicts of interest.

REFERENCES

1. Wampold BE. *The Great Psychotherapy Debate: Models, Methods, and Findings*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2001.
2. Kraus DR, Castonguay L, Boswell JF, et al. Therapist effectiveness: implications for accountability and patient care. *Psychother Res*. 2011; 21(3):267–276.
3. Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *J Consult Clin Psychol*. 2005;73(5):914–923.
4. Elkin I, Falconnier L, Martinovich Z, et al. Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Psychother Res*. 2006;16(2):144–160.
5. Kim D-M, Wampold BE, Bolt DM. Therapist effects in psychotherapy: a random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychother Res*. 2006;16(2):161–172.
6. Crits-Christoph P, Mintz J. Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *J Consult Clin Psychol*. 1991;59(1):20–26.
7. de Jong K, Moerbeek M, van der Leeden R. A priori power analysis in longitudinal three-level multilevel models: an example with therapist effects. *Psychother Res*. 2010;20(3):273–284.
8. Chambless DL, Hollon SD. Defining empirically supported therapies. *J Consult Clin Psychol*. 1998;66(1):7–18.
9. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin; 2002.
10. Elkin I. A major dilemma in psychotherapy outcome research: disentangling therapists from therapies. *Clin Psychol Sci Pract*. 1999; 6(1):10–32.
11. Cottraux J, Note I, Yao SN, et al. Randomized controlled comparison of cognitive behavior therapy with Rogerian supportive therapy in chronic post-traumatic stress disorder: a 2-year follow-up. *Psychother Psychosom*. 2008;77(2):101–110.
12. Staines GL. Comparative outcome evaluations of psychotherapies: guidelines for addressing eight limitations of the gold standard of causal inference. *Psychotherapy (Chic)*. 2007;44(2):161–174.
13. Shapiro DA, Barkham M, Rees A, et al. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *J Consult Clin Psychol*. 1994;62(3):522–534.
14. Luborsky L, Diguier L, Seligman DA, et al. The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. *Clin Psychol Sci Pract*. 1999;6(1):95–106.
15. Luborsky L, Singer B, Luborsky L. Comparative studies of psychotherapies: is it true that "everyone has one and all must have prizes"? *Arch Gen Psychiatry*. 1975;32(8):995–1008.
16. Robinson LA, Berman JS, Neimeyer RA. Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychol Bull*. 1990;108(1):30–49.
17. Kocsis JH, Gerber AJ, Milrod B, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Compr Psychiatry*. 2010;51(3):319–324.
18. Barkham M, Rees A, Shapiro DA, et al. Outcomes of time-limited psychotherapy in applied settings: replicating the Second Sheffield Psychotherapy Project. *J Consult Clin Psychol*. 1996;64(5):1079–1085.
19. Luty SE, Carter JD, McKenzie JM, et al. Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *Br J Psychiatry*. 2007;190(6):496–502.
20. Clark DM, Ehlers A, Hackmann A, et al. Cognitive therapy versus exposure and applied relaxation in social phobia: a randomized controlled trial. *J Consult Clin Psychol*. 2006;74(3):568–578.
21. McIntosh VVW, Jordan J, Carter FA, et al. Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *Am J Psychiatry*. 2005; 162(4):741–747.
22. Taylor S, Thordarson DS, Maxfield L, et al. Comparative efficacy, speed, and adverse effects of three PTSD treatments: exposure therapy, EMDR,

- and relaxation training. *J Consult Clin Psychol.* 2003;71(2):330–338.
23. Agras WS, Walsh T, Fairburn CG, et al. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Arch Gen Psychiatry.* 2000;57(5):459–466.
 24. Barkham M, Shapiro DA, Hardy GE, et al. Psychotherapy in two-plus-one sessions: outcomes of a randomized controlled trial of cognitive-behavioral and psychodynamic-interpersonal therapy for subsyndromal depression. *J Consult Clin Psychol.* 1999;67(2):201–211.
 25. Beck AT, Sokol L, Clark DA, et al. A crossover study of focused cognitive therapy for panic disorder. *Am J Psychiatry.* 1992;149(6):778–783.
 26. Bjornsson AS, Bidwell LC, Brosse AL, et al. Cognitive-behavioral group therapy versus group psychotherapy for social anxiety disorder among college students: a randomized controlled trial. *Depress Anxiety.* 2011; 28(11):1034–1042.
 27. Borkovec TD, Mathews AM, Chambers A, et al. The effects of relaxation training with cognitive or nondirective therapy and the role of relaxation-induced anxiety in the treatment of generalized anxiety. *J Consult Clin Psychol.* 1987;55(6):883–888.
 28. Borkovec TD, Mathews AM. Treatment of nonphobic anxiety disorders: a comparison of nondirective, cognitive, and coping desensitization therapy. *J Consult Clin Psychol.* 1988;56(6):877–884.
 29. Castelnovo G, Manzoni GM, Villa V, et al. Brief strategic therapy vs cognitive behavioral therapy for the inpatient and telephone-based outpatient treatment of binge eating disorder: the STRATOB randomized controlled clinical trial. *Clin Pract Epidemiol Ment Health.* 2011;7:29–37.
 30. Cottraux J, Note ID, Boutitie F, et al. Cognitive therapy versus Rogerian supportive therapy in borderline personality disorder: two-year follow-up of a controlled pilot study. *Psychother Psychosom.* 2009;78(5):307–316.
 31. Dunn NJ, Rehm LP, Schillaci J, et al. A randomized trial of self-management and psychoeducational group therapies for comorbid chronic posttraumatic stress disorder and depressive disorder. *J Trauma Stress.* 2007;20(3):221–237.
 32. Fairburn CG, Kirk J, O'Connor M, et al. A comparison of two psychological treatments for bulimia nervosa. *Behav Res Ther.* 1986; 24(6):629–643.
 33. Fairburn CG, Jones R, Peveler RC, et al. Three psychological treatments for bulimia nervosa: a comparative trial. *Arch Gen Psychiatry.* 1991;48(5): 463–469.
 34. Frank E, Kupfer DJ, Thase ME, et al. Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. *Arch Gen Psychiatry.* 2005;62(9):996–1004.
 35. Goldman RN, Greenberg LS, Angus L. The effects of adding emotion-focused interventions to the client-centered relationship conditions in the treatment of depression. *Psychother Res.* 2006;16(5):537–549.
 36. Gosselin P, Ladouceur R, Morin CM, et al. Benzodiazepine discontinuation among adults with GAD: a randomized trial of cognitive-behavioral therapy. *J Consult Clin Psychol.* 2006;74(5):908–919.
 37. le Grange D, Crosby RD, Rathouz PJ, et al. A randomized controlled comparison of family-based treatment and supportive psychotherapy for adolescent bulimia nervosa. *Arch Gen Psychiatry.* 2007;64(9):1049–1056.
 38. Greenberg L, Watson J. Experiential therapy of depression: differential effects of client-centered relationship conditions and process experiential interventions. *Psychother Res.* 1998;8(2):210–224.
 39. Guthrie E, Creed F, Dawson D, et al. A randomised controlled trial of psychotherapy in patients with refractory irritable bowel syndrome. *Br J Psychiatry.* 1993;163(3):315–321.
 40. Herbert JD, Gaudiano BA, Rheingold AA, et al. Cognitive behavior therapy for generalized social anxiety disorder in adolescents: a randomized controlled trial. *J Anxiety Disord.* 2009;23(2):167–177.
 41. Hopko DR, Armento MEA, Robertson SMC, et al. Brief behavioral activation and problem-solving therapy for depressed breast cancer patients: randomized trial. *J Consult Clin Psychol.* 2011;79(6):834–849.
 42. Hudson JL, Rapee RM, Deveney C, et al. Cognitive-behavioral treatment versus an active control for children and adolescents with anxiety disorders: a randomized trial. *J Am Acad Child Adolesc Psychiatry.* 2009;48(5):533–544.
 43. Lipsitz JD, Gur M, Vermes D, et al. A randomized trial of interpersonal therapy versus supportive therapy for social anxiety disorder. *Depress Anxiety.* 2008;25(6):542–553.
 44. Marchand A, Coutu MF, Dupuis G, et al. Treatment of panic disorder with agoraphobia: randomized placebo-controlled trial of four psychosocial treatments combined with imipramine or placebo. *Cogn Behav Ther.* 2008;37(3):146–159.
 45. Marks I, Lovell K, Noshirvani H, et al. Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: a controlled study. *Arch Gen Psychiatry.* 1998;55(4):317–325.
 46. Masheb RM, Kerns RD, Lozano C, et al. A randomized clinical trial for women with vulvodynia: cognitive-behavioral therapy vs supportive psychotherapy. *Pain.* 2009;141(1–2):31–40.
 47. Miklowitz DJ, Otto MW, Frank E, et al. Psychosocial treatments for bipolar depression: a 1-year randomized trial from the Systematic Treatment Enhancement Program. *Arch Gen Psychiatry.* 2007; 64(4):419–426.
 48. Shapiro DA, Firth J. Prescriptive v. exploratory psychotherapy: outcomes of the Sheffield Psychotherapy Project. *Br J Psychiatry.* 1987;151(6): 790–799.
 49. Snyder DK, Wills RM. Behavioral versus insight-oriented marital therapy: effects on individual and interspousal functioning. *J Consult Clin Psychol.* 1989;57(1):39–46.
 50. Strauman TJ, Vieth AZ, Merrill KA, et al. Self-system therapy as an intervention for self-regulatory dysfunction in depression: a randomized comparison with cognitive therapy. *J Consult Clin Psychol.* 2006;74(2): 367–376.
 51. Tarrrier N, Pilgrim H, Sommerfield C, et al. A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. *J Consult Clin Psychol.* 1999;67(1):13–18.
 52. Walsh BT, Wilson GT, Loeb KL, et al. Medication and psychotherapy in the treatment of bulimia nervosa. *Am J Psychiatry.* 1997;154(4):523–531.
 53. Wilfley DE, Welch RR, Stein RI, 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(8):713–721.
 54. Zettle RD, Rains JC. Group cognitive and contextual therapies in treatment of depression. *J Clin Psychol.* 1989;45(3):436–445.
 55. Gaffan EA, Tsaousis I, Kemp-Wheeler SM. Researcher allegiance and meta-analysis: the case of cognitive therapy for depression. *J Consult Clin Psychol.* 1995;63(6):966–980.
 56. Becker BJ. Synthesizing standardized mean-change measures. *Br J Math Stat Psychol.* 1988;41(2):257–278.
 57. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–159.
 58. Raudenbusch SW, Bryk AS. *Hierarchical Linear Models. Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage Publications; 2002.
 59. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Methods.* 2010;1(1):39–65.
 60. Hedberg EC. *ROBUMETA: Stata Module to Perform Robust Variance Estimation in Meta-Regression With Dependent Effect Size Estimates.*: Boston, MA: Boston College Department of Economics; 2011.
 61. O'Keefe DJ. Colloquy: should familywise alpha be adjusted? *Hum Commun Res.* 2003;29(3):431–447.
 62. Thoma NC, McKay D, Gerber AJ, et al. A quality-based review of randomized controlled trials of cognitive-behavioral therapy for depression: an assessment and metaregression. *Am J Psychiatry.* 2012; 169(1):22–30.
 63. Webb CA, Derubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: a meta-analytic review. *J Consult Clin Psychol.* 2010;78(2):200–211.
 64. Markowitz JC, Kocsis JH, Christos P, et al. Pilot study of interpersonal psychotherapy versus supportive psychotherapy for dysthymic patients with secondary alcohol abuse or dependence. *J Nerv Ment Dis.* 2008; 196(6):468–474.