Open Access Theses   Electronic Theses and Dissertations

2017-08-02

A Meta-Analysis of the Effect of Therapist Experience on Outcome for Clients with Internalizing Disorders

Lucia Walsh

University of Miami, lwalsh@miami.edu

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

UNIVERSITY OF MIAMI

A META-ANALYSIS OF THE EFFECT OF THERAPIST EXPERIENCE ON
OUTCOME FOR CLIENTS WITH INTERNALIZING DISORDERS

By

Lucia M. Walsh

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

August 2017

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

A META-ANALYSIS OF THE EFFECT OF THERAPIST EXPERIENCE ON
OUTCOME FOR CLIENTS WITH INTERNALIZING DISORDERS

Lucia M. Walsh

Approved:

_____
Amanda Jensen-Doss, Ph.D.
Associate Professor of Psychology

_____
Saneya Tawfik, Ph.D.
Clinical Assistant Professor of
Psychology

_____
Debbiesiu Lee, Ph.D.
Associate Professor of Education
and Psychological Studies

_____
Guillermo Prado, Ph.D.
Dean of the Graduate School

WALSH, LUCIA M. (M.S., Psychology)
(August 2017)

A Meta-Analysis of the Effect of Therapist
Experience on Outcome for Clients with
Internalizing Disorders.

Despite having the most prevalent group of psychiatric disorders, many people with internalizing disorders do not receive treatment. One factor related to lack of treatment is a shortage of qualified mental health therapists. Task shifting may be one solution to this shortage, but has been relatively unused due to the idea that less experienced therapists may not be able to attain commiserate client outcomes as more experienced therapists. However, a relationship between therapist experience and client internalizing outcomes has never been found, a critical first step in determining if task shifting is a viable option to getting millions of people in the United States treatment. Through the meta-analysis of 16 articles representing 31 distinct studies, this study provides an updated and more thorough understanding of the relationship between therapist experience and internalizing client outcome. Although a modest relationship between experience and outcomes was found, the clinical significance of this finding is limited. Further investigation of task-shifting as a potential solution to provider shortages is recommended, as well as a call for renewed research in this area using longitudinal within-therapist designs.

TABLE OF CONTENTS

LIST OF FIGURES

vi

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

One area of psychopathology that has received a great deal of attention is internalizing disorders, or those psychological disorders characterized by anxiety, fear, shyness, low self-esteem, sadness, and/or depression (Ollendick & King, 1994). Internalizing disorders are the most common group of mental illnesses in the United States (Anxiety and Depression Association of America, 2014), and are highly comorbid with one another (Kaufman & Charney, 2000; Kessler et al., 2003; Wittchen, Kessler, Pfister, Höfler, & Lieb, 2000). Internalizing disorders can lead to decreased educational attainment and social functioning, and increased substance use and suicide attempts, especially when comorbid (Boden, Fergusson, & Horwood, 2007; Thornicroft & Sartorius, 1993; Van Ameringen, Mancini, & Farvolden, 2003). Poor outcomes are much more likely to occur if no treatment is received. Unfortunately, many people suffering from these disorders never receive treatment (Kessler et al., 2005; Merikangas et al., 2010), and most never receive care consistent with evidence-based practice (EBP), or clinical practice that is informed by evidence about interventions, clinical expertise, and patient needs, values, and preferences (Kazdin, 2008; Young, Klap, Sherbourne, & Wells, 2001). To increase to use of appropriate treatments for people with internalizing disorders, EBP implementation movements have gained speed over the past 20 years by providing training, ongoing consultation, and organizational and system support to communities across the country (Beidas et al., 2013; D'Angelo, Pullmann, & Lyon, 2017; Stirman et al., 2017).

Although these efforts are important to ensure that clients are getting *appropriate* treatment, there is still the issue that many clients with internalizing disorders do not receive any treatment at all. Many factors are associated with limited receipt of services for clients with internalizing disorders, including low identification of people with internalizing disorders (Chavira, Stein, Bailey, & Stein, 2004), living in areas with low number of therapists available (Kessler et al., 2005), and lack of insurance (Young et al., 2001). Limited therapists providing treatment is not unique to clients with internalizing disorders. Our public mental health system is facing a shortage of therapists across client presenting problems, with 96.5 million Americans living in areas considered "mental health care provider shortage areas" (American Psychological Association, 2008; Kaiser Family Foundation, 2014). Additionally, with the advent of the Affordable Care Act and other nation-wide policies, increasing numbers of people are becoming qualified for publicly funded mental health care, likely leading to increased demand for mental health services (Golden & Vail, 2014).

One potential reason for the shortage is the long training period for qualified mental health therapists, due to a basic assumption in the mental health field that the more experience and training a therapist has, the better their client's outcomes will be, especially working with populations they are considered "expert" in. However, in the case of clients with internalizing disorders, the relationship between increased experience and increased client outcomes has never been definitively proven. Over the past 60 years, hundreds of studies have examined the relationship between therapist experience and outcome across diagnostic groups (e.g., symptom improvement, therapeutic alliance, client satisfaction, client dropout), and results have been equivocal (e.g., Franklin,

Abramowitz, Furr, Kalsy, & Riggs, 2003). Furthering our understanding of this relationship using updated methodology, such as meta-analysis, would allow us to determine whether task-shifting, or redistributing treatment tasks from professionally trained workers to those with less training and fewer qualifications (Fulton et al., 2011; Gopalan, 2016), is an appropriate method to address the mental health workforce shortage.

Before examining whether a relationship between experience and outcomes exists, we need a better understanding of why equivocal findings exist. A contributor to confusion in the literature is the diversity of ways experience has been defined. Therapist experience has been operationalized with comparisons between "professional" (those that had attained some level of specialized training in mental health) versus "paraprofessional" groups (those who had no formalizing training in mental health past a bachelor's degree); different degree types; status in training program, and years, months, or even days practicing therapy (Bright, Baker, & Neimeyer, 1999; Budge et al., 2013; Propst, Paris, & Rosberger, 1994). Inconsistencies like these make it hard to get a clear picture of which therapists attain better client outcomes.

Beyond definitional discrepancies, researchers have argued that definitions of therapist experience miss the mark in terms of actually measuring concrete and meaningful clinical experience across therapist groups (Beutler, 1997). These arguments focus on definitions of experience being too broad and the length of time it takes to become proficient depending on the task of interest. To illustrate these arguments, a comparison to training in the medical field is helpful. First, if attempting to decide how much experience a brain surgeon had, we mostly likely would not be interested in the

number of urological surgeries she had performed, because the skills needed in that type of surgery are not representative of the skills needed to be a proficient brain surgeon. Second, it is likely that it takes less experience to be proficient at stitching a wound closed than at completing a heart transplant. Bringing these arguments to therapist experience, researchers have argued it would be better to define experience more specifically (e.g., types of client problems treated, number of times using a specific therapy protocol), and to examine whether it varies as a function of client (e.g., motivation to change, gender, age) or therapist (e.g., level of empathy, flexibility) factors (Beutler, 1997).

Thankfully, there is a growing body of literature with researchers who are studying therapist experience in specific populations (Blatt, Sanislow III, Zuroff, & Pilkonis, 1996) and/or using fine-grained indices of therapist experiences (Huppert et al., 2001; Podell et al., 2013). With the influx of these new studies, it may be an appropriate time to reexamine the issue of whether therapist experience matters to client outcomes while heeding Beutler's advice to examine this in a more specific client population, clients with internalizing disorders.

One way to examine the question of whether therapist experience is related to internalizing client outcomes is meta-analysis. Meta-analysis is the analysis of quantitative effect sizes drawn from multiple studies, and often includes examinations of moderators of effect size (Uman, 2011). Meta-analyses are thought to be more transparent, more replicable, reduce some of the inherent subjectivity, and deliver a more interpretable message than review papers and individual studies alone (Lipsey & Wilson, 2001). Several large scale meta-analyses have examined whether therapist experience has

a significant effect on client outcome (Berman & Norton, 1985; Durlak, 1979; Stein & Lambert, 1995; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995). Examination of this question has included analyses focused on comparisons across studies, where outcomes from studies utilizing more experienced therapists are compared to outcomes from studies with less experienced therapists (i.e., "between studies meta-analyses"), and analyses focused on pooling results of analyses conducted within studies, where therapist experience level is examined as a predictor of outcomes within a single sample (i.e., "within studies meta-analyses"). Relevant findings from both types of meta-analyses are reviewed below to identify potentially useful outcomes and moderators to examine within this relationship, and provide additional justification for our focus on internalizing populations.

Past meta-analyses have found a modest positive relationship between therapist experience and client outcomes when examining certain types of outcomes, treating certain populations, and specific study characteristics (Stein & Lambert, 1995; Weisz et al., 1987; Weisz et al., 1995). In terms of outcome characteristics. the relationship between therapist experience and client outcomes was stronger when outcomes were based on client satisfaction with treatment and change on psychological test measures, as well as rated by independent evaluators (Stein & Lambert, 1995). Differences in the relationship between therapist experience and client outcome were also found between clients with internalizing and externalizing disorders (Weisz et al., 1987; Weisz et al., 1995), where youth with internalizing disorders achieved better reduction in symptoms when treated by professional therapists, while paraprofessionals attained better or commiserate outcomes in youth with externalizing disorders than graduate-level or

professional therapists. Finally, studies with more clients tended to find more positive effects between experience and outcome than smaller studies (Stein & Lambert, 1995).

Although meta-analyses focused on broad client populations are helpful for identifying potentially relevant variables, meta-analyses focused on solely internalizing client populations are also important to understand the relationship between therapist experience and client outcomes. Unfortunately, only two meta-analyses on the effect of therapist experience on outcomes specifically in clients with depressive disorders have been completed, both using between study meta-analysis (Johnsen & Friborg, 2015; Michael, Huelsman, & Crowley, 2005). Michael et al. (2005) investigated this question in treatment studies for youth depression. They found that professional therapists did not produce significantly higher effect sizes than graduate students, with both types of therapists producing large treatment effects. This finding was inconsistent with Weisz et al. (1987) and Weisz et al. (1995) findings, perhaps due to the broad definition of internalizing disorders in the former meta-analysis. In secondary analyses of a meta-analysis of CBT for depression, Johnsen and Friborg (2015) compared psychologists and graduate student trainees' effectiveness treating adults with depression. They found that professional psychologists achieved significantly better outcomes than graduate students (Johnsen & Friborg, 2015).

The Michael et al. (2005) and Johnsen and Friborg (2015) meta-analyses provided differing evidence regarding the effect of therapist experience on depressed client outcomes, potentially because of the different client age groups examined in each study. In addition, these meta-analyses were both conducted using data from large-scale randomized control trials (RCTs). Thus, many of the graduate students in these studies

were rigorously trained using manualized treatments, potentially obscuring differences that may have been found in less trained and supervised student or paraprofessional populations. These studies did not compare differing therapist groups within the same study. This is problematic as all therapists in each study had to be assigned to one experience level based on the majority experience level of the therapist group, potentially ignoring important and interesting heterogeneity within the experience levels of study therapists. In addition, comparing relatively inexperienced therapists from one study to experienced therapists in another study ignores a variety of potential confounding differences between the two studies, such as differences associated with treatment site like general severity of client populations or quality of supervision provided. Therapist experience might also have a differential effect on client outcomes depending on internalizing diagnoses (i.e., anxiety only, depression only, comorbid anxiety and depression). To our knowledge, there have been no meta-analyses focused on anxious populations alone, any comparing the effect of therapist experience across different internalizing diagnoses, and no within-study meta-analysis specifically in this population.

Elucidating this relationship has real world implications for the millions of people with internalizing disorders in the United States. If experience does matter for internalizing client outcome, we may need to wait to train novice therapists in an internalizing EBP until they have gained more general experience, or provide increased supervision of and consultation with novice therapists when they are working internalizing clients. If experience does not affect outcomes for clients with internalizing disorders, this may give hope to those implementing EBPs that the training of lay

therapists can lead to effective outcomes (Gopalan, 2016; Murray et al., 2015) and potentially fill the demand for mental health care therapists in the United States.

With this gap in the literature and the relevance of therapist experience for implementation efforts, a meta-analysis of the literature is merited. This study aims to better understand the relationship between therapist experience and outcomes for internalizing clients. To do this, we have four overarching aims.

Aim 1 examined the overall effect of therapist experience on internalizing client outcomes using one aggregated outcome for each study, as we expected many of our studies would use multiple outcome measures. Aggregating measures by averaging across all effects in a study removes dependency between measures, and allows for easier interpretation of findings. From past research, we hypothesized that there will be a small, but significant relationship between therapist experience and internalizing client outcomes, where therapists with more experience will attain better outcomes.

Testing Beutler's (1997) hypothesis that different definitions of therapist experience may show different relationships to client outcomes, aim 2 was to conduct several sub-group meta-analyses estimating the effect of therapist experience on client outcome using different definitions of therapist experience. Consistent with Beutler (1997), we hypothesized that studies using more fine-grained definitions of therapist experience (e.g., total number of client hours, number of times using a specific manual) would find more positive relationships between therapist experience and internalizing client outcome. We hypothesized that studies with broader definitions of therapist experience (e.g., years of experience, degree attained) would find negative or null relationships between therapist experience and internalizing client outcome.

Due to findings from past meta-analyses suggesting certain characteristics of outcome measures may be related to a differential relationship between therapist experience and internalizing client outcomes (Stein & Lambert, 1995), aim 3 focused on conducting several sub-group meta-analyses on different outcome measure characteristics. This included sub-group meta-analyses exploring different measure domains (e.g., anxiety symptoms, functioning), rater of outcomes (e.g., self-rated, independent-evaluator rated), and types of measures (e.g., rating scales, semi-structured interviews). First, we hypothesized that the relationship between therapist experience and client outcome would differ by the rater of client symptom change. When outcomes were rated by objective raters (e.g., independent evaluators), we expected to find a more positive relationship between therapist experience and outcome (Stein & Lambert, 1995). In contrast, we expected the relationship between experience and outcome would be equivalent when client outcomes were rated by therapists and clients (Stein & Lambert, 1995). In terms of outcome domain (e.g., internalizing) and measure type (e.g., rating scale, semi-structured interview), limited research has focused on these outcome measure characteristics, thus no a priori hypotheses were made.

Aim 4 was to examine many of the significant moderators found in between study meta-analyses (e.g., Weisz et al., 1995) and within study meta-analyses (e.g., Stein & Lambert, 1995), as well as potentially important study-design factors (e.g., randomization of clients to therapist, treatment type). Moderators were examined for both the aggregated effect sizes from Aim 1 and the sub-group meta analyses from Aims 2 and 3. Within aim 4, there were several distinct research questions. First, did any client-level factors moderate the relationship between therapist experience and client outcomes? We

hypothesized that client comorbidity (i.e., whether clients with comorbid disorders allowed in the study) would moderate the relationship between therapist experience and client outcomes, with experienced therapists achieving better outcomes with these more complex clients, but no differences based on therapist experience in samples without comorbidities. In terms of client age, we hypothesized that more experienced therapists would achieve better outcomes with adolescents than inexperienced therapists (Weisz et al., 1995), but that outcomes for children would be equivalent across therapist groups (Michael et al., 2005; Weisz et al., 1995). There has been limited work examining if experience matters to adult client outcomes (Johnsen & Friborg, 2015), thus we had no a priori hypothesis regarding this population. For client diagnoses, we hypothesized that equivalent outcomes would be attained across therapist groups when treating anxiety disorders, as the construct of anxiety disorders as a whole contains many easily treated disorders like specific phobias (Weisz et al., 1995). For depressive disorders, we hypothesized better client outcomes would be attained by more experienced therapists, as depressive disorders contain more complex disorders than the anxiety disorder cluster (Johnsen & Friborg, 2015; Weisz et al., 1995)

Next, did any treatment characteristics moderate the relationship between therapist experience and client outcomes? We had no a priori hypotheses regarding moderation by treatment modality (e.g., CBT, psychodynamic therapy, interpersonal therapy) due to the mixed literature in this area (Johnsen & Friborg, 2015; Michael et al., 2005; Weisz et al., 1995).

Finally, did study characteristics moderate the relationship between therapist experience and client outcomes? With many studies assigning more severe clients to

more experienced therapists, we hypothesized that randomization would moderate the relationship between therapist experience and client outcomes. Specifically, studies who randomized clients would find stronger positive associations between therapist experience and client outcomes, and studies without randomization would find equivalent outcomes across therapist experience groups.

CHAPTER TWO

METHOD

Search procedure

We conducted an exhaustive search for articles examining the relationship between therapist experience and internalizing client outcomes. We identified potential studies using the following search procedures. A search was conducted in *PsycINFO* and *Web of Science*. We used combinations of any of the following key words to identify relevant studies: "Therapist OR provider OR clinician OR counselor OR psychologist OR mental health professional OR psychiatrist," "paraprofessional OR layperson," "experience OR training," "client OR patient OR participant," "symptom OR outcome," "internalizing OR depression OR anxiety," "withdrawal OR worry OR sadness OR irritability OR nervousness OR fear OR neuroticism." Wildcard terms were used to allow for results to include all possible versions of a word (e.g., using depress* to include "depression," "depressive," and "depressed"). In addition, "NOT" statements were used to exclude inappropriate articles (e.g., "NOT speech therapist" when searching for "therapist"). We identified additional appropriate articles from past reviews and meta-analyses. The reference list of identified studies were searched to find additional relevant studies. To minimize publication bias, we also solicited for data from relevant unpublished studies via professional listservs including American Psychological Association Divisions 12, 17, 53, 54 and the Association for Behavioral and Cognitive Therapies Special Interest Groups including Anxiety, Child and Adolescent Anxiety, Child & Adolescent Depression, and Couples Research and Treatment. Finally, we contacted authors involved in seminal research in this area to access potential in-progress

studies. All citations were exported and saved into an EndNote X8 database and analyzed for duplicate references.

Inclusion criteria for articles were: (1) written in English, (2) included mental health professionals (i.e., those people who have received specific training in providing psychotherapy to clients, e.g., social workers, psychologists, psychiatrists, marriage and family therapists), and (3) examined the impact of therapist experience specifically on client outcomes. Articles were excluded if: (1) the treatment group was not relevant for psychological services (e.g., college students seeking academic counseling) and (2) the treatment therapists were inappropriate (e.g., parents of clients). Studies that meet all other inclusion criteria but did not include necessary statistical information to compute effect sizes were not initially excluded, and study authors were contacted for this information. We included studies published in any year.

This search process initially identified 2248 articles. We removed 183 duplicates, and 2065 were included in the first screen. From the first screen, 1933 studies were removed for relevance. Studies were removed for several reasons, including focus on non-internalizing psychiatric disorders, focus on medical conditions, focus on therapeutic processes, no use of psychotherapy, and use of non-mental health therapists. We included 132 articles in a detailed screen of the entire manuscript.

To ensure consistent inclusion and exclusion of recovered studies, a graduate student and a trained undergraduate research assistant coded the remaining 132 articles as "included" or "excluded," as well as the specific reason that they were excluded (i.e., client, therapist, treatment, or design factors). To examine interrater reliability for categorical codes, kappas were calculated between the two students. Interrater reliability

was considered acceptable if it is greater than 0.70. Interrater reliability ranged from .80-.98. After reliability was met, the graduate student conducted random checks on the inclusion and exclusion of studies to combat drift. We excluded a further 110 studies from the sample, leaving 21 studies to review and code. For further information regarding our screening process, please see Figure 1.

Study coding

We coded included articles to calculate effect size, characterize their design, and identify moderators of interest. If an article contained multiple distinct samples or studies, each sample or study was coded as a separate "study." A study codebook was developed for this study (included in Supplemental Materials). Study screening and coding was conducted by three graduate student coders. Interrater reliability was established by randomly sampling 50% of the articles to be coded by all three authors. Any disagreements between coders were discussed by all three coders and brought to a supervising faculty member for a final decision, if needed. Reliability for nominal coding categories (e.g., treatment modality, definition of therapist experience) was calculated using Cohen's kappa coefficient. Reliability of continuous coding categories (e.g., client age, percentage of clients identifying as Hispanic/Latino) were calculated using intraclass correlation coefficients (ICC). Both forms of reliability were considered acceptable if they were greater than .70. Reliability ranged from .72 to .98 across codes. Once reliability was established, the remaining articles continued to be coded by pairs of coders to combat drift.

Each article was coded based on four general areas of interest. First, study level codes were used to characterize the overall study. Examples of study level codes included

whether the clients included in the study were diagnosed with anxiety, depression, or internalizing disorders more broadly, whether randomization of clients to therapist was conducted, and if all therapists received equal supervision. Second, group level codes characterized important variables in therapist experience groups (e.g., What was this group's general level of experience?). Measure level codes characterized each outcome measure on what domain it assessed, who completed the measure and what type of measure was used (e.g., rating scale, semi-structured interview). Finally, information needed to calculate effect sizes (see below) was recorded. The codebook is included in Supplemental Materials. After consensus was reached on all codes, effect size and moderator data were entered into Comprehensive Meta-Analysis Software®.

Power statement

In order to determine if the power of our statistical tests are sufficient to determine practical significance, we considered our overall effect size calculation to be powered if we have at least four studies included for each analysis of interest (Pigott, 2012).

Calculation of effect size

Using statistical information included from each appropriate study, or sub-sample within each study, an effect size was calculated initially using the Practical Meta-Analysis Effect Size Calculator, an online calculator associated with Lipsey and Wilson (2001) book, *Practical Meta-Analysis*. Effect sizes are a standardized way to demonstrate evidence of a result across different studies (Cooper, Hedges, & Valentine, 2009). Some included studies used between-group designs (i.e. mean differences in outcomes between

two or more experience groups) and others used within-group designs, so analyses involved both Cohen's d and Pearson's product moment coefficients. We converted effect sizes into the metric that aligned with the majority of studies. In our sample, the majority of studies used Cohen's d. Thus, we transformed all Pearson's product moment correlations into Cohen's d using the following formula:

$$\sqrt{\frac{4r^2}{1-r^2}}$$

In smaller sample sizes, variance estimates tend to be larger, causing the distribution of Cohen's d to become skewed. To correct for this, all effect sizes were adjusted using Hedge's g to attain an unbiased estimator to be used in analyses (Hedges & Olkin, 1985). After consensus on effect size calculations was reached, all effect sizes were entered into the Comprehensive Meta-Analysis Software® to check our calculations.

Addressing publication bias

One concern to conducting a valid meta-analysis is publication bias, or the decreased likelihood of studies that find negative or null effects to be published and/or widely disseminated (Sutton, 2009). Publication bias can affect the studies identified and included in a meta-analysis, therefore affecting conclusions drawn from overall effect sizes. We examined whether there was evidence of a publication bias in our meta-analysis using a Funnel plot. These plots are a scatterplot of the effect sizes found in the included studies relative to their individual standard error (Greenhouse & Inyengar, 2009). The presence of many studies with large standard error and large effect sizes in combination with few or no small studies with large standard error and small effect sizes

may indicate publication bias (Greenhouse & Inyengar, 2009). Interpreting funnel plot symmetry can be subjective (Sutton, 2009), so we also calculated Egger's linear regression test. Egger's regression test regresses the standard normal deviate of ES of each study from zero onto precision, where slope is the average ES, and where the intercept is expected to be zero. In Egger's regression test, a nonzero intercept indicates asymmetry in the funnel plot, or the possibility of publication bias. For significant models, we also calculated a fail-safe N, a calculation that estimates the number of studies needed for the p-value to become insignificant (Rosenberg, 2005). A Forest plot was also used to represent uncertainty in the estimate and the summary effect, and indicate the extent to which each study contributed to the overall result. It was also used to identify outliers. Funnel plots, Egger's test, fail-safe N, and Forest plots were conducted using the *metafor* package in R (Viechtbauer, 2010). In sub-group meta-analyses where there were less than 10 effect sizes, we did not create funnel plots or use Egger's Regression Test, as the power of the regression test can be too low to distinguish chance from true asymmetry (Higgins & Green, 2011).

Analysis

All analyses were run using the *metafor* package in R (Viechtbauer, 2010). To conduct a meta-analysis, one must first decide which model (i.e., fixed-effect, random-effect, or mixed-effect models) is appropriate to synthesize the data from all studies to calculate the mean effect, or our overall analysis (Hedges & Olkin, 1985). We decided a priori to use a random-effect model, allowing the true effect size to vary from study to study, due to the heterogeneity in methods, client sample, and treatment techniques included in this meta-analysis. In other words, we believed that the effect sizes across

these studies are similar enough to be synthesized, but did not believe that the true effect size is exactly the same in all studies, as we would in a fixed-effect model (Pigott, 2012).

We calculated the overall effect size of therapist experience on client outcome as a weighted mean, in which the weight associated with each study is the inverse of that study's variance. However, multiple measures of outcomes (e.g. global improvement, decrease in social anxiety severity) rated by multiple raters (e.g., therapist, client) per study were the norm in our sample. One potential way to address this is to allow each study to contribute multiple effect sizes to the calculation of the overall effect size. Unfortunately, doing so would assign more weight to studies with more outcome measurements (Borenstein, Hedges, Higgins, & Rothstein, 2009). This approach also treats each outcome from a study as independent from one another. This underestimates the error and overestimates the precision of the calculation, potentially leading to a biased estimate of this summary effect (Borenstein et al., 2009). The issue of dependency was first handled by choosing effect sizes from the total score if it existed. If no total score existed for a measure, we averaged effect sizes from subtest scores. Effect sizes were then grouped into subcategories of outcome measures. In our analyses, we created several aggregated outcome groups. First, we aggregated all outcome measures for each study so each study only had one overall outcome represented in one effect size. We aggregated all outcome measures using the *MAd* package in R, which averages all within-study effect sizes and variances, taking into account the correlations among the within-study outcome measures consistent with Borenstein, Hedges, Higgins, and Rothstein procedures (BHHR; Cooper et al., 2009). The default correlation between within-study effect sizes is .50 (Wampold et al., 1997). When we aggregated all outcome measures together, we kept

the default correlation. After examining the overall relationship between therapist

experience and internalizing client outcome (aim 1), we conducted several sub-group

meta-analyses to investigate how this relationship differed depending on definition of

therapist experience with all outcomes aggregated for each study included (aim 2). To

address aim 3 of this study, we returned to our original dataset, and aggregated outcomes

according to three different measure characteristics. Outcomes were aggregated by

measure domain (where outcomes measuring similar domains (e.g., anxiety) were

averaged), by rater of outcome (where outcomes rated by the same person (e.g., self-

reported) were averaged) or by type of measure (where outcomes using the same measure

format (e.g., rating scale) were averaged). In this round of aggregation, we increased the

correlation between within-study effect sizes to .70 to represent the increase in similarity

between measures within each aggregating group.

For all analyses, we chose to use the Borenstein, Hedges, Higgins, and Rothstein

procedures (BHHR; Cooper et al., 2009), as it is the univariate method found to be least

biased and most precise in a large simulation study of such methods (Del Re, 2015). We

chose this method due to its simplicity and feasibility as compared to multivariate

methods, which require a full covariance matrix, and often lends limited increases in

accuracy (Scammacca, Roberts, & Stuebing, 2014; Wei & Higgins, 2012) .

To accomplish aim 4, we also conducted moderator analyses to search for sources

of heterogeneity in our effect sizes. We examined our moderators using mixed effect

models via the calculation of the Q-statistic. Somewhat similar to the random-effects

model, mixed-effect models assume that there is some variation of effects within each

moderator group. Our included studies were heterogeneous enough even when divided

into different moderator groups that a mixed-effect model was most appropriate to incorporate these between-study differences within each subgroup of studies. To confirm the use of a mixed-effect model, we calculated the Q-statistic, In the case of a large Q-statistic, we first estimated the between-study variance in effect within subgroups of our study using Restricted Maximum Likelihood Estimation (REML). A large Q-statistic suggests that effect sizes are not consistent, so moderator analyses may be indicated. The Q-statistic only tells us that we have different effect sizes in our population, but not what the magnitude of this dispersion is, and if the difference is due to true differences in the ES (Shadish & Haddock, 2009). Thus, we also examined the descriptive statistic I^2, or the proportion of total variation in the estimates of treatment effects that is due to heterogeneity rather than chance (Shadish & Haddock, 2009). The following benchmarks are often used when interpreting I^2: 25% (small heterogeneity), 50% (medium heterogeneity), and 75% (large heterogeneity; Higgins & Thompson, 2002). I^2 of 0% suggests that all of the heterogeneity in the model is due to sampling error. Even when we obtained an I^2 of less than 25% with a significant Q-statistic, we still examined potential moderators, as some researchers have cautioned against using I^2 as definitive proof of meaningful heterogeneity in effect sizes (Cooper et al., 2009). All of our moderators of interest were categorical, so we computed weighted means for each group (formula 1 in Appendix A), variances and standard errors of the group mean effect estimates (formula 2 in Appendix A), tested the null hypothesis that each group is equal to zero (formula 3 in Appendix A), and created confidence intervals around the weighted mean for each group (formula 4 in Appendix A). Of note, our meta-analysis was small, and thus the potential for committing a Type 1 error while conducting moderator analyses

is a concern (Pigott, 2012). Thus, moderators were chosen in advance based on theory and past studies. In addition, we did not examine moderators in smaller sub-group meta-analyses (e.g., k<10), and did not examine moderators if there was not variability in moderator variables of interest.

CHAPTER THREE

RESULTS

Systematic review

A full list of study characteristics is included in Table 1. Studies that included more than one active treatment arm have a single row in the table for each treatment arm (e.g., Bisbey, 1995). Included studies ranged in publication date from 1976 to 2015, and sample sizes ranged from 19 to 416 clients and from 4 to 52 therapists. All studies used reduction in symptoms or diagnostic remission as an outcome variable. Few examined functional outcomes (e.g., life satisfaction; six studies), therapeutic outcomes (e.g., satisfaction with treatment, one study), or service utilization/drop out (four studies). Of the 21 studies, only four collected measures after treatment completed in a follow-up period (Andersson, Carlbring, Furmark, & Group, 2012; Hahlweg, Fiegenbaum, Frank, Schroeder, & von Witzleben, 2001; Propst et al., 1994; Shelton & Madrazo-Peterson, 1978). Findings favored therapists with more experience in 10 comparisons, favored therapists with less experience in four comparisons, and did not significantly favor either group in 22 comparisons.

Final study sample for meta-analytic coding

From the 21 studies included in the systematic review, a further five studies were excluded in overall effect size calculations (Hahlweg et al., 2001; Lewis, 2011; Mason, Grey, & Veale, 2016; Norton, Little, & Wetterneck, 2014; Shelton & Madrazo-Peterson, 1978). In these studies, not enough information was given in order to calculate effect sizes. For further information on reasons for exclusion, please see Table 1. This left 16

22

www.manaraa.com

studies included in final effect size calculations, two of which were unpublished dissertations (Bisbey, 1995; Lewis, 2011).

Of note, many of the included studies compared multiple therapist groups within the same study (Franklin et al., 2003; Nyman, Nafziger, & Smith, 2010; Propst et al., 1994) or examined the same group of therapists according to two different definitions of experience (Huppert et al., 2001; Norton et al., 2014; Podell et al., 2013), creating several comparisons of interest per study. For example, Propst et al. (1994) compared four therapist groups, psychiatrists, psychiatry residents, family medicine residents, and medical students. When all therapist groups were compared, six unique comparisons resulted from one study (see Table 1). For parsimony, we will use the term "study" or k to refer to unique comparisons, even though they may have been included in the same article. From this definition, this meta-analysis contained 31 studies from 16 unique articles. Study analyses yielded a total of 164 effect sizes, representing 1,849 clients and 285 therapists. Five studies from three articles focused on youth clients; the rest of the studies used adult clients only. The client sample was primary female (63.5%) and Caucasian (87.52%). Studies most often used CBT or CBT variants as treatment (k=17), treated anxious clients (k=16), included clients with comorbid diagnoses in the sample (k=16), and used categorical definitions of therapist experience (k=26). Most studies (k=18) relied on multiple types of measures, such as rating scales, behavioral measures, and semi-structured/structured interviews, to examine client outcomes. Similarly, fifteen studies used multiple raters of outcome, with the nine of the remaining studies relying solely on self-rated measures, and seven relying completely on independent evaluator rated measures. Studies were evenly split between those that randomized clients to

therapists (k=11) and those who did not (k=13). This was also the case with the level of supervision provided to study therapists, where 14 studies provided equal amounts of supervision to all therapists, and 15 provided more supervision to less experienced therapists.

Relationship between moderator variables

We examined the relationship between moderator variables in order to better understand our results, and identify any instances of confounding moderators (Pigott, 2012). Comparisons between categorical moderator variables were conducted using Fisher's exact test for count data, as many of the number of comparison in each cell was too low, thus violating chi square assumptions (Agresti, 2002). A comparison between each categorical moderator and client mean age, our only continuous moderator variable, was examined via multiple linear regression. Results from these comparisons can be found in Table 2. From an examination of Table 2, several important relationships between moderator variables emerged. The majority of differential relationships between moderators occurred between studies that treated clients with and without anxiety disorders. Most studies treating clients with anxiety disorders did not randomize clients to therapists (k=10 versus k=5), whereas studies treating clients with depression and clients with mixed internalizing disorders were more likely to randomize clients (k=8 versus k=3). Studies treating clients with anxiety disorders were also more likely to provide equal supervision for all therapists, use CBT over other treatment modalities, and treat youth than studies treating depression or mixed internalizing diagnoses.

Examination of the relationship between experience and the average effect size across all comparison measures

Publication bias

The scatter plot of 31 Hedge's gs against standard error appeared relatively symmetrical and resembled a funnel shape (see Figure 2), suggesting publication bias was unlikely. Egger's regression test was found to be nonsignificant for the model (z=.42, p=.68), also suggesting that publication bias is unlikely. Of note, one Hedge's g from the Traumatic Incident Reduction arm of Bisbey et al. (1995) was an outlier, representing a Hedge's g of -1.50. However, this study had a very small number of clients (N=19) and therapists (N=4), thus contributed little to the overall effect size when corrected using Hedge's G (see Figure 3 for further information). Effect sizes from this study were retained in the model, and all subsequent models, where it was also an outlier. We obtained a significant finding in the random effect model. Thus, we calculated Rosenthal's Fail Safe N to determine the number of unpublished/unidentified studies that would have be found to make our results insignificant. This test returned a value of 23 studies.

Overall estimate of effect size

When all measures were averaged within each study, therapist experience was found to be significantly related to internalizing client outcomes (Hedge's g=.08, p=.03), suggesting that more therapist experience leads to better client outcomes. However, this effect size is very small, so caution is warranted, as the relationship may not be clinically meaningful. A significant Q-statistic was obtained (Q(30)=49.63, p=.01), but I^2 was very

small (4.57%). Moderator analyses were conducted due to the significant Q-statistic, see

Table 3 for further information regarding this model.

Moderators of effect size

Table 4 shows results from the moderator analyses of the aggregated outcomes

model. No variables examined significantly moderated the relationship between therapist

experience and client outcomes.

Examination of the relationship between experience and average effect sizes within

therapist experience definition

We ran four different models, each representing one definition of therapist

experience. These included professional versus paraprofessional, general clinical

experience (e.g., years conducting therapy), degree/schooling level, and experience with

a specific treatment. Experience with a specific client population and professional versus

trainee were only represented in one study each, so no analyses were conducted.

Publication bias

Funnel plots and Egger's regression tests for these sub-group meta-analyses were

not conducted due to a small number of studies. Forest plots for these sub-group meta-

analyses are included in Figures 4-7.

Overall estimate of effect size

Results of random effect models for the four therapist experience definition

categories are included in Table 3. No other experience definition models showed a

relationship between therapist experience and client outcome. No moderator analyses

were conducted, despite a significant Q-statistic obtained in the experience with specific

treatment model ($Q_{within}(7)=22.11$, p=.002), due to the small number of studies and homogeneity of the moderator variables within these studies.

Examination of the relationship between experience and the average effect sizes within measure domains

We ran seven different models, each representing one domain of measures. These included measures focused on anxious symptoms, depressive symptoms, general internalizing symptoms, functioning, satisfaction with treatment, other measures (e.g., number of sessions, relapse rate), and both internalizing and externalizing symptoms.

Publication bias

From a visual inspection of the funnel plots, depressive symptoms and functioning domains appeared to be asymmetrical, suggesting possible publication bias, whereas anxious symptom funnel plots appeared relatively symmetrical (see Figures 8-10). To follow up on models whose funnel plots appeared to asymmetrical, Egger's regression test was used. The Egger's regression tests for functioning measures was significant (k=17, z=2.11, p=.03), suggesting possible publication bias. Egger's regression tests for the depressive symptom domain was nonsignificant. Forest plots for all sub-group meta-analyses are included in Figures 11-17.

Overall estimate of effect size

Results of random effect models for all seven outcome domains are included in Table 3. The relationship between therapist experience and client outcome was significant when measured via "other" measure domains (k=4, Hedge's g=.31, p=.03). The Rosenthal's Fail Safe N returned a value of 1 study. No other outcome domain models showed significant relationships between therapist experience and client outcome.

A significant Q-statistic was obtained in the anxiety symptoms outcome domain ($Q_{within}(18)=50.26$, p<0.001), as well as a large I^2 value (76.79%), so moderator analyses were conducted.

Moderators of effect size

We examined three moderators, client age, inclusion of comorbid disorders, and supervision, to explain heterogeneity within the anxiety symptom outcome domain model. None of the variables examined were significant moderators. Further information regarding these analyses can be found in Table 5.

Examination of the relationship between experience and the average effect sizes within rater

We conducted four different random effect models, each representing one rater of outcome. These included self-rated, independent evaluator rated, caregiver rated, and other rated (e.g., chart review) outcomes.

Publication bias

We examined funnel plots for self and independent evaluator rated models only, as caregiver and other rated models included only two studies each. From visual inspection of the funnel plots, self-rated outcomes appeared to be asymmetrical, where comparisons with high standard error appeared more likely to have positive findings than neutral or negative findings. The funnel plot for independent evaluator rated outcomes appeared relatively symmetrical (Figures 18-19). Neither of the Egger's regression tests for these models were significant (p>.25), suggesting publication bias is unlikely. Forest plots for self-rated and independent evaluator rated outcomes are included in Figures 20-21.

Overall estimate of effect size

We included the results of the random effect model for outcomes aggregated by rater in Table 3. The relationship between therapist experience and outcome was significant when measures via self-rated outcomes (Hedge's g=.12, p=.003), suggesting that when outcomes were rated by the clients themselves, therapists with more experience attained better outcomes than therapists with less experience. The Rosenthal's Fail Safe N returned a value of 12 studies. No other model showed a significant relationship between therapist experience and outcomes. We obtained a significant Q-statistic in the self-rated (Q_{within}(22)=39.02, p=.01)) and independent evaluator rated models (Q_{within}(20)=44.83, p=.001)). In the self-rated outcome model, we obtained a very small I^2 value (0.02%), while in the independent evaluator rated outcome model, we obtained a medium to large I^2 value (58.86%)

Moderators of effect size

We examined four and five moderators to explain heterogeneity within the self-rated and independent rated outcome models, respectively (Tables 6-7). None of the moderators were significant.

Examination of the relationship between experience and the average effect sizes within measure types

We conducted three random effect models, each representing a type of measure. These included rating scales, semi-structured/structured interviews, and behavioral tasks/measures.

Publication bias

We examined funnel plots for rating scale and semi-structured/structured models only, as behavioral tasks/measures models included only two comparisons. From visual inspection of the funnel plots, rating scale outcomes appeared to be asymmetrical, where comparisons with high standard error appeared more likely to have positive findings than neutral or negative findings (Figure 22). The funnel plot for independent evaluator rated outcomes appeared relatively symmetrical, though several comparisons fell outside of the funnel plot (Figure 23). Neither of the Egger's regression tests for these models were significant (p>.40), suggesting publication bias is unlikely. Forest plots for the rating scale and semi-structured/structured models are included in Figures 24-25.

Overall estimate of effect size

We included the results of the random effect model for outcomes aggregated by type of measure in Table 3. None of the models showed significant relationships between therapist experience and outcomes. We obtained a significant Q-statistic in the rating scale ($Q_{within}(26)=44.09$, p=.01)) and semi-structured/structured interview models ($Q_{within}(19)=55.27$, p<.001)). In the rating scale outcome model, we obtained a very small I^2 value (1.56%), while in the semi-structured/structured interview outcome model, we obtained a large I^2 value (77.07%).

Moderators of effect size

We examined four and six moderators to explain heterogeneity within the rating scale and semi-structured/structured interview outcome models, respectively (Tables 8-9). None of the moderators examined were significant.

CHAPTER FOUR

DISCUSSION

This meta-analysis sought to better understand the relationship between therapist experience and internalizing client outcomes. As hypothesized, there was a significant positive relationship between therapist experience and internalizing client outcome when all outcomes were aggregated. This suggests therapists with more experience attain better outcomes with internalizing clients than less experienced therapists. Still, this relationship represented a very small effect size (Hedge's g=.079), calling into question the clinical importance of this finding. In addition, Rosenthal's Fail Safe N was relatively low (n=23), suggesting that this finding could easily become nonsignificant with the addition of unpublished studies. None of the hypothesized variables moderated the relationship between therapist experience and internalizing client outcomes when all outcomes were aggregated.

After conducting sub-group analyses focused on the relationship between experience and outcome within different therapist experience definitions and outcome measure characteristics, the positive relationship between therapist experience and client internalizing outcomes emerged in "other measure" domains and self-reported outcomes. Findings regarding the "other" measure domains were consistent with previous findings by Stein and Lambert (1995). Outcome measures coded under this category included client drop out from treatment (Bright et al., 1999; Thompson, Gallagher, Nies, & Epstein, 1983), total number of sessions to reach completion (Howard, 1999), and relapse rates (Howard, 1999). In some of these studies, clients of therapists with more experience

dropped out less and took less time to complete treatment (Bright et al., 1999; Howard, 1999). However, the number of comparisons included in this sub-group meta-analysis was low (k=4) and Rosenthal's Fail Safe N was one study, so caution is definitely indicated when interpreting these results. The relationship between therapist experience and self-rated outcomes was contrary to our hypothesis, and previous findings by Stein and Lambert (1995) where a relationship between therapist experience and outcomes was only found when researchers used independent evaluators rather than client self-report. There are several potential explanations or this. First, clients with internalizing disorders in particular may be more accurate reporters of outcome as they live with these internalized symptoms daily, some of which may not present in a relatively short interview with an independent evaluator. Clients of experienced therapists might have known that their therapist was highly experienced (e.g., if the therapist were the principal investigator of the study), thus went into treatment expecting better results and rated themselves as improving more. Finally, independent evaluators may have been more conservative in their ratings of improvement than clients.

Limitations of the current study and therapist experience literature

Despite over half of our studies comprising of new data since previous meta-analyses were conducted, we found limited support for increased therapist experience leading to better internalizing client outcomes. Importantly, methodological concerns identified in previous meta-analyses were still present, even in newer studies. Over half of the studies did not randomize clients to therapists, a vital step in truly determining whether therapist experience matters. In the studies that did not randomize clients to therapists, more experienced therapist often received more complex clients, possibly

obscuring any additional improvements in outcome experience may provide. Conversely, clients may not have been blinded to therapist experience level, potentially providing an unfair advantage to experienced therapists. Several studies containing small samples of clients (<50 per group) and/or small samples of therapists (<5 therapists per group) were also included in this meta-analysis (Bisbey, 1995; Propst et al., 1994; Russell & Wise, 1976). These studies also did not conduct a priori power analyses, and thus were potentially underpowered to detect meaningful differences between therapist experience levels. Studies rarely examined client outcomes beyond discharge from treatment, leaving unknown the possibility that the maintenance of treatment gains may differ by therapist experience level.

Beyond design limitations of the included studies, we could not investigate many moderators of interest due to limited information provided. More than half of the studies did not provide basic race/ethnicity information for the clients and therapists included in the sample. Of the remaining studies, less than half identified clients and therapists as anything but "Caucasian" or "Other." Our sample of studies also lacked heterogeneity in terms of client age, so examining the differential effect of therapist experience on internalizing client outcomes between children and adolescents, as identified in Weisz et al. (1995), was not possible. Other variables were initially of interest but could not be coded due to limited reporting or lack of heterogeneity in the included studies, including treatment setting (e.g., college counseling center, outpatient medical center), treatment delivery (e.g., individual versus group treatment), whether treatment was manualized, client recruitment procedures, and therapist caseload.

Operationalization of therapist experience remained inconsistent, even within definitions that were purportedly the same, potentially leading to uninterpretable results. In many of the early studies of therapist experience, therapists were often classified as either "professionals," meaning that they had attained some level of specialized training in mental health, whether that be a Master's or Doctoral degree, or "paraprofessionals," those who had no formalizing training in mental health past a bachelor's degree (e.g., Bright et al., 1999; Russell & Wise, 1976). However, there was inconsistent categorization of these two groups across studies, where some therapists considered paraprofessionals in one study would likely be considered professionals in another. Our choice to treat these studies as equivalent may have masked important differences between groups.

Finally, we obtained significant Egger's regression test values for the functioning measures sub-group meta-analyses raising concerns of publication bias. Despite efforts to combat publication bias, we only identified three unpublished studies, all dissertations (Bisbey, 1995; Lewis, 2011; Podell et al., 2013), one of which was not included in our final sample (Lewis, 2011). Thus, publication bias is certainly possible. In addition, although Egger's regression test is more powerful than nonparametric tests such as rank correlation tests, it can be underpowered to detect publication bias in a smaller sub-group meta-analysis like this one (Sutton, 2009). It is also important to note that Egger's regression test does not specifically test for publication bias; it tests for funnel plot asymmetry, which may be due to poor methodology in included studies or small sample sizes (Higgins & Green, 2011) so caution is indicated when interpreting these results as evidence of publication bias.

Implications and future directions

Although this meta-analysis heeded the advice of previous meta-analysts and researchers, the relationship between therapist experience and internalizing client outcomes specifically appears to be modest at best. Thus, the current literature suggests it would be appropriate to consider increasing the role of lay therapists in treating clients with internalizing disorders. Several studies of task-shifting mental health services from professionals to laypeople in low and middle income countries with depressed and traumatized populations have found promising outcomes (Chibanda et al., 2011; Murray et al., 2013; Petersen, Bhana, Baillie, & Consortium, 2012; Petersen, Hancock, Bhana, & Govender, 2014). However, task-shifting may be a natural fit in these settings due to the lack of essentially any mental health workers, limiting consumer options. Extension of this work to these and other internalizing populations in higher income countries are needed to determine the feasibility, acceptability, and effectiveness of lay therapists in mental health systems where qualified mental health workers are present.

Work to address the methodological concerns in this literature is clearly needed. Although comparisons between therapists at different levels of experience is valuable, there are several confounding factors that can make the relationship difficult to examine. Within all levels of training and experience, some therapists still remain more effective than others (Nissen-Lie et al., 2016). While the cause of these differences is still debated (see Blow, Sprenkle, & Davis, 2007; Kazdin, 2005; Lambert, 2005), several examinations of whether the outcomes achieved by specific therapists improve as those therapists gain experience have been conducted (e.g., Goldberg et al., 2016; Leon, Martinovich, Lutz, & Lyons, 2005; Owen, Wampold, Kopta, Rousmaniere, & Miller,

2016). Initial findings are promising, suggesting that as therapists gain more experience treating clients, they attain better client outcomes and become more efficient. Future work that examines this question using this methodology within internalizing client populations would be invaluable to determine appropriate strategies to get more clients treatment, and inform training methods within Masters and Doctoral programs in psychology, social work, and allied mental health fields. As more studies emerge in this area, meta-analysis might also be an appropriate method to synthesize these findings, and determine whether this methodology provides more evidence for the role of experience on client outcomes.

In addition to the use of longitudinal designs within the same therapist, more studies using experimental designs (i.e., randomizing clients to therapists), treating youth, treating diverse populations, and using newer treatment methodologies (e.g., mindfulness-based practices) are needed to understand the nuanced relationship between therapist experience and internalizing client outcomes. Nonetheless, internalizing clients are not the only client population at risk of never receiving treatment due to a lack of available qualified therapists. Meta-analyses examining the relationship between therapist experience and client outcomes in other subpopulations such as substance abuse, marital difficulties, and personality disorders might provide differing evidence for task shifting as a viable solution for this shortage.

Conclusion

Any definitive verdict regarding the role of therapist experience and internalizing client outcomes is difficult due to the small number of high-quality studies examining this question, pervasive methodological issues, and limited moderator variables of

interest, but there currently remains very little data to support the case that more experienced clinicians gain better client outcomes. Despite limitations, the present study included newer, more methodologically sound studies, and is timely for implementation and service provision concerns. Renewed research in this basic area of psychotherapy research using longitudinal within-therapist designs is necessary to address many concerns in the literature, and to best serve the millions of Americans in need of mental health services yearly.

FIGURES
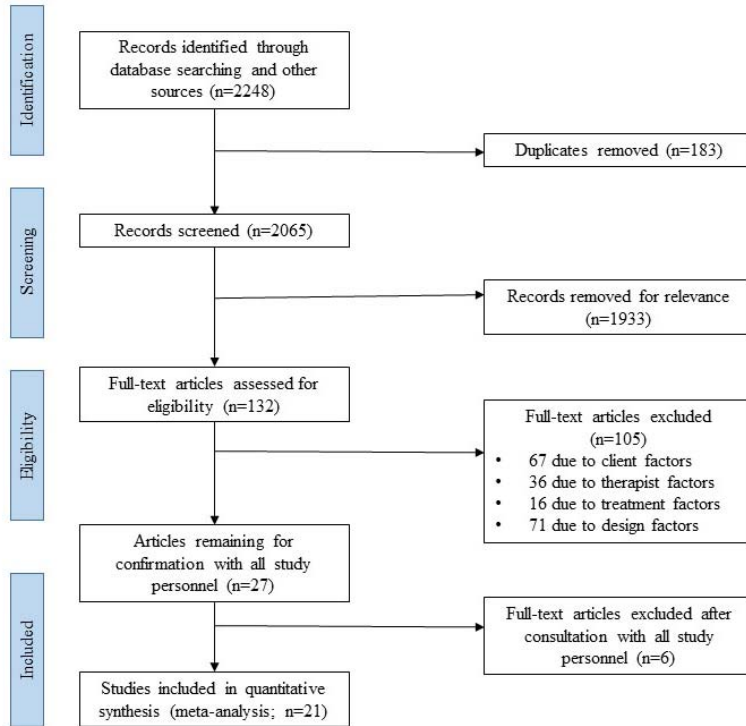
Figure 1. Article screening, Numbers add up to more than 104 because articles could meet exclusion for several factors.
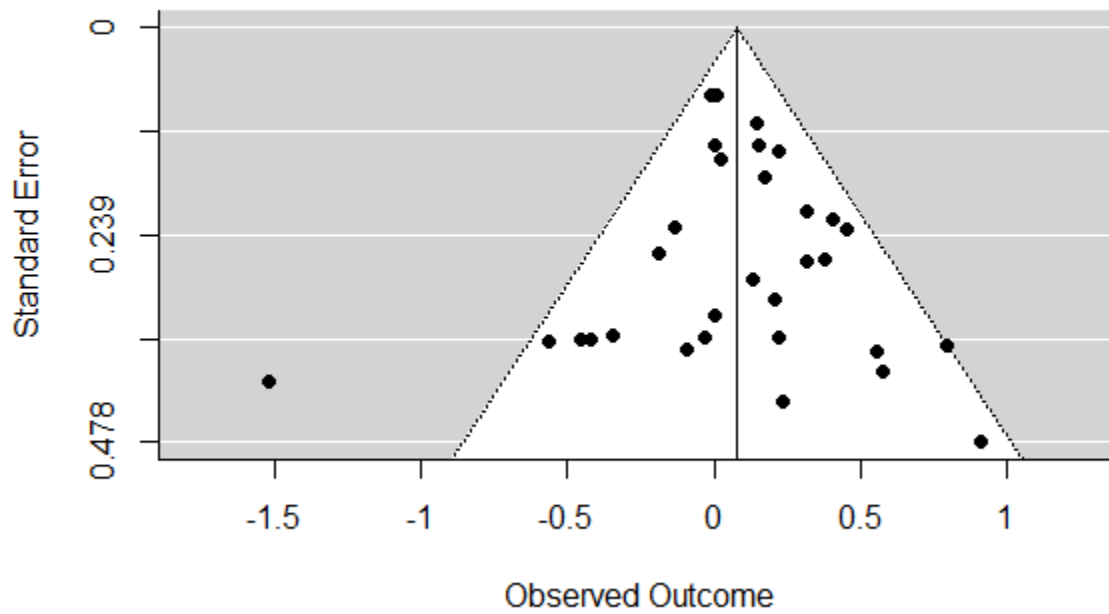
Figure 2. Funnel plot of all aggregated outcomes.
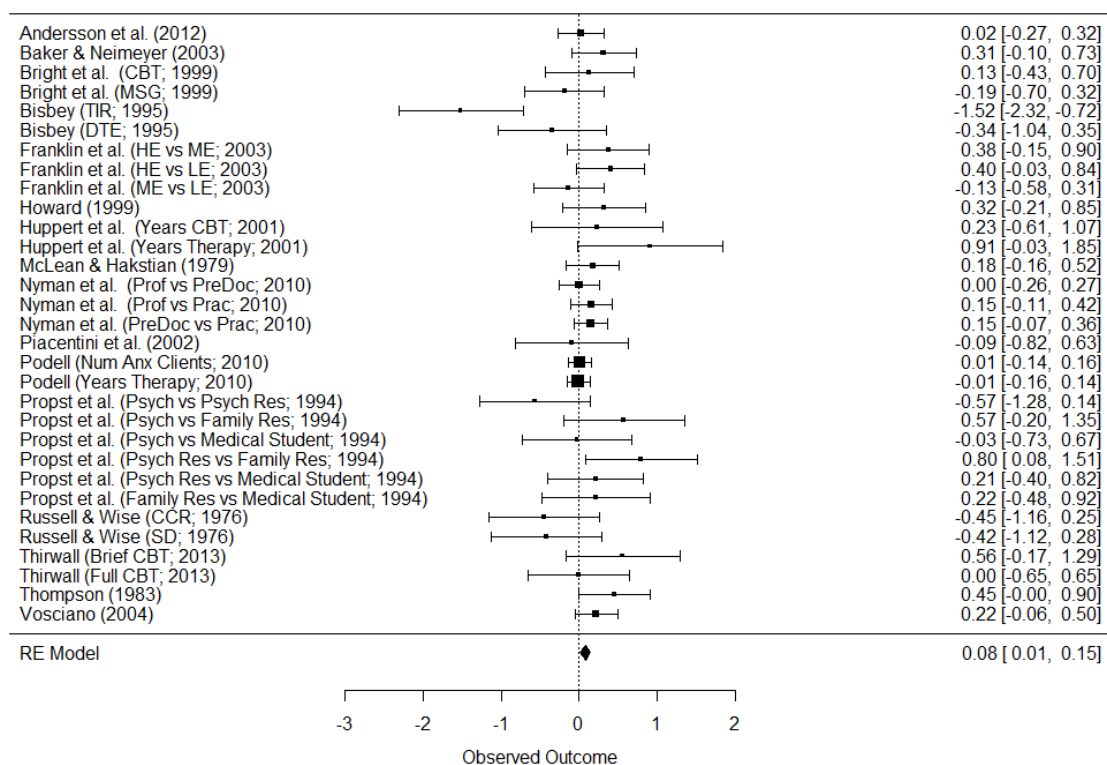
Figure 3. Forest plot of all aggregated outcomes. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident, CCR= Cue-Controlled Relaxation, SD=Systematic Desensitization.
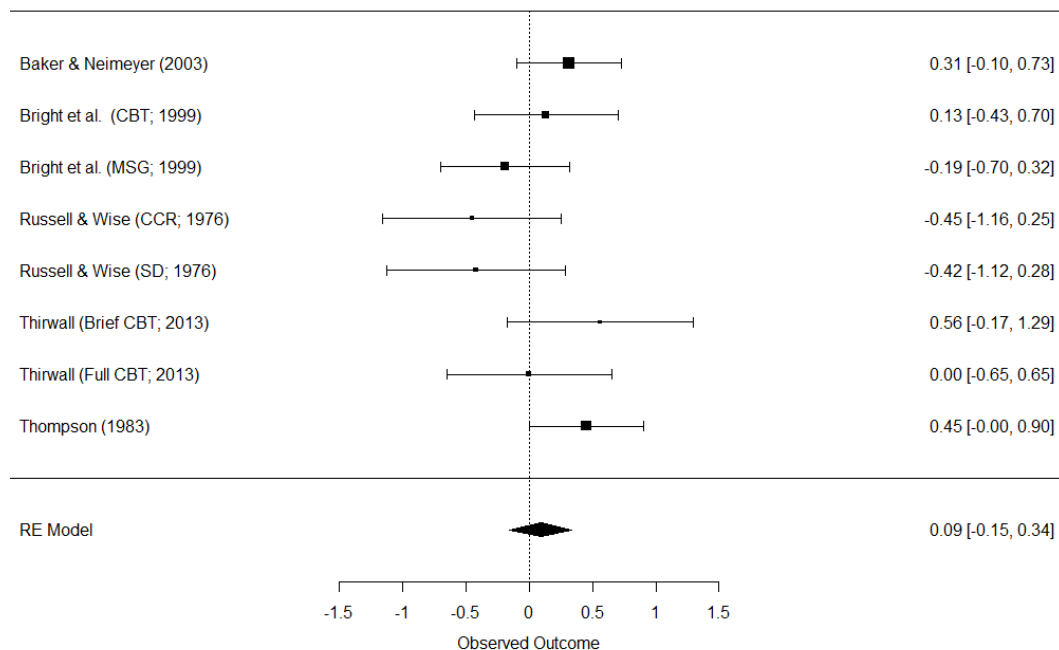
Figure 4. Forest plot of professional versus paraprofessional experience definitions. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, CCR= Cue-Controlled Relaxation, SD=Systematic Desensitization.
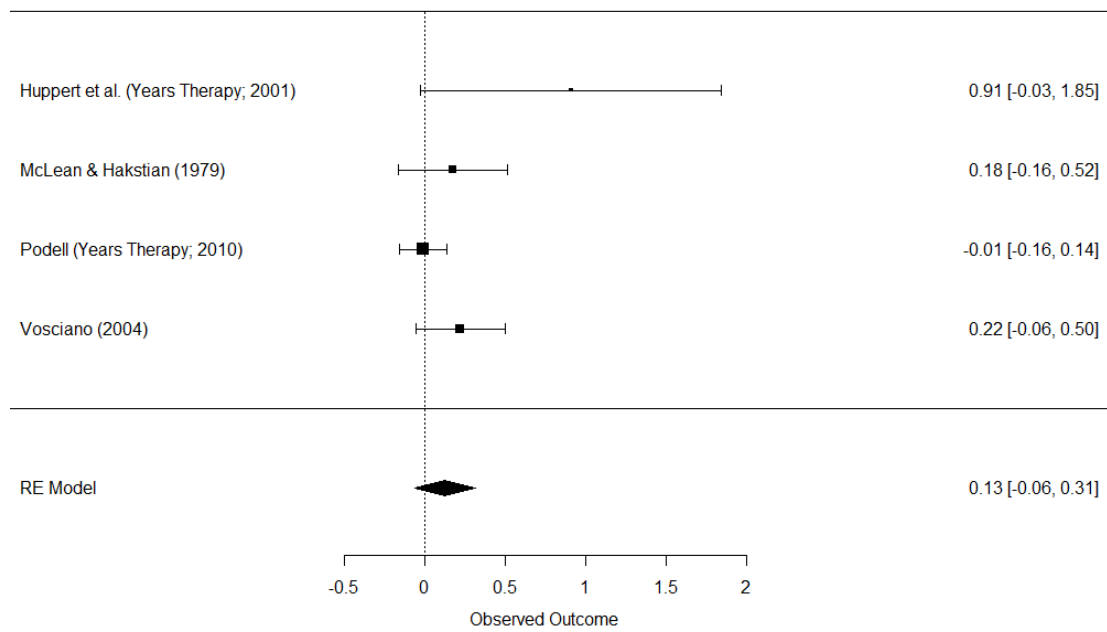
Figure 5. Forest plot of general clinical experience definitions.
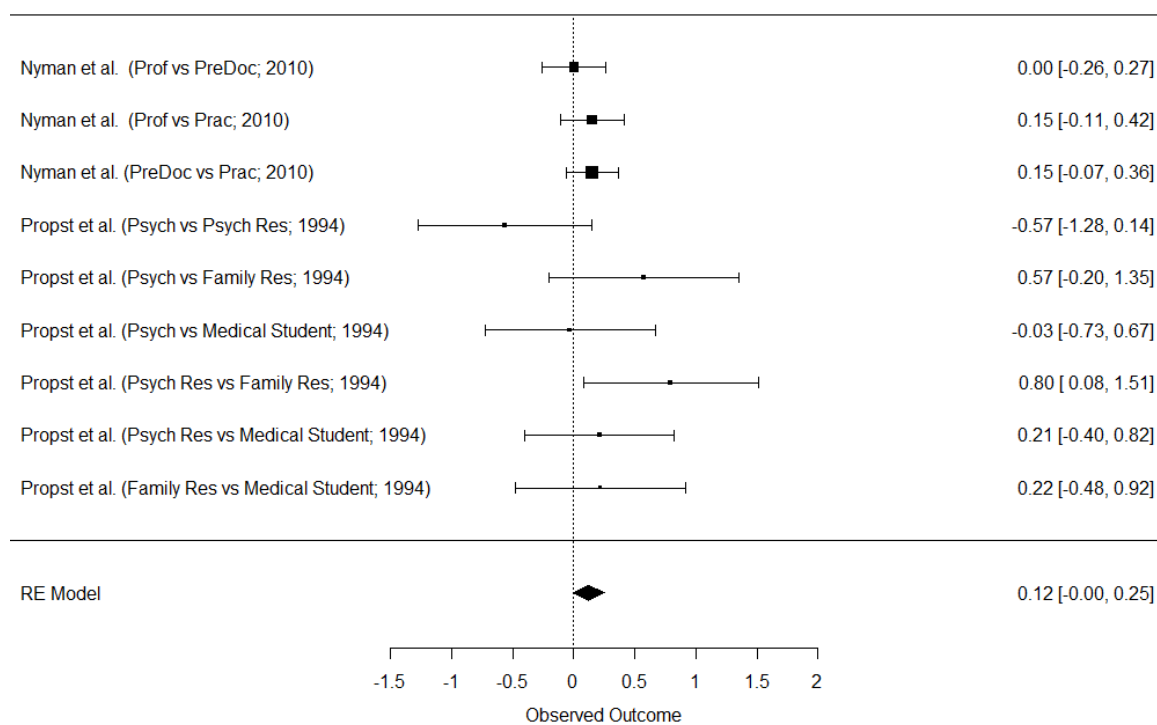
Figure 6. Forest plot of degree/schooling level definitions. Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.
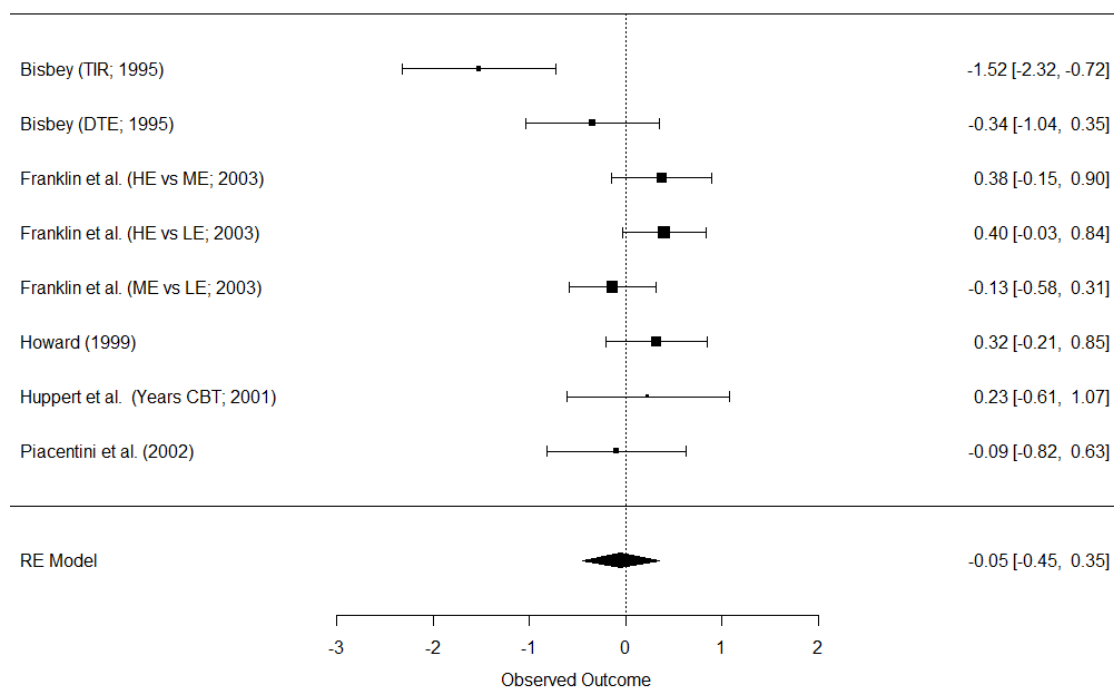
Figure 7. Forest plot of experience with a specific treatment definitions. TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience
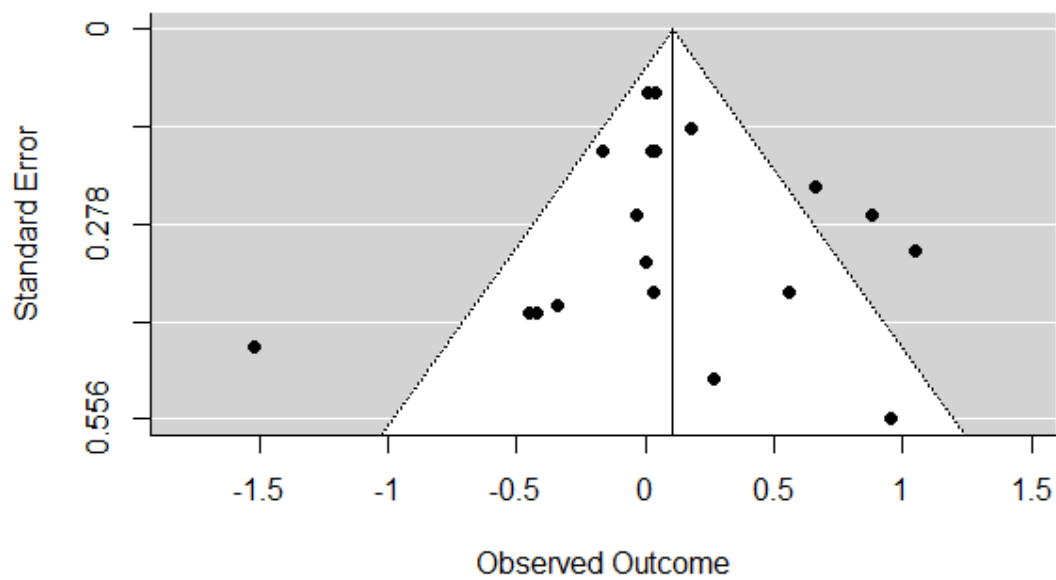
Figure 8. Funnel plot of aggregated anxiety measures.
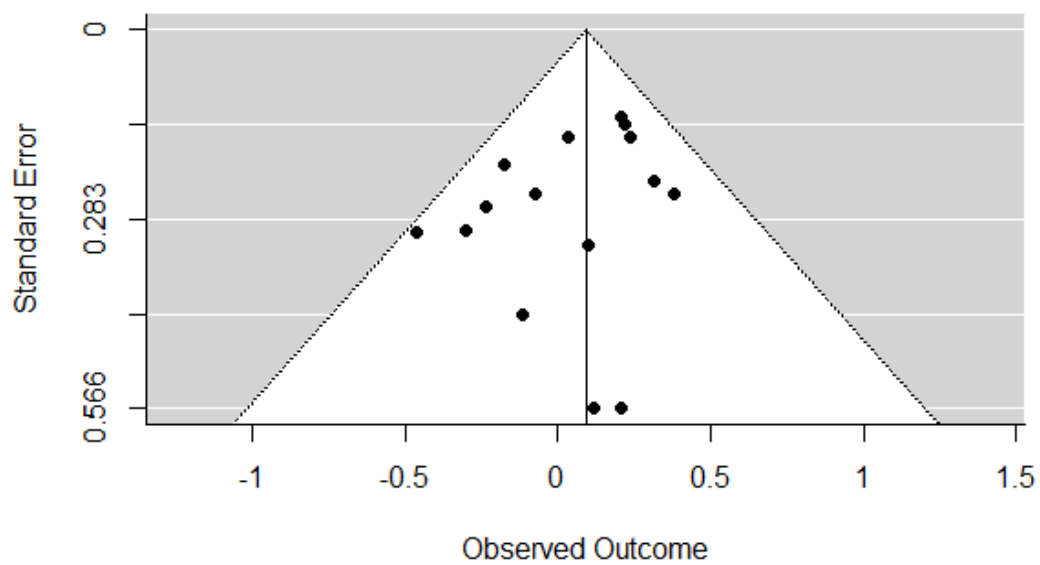
Figure 9. Funnel plot of aggregated depression measures.
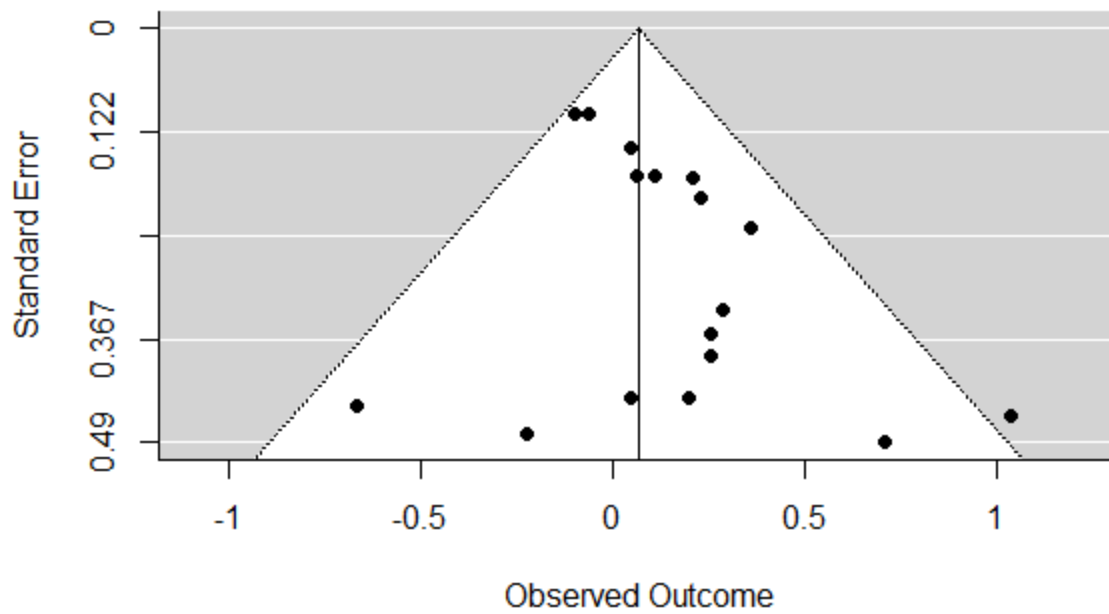
Figure 10. Funnel plot of aggregated functioning measures.
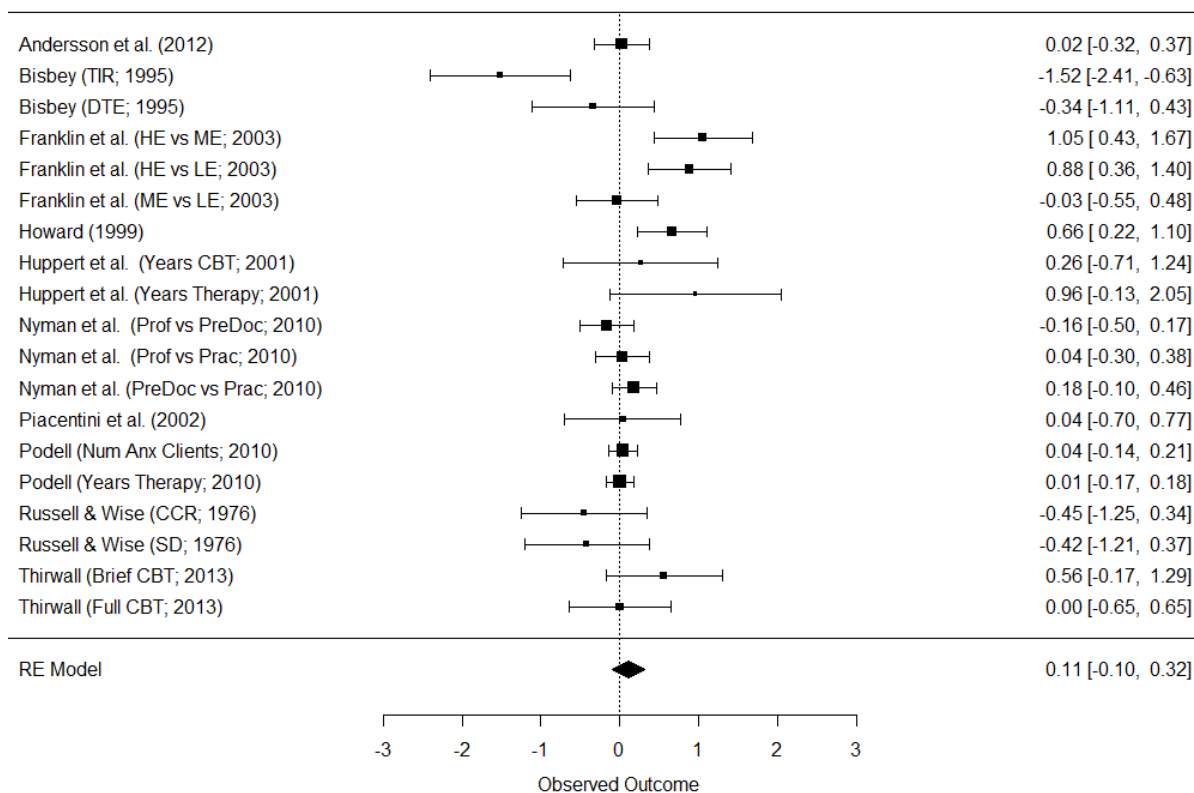
Figure 11. Forest plot of aggregated anxiety measures. CBT=Cognitive Behavioral Therapy, TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients, CCR= Cue-Controlled Relaxation, SD=Systematic Desensitization.
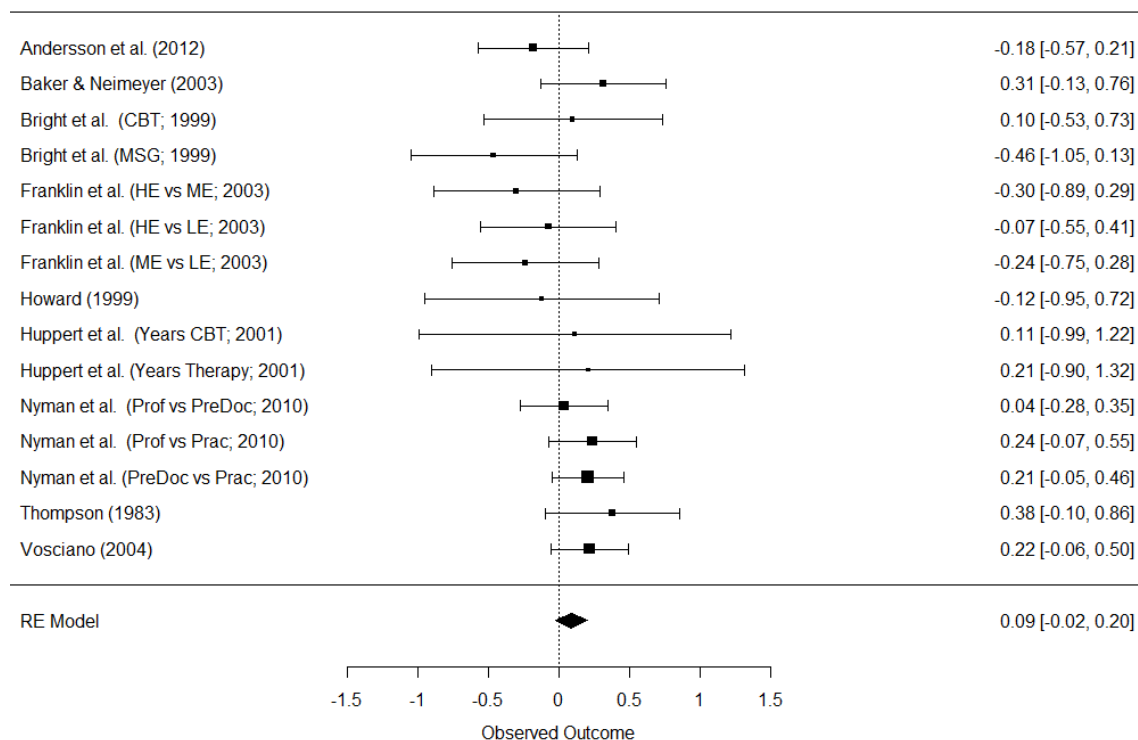
Figure 12. Forest plot of aggregated depression measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, HE=High Experience, ME=Medium Experience, LE=Less Experience, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student.
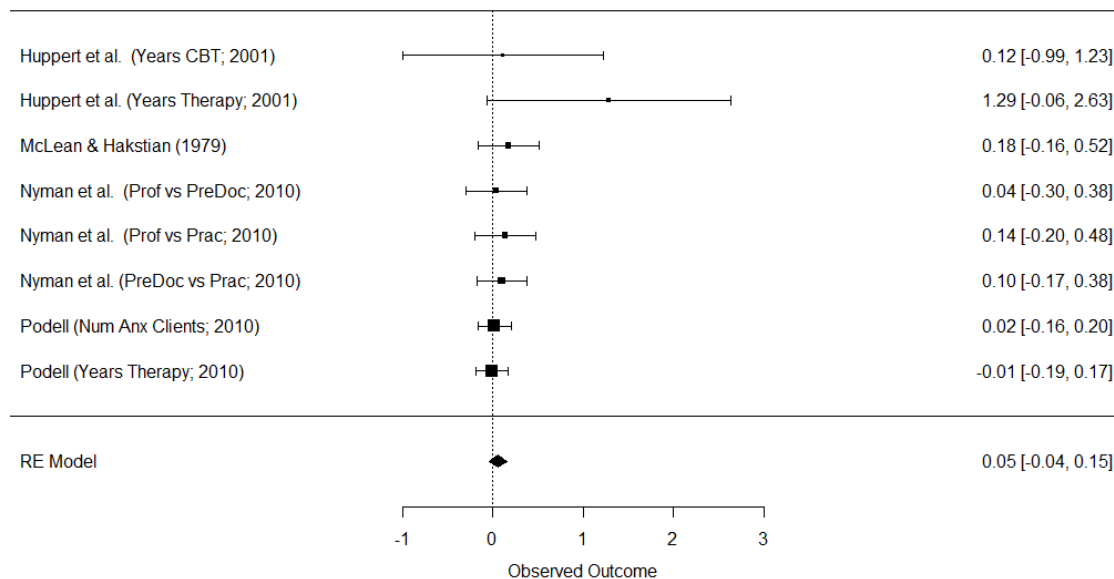
Figure 13. Forest plot of aggregated internalizing measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients.
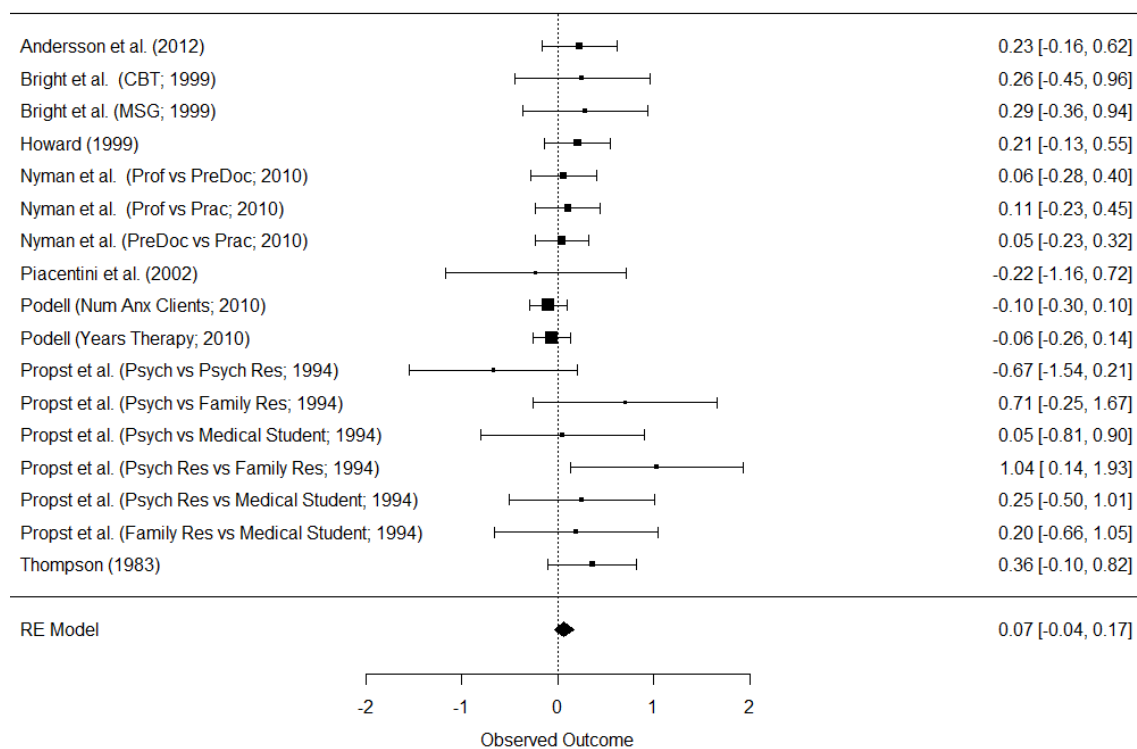
Figure 14. Forest plot of aggregated functioning measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.
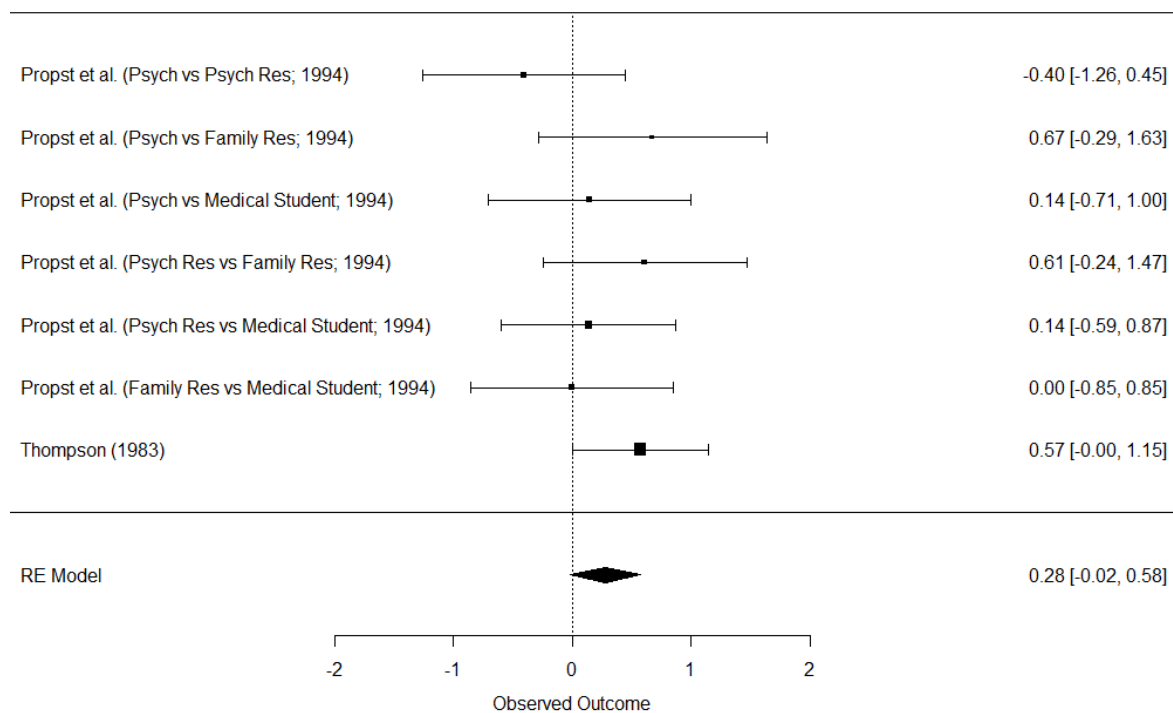
Figure 15. Forest plot of aggregated satisfaction measures. Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.
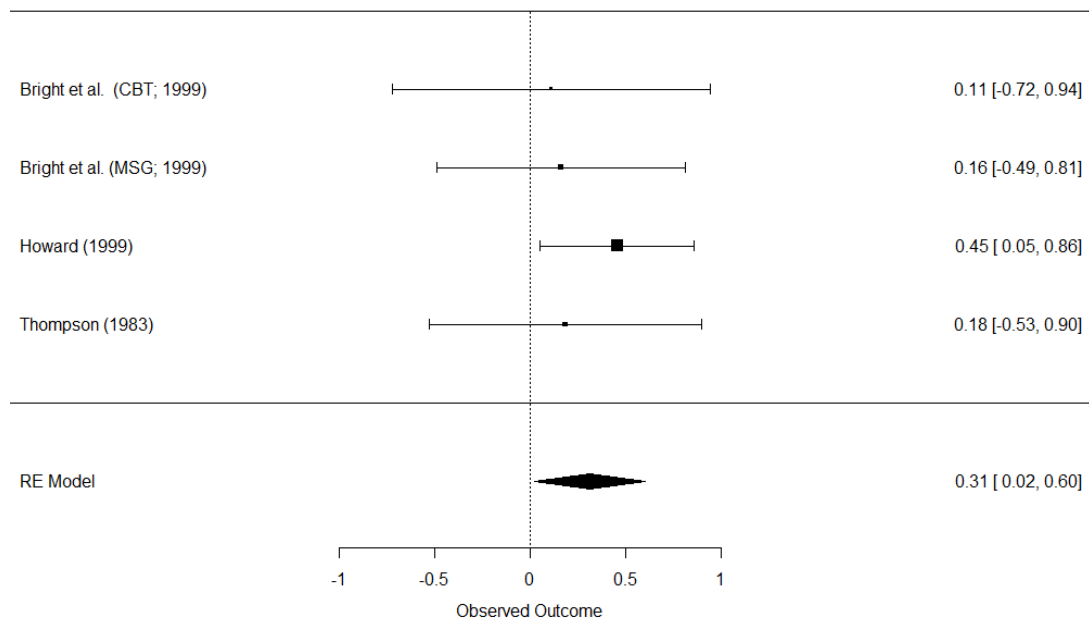
Figure 16. Forest plot of aggregated other measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group.
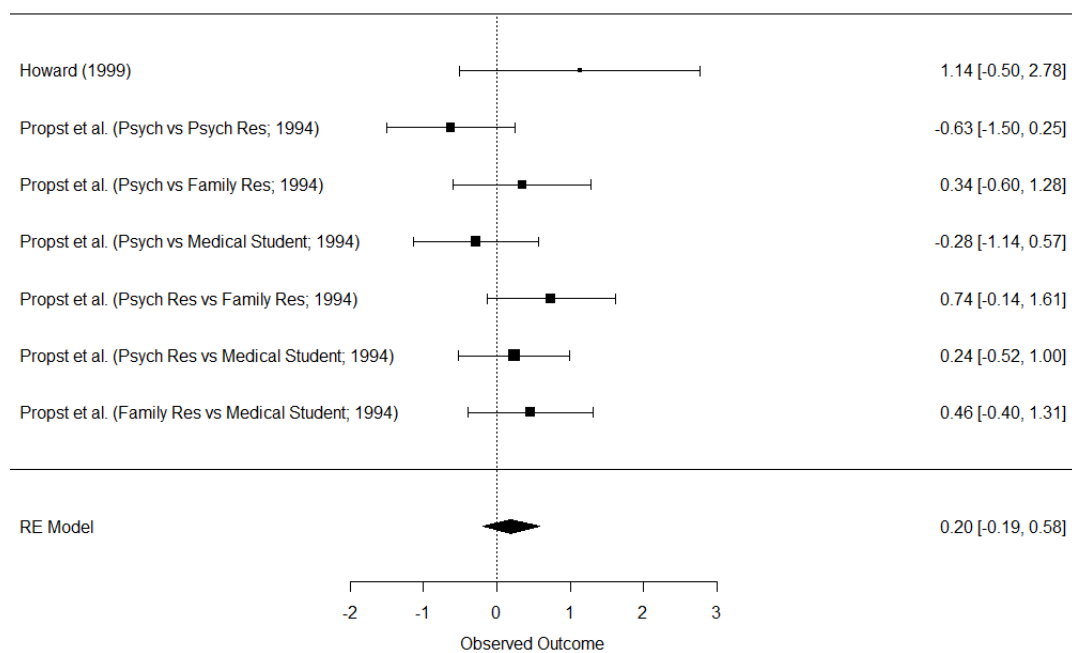
Figure 17. Forest plot of aggregated internalizing and externalizing measures. Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.
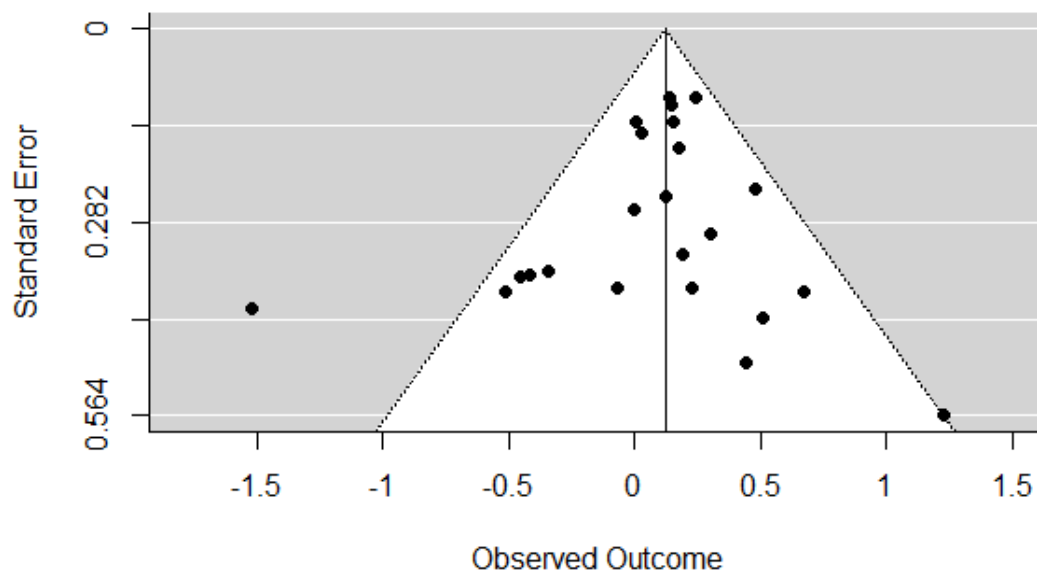
Figure 18. Funnel plot of aggregated self-rated measures.
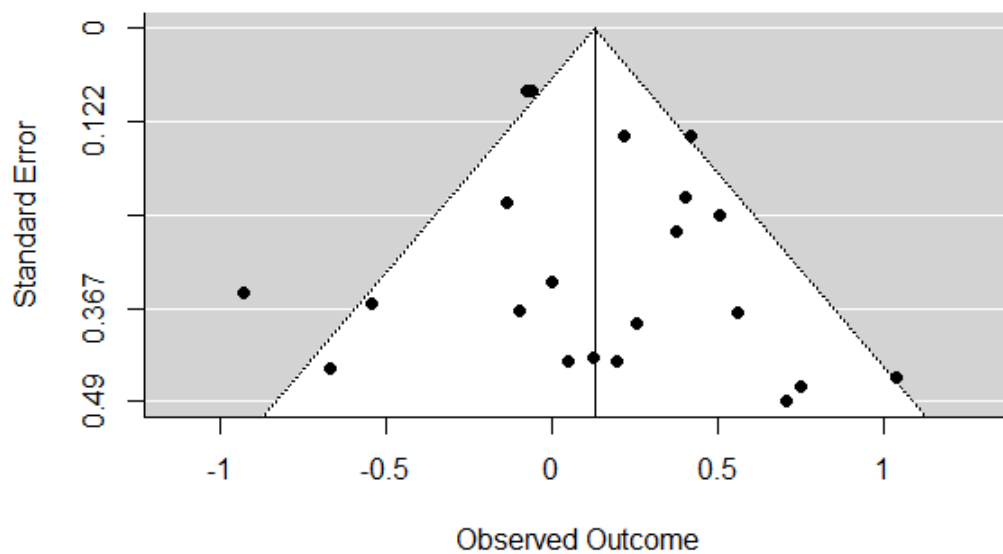
Figure 19. Funnel plot of aggregated independent evaluator rated measures.
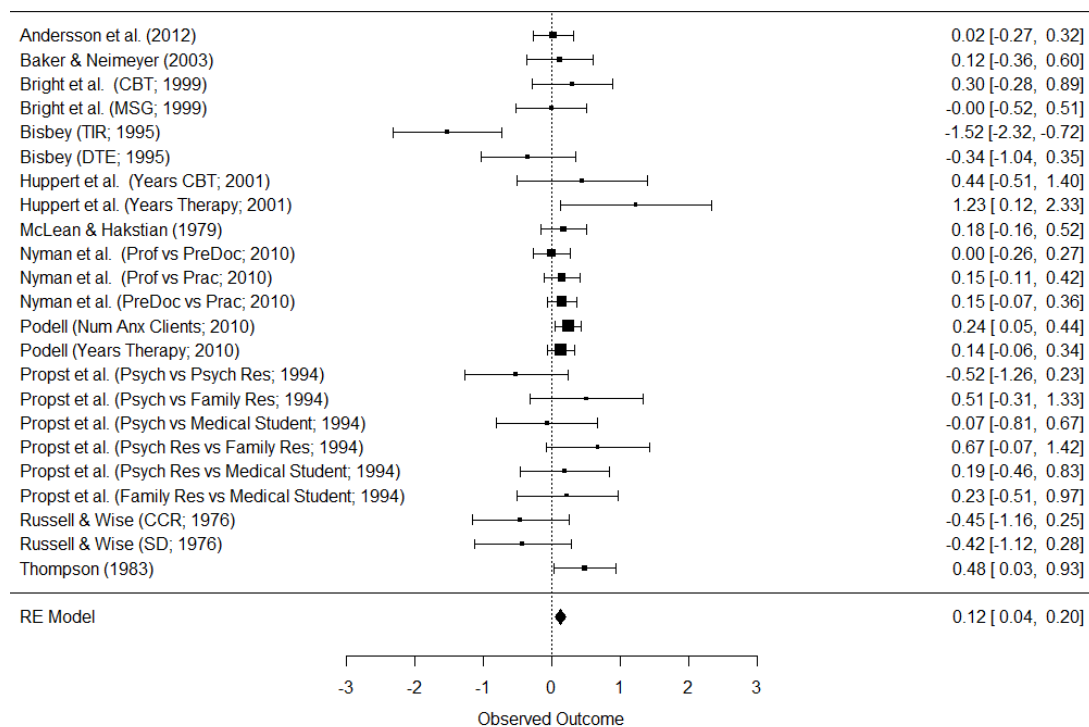
Figure 20. Forest plot of aggregated self-rated measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident, CCR= Cue-Controlled Relaxation, SD=Systematic Desensitization.
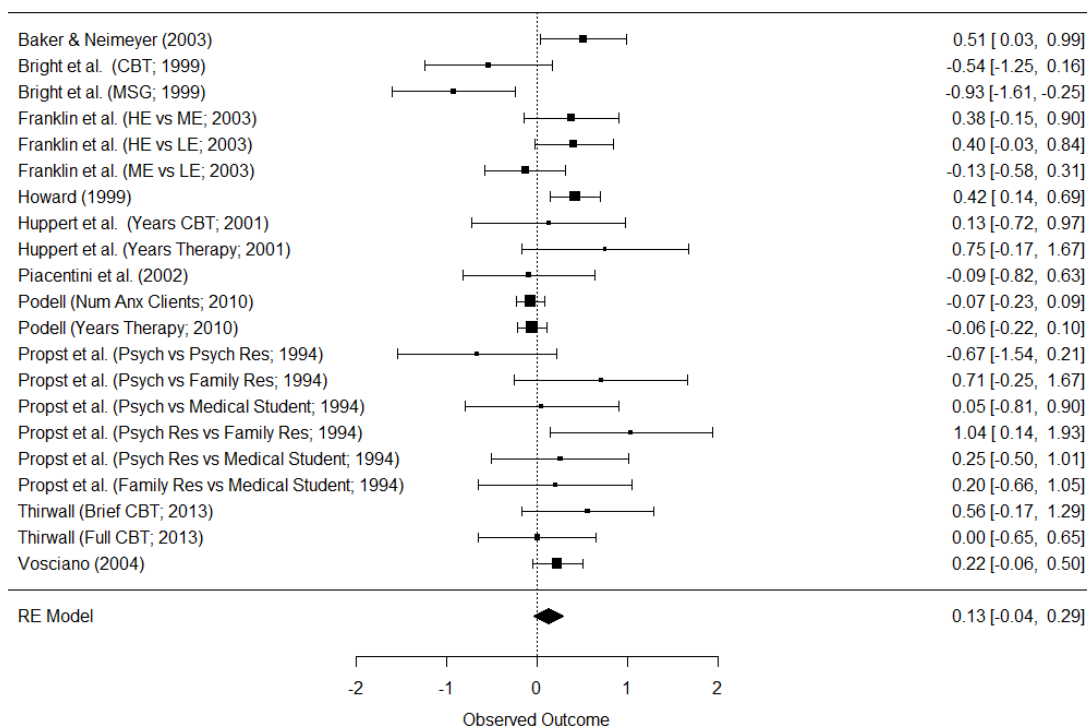
Figure 21. Forest plot of aggregated independent evaluator rated measures. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.
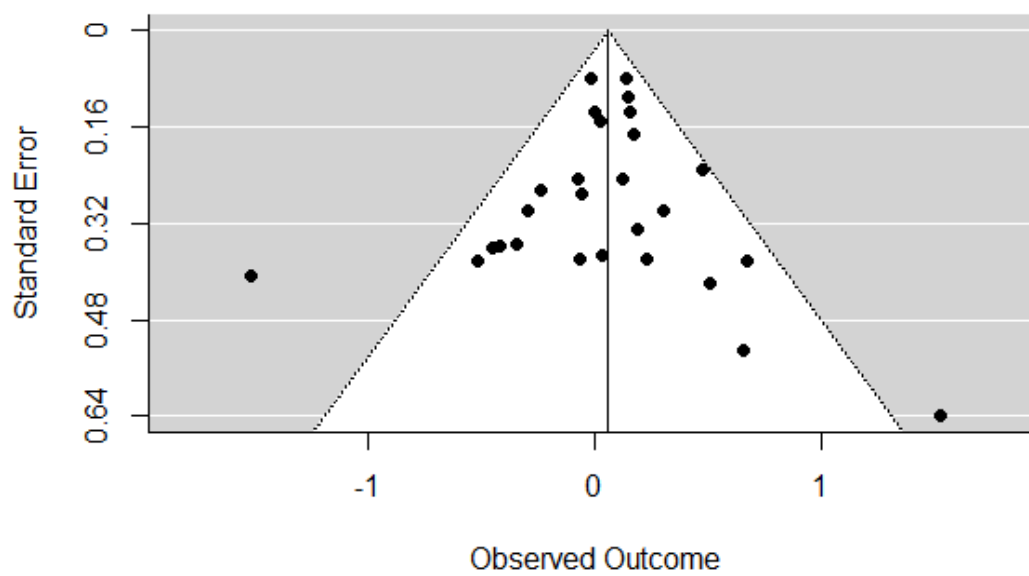
Figure 22. Funnel plot of aggregated rating scales.
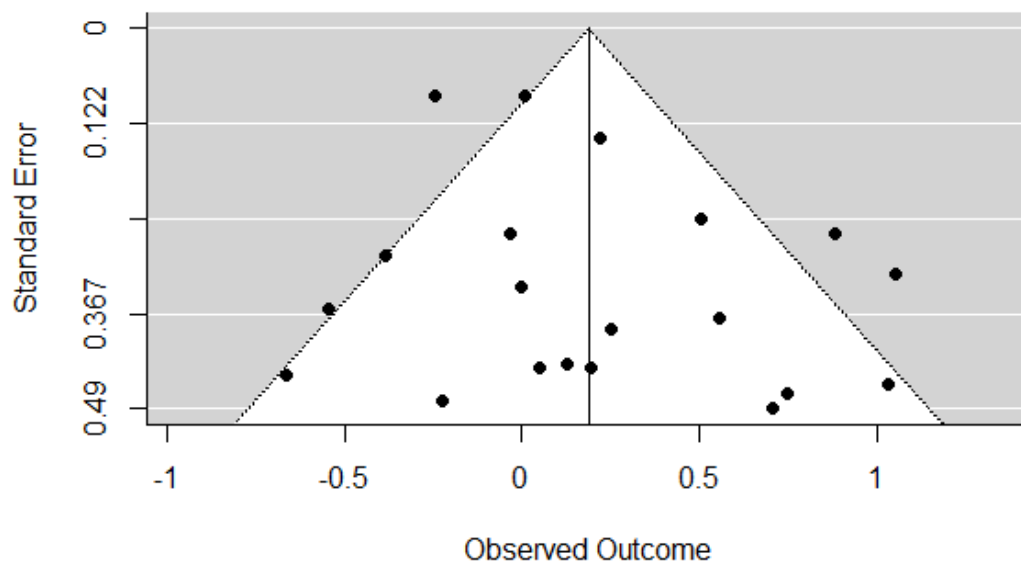
Figure 23. Funnel plot of aggregated semi-structured/structured interviews.

Andersson et al. (2012)	0.02 [-0.27, 0.32]
Baker & Neimeyer (2003)	0.12 [-0.36, 0.60]
Bright et al. (CBT; 1999)	0.30 [-0.28, 0.89]
Bright et al. (MSG; 1999)	-0.06 [-0.59, 0.48]
Bisbey (TIR; 1995)	-1.52 [-2.32, -0.72]
Bisbey (DTE; 1995)	-0.34 [-1.04, 0.35]
Franklin et al. (HE vs ME; 2003)	-0.30 [-0.89, 0.29]
Franklin et al. (HE vs LE; 2003)	-0.07 [-0.55, 0.41]
Franklin et al. (ME vs LE; 2003)	-0.24 [-0.75, 0.28]
Huppert et al. (Years CBT; 2001)	0.66 [-0.38, 1.69]
Huppert et al. (Years Therapy; 2001)	1.53 [0.28, 2.78]
McLean & Hakstian (1979)	0.18 [-0.16, 0.52]
Nyman et al. (Prof vs PreDoc; 2010)	0.00 [-0.26, 0.27]
Nyman et al. (Prof vs Prac; 2010)	0.15 [-0.11, 0.42]
Nyman et al. (PreDoc vs Prac; 2010)	0.15 [-0.07, 0.36]
Piacentini et al. (2002)	0.04 [-0.70, 0.77]
Podell (Num Anx Clients; 2010)	0.14 [-0.02, 0.29]
Podell (Years Therapy; 2010)	-0.02 [-0.18, 0.13]
Propst et al. (Psych vs Psych Res; 1994)	-0.52 [-1.26, 0.23]
Propst et al. (Psych vs Family Res; 1994)	0.51 [-0.31, 1.33]
Propst et al. (Psych vs Medical Student; 1994)	-0.07 [-0.81, 0.67]
Propst et al. (Psych Res vs Family Res; 1994)	0.67 [-0.07, 1.42]
Propst et al. (Psych Res vs Medical Student; 1994)	0.19 [-0.46, 0.83]
Propst et al. (Family Res vs Medical Student; 1994)	0.23 [-0.51, 0.97]
Russell & Wise (CCR; 1976)	-0.45 [-1.16, 0.25]
Russell & Wise (SD; 1976)	-0.42 [-1.12, 0.28]
Thompson (1983)	0.48 [0.03, 0.93]
RE Model	0.06 [-0.01, 0.13]

Figure 24. Forest plot of aggregated rating scales. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, TIR=Traumatic Incidence Reduction, DTE=Direct Therapeutic Exposure, HE=High Experience, ME=Medium Experience, LE=Less Experience, Prof=Professional, PreDoc=Predoctoral Intern, Prac=Practicum Student, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident, CCR= Cue-Controlled Relaxation, SD=Systematic Desensitization.
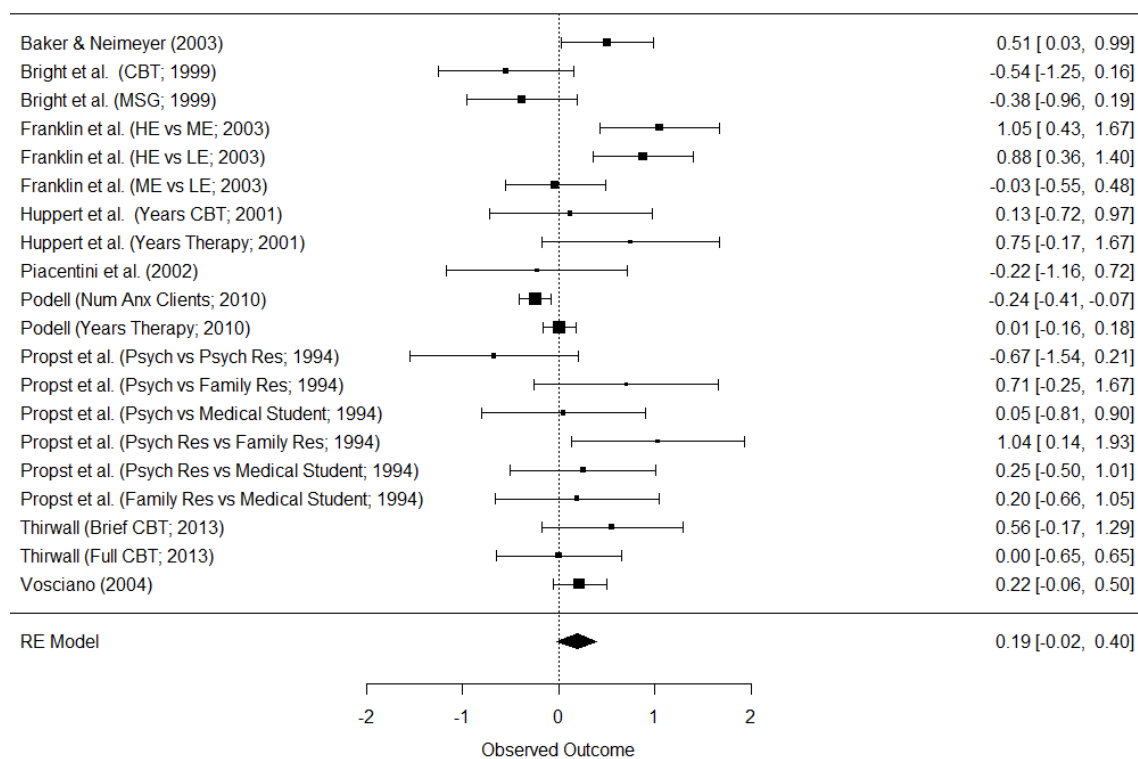
Figure 25. Forest plot of semi-structured/structured interviews. CBT=Cognitive Behavioral Therapy, MSG=Mutual Support Group, HE=High Experience, ME=Medium Experience, LE=Less Experience, Num Anx Client=Number of Anxious Clients, Psych=Psychiatrist, Psych Res=Psychiatry Resident, Family Res=Family Practice Resident.

TABLES

Table 1
Results of systematic review

Authors	Year	Subgroup used in analyses	Experience definition category	# Clients (# Therapists)	Client Age Group	Client Disorder	Treatment	Outcome Measures	Time points	Number of Effect Sizes Contributed	Information used to calculate effect size	Effect Size*
Andersson, Calbring, Furmark, & SOFIE Research Group	2012	Licensed psychologists versus advanced student trainees	Professional versus Trainee	102(13)	Adult	Social Anxiety Disorder	Internet-delivered CBT	•Symptoms •Functioning	•Post •1 year follow-up	16	Change in M/SD over time	.095
Baker & Neimeyer[1]	2003	Professionals (Master's level or higher) versus paraprofessionals	Professional versus Paraprofessional	98 (14)[¥]	Adult	Major Depression, Dysthymia, or Depression NOS	Combined MSG and CBT groups	•Symptoms	Post	2	F test for difference in change	.314
Bright, Baker, & Neimeyer (CBT arm)[1]	1999	Professionals (Master's level or higher) versus paraprofessionals	Professional vs Paraprofessional	31(14)[¥]	Adult	Major Depression, Dysthymia, or Depression NOS	CBT	•Symptoms •Mechanisms of change	Post	5	Change in M/SD over time; M/SD at post	.143
Bright, Baker, & Neimeyer (MSG arm)[1]	1999	Professionals (Master's level or higher) versus paraprofessionals	Professional vs Paraprofessional	36(14)[¥]	Adult	Major Depression, Dysthymia, or Depression NOS	MSG	•Symptoms •Mechanisms of change	Post	5	Change in M/SD over time; M/SD at post	-.197
Bisbey (TIR arm)	1995	Therapists with more or less experience with TIR	Experience with specific treatment	19(4)[¥]	Adult	Posttraumatic Stress Disorder	TIR	•Symptoms	Post	4	Change in M/SD over time	-1.519
Bisbey (DTE arm)	1995	Therapists with more or less experience with TIR	Experience with specific treatment	19(4)[¥]	Adult	Posttraumatic Stress Disorder	DTE	•Symptoms	Post	4	Change in M/SD over time	-.343

63

www.manaraa.com

Author(s)	Year	Moderator	N(k)	Population	Disorder	Treatment	Outcome	Timepoint	k	Analysis	Effect size
Franklin, Abramowitz, Furr, Kalsy, & Riggs	2003	High experience versus Medium experience High experience versus Low experience Medium experience versus Low experience	86(30)	Adult	Obsessive-Compulsive Disorder	Exposure and Response Prevention	•Symptoms	Post	2 2 2	Change in M/SD over time	.404 .376 -.135
Hahlweg, Fiegenbaum, Frank, Schroeder, & von Witzleben	2001	Experienced versus medium experience Experienced versus novice Medium experience versus novice	416(52)	Adult	Agoraphobia	High-Density Exposure	•Symptoms	•Post •1 year follow-up	0	ANOVAs	Paper reported no significant effects Original dataset also misplaced. Article excluded.
Howard	1999	Specialist versus nonspecialists	165(47)	Adult	Anxiety disorders	CBT	•Service utilization •Diagnostic remission	Post	9	Post M/SD, 2x2 events, 2x2 rate	.320
Huppert, Bufka, Barlow, Gorman, Shear, & Woods	2001	Years conducting CBT Years conducting general psychotherapy	183(14)	Adult	Panic Disorder	CBT	•Symptoms •Responder status	Post	9 9	Correlation coefficient	.232 .910
Lewis	2011	Months conducting therapy	222(33)	Adult	Internalizing Disorders	CBT or 3rd wave CBT variants	•Symptoms	•5 sessions into treatment •Post	0	Latent growth curve modeling	Could not calculate effect sizes from available data. Excluded from the study
Mason, Grey, & Veale	2015	Professional versus trainee Qualified versus trainees	196(34)	Adult	Anxiety disorders	CBT	•Symptoms •Reliable change •Clinically significant change •Service utilization •Drop out	Post	0	F tests	No analyses presented with main effect of therapist experience on outcome. Excluded from study
McLean & Hakstian	1979	Years conducting general psychotherapy	129(4)	Adult	Depression	Multiple treatment modalities	•Symptoms	Post	1	F test for difference in change	.176
Norton, Little, & Wetterneck	2014	Years of graduate level training in clinical psychology Number of previously led transdiagnostic CBT groups	319(33)	Adult	Anxiety Disorders	CBT	•Symptoms •Overall improvement •Drop out	Post	0	ANCOVA Regression	Only provided slopes and intercept scores. Not able to calculate effect sizes from information. Excluded from study.

Author	Year	Comparison	Moderator	N(k)	Age	Disorder	Treatment	Outcome	Timepoint	Code	Method	Effect size
		Professional versus Pre-doctoral intern	Degree/Schooling levels	264(32)	Adult	Internalizing Disorders	Eclectic	•Symptoms	Post	5	Change in M/SD over time	.149
Nyman, Nafziger, & Smith	2010	Professional versus Practicum student								5	Change in M/SD over time	.152
		Pre-doctoral intern versus Practicum student								5		.003
Piacentini, Bergman, Jacobs, McCracken, & Kretchman	2002	Experienced CBT therapists versus trainees	Experience with specific treatment	42(Unknown)	Youth	Obsessive-Compulsive Disorder	CBT	•Symptoms •Improvement	Post	2	Chi square	-.095
Podell	2010	Number of anxious clients seen	Experience with client population	279(38)	Youth	Anxiety Disorders	CBT	•Symptoms	Post	6	Correlation coefficient	.009
		Years conducting therapy	General clinical experience							6		-.010
		Psychiatrist versus Psychiatry Resident								6		.092
		Psychiatrist versus Family Medicine Resident								6		.239
		Psychiatrist versus Medical Student								6		.094
Propst, Paris, & Rosberger	1994	Psychiatry Resident versus Family Medicine Resident	Degree/Schooling levels	42(Unknown)	Adult	Internalizing Disorders	Psychodynamic	•Symptoms •Functioning	•Post •Follow-up	6	Change in M/SD over time; M/SD at post	.102
		Psychiatry Resident versus Medical Student								6		-.023
		Family Medicine Resident versus Medical Student								6		-.133
Russell & Wise (Systematic Desensitization arm)	1976	Experienced versus paraprofessionals	Professional versus paraprofessional	35(6)¥	Adult	Speech Anxiety	Systematic Desensitization	•Symptoms	Post	6	Change in M/SD over time	-.420
Russell & Wise (Cue-Controlled Relaxation arm)	1976	Experienced versus paraprofessionals	Professional versus paraprofessional	35(6)¥	Adult	Speech Anxiety	Cue-Controlled Relaxation	•Symptoms	Post	6	Change in M/SD over time	-.453
Shelton & Madrazo-Peterson	1978	Professional versus Paraprofessional	Professional versus paraprofesional	30(6)	Adult	Test Anxiety	Systematic Desensitization	•Symptoms •Grade point average	•Post •5 month follow-up	0	Means only	Did not provide enough information to calculate effect size. Excluded from study.

Study	Year			n	Age	Disorder	Treatment	Outcomes	Time	k	Effect size metric	Value
Thirlwall, Cooper, Karalus, Voysey, Willetts, & Creswell (Brief CBT arm)	2013	Some clinical experience versus Novices	Professional versus paraprofessional	61(19)¥	Youth	Anxiety disorders	Brief CBT	•Diagnostic remission	Post	1	Relative Risk Ratio	.559
Thirlwall, Cooper, Karalus, Voysey, Willetts, & Creswell (Full CBT arm)	2013	Some clinical experience versus Novices	Professional versus paraprofessional	64(19)¥	Youth	Anxiety disorders	Full CBT	•Diagnostic remission	Post	1	Relative Risk Ratio	.00
Thompson, Gallagher, Nies, & Epstein	1983	Professional versus paraprofessional	Professional versus paraprofessional	56(16)	Adult (Elderly)	Depression	Behavior therapy	•Symptoms •Life satisfaction •Client Satisfaction with treatment	Post	20	F test for difference in change, chi square, means and t tests	.452
Vosciano et al.	2004	Years conducting therapy	General clinical experience	173(48)	Adult	Depression	Cognitive-Behavioral Analysis System of Psychotherapy	•Symptoms	Post	1	Correlation coefficients	.220

Note. CBT=cognitive behavioral therapy, MSG=Mutual support group, TIR=Traumatic Incident Reduction, DTE=Direct Therapeutic Exposure, EX/RP=Exposure and Response Prevention, CBSAP=Cognitive-Behavioral Analysis System of Psychotherapy

*All effect sizes reported in Hedge's g. All effect sizes that were originally calculated as Fisher's z or Odds Ratio were converted to Hedge's g. Positive effect sizes represent a positive relationship between higher therapist experience and client outcomes. Effect sizes represent an average of all outcome measures for each subgroup used in the study. In the meta-analysis portion of this study, multilevel modeling was used to better account for dependency between outcome measures in the same study.

ᵇData in both studies are from the same randomized control trial.

¥Same group of therapists delivered both arms of treatment. Number of therapists counted once towards total sample size.

Table 2

Relationship between moderator variables

	Clients with anxiety disorders versus clients without anxiety disorders	Equal supervision of therapists versus more supervision for less experienced therapists	No randomization of clients versus Randomization of clients	CBT versus non-CBT treatment	Comorbidity allowed versus Not allowed	Adult versus Youth clients	Multiple types of measurement versus Single type of measurement	Multiple raters of outcome versus Single rater of outcome	Multiple domains of measurement versus Internalizing domain of measurement	Client mean age
Clients with anxiety disorders versus clients without anxiety disorders		4.70†	7.96*	34.05**	.53	0*	.51	.24†	8.10*	-7.78
Equal supervision of therapists versus more supervision for less experienced therapists	4.70†		7.18*	.41	.67	.19	.51	1.48	6.36*	2.77
No randomization of clients versus Randomization of clients	7.96*	7.18*		.12*	.61	1.21	.61	.34	12.92*	-6.99
CBT versus non-CBT treatment	34.05**	.41	.12*		4.25	∞	.25	.10†	9.91†	-5.42
Comorbidity allowed versus Not allowed	.53	.67	.61	4.25		.22	16.91**	5.66*	.54	-1.66
Adult versus Youth clients	0*	.19	1.21	∞	.22		.91	1.48	1.28	26.60**
Multiple types of measurement versus Single type of measurement	.51	.51	.61	.25	16.91**	.91		35.47**	.33	.51
Multiple raters of outcome versus Single rater of outcome	.24†	1.48	.34	.10†	5.66*	1.48	35.47**		.12*	5.01
Multiple domains of measurement versus Internalizing domain of measurement	8.10*	6.36*	12.92*	9.91†	.54	1.28	.33	.12*		-5.73
Client mean age	-7.78	2.77	-6.99	-5.42	-1.66	26.60**	.51	5.01	-5.73	

Table 3
Random effects models

Model	k	Hedge's g	95% CI		Q_{within}	I^2
			Lower	Upper		
All outcomes aggregated	31	.079*	.0095	.15	49.63*	4.57%
Therapist experience definitions						
Professional versus Paraprofessional	8	.091	-.15	.34	10.59	32.49%
General clinical experience	4	.13	-.06	.31	5.69	36.71%
Degree/Schooling Levels	9	.12[†]	-.0020	.25	9.54	0.01%
Experience with specific client population	1	.0095	--	--	--	--
Experience with specific treatment	8	-.051	-.45	.35	22.11**	72.41%
Professional versus Trainee	1	.024	--	--	--	--
Outcomes aggregated by domain						
Anxiety	19	.11	-.10	.32	50.26**	76.79%
Depression	15	.092	-.05	.20	14.12	3.10%
Internalizing Symptoms	8	.055	-.044	.15	4.73	0.00%
Functioning	17	.067	-.037	.17	17.48	9.82%
Satisfaction with treatment	7	.28[†]	-.015	.58	5.34	0.00%
Other measure domains (e.g., number of sessions, relapse rate)	4	.31*	.020	.60	1.05	0.00%
Combination of internalizing and other symptom domains	7	.20	-.19	.58	7.77	20.11%
Outcomes aggregated by rater						
Self-rated	23	.12**	.042	.20	39.016*	0.02%
Independent evaluator rated	21	.13	-.035	.29	44.82**	58.86%
Caregiver rated	2	.0048	-.12	.12	0.061	0.00%
Other (e.g., chart review)	2	.12	-.32	.56	0.94	0.00%
Outcomes aggregated by type of measure						
Rating scales	27	.059	-.012	.13	44.09*	1.56%
Semi-structured/Structured interviews	20	.19[†]	-.022	.40	55.27**	77.07%
Behavioral tasks/measures	2	.31	-.49	1.11	0.56	0.00%

Note. [†]p<.1*p<.05, **p<0.01

Table 4
Moderator analyses for all outcome measures aggregated

Model	k	Hedge's g	95% CI		Q$_{moderators}$
			Lower	Upper	
Treatment					
Intercept	27	.13	-.19	.44	.17
Cognitive Behavioral Therapy versus Supportive/Humanistic Therapy		-.06	-.41	.29	
Psychodynamic Therapy versus Supportive/Humanistic Therapy		-.014	-.45	.42	
Client age					
Intercept	31	.0087	-.094	.11	2.62
Adult versus youth clients		.11	-.023	.24	
Client diagnostic category					
Intercept	31	.12†	-.0020	.25	5.21†
Anxiety versus internalizing		-.11	-.26	.041	
Depression versus internalizing		.082	-.12	.29	
Comorbidity					
Intercept	31	.09†	-.0066	.19	.20
Comorbidity allowed versus not allowed		-.03	-.16	.10	
Supervision					
Intercept	29	.13**	.034	.23	3.55†
Equal supervision versus more supervision for less experienced therapists		-.13†	-.26	.0052	
Randomization					
Intercept	24	.15†	-.0060	.31	1.16
No randomization of clients versus randomization of clients		-.10	-.28	.08	
Rater of outcomes					
Intercept	31	.13†	-.0048	.27	3.25
Self-rated versus multiple raters		-.15	-.36	.056	
Independent evaluator rated versus multiple raters		.063	-.19	.31	

Note. †p<.1*p<.05, **p<0.01

Table 5

Moderator analyses for aggregated anxiety measures

Model	k	Hedge's g	95% CI		$Q_{moderators}$
			Lower	Upper	
Client age					
Intercept	19	.10	-.31	.51	.0013
.Adult versus youth clients		.0089	-.48	.49	
Comorbidity					
Intercept	19	-.029	-.30	.24	2.38
Comorbidity allowed versus comorbidity not allowed		.34	-.091	.76	
Supervision					
Intercept	18	.22	-.069	.50	2.02
Equal supervision versus more supervision for less experienced therapists		-.30	-.71	.11	

Note. ‡p<.1*p<.05, **p<0.01

Table 6

Moderator analyses for aggregated self-rated measures

Model	k	Hedge's g	95% CI		$Q_{moderators}$
			Lower	Upper	
Treatment					
Intercept	20	.11	-.32	.53	.041
Cognitive Behavioral Therapy versus Supportive/Humanistic Therapy		-.048	-.54	.45	
Psychodynamic Therapy versus Supportive/Humanistic Therapy		-.019	-.58	.54	
Client diagnostic category					
Intercept	23	.12	-.049	.29	1.68
Anxiety versus internalizing		-.093	-.33	.15	
Depression versus internalizing		.094	-.20	.38	
Comorbidity					
Intercept	23	.074	-.035	.18	1.62
Comorbidity allowed versus comorbidity not allowed		.10	-.056	.26	
Supervision					
Intercept	22	$.10^{t}$	-.014	.22	.027
Equal supervision versus more supervision for less experienced therapists		.013	-.15	.17	

Note. $^{t}p<.1$ $^{*}p<.05$, $^{**}p<0.01$

Table 7

Moderator analyses for aggregated independent evaluator rated measures

Model	k	Hedge's g	95% CI		Q$_{moderators}$
			Lower	Upper	
Treatment					
Intercept	20	-.083	-.64	.48	.89
Cognitive Behavioral Therapy versus Supportive/Humanistic Therapy		.18	-.42	.78	
Psychodynamic Therapy versus Supportive/Humanistic Therapy		.31	-.37	1.02	
Client diagnostic category					
Intercept	21	.25	-.17	.67	1.28
Anxiety versus internalizing		-.085	-.57	.39	
Depression versus internalizing		-.30	-.88	.27	
Client age					
Intercept	21	.0001	-.27	.27	1.26
Adult clients versus youth clients		.19	-.14	.51	
Comorbidity					
Intercept	21	.28‡	-.024	.59	1.34
Comorbidity allowed versus comorbidity not allowed		-.21	-.57	.15	
Supervision					
Intercept	20	.20‡	-.024	.43	1.61
Equal supervision versus more supervision for less experienced therapists		-.20	-.52	.11	
Randomization					
Intercept	20	.16	-.074	.40	.098
No randomization of clients versus randomization of clients		-.056	-.41	.30	

Note. ‡p<.1*p<.05, **p<0.01

Table 8

Moderator analyses for aggregated rating scales Moderator analyses for semi-structured/structured interviews

Model	k	Hedge's g	95% CI		$Q_{moderators}$
			Lower	Upper	
Treatment					
Intercept	24	.098	-.24	.43	.50
Cognitive Behavioral Therapy versus Supportive/Humanistic Therapy		-.11	-.48	.26	
Psychodynamic Therapy versus Supportive/Humanistic Therapy		-.015	-.47	.44	
Client diagnostic category					
Intercept	24	.12	-.027	.26	4.75[t]
Anxiety versus internalizing		-.15	-.34	.037	
Depression versus internalizing		.089	-.17	.35	
Comorbidity					
Intercept	27	.071	-.042	.18	.084
Comorbidity allowed versus comorbidity not allowed		-.023	-.18	.13	
Supervision					
Intercept	26	.062	-.050	.17	.13
Equal supervision versus more supervision for less experienced therapists		-.028	-.18	.13	

Note. [t]$p<.1$ *$p<.05$, **$p<0.01$

Table 9

Moderator analyses for semi-structured/structured interviews

Model	k	Hedge's g	95% CI		$Q_{moderators}$
			Lower	Upper	
Treatment					
Intercept	20	.087	-.57	.74	.16
Cognitive Behavioral Therapy versus Supportive/Humanistic Therapy		.10	-.61	.81	
Psychodynamic Therapy versus Supportive/Humanistic Therapy		.16	-.64	.97	
Client diagnostic category					
Intercept	20	.25	-.22	.72	.82
Anxiety versus internalizing		-.010	-.56	.56	
Depression versus internalizing		-.24	-.89	.41	
Client age					
Intercept	20	-.014	-.38	.35	1.65
Adult versus youth clients		.29	-.15	.73	
Comorbidity					
Intercept	20	.22	-.29	.73	.014
Comorbidity allowed versus comorbidity not allowed		-.035	-.60	.53	
Supervision					
Intercept	20	.34*	.058	.62	2.31
Equal supervision versus more supervision for less experienced therapists		-.31	-.70	.089	
Domain of measure					
Intercept	20	-.18	-.56	.21	6.20
Anxious symptoms versus multiple domains		.069	-.78	.91	
Depressive symptoms versus multiple domains		-.28	-1.15	.60	
Internalizing symptoms versus multiple domains		.60	-.25	1.44	
Functioning versus multiple domains		.36	-.19	.91	

Note. $^\ddagger p<.1$ *$p<.05$, **$p<0.01$

Formulas for calculating categorical moderators using the Mixed-Effect Model

Formula 1: Weighted means for each group

$$T_{\bullet j}^{M} = \frac{\sum_{i=1}^{mj} w_{ij}^{M} T_{ij}}{\sum_{i=1}^{mj} w_{ij}^{M}}, i = 1, 2, ..., m_{j}$$

Formula 2: Variance and standard errors of the group mean effect estimates

$$v_{\bullet j}^{M} = \frac{1}{\sum_{i=1}^{mj} w_{ij}^{M}}, i = 1, 2, ..., m_{j}$$

Formula 3: Testing the null hypothesis

$$H_{0} : \theta_{\bullet j} = 0$$

$$Z = \frac{T_{\bullet j}^{M}}{\sqrt{v_{\bullet j}^{M}}}$$

Formula 4: Constructing confidence intervals around the weighted mean

$$T_{\bullet j}^{M} \pm z_{1-\alpha} v_{\bullet j}^{M}$$

REFERENCES

*Articles included in study

Agresti, A. (2002). *Categorical data analysis* (Second ed.). New York, NY: Wiley.

American Psychological Association (2008). *Health disparities & mental/behavior health workforce*. Retrieved from http://www.apa.org/about/gr/issues/workforce/disparity.aspx

*Andersson, G., Carlbring, P., Furmark, T., & Group, S. R. (2012). Therapist experience and knowledge acquisition in internet-delivered CBT for social anxiety disorder: a randomized controlled trial. *PloS One, 7*(5), 1-10.

Anxiety and Depression Association of America (2014). *Facts and statistics*. Retrieved from http://www.adaa.org/about-adaa/press-room/facts-statistics

Beidas, R. S., Aarons, G., Barg, F., Evans, A., Hadley, T., Hoagwood, K., . . . Mandell, D. S. (2013). Policy to implementation: evidence-based practice in community mental health–study protocol. *Implementation Science, 8*(1), 38.

Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98*(2), 401.

Beutler, L. E. (1997). The psychotherapist as a neglected variable in psychotherapy: An illustration by reference to the role of therapist experience and training. *Clinical Psychology: Science and Practice, 4*(1), 44-52.

*Bisbey, L. B. (1995). *No longer a victim: A Treatment Outcome Study for Crime Victims with post-traumatic stress disorder*: UMI Proquest.

Blatt, S. J., Sanislow III, C. A., Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*(6), 1276.

Blow, A. J., Sprenkle, D. H., & Davis, S. D. (2007). Is who delivers the treatment more important than the treatment itself? The role of the therapist in common factors. *Journal of Marital and Family Therapy, 33*(3), 298-317.

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2007). Anxiety disorders and suicidal behaviours in adolescence and young adulthood: Findings from a longitudinal study. *Psychological Medicine, 37*(3), 431-440.

76

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). Subgroup analyses. *Introduction to Meta-analysis*, 149-186.

*Bright, J. I., Baker, K. D., & Neimeyer, R. A. (1999). Professional and paraprofessional group treatments for depression: A comparison of cognitive-behavioral and mutual support interventions. *Journal of Consulting and Clinical Psychology, 67*(4), 491.

Budge, S. L., Owen, J. J., Kopta, S. M., Minami, T., Hanson, M. R., & Hirsch, G. (2013). Differences among trainees in client outcomes associated with the phase model of change. *Psychotherapy, 50*(2), 150.

Chavira, D. A., Stein, M. B., Bailey, K., & Stein, M. T. (2004). Child anxiety in primary care: Prevalent but untreated. *Depression and Anxiety, 20*(4), 155-164.

Chibanda, D., Mesu, P., Kajawu, L., Cowan, F., Araya, R., & Abas, M. A. (2011). Problem-solving therapy for depression and common mental disorders in Zimbabwe: piloting a task-shifting primary mental health care intervention in a population with a high prevalence of people living with HIV. *BMC Public Health, 11*(1), 828.

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

D'Angelo, G., Pullmann, M. D., & Lyon, A. R. (2017). Community engagement strategies for implementation of a policy supporting evidence-based practices: A case study of Washington State. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(1), 6-15.

Del Re, A. C. (2015). A practical tutorial on conducting meta-analysis in R. *The Quantitative Methods for Psychology, 11*(1), 37-50.

Durlak, J. A. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin, 86*(1), 80.

Kaiser Family Foundation (2014). *Mental health care Health Professional Shortage Areas (HPSAs)*.   Retrieved from http://kff.org/other/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas/

*Franklin, M. E., Abramowitz, J. S., Furr, J. M., Kalsy, S., & Riggs, D. S. (2003). A naturalistic examination of therapist experience and outcome of exposure and ritual prevention for OCD. *Psychotherapy Research, 13*(2), 153-167.

Fulton, B. D., Scheffler, R. M., Sparkes, S. P., Auh, E. Y., Vujicic, M., & Soucat, A. (2011). Health workforce skill mix and task shifting in low income countries: A review of recent evidence. *Human Resources for Health, 9*(1), 1.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1.

Golden, R. L., & Vail, M. R. (2014). *Implications of the Affordable Care Act for mental health care*. Retrieved from http://www.asaging.org/blog/implications-affordable-care-act-mental-health-care

Gopalan, G. (2016). Feasibility of improving child behavioral health using task-shifting to implement the 4Rs and 2Ss program for strengthening families in child welfare. *Pilot and Feasibility Studies, 2*, 21.

Greenhouse, J. B., & Inyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hegdes, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 417-433). New York, NY: Russell Sage Foundation.

*Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B., & von Witzleben, I. (2001). Short-and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology, 69*(3), 375.

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

Higgins, J., & Green, S. (2011, March 2011). Recommendations on testing for funnel plot asymmetry. *Cochrane handbook for systematic reviews of interventions.* Retrieved from www.handbook.cochrane.org

*Howard, R. C. (1999). Treatment of anxiety disorders: Does specialty training help? *Professional Psychology: Research and Practice, 30*(5), 470.

*Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*(5), 747.

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*(4), 747-768.

Kaufman, J., & Charney, D. (2000). Comorbidity of mood and anxiety disorders. *Depression and Anxiety, 12*(1), 69-76.

Kazdin, A. E. (2005). Treatment outcomes, common factors, and continued neglect of mechanisms of change. *Clinical Psychology: Science and Practice, 12*(2), 184-188.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289*(23), 3095-3105.

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., . . . Zaslavsky, A. M. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine, 352*(24), 2515-2523.

Lambert, M. J. (2005). Early response in psychotherapy: Further evidence for the importance of common factors rather than "placebo effects". *Journal of Clinical Psychology, 61*(7), 855-869.

Leon, S. C., Martinovich, Z., Lutz, W., & Lyons, J. S. (2005). The effect of therapist experience on psychotherapy outcomes. *Clinical Psychology and Psychotherapy, 12*(6), 417.

*Lewis, C. C. (2011). *Understanding patterns of change: Predictors of response profiles for clients treated in a CBT training clinic.* (Doctor of Philosophy), University of Oregon.

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

*Mason, L., Grey, N., & Veale, D. (2016). My therapist is a student? The impact of therapist experience and client severity on cognitive behavioural therapy outcomes for people with anxiety disorders. *Behavioural and Cognitive Psychotherapy, 44*(2), 193-202.

Merikangas, K. R., He, J.-P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics, 125*(1), 75-81.

Michael, K. D., Huelsman, T. J., & Crowley, S. L. (2005). Interventions for child and adolescent depression: Do professional therapists produce better results? *Journal of Child and Family Studies, 14*(2), 223-236.

Murray, L. K., Dorsey, S., Skavenski, S., Kasoma, M., Imasiku, M., Bolton, P., . . . Cohen, J. A. (2013). Identification, modification, and implementation of an evidence-based psychotherapy for children in a low-income country: The use of TF-CBT in Zambia. *International Journal of Mental Health Systems, 7*(1), 24.

Murray, L. K., Skavenski, S., Kane, J. C., Mayeya, J., Dorsey, S., Cohen, J. A., . . . Bolton, P. A. (2015). Effectiveness of trauma-focused cognitive behavioral therapy among trauma-affected children in lusaka, zambia: A randomized clinical trial. *JAMA Pediatrics, 169*(8), 761-769.

Nissen-Lie, H. A., Goldberg, S. B., Hoyt, W. T., Falkenström, F., Holmqvist, R., Nielsen, S. L., & Wampold, B. E. (2016). Are therapists uniformly effective across patient outcome domains? A study on therapist effectiveness in two different treatment contexts. *Journal of Counseling Psychology, 63*(4), 367.

*Norton, P. J., Little, T. E., & Wetterneck, C. T. (2014). Does experience matter? Trainee experience and outcomes during transdiagnostic cognitive-behavioral group therapy for anxiety. *Cognitive Behaviour Therapy, 43*(3), 230-238.

*Nyman, S. J., Nafziger, M. A., & Smith, T. B. (2010). Client outcomes across counselor training level within a multitiered supervision model. *Journal of Counseling & Development, 88*(2), 204-209.

Ollendick, T. H., & King, N. J. (1994). Diagnosis, assessment, and treatment of internalizing problems in children: The role of longitudinal data. *Journal of Consulting and Clinical Psychology, 62*(5), 918.

Owen, J., Wampold, B. E., Kopta, M., Rousmaniere, T., & Miller, S. D. (2016). As good as it gets? Therapy outcomes of trainees over time. *Journal of Counseling Psychology, 63*(1), 12.

Petersen, I., Bhana, A., Baillie, K., & Consortium, M. R. P. (2012). The feasibility of adapted group-based interpersonal therapy (IPT) for the treatment of depression by community health workers within the context of task shifting in South Africa. *Community Mental Health Journal, 48*(3), 336-341.

Petersen, I., Hancock, J. H., Bhana, A., & Govender, K. (2014). A group-based counselling intervention for depression comorbid with HIV/AIDS using a task shifting approach in South Africa: A randomized controlled pilot study. *Journal of Affective Disorders, 158*, 78-84.

Pigott, T. D. (2012). Planning a meta-analysis in a systematic revie. In T. D. Pigott (Ed.), *Advances in meta-analysis* (pp. 13-32). New York, NY: Springer.

*Podell, J. L., Kendall, P. C., Gosch, E. A., Compton, S. N., March, J. S., Albano, A.-M., . . . Ginsburg, G. S. (2013). Therapist factors and outcomes in CBT for anxiety in youth. *Professional Psychology: Research and Practice, 44*(2), 89.

*Propst, A., Paris, J., & Rosberger, Z. (1994). Do therapist experience, diagnosis and functional level predict outcome in short term psychotherapy? *The Canadian Journal of Psychiatry, 39*(3), 168-176.

Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*(2), 464-468.

*Russell, R. K., & Wise, F. (1976). Treatment of speech anxiety by cue-controlled relaxation and desensitization with professional and paraprofessional counselors. *Journal of Counseling Psychology, 23*(6), 583.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*(3), 328-364.

*Shelton, J. L., & Madrazo-Peterson, R. (1978). Treatment outcome and maintenance in systematic desensitization: Professional versus paraprofessional effectiveness. *Journal of Counseling Psychology, 25*(4), 331.

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63*(2), 182-196.

Stirman, S. W., Finley, E. P., Shields, N., Cook, J., Haine-Schlagel, R., Burgess, J. F., . . . Gutner, C. A. (2017). Improving and sustaining delivery of CPT for PTSD in mental health systems: A cluster randomized trial. *Implementation Science, 12*(1), 32.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435-454). New York, NY: Russell Sage Foundation

*Thompson, L. W., Gallagher, D., Nies, G., & Epstein, D. (1983). Evaluation of the effectiveness of professionals and nonprofessionals as instructors of "coping with depression" classes for elders. *The Gerontologist, 23*(4), 390-396.

Thornicroft, G., & Sartorius, N. (1993). The course and outcome of depression in different cultures: 10-year follow-up of the WHO Collaborative Study on the Assessment of Depressive Disorders. *Psychological Medicine, 23*(4), 1023-1032.

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20*(1), 57.

Van Ameringen, M., Mancini, C., & Farvolden, P. (2003). The impact of anxiety disorders on educational achievement. *Journal of Anxiety Disorders, 17*(5), 561-571.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially, "All must have prizes." *Psychological Bulletin, 122*(3), 203-215.

Wei, Y., & Higgins, J. P. (2012). Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine, 32*, 1191-1205.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55*(4), 542-549.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*(3), 450.

Wittchen, H. U., Kessler, R., Pfister, H., Höfler, M., & Lieb, R. (2000). Why do people with anxiety disorders become depressed? A prospective longitudinal community study. *Acta Psychiatrica Scandinavica, 102*(406), 14-23.

Young, A. S., Klap, R., Sherbourne, C. D., & Wells, K. B. (2001). The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry, 58*(1), 55-61.