# A meta-analytic review of psychodynamic therapies for anxiety disorders

John R. Keefe [a], Kevin S. McCarthy [a,b], Ulrike Dinger [c,d], Sigal Zilcha-Mano [d], Jacques P. Barber [d,*]

[a] University of Pennsylvania, USA
[b] Chestnut Hill College, USA
[c] University of Heidelberg, Germany
[d] Adelphi University, USA

## HIGHLIGHTS

- Anxiety disorders are mental-health burdens that are sometimes difficult to treat.
- Psychodynamic therapy is commonly used to treat anxiety.
- Psychodynamic therapies may not differ overall in efficacy from other treatments.
- For most disorders investigated, efficacy may continue over a year post-treatment.
- Research should identify who may uniquely benefit from psychodynamic therapy.

## ARTICLE INFO

## ABSTRACT

Recent randomized controlled trials (RCTs) suggest that psychodynamic therapy (PDT) may be useful in the treatment of anxiety disorders. This paper presents the most comprehensive meta-analysis to date examining the controlled effects of PDT for anxiety disorders. 14 RCTs totaling 1073 patients were included. PDT was found to be significantly more effective than control conditions ($g = 0.64$). PDT did not differ significantly from alternative treatments at post-treatment ($g = 0.02$), follow-up (FU) up to a year ($g = -0.11$), and FU past a year ($g = -0.26$). Medium-to-high levels of heterogeneity were detected, indicating significant differences between studies. Nevertheless, our findings remained unchanged when heterogeneity outliers were removed (termination $g = -0.06$/short FU $g = -0.01$/long FU $g = -0.10$). Power analyses indicated that large or medium effect size differences between PDT and other active treatments could be detected even with high heterogeneity. Exploratory moderator analyses found few significant predictors of effect (e.g., relative risk of dropout). No differences were found examining remission rates or relative risk of dropout. Overall, PDT was shown to be as efficacious as other active treatments that have been studied for anxiety disorders.

© 2014 Elsevier Ltd. All rights reserved.

## Contents

* Corresponding author at: Derner Institute of Advanced Psychological Studies, Adelphi University, Hy Weinberg Center, Room 301, 1 South Avenue, P.O. Box 701, Garden City, NY 11530-0701, USA. Tel.: +1 516 877 4807 (Office); fax: +1 516 877 4805.
　E-mail address: jbarber@adelphi.edu (J.P. Barber).

## 1. Introduction

Anxiety disorders are among the most prevalent psychiatric conditions, with combined lifetime prevalence near 17% (Somers, Goldner, Waraich, & Hsu, 2006). Anxiety disorders have high rates of co-morbidity with other Axis I and II psychiatric disorders (Andrews, Slade, & Issakidis, 2002), and are associated with substantial physical and mental health liabilities that are further aggravated by comorbidity (Andrews, Henderson, & Hall, 2001; Bruce et al., 2005; Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007).

Several well-established treatments for anxiety disorders exist, including cognitive-behavioral therapies (CBT; Olatunji, Cisler, & Deacon, 2010) and psychopharmacological treatments (Hoffman & Mathew, 2008; Koen & Stein, 2011). However, as current treatments have incomplete efficacy and tolerability, it is valuable to explore other treatment options such as the widely used psychodynamic therapies (PDTs) (Fonagy, Roth, & Higgitt, 2005; Goisman, Warshaw, & Keller, 1999). PDTs have been studied and found to be efficacious for other types of disorders (Abbass, Hancock, Henderson, & Kisley, 2006 for short-term PDTs; Driessen et al., 2010 for PDTs for depression; Leichsenring & Rabung, 2011b for long-term PDTs for complex conditions; for a review of different disorders including anxiety disorders, see Barber, Muran, McCarthy, & Keefe, 2013) and have a rich theoretical literature concerning the nature of anxiety symptoms and their resolution (e.g., Busch, Milrod, Singer, & Aron, 2011 for panic; Slavin-Mulford & Hilsenroth, 2012 for a general review). However, PDT awaits a meta-analytic validation of its efficacy for anxiety disorders.

Broadly speaking, PDT is distinguished from CBT on the basis of different theoretical assumptions about the psychological processes underlying anxiety disorders, which result in different approaches to treatment (e.g., Busch et al., 2011; Crits-Christoph, Wolf-Palacio, Ficher, & Rudick, 1995; Leichsenring, Beutel, & Leibing, 2007; for a review see Slavin-Mulford & Hilsenroth, 2012). In psychodynamic theory, anxiety symptoms are often assumed to originate from relationship experiences in which certain feelings or wishes were experienced by the patient as painful, dangerous, or unacceptable (e.g., feelings of loss or abandonment, a wish to express anger or assert oneself). The patient learns to disavow these intense, negative feelings and desires, avoids their experiences, and develops anxiety symptoms (e.g., having a panic attack when triggered by sensations of loss or anger; Busch et al., 2011). Psychodynamic therapists encourage the patient to discuss the contexts in which their symptoms arise in order to understand the experiences surrounding the occurrence of symptoms. Therapists help the patient make connections between the experience of their current symptoms and the prior interpersonal and intrapsychic events from which these anxiety-producing defenses and dynamics may have originated, with the aim of reaching emotional insight. This may be especially helpful when the anxiety symptoms emerge in the therapeutic setting. Making such connections helps the patient to become more aware of and tolerate their own affect and wishes (i.e., lowering experiential avoidance; Kashdan, Barrios, Forsyth, & Steger, 2006), less rigid in interpersonal perceptions and behaviors, and allows the patient to try new ways of getting their needs met without anxiety while using more adaptive psychological defenses (Summers & Barber, 2009). Other PDT theories of anxiety emphasize object-relations theory and ambivalent feelings about significant others, attachment, and self-psychology concepts of esteem-regulation.

Unlike CBTs, psychodynamic therapists do not usually give out homework exercises to be performed outside of the therapeutic hour (e.g., in vivo self-exposure) nor do they provide adjunctive sessions (e.g., additional hours of guided exposure). However, encouraging patients to try new behaviors, especially those relevant to their own fears, has been within the aims of psychodynamic therapy since its inception, and the feelings and conflicts evoked by exposure may be useful material for psychodynamic work (Barber & Luborsky, 1991; Freud, 1926/1990; Summers & Barber, 2009; Wachtel, 1977). Nevertheless, it is unclear how often these recommendations are implemented in PDT (see Leichsenring et al., 2007 for a prominent exception concerning supportive–expressive therapy for social phobia, which includes a recommendation for exposure framed psychodynamically). Even if PDT does sometimes entail exposure, there is reason to suspect that it is often performed in a less directive and systematic manner as a consequence of other therapeutic foci (e.g., intense exploration of emotionally charged issues). By contrast, a recent survey study of CBT practitioners treating anxiety disorders identified the directive nature of CBTs and issues with behavioral assignments (e.g., exposure) as substantial barriers to treatment success for some patients with diagnoses of GAD, panic, and social anxiety disorders (McAleavey, Castonguay, & Goldfried, 2014; Szkodny, Newman, & Goldfried, 2014; Wolf & Goldfried, 2014). Thus, the less directive PDTs could conceivably provide an efficacious treatment frame for these patients.

At present, no PDT for any anxiety disorder qualifies as a well-established "empirically supported therapy" (EST) as per American Psychological Association (APA) Division 12 criteria (Chambless & Hollon,

1998). However, APA Division 12 has determined that panic focused psychodynamic psychotherapy (Busch et al., 2011) has "modest research support" (Milrod, Leon, Busch, et al., 2007). A number of quasi-experimental and naturalistic studies further suggest that psychodynamic therapy may successfully treat specific anxiety disorders (see review in Slavin-Mulford & Hilsenroth, 2012).

A recent preliminary meta-analysis of the anxiety trials included in the Gerber et al. (2011) qualitative review of RCTs of PDT found no significant outcome differences between PDTs and other active treatments at termination or short-term follow-up and a large controlled effect size of PDT treatment ($g = 0.775$, $K = 3$; Barber, Muran, McCarthy, & Keefe, 2013). However, this meta-analysis did not include recently published large scale trials, did not include group therapies, did not explore long-term follow-up, did not examine issues of power, and did not analyze secondary outcomes (e.g., dropout, remission). Other previous meta-analytic investigations have included psychodynamic treatment of anxiety disorders or symptomatology as a secondary, yet incomplete undertaking to their primary investigations. In a Cochrane Collaboration meta-analysis, PDT was found to be superior to control conditions (e.g., wait-list, treatment-as-usual) on measures of anxiety at both termination and follow-up (Abbass et al., 2006). However, this result was folded across diagnoses (i.e. included change in anxiety in non-anxiety patients), only included short-term therapies (i.e., treatments less than 40 sessions), and did not investigate active treatment comparisons (e.g., CBT). Another meta-analysis of a small number of CBT studies ($\kappa = 5$) suggested that CBTs were superior to PDTs on measures of anxiety with a small-to-medium effect size advantage (Tolin, 2010). Yet, this meta-analysis had several limitations, including (a) designating as a bona fide psychodynamic therapy what was in fact a control treatment explicitly designated as a control (Shear, Houck, Greeno, & Masters, 2001) and otherwise including predominantly older studies with nonspecific psychodynamic therapies that arguably did not meet bona fide criteria as per Wampold et al., 1997 (used as minimal criterion for inclusion in analyses); (b) mixing together pediatric and adult patient samples; and (c) not including studies conceivably relevant for meta-analysis (e.g., Leichsenring et al., 2009). A recent comprehensive re-analysis of Tolin's (2010) meta-analytic question by Baardseth et al. (2013) that used bona fide CBT trials of anxiety disorders identified by an Association for Behavioral and Cognitive Therapies survey argued that there was no significant difference in anxiety-specific or general effect between bona fide CBTs and other bona fide therapies. Thus, there is still need for an updated, comprehensive meta-analysis examining the efficacy of PDT of anxiety disorders.

In the present study, we aimed to synthesize the current empirical literature on the PDT of anxiety disorders by conducting a meta-analysis of randomized controlled trials (RCTs). Rich data and theory suggest that anxiety disorders share substantial environmental and genetic vulnerabilities that relate to a nontrivial degree of shared mechanisms of disorder development, in addition to disorder-specific mechanisms (e.g., Hettema, Prescott, Myers, Neale, & Kendler, 2005; McTeague & Lang, 2012; Tambs et al., 2009). In light of this evidence, both cognitive-behavioral (Barlow et al., 2011; Boswell et al., 2013) and psychodynamic researchers (Johansson et al., 2013; Leichsenring & Salzer, in press) have begun moving toward developing treatment protocols that address transdiagnostic (e.g., for all anxiety disorders), empirically-supported principles of change. Given the overlapping conceptualizations between anxiety diagnoses and similarities between PDTs, we conducted our primary analyses on anxiety disorders as a group. As it has been asserted that PDTs are unlikely to be efficacious for any anxiety disorders, the scope of our meta-analysis is revelatory insofar as it probes whether psychodynamic therapies—from generic psychodynamic therapies to adherence-checked, manualized, disorder-specific therapies—tend to produce effects comparable to other commonly tested active treatments for these disorders. Such a meta-analysis may aid clinicians to decide whether PDTs may be worthwhile treatments for anxiety patients. We then proceeded to conduct disorder-specific

moderation analyses to see if particular anxiety disorders differed significantly from primary estimates of effect size (i.e., whether PDT tended to treat a given disorder substantially better or worse than it did for remaining anxiety disorders), based on the study samples available.

## 2. Methods

### 2.1. Study selection

Pertinent studies were identified through searches of relevant databases through January, 2013, including a comprehensive search of PubMed/MEDLINE and PsycInfo. English language was not an explicit qualification of the search, but no translated abstract appeared eligible for inclusion. The following terms were used as descriptors: (psychodynamic OR dynamic OR dynamically OR psychoanalytic OR psychoanalysis OR analytic OR insight OR interpretive) AND (therapy OR psychotherapy OR treatment OR counseling) AND (anxiety OR panic OR phobia OR phobic OR agoraphobia OR agoraphobic OR stress OR trauma OR posttraumatic OR traumatic OR PTSD OR obsessive–compulsive) AND (study OR trial). The references of existing relevant meta-analyses, reviews, chapters, and articles were inspected to find further relevant studies: in particular, the quality-based review of the majority of published psychodynamic RCTs by Gerber et al. (2011), the Cochrane Collaboration meta-analysis of short-term psychodynamic psychotherapy by Abbass et al. (2006), and the qualitative review of PDT for anxiety disorders by Slavin-Mulford and Hilsenroth (2012).

To be included in the meta-analysis, studies had to meet the following criteria (more detailed explanation for each criterion can be found in the Online Appendix Supplement):

1. Published after 1970.
2. Investigated one or more specific class of anxiety disorder (e.g., GAD, panic disorder). Studies investigating patients of mixed disorder groups (e.g., treatment for both depressive patients and anxious patients with data undifferentiated between the two) were excluded, unless data for the patients with a primary anxiety diagnosis were reported separately from other data or could be provided. When possible, we contacted study authors for this information.
3. Treatment groups included an individual or group PDT (for further description of defining characteristics of PDTs see Blagys & Hilsenroth, 2000; Summers & Barber, 2009). Both short-term and long-term PDTs were eligible for inclusion. Theoretically integrative or eclectic treatments adding PDT interventions to a different primary theory/modality of treatment were not included. Some examples of such a treatment would be Gerson's brief eclectic therapy for the treatment of PTSD (Gerson & Carlier, 1994) or psychodynamic body therapy (Monsen, 1989). PDT could also not have been delivered as a combined therapy with psychopharmacology (e.g., Klein, Zitrin, Woerner, & Ross, 1983; Martini et al., 2011), though scattered concurrent psychopharmacological use by patients was acceptable.
4. PDT took place in the context of a RCT, neither a naturalistic nor a quasi-experimental study. The RCT must have compared the PDT against another, non-psychodynamic active treatment intended to produce a therapeutic effect over and above generic attention and support (e.g., CBT, relaxation training, psychopharmacology) or a control condition that was intended to and would be expected to underperform any uniquely therapeutic treatment (e.g., a wait-list, a generic non-bona fide supportive counseling condition).
5. Patients were taken from an adult rather than a pediatric sample (defined as age less than 18 years old).

### 2.2. Coding

To describe the meta-analytic sample and to provide descriptive data for effect size moderator analyses (e.g., studies that treated GAD versus other studies), RCT reports were coded for PDT format (individual vs.

group), comparison group type (active vs. control), disorder(s) investigated, number of therapy sessions, attrition rates for both therapies, average age of patients, percentage of female patients, number of therapists, years of therapist experience, use of a standardized/validated method for diagnosing target anxiety disorders, manualization of therapies (use of a specific manual for both psychodynamic and/or comparison treatments), presence of an adherence check for treatment integrity, whether the therapy described a disorder-specific theory of treatment, and the manner in which patient attrition was accounted for in outcome analyses (e.g., completer; last observation carried forward; mixed model; multiple imputation). Study characteristic coding was undertaken by the primary study author (JRK) and checked by study quality coders (see below).

### 2.2.1. Study quality

Studies were also rated on the Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010) by three of the authors (JRK, KSM, UD). This scale is used to assess the methodological quality of a comparative trial on 24 indices (e.g., use of an appropriate sample size to answer the questions posed by the trial, presence of a full treatment description for all conditions) that are each rated on a 0–2 scale. In addition, raters assess a separate item of overall quality rated on a scale of 1 (exceptionally poor) to 4 (average) to 7 (exceptionally good). The full scale or "Total" score is the sum of the 24 indices and has a range of 0–48. Gerber et al. (2011) posed a full-scale cutoff score of 24 for a study of adequate quality, corresponding to an average score of 1 (scale of 0–2) on each individual item, though this was noted by the authors as provisional as it is unclear whether other categorical divisions would be more appropriate. Nevertheless, we used this as a dichotomous cutoff to see if approximately-defined "higher" quality studies differed systematically from "lower" quality studies in terms of effect sizes. Scores were averaged across all three raters. Interrater reliability (calculated as ICC[2,3], Shrout & Fleiss, 1979) was excellent ($\rho_I > .90$; Shrout, 1995) for both the Total score ($\rho_I = .975$) and the Overall quality item score ($\rho_I = .980$).

### 2.3. Analysis strategy

### 2.3.1. Calculation of effect sizes

To calculate effect sizes we used the author-identified primary anxiety outcome measure for the anxiety disorder(s) treated in a particular study. When primary outcome was not explicitly indicated, the judgment of meta-analysis authors (JRK, JPB) was used to select symptomatic outcome measures for analysis based on the specific psychopathology of the disorder in question (reported in Table 1). To address the construct of general anxiety, we chose to use the Hamilton Rating Scale for Anxiety (Hamilton, 1959) if there was no author-specified primary anxiety outcome due to its ubiquity in the research literature. When scores were not reported, ESs were imputed or estimated from available data and reported statistical tests as per Lipsey and Wilson (2001). Authors were also contacted for missing or incompletely reported data.[1] For two studies (Beutel et al., 2013; Knekt et al., 2008), only remission rates were available and were used as the primary outcome measure.

Calculations of weighted mean effect sizes, as well as the heterogeneity and moderator analyses described below, were conducted using the statistical package *metafor* version 1.9-1 (Viechtbauer, 2010) as implemented in the R statistical computing language version 3.0.1 (R Core Team, 2012). As we expected significant heterogeneity of effect sizes due to between-study design differences (e.g., different anxiety disorders, different exemplars of psychodynamic therapy, different comparison treatments), we chose the random effects model for its robust estimation of effect size (Hedges & Vevea, 1998). The Sidik–Jonkman estimator of heterogeneity variance was used as it has shown reduced bias and more reliable estimates compared to DerSimonian–Laird or

restricted maximum likelihood approaches, particularly in the context of moderate to high heterogeneity (Sidik & Jonkman, 2005b) and publication bias (Henmi & Copas, 2010). We additionally applied the Knapp and Hartung (2003) adjustment to the standard errors of the estimated coefficients to account for uncertainty in the estimated residual heterogeneity (Sidik & Jonkman, 2005a), which produces confidence intervals and *p*-value estimates much closer to nominal significance under common meta-analytic conditions. In a recent replication, combining the Sidik and Jonkman (2005b) and Knapp and Hartung (2003) methods was shown to calculate more accurate estimates compared to DerSimonian-Laird estimation in both simulated and real meta-analytic data (IntHout, Ioannidis, & Borm, 2014).

Three separate meta-analyses were conducted: (a) an estimation of the uncontrolled effect size (i.e., pre-post within-groups effect size) of PDT; (b) an estimation of the controlled effect size (i.e. between-groups effect sizes against a control group) of PDT; and (c) an estimate of between-group effect sizes between PDT and alternative active treatments (e.g., CBT). We also meta-analyzed available data on PDT versus active treatment comparisons from post-therapy follow-up, divided into "short-term" (<1 year of post-termination follow-up) and "long-term" (≥1 year follow-up) periods. Between-groups effect sizes were based on termination or follow-up score comparisons between PDT and the comparison group (i.e., either control or active treatments).

We used Cohen's (1992) interpretive framework for describing the magnitude of effect sizes, wherein values of 0.20, 0.50, and 0.80 are considered respectively small, medium, and large effects. It should be noted that these effect size conventions only apply to between-groups ESs, not uncontrolled ESs (cf. Barber, Barrett, Gallop, Rynn, & Rickels, 2012) and are in some sense arbitrary (see Durlak, 2009, for a discussion of this point specifically and effect sizes generally). For the purposes of interpretability: assuming a normal distribution, small, medium, and large between group effect sizes correspond respectively to the experimental distribution mean being at the 58th, 69th, and 79th percentiles of the comparison distribution. For further information on the calculation of effect sizes, see the Online Supplement Appendix.

### 2.3.2. Heterogeneity

Heterogeneity of ESs was examined using the Cochrane's Q statistic and the $I^2$ index (Higgins & Thompson, 2002). Heterogeneity refers to substantial differences in effect sizes between studies indicating that some studies may belong to a different effect size distribution, as effects differ more than would be expected based on within-study variances. Significant Q statistics indicate that the observed range of ES is significantly larger than what would otherwise be expected based on within-study variances; the $I^2$ index is a quantification of this heterogeneity, with 25%, 50%, and 75% percent reflecting respectively low, medium, and high heterogeneity. As the Cochrane's Q is known to be underpowered even in cases of higher heterogeneity (Higgins & Thompson, 2002), we report any $I^2$ indices at or above medium heterogeneity even in the case of an insignificant Q finding.

### 2.3.3. Sensitivity and publication bias

We performed sensitivity (i.e., "leave-one-out") analyses running each main meta-analysis (uncontrolled effect size, controlled effect size, active group comparisons) with each study removed once to determine whether a given finding of significant effect (or lack thereof) or heterogeneity was driven by a single study's inclusion.

Publication bias was assessed by examination of publication bias funnel plots and Duval and Tweedie's (2000) trim-and-fill procedure. Bias funnel plots plot a measure of study size (standard error) on the vertical axis as a function of study effect size on the horizontal axis. When asymmetry was evident in the funnel plot as per Egger's regression test for funnel plot asymmetry (Egger, Smith, Schneider, & Minder, 1997), we planned to apply Duval and Tweedie's (2000) trim-and-fill procedure to provide an adjusted ES estimate that corrects for the number and assumed location of the missing studies. Resultant

---

[1] Crits-Christoph et al. (2005) provided unpublished data for their study of supportive–expressive dynamic therapy and a manualized supportive therapy for GAD.

**Table 1**
Characteristics of studies included in meta-analysis, part I.

| Study | Anxiety disorder(s) | Psychodynamic therapy groups | Comparison groups | Follow-up periods used (mos.) | Primary findings | Total/overall quality score | Outcome measure | Secondary outcomes |
|---|---|---|---|---|---|---|---|---|
| Alstrom et al. (1984a) | Agoraphobia | Avg. 8.5 sessions DST + 0.7 during FU, UNMAN (n = 14 at T, 13 at FU) | 8.8 sessions PE + 5.6 during FU (n = 11 at T, 9 at FU); 9.1 sessions AR + 3.2 during FU (n = 17 at T and FU); BTC (n = 19 at T, 12 at FU) | 9 | Termination: DST = PE; DST > AR & BTC Short-Term FU: DST > PE, AR, & BTC | 20/3 | AIPS global rating | None |
| Alstrom et al. (1984b) | Specific social phobias | Avg. 9.4 sessions DST + 2.6 during FU, UNMAN (n = 16 at T, 13 at FU) | 9.1 sessions PE + 1.7 during FU (n = 7 at T and FU); 9.4 sessions AR + 1.7 during FU (n = 9 at T, 6 at FU); BTC (n = 10 at T, 8 at FU) | 9 | Termination: DST < PE; DST > AR & BTC Short-Term FU: Maintained | 18.67/2.67 | AIPS global rating | None |
| Beutel et al. (2013) | Panic with and without agoraphobia | 24 sessions PFPP, MAN (n = 36) | 24 sessions CBT with exposure sessions (n = 18) | 6 | Termination: PFPP = CBT Short-Term FU: Maintained | 35/5 | PDSS remission | Remission |
| Bögels et al. (in press) | Social anxiety | Avg. 31.1 sessions BDT, UNMAN (n = 22) | Avg. 19.8 sessions CBT (n = 27) | 3, 12 | Termination: DT = CBT Short-Term FU: Maintained Long-Term FU: Maintained | 32.67/5 | Author-Defined Social Anxiety Composite Factor | Remission |
| Bressi et al. (2010) | GAD (41.0%), panic disorder (38.5%), social anxiety (20.5%) | 40 sessions of BDT, MAN (n = 20) | ADT (n = 19) | None | Termination: BDT = ADT | 37/5.67 | Combined: SCL-90-R Phobic Anxiety + SCL-90-R Anxiety | Depression, Interpersonal Problems |
| Brom, Kleber, and Defares (1989) | PTSD | Avg. 18.8 sessions TBDT, MAN (n = 26) | 15.0 sessions TD (n = 28); 14.4 sessions BHT (n = 26); WLC (n = 20) | 3 | Termination: TFDT = BHT, TD; TFDT > WLC Short-Term FU: Maintained | 23/5 | Combined: TSS + IES | None |
| Crits-Christoph et al. (2005) | GAD | 16 sessions SE therapy adapted to GAD, MAN (n = 14) | 16 sessions SP (n = 14) | None | SE = SP on scores, SE > SP on remission | 32/5 | HAM-A | Remission |
| Durham et al. (1994)/ Durham et al. (1999) | GAD | 8 or 16 sessions of an unspecified model (11.86 avg), UNMAN (n = 29 at T/Short FU, 22 at Long FU) | 8 sessions CT (n = 20 at T/Short FU, 14 at Long FU); 16 session CT (n = 15 at T/Short FU, 13 at Long FU); 8 sessions AMT (n = 16 at T/Short FU, 12 at Long FU) | 6, 12 | Termination: DT < AMT, CT-8 & CT-16 Short-Term FU: Maintained Long-Term FU: Maintained | 29/5 | HAM-A for Termination, STAI-T for FU | Depression, Remission |
| Knekt et al. (2004)/ Knekt et al. (2008) | GAD (22.1%), Anxiety NOS (22.1%), OCD (8%), panic disorder (18.9%), social anxiety (37.9%), specific phobia (8.4%) | Avg. 18.5 sessions BDT, UNMAN (n = 50) | 9.8 sessions SFT, UNMAN (n = 45) | 5, 29 | Termination: BDT = SFT Short-Term FU: Maintained Long-Term FU: Maintained | 36/5 | Diagnostic remission | Remission |
| Knijnik, Kapczinski, Chachamovich, Margis, and Eizirik (2004) | Social anxiety | 12 sessions PGT, MAN (n = 15) | 12 sessions CGT (n = 15) | None | PGT > CGT | 24/3.67 | LSAS | None |
| Leichsenring et al. (2009)/ Salzer, Winkelbach, Leweke, Leibing, and Leichsenring (2011) | GAD | Avg. 29.1 sessions SE for GAD, MAN (n = 28) | 28.8 sessions CBT (n = 29) | 6, 12 | Termination: SE = CBT for primary outcome, but CBT > SE for some secondary outcomes Short-Term FU: Maintained Long-Term FU: Maintained | 38/6 | HAM-A | Depression, Interpersonal Problems |
| Leichsenring et al. (2013) | Social anxiety | Avg. 25.7 sessions SE therapy adapted to social anxiety, MAN (n = 207) | 25.8 sessions CBT (n = 209), WLC (n = 79) | 6, 24[a] | Termination: CT > SE > WLC Short-Term FU: SE = CT Long-Term FU: Maintained | 44.33/7 | LSAS | Depression, Interpersonal Problems, Remission |
| Milrod, Leon, Busch, et al. (2007) | Panic disorder with and without agoraphobia | 24 sessions PFPP, MAN (n = 26) | 24 sessions AR (n = 23) | None | PFPP > AR | 42/6 | PDSS | Depression, Remission |
| Pierloot and Vinck (1978) | General anxiety (GAD-like) | Avg. 19.8 sessions FDT, UNMAN (n = 9) | Avg. 19.9 sessions SD (n = 13) | Mean of 3.95 months | Termination: FDT < SD Short-Term FU: FDT = SD | 8/2 | Combined: PSS-A + TMAS + STAI-T | None |

*Note.* ADT = Antidepressant Therapy, AIPS = Alstrom Interview of Phobic Symptoms, AMT = Anxiety Management Training, AR = Applied Relaxation, BTC = Basal Therapy Control, BHT = Behavioral Hypnotherapy, BDT = Brief Dynamic Therapy, CT = Cognitive Therapy, CBT = Cognitive-Behavioral Therapy, CGT = Control Group Therapy, DST = Dynamic Supportive Therapy, FDT = Focal Dynamic Therapy, GAD = Generalized Anxiety Disorder, HAM-A = Hamilton Anxiety Rating Scale, IES = Impact of Event Scale, LSAS = Liebowitz Social Anxiety Scale, MAN = Manualized Therapy, PDSS = Panic Disorder Severity Scale, PE = Prolonged Exposure, PFPP = Panic Focused Psychodynamic Psychotherapy, PGT = Psychodynamic Group Therapy, PSS-A = Psychiatric Status Schedule-Anxiety Subsection, PTSD = Post-Traumatic Stress Disorder, SCL-90-R = Symptom Checklist-90-Revised, SD = Systematic Desensitization, SE = Supportive–Expressive Therapy, SFT = Solution-Focused Therapy, SP = Manualized Supportive Psychotherapy, STAI-T = State-Trait Anxiety Inventory (Trait), TBDT = Trauma-based Dynamic Therapy, TD = Trauma Desensitization, TSS = Trauma Symptoms Scale, TMAS = Taylor Manifest Anxiety Scale, UNMAN = Unmanualized Therapy, WLC = Wait-List Control.

[a] Leichsenring et al. (2013) follow-up based on unpublished data.

funnel plots can be observed in the Online Supplement (Supplemental Figs. 4–6).

To further probe for possible small-study effects related to publication bias (insofar as smaller studies with unfavorable results may not be published depending on authorial bias; Kraemer, Gardner, Brooks, & Yesavage, 1998), we ran an additional sensitivity check using the modified meta-analytic technique of Henmi and Copas (2010). This is a hybrid method using a fixed effects estimate to give absolute greater effect weighting to larger studies (that are more likely to be published regardless of result), while utilizing random effects meta-analysis in incorporating information about heterogeneity into the standard error of the overall estimate. We defined a significant difference as an absolute ES difference ≥ 0.20, or a shift from significant to nonsignificant confidence intervals or vice-versa.

### 2.3.4. Study-level moderators of effect size

In meta-analysis, a moderator is a study characteristic (either categorical or continuous) that may conceivably predict the magnitude of an effect size. Categorical moderators can be thought of as splitting effect sizes into different distributions, analyzing whether studies with one feature cluster around a significantly different effect size estimate than studies without that feature, while continuous moderators can be thought of as akin to a continuous regression predictor. Moderator analyses allow for the probing of dispersion of effect sizes to develop hypotheses disambiguating the heterogeneity already extant in the literature, which Borenstein, Hedges, Higgins, and Rothstein (2009) view as a primary objective and advantage of meta-analysis. A given moderator finding does not necessarily reflect the presence or lack of a "true" study-level effect, especially when used with smaller samples, and must be evaluated critically. Exploratory analyses of moderators of effect sizes between types of studies were performed in two ways:

(1) For categorical moderators (e.g., use of manualized vs. unmanualized PDT), we performed subgroup analyses to test for significant differences between effect sizes in different categories of studies. We used a mixed-effects model that pooled studies within subgroups with the random-effects model, but tested for significant differences between subgroups with the fixed-effects model. If any subgroup had less than two studies, it was not included in a given moderator analysis.

(2) For continuous moderators (e.g., Total study quality score), we performed random-effects metaregressions. Metaregression is a weighted regression that gives studies with larger sample size more weight (Borenstein et al., 2009). Effect size (either uncontrolled or between-groups) was used as a dependent variable, and the continuous moderator was used as the predictor.

All moderator results can be interpreted as unstandardized beta coefficients. Due to the relatively small sample size ($K = 14$ studies) for our meta-analysis, significant moderator findings are considered exploratory (Hedges & Pigott, 2004; Ioannidis, 2008) and should be interpreted in context of their samples (Tables 2 & 3). Based on several test estimates of power from Hedges and Pigott (2004), for the larger comparisons (uncontrolled ES, between-groups at termination) we would expect approximately adequate power (around .80) to detect moderators of medium effect size. Furthermore, we performed permutation tests of robustness for any moderator findings that were $p < .20$ or below in the original analyses (10,000 iterations; Higgins & Thompson, 2004). For the sake of space, we note here that unless otherwise noted no permutation tests provided estimates of significance that were substantively different from our original estimates (i.e., significance crossing over the $p < .10$ or $p < .05$ lines).

### 2.3.5. Secondary outcome calculations

We also performed several separate secondary analyses using data other than continuous anxiety symptom scores. First, we examined dichotomous response rates to treatment as defined by the study (analyzed as odds ratios [OR]).[2]

Second, relative risk (RR) of drop-out (i.e., the ratio of the probabilities of dropping out of the two treatments) was also explored to see if psychodynamic therapies differed significantly from other treatments in patient retention. OR and RR were calculated such that values above 1.00 represent greater chance of treatment response and treatment dropout for psychodynamic therapy, respectively. Last, depression (preferably using assessor-rated scales) and interpersonal problems (Inventory of Interpersonal Problems [Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988]) were also analyzed at termination. These secondary outcomes (with the exception of dropout) should be considered exploratory, as different types of secondary outcomes were inconsistently and possibly nonrandomly collected between studies.

## 3. Results

### 3.1. Study characteristics

A combined search using PubMed/Medline and PsycInfo in January 2013 procured 7432 titles and abstracts. 7001 studies were eliminated via title and abstract because of clear irrelevance to the present research question, most of which appeared to be due to the generic nature of certain search terms (e.g., stress, dynamic, anxiety). The remaining 418 studies were eliminated from review of abstract and full-text article (see Fig. 1 for reasons for elimination). All remaining articles were target articles and follow-up papers on main effects, summing to 14 RCTs meeting inclusion criteria. One RCT (Bögels, Wijts, Oort, & Sallaerts, in press) was originally found through the citation of a conference presentation in a previous meta-analysis (Leichsenring, Rabung, & Leibing, 2004) but is now in press, and 2 were published during manuscript preparation (Beutel et al., 2013; Leichsenring et al., 2013).

The 14 RCTs included in this analysis were from articles between 1978 and 2014. The majority of studies (71.4%) were of adequate quality or above as per Gerber et al. (2011) study quality norms (mean = 30.0, SD = 10.1). Social anxieties/phobias were the most frequently studied disorder group ($K = 5$), followed by GAD ($K = 4$) and panic disorder ($K = 2$). One study treated patients with a primary PTSD diagnosis (Brom, Kleber, & Defares, 1989). The remaining two studies (Bressi, Porcellana, Marinaccio, Nocito, & Magri, 2010; Knekt et al., 2008) used a sample of anxiety patients with multiple types of anxiety disorders: Bressi et al. (2010) included patients with either GAD, panic disorder, or social anxiety as a primary diagnosis, and Knekt et al. (2008) included patients with GAD, OCD, panic disorder, social anxiety, specific phobias, and anxiety disorder-NOS. Five studies compared PDT to a control condition, whereas 13 studies compared PDT to an active treatment group. Ten studies of active treatment comparisons included a follow-up period, with 10 studies having at least short-term follow-up (i.e., <1 year after termination) and 5 studies having long-term follow-up (i.e., ≥1 year after termination). Too few studies ($K = 2$) provided follow-up data for control treatment comparisons to analyze meta-analytically.

On average, a study compared PDT to 1.38 active treatment groups. The large majority of active treatment groups were CBTs ($n = 13$, 76.5%), including anxiety management training ($n = 1$), applied relaxation training ($n = 3$), CBT ($n = 2$), CBT plus exposure sessions ($n = 1$), cognitive therapy ($n = 2$), prolonged exposure ($n = 2$), and trauma-based/systematic desensitization ($n = 2$). Remaining active comparisons

---

[2] ORs were examined because considering ORs and numerical score outcomes in tandem could tease out different patterns of change across the treatments—whether there is a difference in how many patients meet a cut-off for symptomatic improvement versus their average score outcomes (which could be a mixture of any kind between high, middling, and low responders; see Crits-Christoph et al., 2005 for further discussion).
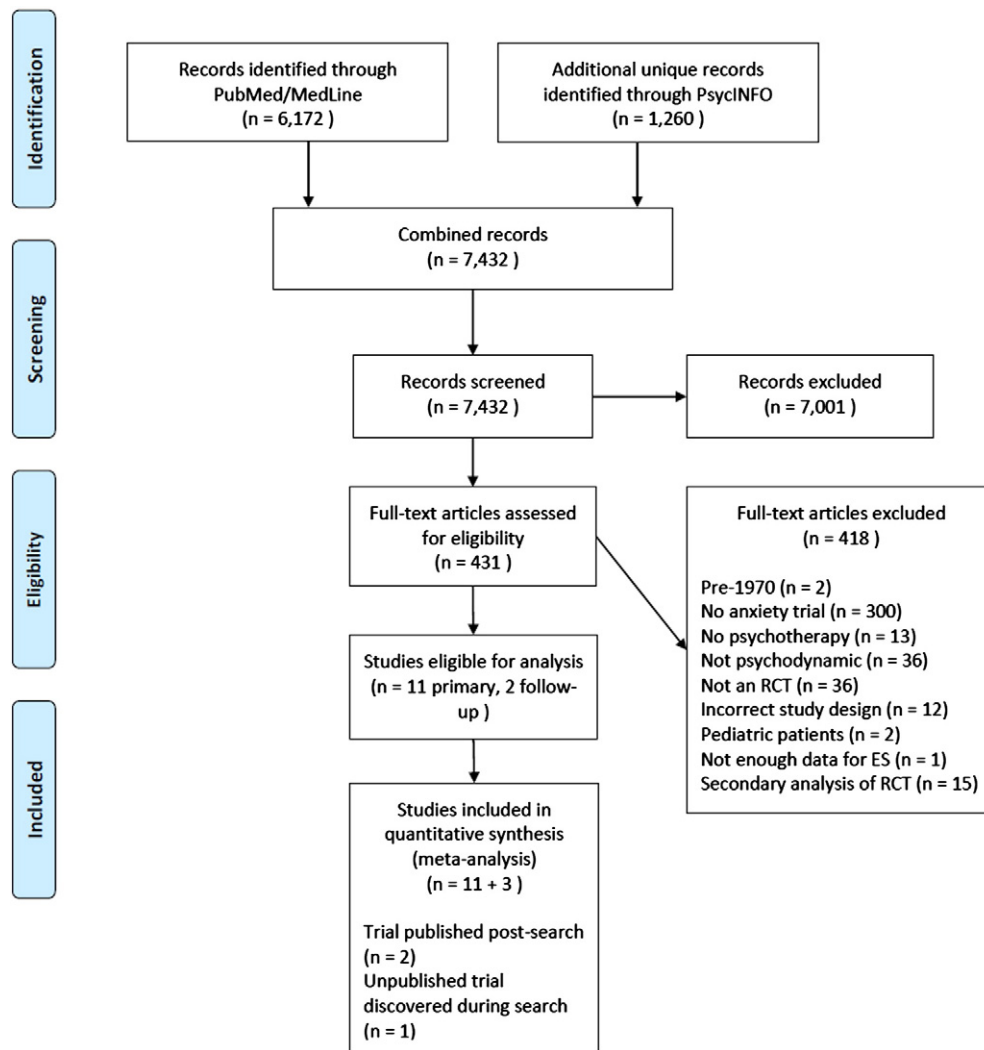
**Fig. 1.** Flow chart of the selection of studies of psychodynamic therapy (PDT) for the treatment of anxiety disorders.

included a behaviorally-aligned hypnotherapy treatment ($n = 1$), solution-focused therapy ($n = 1$), selective serotonin or norepinephrine reuptake inhibitor (SSRI/SNRI) psychopharmacological treatment ($n = 1$), and a manualized supportive therapy ($n = 1$). The majority of treatment comparisons were balanced for dose, with some exceptions (see Table 1; e.g., Knekt et al., 2008).

Control comparisons included an attention placebo group therapy control ($n = 1$), minimal treatment groups ($n = 2$; patients were assessed and given both psychoeducational literature and instructions as to how to self-expose), and wait-lists ($n = 2$). Further information on study characteristics can be found in Table 1 in this report, and the Online Table Supplement (Supplemental Table 1).

### 3.2. Power analysis

We performed a prospective power analysis using final study sample sizes to estimate power for detecting small ($d = 0.20$), medium ($d = 0.50$), and large effects ($d = 0.80$) in the primary random effects meta-analytic estimates comparing PDT to alternative active treatments (e.g., CBT). This was tested in conditions of low (25%), medium (50%), and high (75%) between-study heterogeneity. The method of Hedges and Pigott (2001) was employed in SAS version 9.2 (SAS Institute, Cary, NC) using a macro developed by Cafri, Kromrey, and Brannick (2009). At termination, large and medium effect sizes were essentially certain to be detected regardless of heterogeneity level as per Cafri's

implementation (power $\approx 1.00$). Small effect sizes had power of 0.83, 0.75, and 0.69 respectively in cases of low, medium, and high between-study heterogeneity. Typically, adequate power is defined as 0.80 (Cohen, 1992).

At short-term follow-up (entailing fewer studies), power was estimated to be lower for small effect sizes (though not for medium or large), at 0.77, 0.69, and 0.62 for low, medium, and high heterogeneity. Thus, we expected reasonable power to detect large, medium, and to a lesser certainty small effects at termination, but did not expect to properly detect small effects at short-term follow-up.

### 3.3. Uncontrolled effect sizes

#### 3.3.1. Anxiety outcome analyses

First, we calculated the within-group effect size for PDT using the pre-post scores at termination. Overall, a significant pre-post effect size was calculated for anxiety outcomes for patients in PDTs ($g = 1.063$ [0.791 to 1.334], $p < 0.001$, subject $n = 455$, study $K = 13$).[3] A large amount of heterogeneity was found between study effect sizes ($Q$-test $p = 0.003$, $I^2 = 72.85$), though with no finding of publication bias as per Egger's regression test ($t = 1.071$, $df = 11$, $p = 0.307$) or Henmi and Copas' bias check. Using a leave-one-out sensitivity analysis,

---

[3] See Online Supplemental Fig. 1 for the forest plot showing these results.

we found that no single study drove the finding of significant heterogeneity and that no single study skewed the effect size estimate in either direction. Due to the observed heterogeneity, interpretation of the precise magnitude of the ES estimate should be cautious.

### 3.3.2. Secondary outcome measure analyses

Our analysis estimated that about a sixth of patients dropped out of PDT (weighted average = 17.1% [9.61% to 25.4%], $p < 0.001$, subject $n = 512$, study $K = 14$), though there was substantial heterogeneity among dropout rates from different studies (Q-test p < .001; $I^2 = 80.04$; range = 0% to 59.1%; median = 12.5%). Removing Pierloot & Vinck (1978) notably reduced the observed heterogeneity among dropouts, although the amount of heterogeneity was still significant (Q-test $p < 0.001$, $I^2 = 56.12$). Almost half of individuals experienced clinical remission at termination (as defined per study) after receiving psychodynamic treatment (weighted average = 44.7% [24.4% to 65.1%], $p = 0.002$, subject $n = 389$, study $K = 7$), but a very large amount of heterogeneity was detected among study outcomes (Q-test $p < .001$, $I^2 = 90.31$). No single study drove the effect or the finding of heterogeneity.

### 3.3.3. Moderators of effect size

As there was substantial heterogeneity in the primary effect size estimate, we performed exploratory moderator analyses to find associations between study characteristics and effect sizes (see Online Supplemental Table 2). First, neither the Total RCT-PQRS quality score ($p = 0.106$) nor the Overall/Item 25 score ($p = 0.160$) nor a dichotomous criterion of being above or below Gerber et al.'s (2011) cutoff for adequate study quality (total score ≥24; $p = 0.212$) were significantly related to the uncontrolled effect size. Using the permutation test (see Methods), however, total RCT-PQRS quality scores predicted higher uncontrolled effect sizes at the level of a statistical trend (β = 0.020, $p = 0.087$).

Neither treatment of social phobias defined broadly ($p = 0.696$), nor specifically social anxiety disorder ($p = 0.935$), nor GAD ($p = 0.800$), nor panic ($p = 0.174$) was significantly associated with uncontrolled effect size. In addition, differences in remission rates between disorder groups did not explain heterogeneity among trials. For example, the social anxiety disorder studies sharply differed in percentage remission: the Bögels et al. (in press) trial reported a remission rate of about 59% for PDT (average between two measures of remission), while the larger-scale Leichsenring et al. (2013) trial found a remission rate of around 26%. Furthermore, Crits-Christoph, Gibbons, Narducci, Schamberger, and Gallop (2005) reported that approximately 50% of GAD patients remitted under supportive–expressive PDT, compared to only 10% of GAD patients under the more generic Durham et al. (1994) psychodynamic treatment protocol.

One moderator was found to be significant. Using meta-regression, we found that the greater the average experience of psychodynamic therapists in the trial, the larger the observed effect of PDT ($p = 0.029$, β = 0.058), though the addition of this term still left significant unexplained heterogeneity in the model (Q-test $p = 0.023$). There was also a trend for studies with more dropout in PDT to have a lower uncontrolled effect size ($p = 0.080$, β = −0.016).

### 3.4. Controlled effect sizes

### 3.4.1. Anxiety outcome analyses

At termination, PDT was found to be significantly superior to control conditions with a medium effect size ($g = 0.643$ [0.346 to 0.941], $p = 0.004$, subject $n = 421$, study $K = 5$).[4] No significant heterogeneity was found (Q-test $p = 0.630$), nor was there indication of publication bias as per Henmi & Copas' publication bias check, nor sensitivity to the removal of any one study. Precise magnitude of effect size should

---

[4] See Online Supplemental Fig. 2 for the forest plot showing these results.

be interpreted cautiously due the limited types of comparisons available in this particular analysis—almost exclusively treating social phobias and using different control conditions.

### 3.4.2. Moderators of effect size

We did not find any significant difference ($p = 0.976$) in controlled effect size between trials employing inactive controls (i.e., wait-list) and more active controls (e.g., control therapy group), though this may be an issue of power considering the small number of comparisons. Dichotomous quality ($p = 0.336$) was not significantly related to effect sizes. Given the small number of studies available to analyze no further moderator analyses were run.

### 3.5. Active treatment comparisons

Lastly, we estimated the effect sizes for PDT relative to active treatments conditions at termination, short-term follow-up, and long-term follow-up.

### 3.5.1. Termination

First, we analyzed the effect of PDT compared to active conditions at the time of therapy termination.

*Anxiety outcome analyses.* PDT did not differ from other active treatments ($g = 0.024$ [−0.212 to 0.259], $p = 0.831$, subject $n = 1043$, study $K = 13$; see Fig. 2). Medium amounts of heterogeneity were found (Q-test $p = 0.005$, $I^2 = 61.95$), suggesting the existence of variability between studies unexplained by within-study variance. Removal of the Milrod, Leon, Busch, et al. (2007) trial caused a detectable drop in heterogeneity (Q-test $p = 0.111$, $I^2 = 46.36$, removing 25.2% of previous heterogeneity as per the I² index) suggesting the Milrod trial may be introducing heterogeneity into the analysis with an atypically high effect size ($g = 0.89$) relative to other included studies. However, our primary effect size finding was unaffected by any single study's removal, including the Milrod trial ($g = -0.056$, $p = 0.537$ when removed). Egger's test for publication bias was insignificant ($t = 0.983$, $df = 10$, $p = 0.349$), as was Henmi & Copas' sensitivity check for publication bias.

*Secondary outcome measure analyses.* No significant differences for secondary outcomes were found between psychodynamic therapy and other active treatments, including: depressive symptoms ($g = -0.140$ [−0.792 to 0.512], $p = 0.584$, subject $n = 641$, study $K = 5$), interpersonal problems ($g = 0.028$ [−1.135 to 1.192], $p = 0.918$, subject $n = 512$, study $K = 3$), rates of remission or response (log OR = 0.133 [−0.940 to 1.205], $p = 0.772$, subject $n = 771$, study $K = 7$), or drop-out (log RR = −0.010 [−0.255 to 0.235], $p = 0.931$, subject $n = 1043$, $K = 13$). However, there were significant and high levels of between-study heterogeneity for depression (Q-test $p = 0.005$, $I^2 = 81.93$), interpersonal problems (Q-test $p = 0.017$, $I^2 = 76.57$), and remission (Q-test $p < .001$, $I^2 = 82.18$), though not for relative risk of dropout (Q-test $p = 0.862$). When examining remission from anxiety disorder and depressive symptomatology, the removal of any single study did not change the significance of the results, nor heterogeneity findings. However, there were not enough studies examining interpersonal problems as an outcome ($K = 3$) to make this a coherent check.

*Moderators of effect size.* There was no significant relation between effect size and either Total ($p = 0.887$) or Overall ($p = 0.675$) RCT-PQRS quality score, or having a quality score above or below the 24-point adequate quality cutoff ($p = 0.702$). There was no association of combined sample size and between-groups effect ($p = 0.380$), nor was there any association of between-groups disparity in sessions received (i.e., one therapy receiving a higher "dose" than the other) and effect size ($p = 0.459$). Furthermore, we did not find any significant differences in outcome between studies primarily treating social phobias and remaining studies ($p = 0.802$), nor between studies primarily

| Author(s) and Year | | Hedge's g [95% CI] |
|---|---|---|
| Alstrom et al. (1984a) | | 0.53 [ -0.11 , 1.17 ] |
| Alstrom et al. (1984b) | | 0.08 [ -0.60 , 0.76 ] |
| Beutel et al. (2013) | | -0.31 [ -0.88 , 0.26 ] |
| Bögels et al. (in press) | | 0.23 [ -0.33 , 0.78 ] |
| Bressi et al. (2010) | | -0.09 [ -0.71 , 0.52 ] |
| Brom et al. (1989) | | 0.02 [ -0.44 , 0.49 ] |
| Crits-Christoph et al. (2005) | | 0.24 [ -0.48 , 0.97 ] |
| Durham et al. (1994) | | -0.47 [ -0.93 , -0.01 ] |
| Knekt et al. (2008) | | 0.31 [ -0.14 , 0.75 ] |
| Leichsenring et al. (2009) | | -0.23 [ -0.74 , 0.28 ] |
| Leichsenring et al. (2013) | | -0.29 [ -0.49 , -0.10 ] |
| Milrod et al. (2007b) | | 0.89 [ 0.31 , 1.47 ] |
| Pierloot & Vinck (1978) | | -0.32 [ -1.15 , 0.50 ] |
| RE Model / p=0.831 | | 0.02 [ -0.21 , 0.26 ] |

-2.00   -1.00   0.00   1.00   2.00

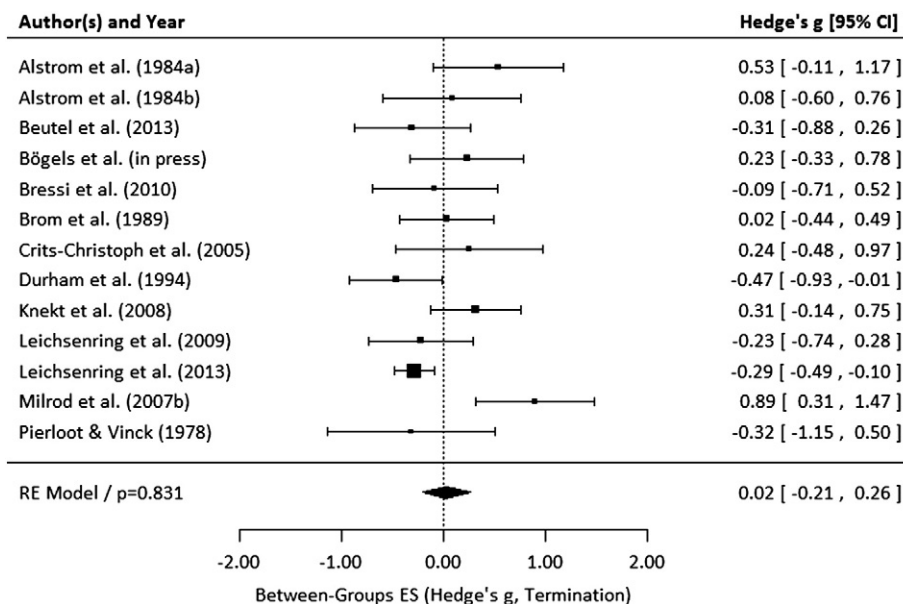Between-Groups ES (Hedge's g, Termination)

**Fig. 2.** Forest plot of between-group effect sizes of anxiety symptom outcomes of PDT as compared to active treatments at termination. *Note.* The squares represent the effect sizes for each study, the size of the square the relative weighing of the study in the analysis, and the errors bars indicate a 95% confidence interval. Positive g represents advantage for PDT relative to other active treatments.

treating social anxiety disorder specifically ($p = 0.604$), GAD[5] ($p = 0.166$), or panic ($p = 0.337$) and remaining studies.

Three trials compared PDT to an active treatment that was not a CBT (Bressi et al., 2010; Crits-Christoph et al., 2005; Knekt et al., 2008), while all other studies used active treatments with predominantly cognitive-behavioral features (e.g., CBT, CT, exposure therapy, applied relaxation, behavioral hypnotherapy). However, we did not find any significant difference in between-groups effect sizes in trials comparing PDT to a CBT versus a non-CBT active treatment group ($p = 0.521$).

A subset of trials ($K = 5$) which investigated PDTs included a secondary exposure component within the therapeutic protocol in addition to more typically-defined psychodynamic techniques (Alstrom, Norlund, Persson, Harding, & Ljungqvist, 1984a,b provided all patients with instructions as to how to self-expose and psychoeducation supporting its use; Crits-Christoph et al., 2005 and Leichsenring et al., 2009, 2013 use therapy manuals indicating exercises for patients to self-expose or confront threatening situations and then discuss within a psychodynamic frame). However, presence of an exposure component was not related to effect sizes ($p = 0.866$).

Studies in which fewer patients dropped out of PDT had significantly better outcomes against comparison treatments compared to studies in which more PDT patients dropped out (meta-regression $p = 0.038, \beta = -0.018$). Convergently, we also found a significant relation between the log relative risk (RR) of dropout and PDT performance. As fewer psychodynamic patients relative to comparison treatment patients dropped out of treatment, effect sizes increasingly favored PDTs (meta-regression $p < 0.001, \beta = -0.698$). Both moderators also substantively explained residual heterogeneity in the model such that remaining heterogeneity did not meet significance; furthermore, when entered into the same meta-regression, both RR ($p < .001$) and percentage dropout in the PDT group ($p = 0.038$) independently explained variability in effect size in the same direction (i.e., more dropout and

more RR for psychodynamic therapy predict worse effect sizes). There was also a significant finding wherein studies in which psychodynamic clinicians were on average more experienced had more favorable effect sizes for the PDTs ($p = 0.037, \beta = 0.048$). No remaining moderators examined were significant (see Table 2).

*3.5.2. Short-term follow-up*

We then examined the effect size of PDT compared to active conditions at follow-up occurring less than a year from termination.

*Anxiety outcome analyses.* At short-term follow-up, no significant outcome differences between PDT and active treatment groups emerged ($g = -0.109$ [$-0.447$ to $0.229$], $p = 0.484$, subject $n = 918$, study $K = 10$; see Fig. 3). The finding was not sensitive to the removal of any single study, and both Egger's test of publication bias ($t = 0.527$, $df = 8$, $p = 0.612$) and Henmi & Copas' publication bias sensitivity check were insignificant. We found significant, high levels of heterogeneity among study outcomes ($Q$-test $p < .001$, $I^2 = 74.33$). Overall heterogeneity was significantly reduced ($Q$-test $p = 0.099$, $I^2 = 50.12$, removing 32.6% of previous heterogeneity) for removing Durham et al. (1994; negative outlier) but not Alstrom et al. (1984a; positive outlier; $Q$-test $p = 0.007$, $I^2 = 67.38$), suggesting that Durham et al. (1994) may be especially unrepresentative of the overall effect size distribution. Interestingly, removing Durham et al. (1994) also brought the estimate of weighted ES to nearly zero ($g = -0.005$, $p = 0.963$). Removing both trials substantially reduced heterogeneity ($Q$-test $p = 0.443$, $I^2 = 25.19$) while resulting in a similarly nonsignificant effect estimate ($g = -0.092$, $p = 0.272$).

*Secondary outcome measure analyses.* A minority of trials published remission data for follow-up, which did not evidence significantly different rates of remission between PDT and other active treatments (log OR $= -0.476$ [$-1.770$ to $0.819$], $p = 0.365$, subject $n = 694$, $K = 5$). A high level of heterogeneity emerged in this comparison ($Q$-test $p = 0.024$, $I^2 = 81.94$). This heterogeneity dissipated ($Q$-test $p = 0.369$, $I^2 = 22.46$, removing 72.6% of previous heterogeneity) when removing from the analysis the negative outlier of Durham et al., 1994 (log OR $= -2.56$), resulting in another overall null finding (log OR $= -0.209$ [$-0.794$ to $0.376$], $p = 0.338$, $K = 4$).

---

[5] One counterargument to this moderator finding at termination might be that one of the GAD comparisons used a manualized supportive therapy (Crits-Christoph et al., 2005) that may be a relatively weaker therapy for GAD compared to a CBT. There was indication that this may be a reasonable hypothesis—if comparing at termination just the three studies utilizing CBT as a comparator for GAD and excluding the Crits-Christoph trial, PDT exhibits a small-to-medium effect size disadvantage at termination ($\beta = -0.471$, $p = 0.069$; CBT as active comparison for GAD studies $g = -0.35$, $p = 0.119$; remaining studies $g = 0.12$, $p = 0.291$).

**Table 2**
Moderator findings for active treatment comparisons (treatment termination).

| Moderator | Significance of moderator | β of effect | df or Subgroup sizes (Yes/No) |
|---|---|---|---|
| Year of publication | p = .871 | −0.002 | df = 11 |
| Average age (years) | p = .230 | −0.031 | df = 11 |
| Percent female | p = .230 | 0.009 | df = 11 |
| Total sample size | p = .380 | −.001 | df = 11 |
| Number of psychodynamic sessions | p = .690 | −0.006 | df = 11 |
| Dose difference between therapies | p = .459 | 0.023 per session advantage of PDT | df = 11 |
| Avg. # years psychodynamic therapy experience | p = .037 | 0.048 | df = 11 |
| # of Psychodynamic therapists | p = .344 | −0.007 | df = 9 |
| Psychodynamic therapy manualization | p = .770 | −0.068 | 7/6 |
| Psychodynamic therapy disorder-specific model | p = .906 | 0.028 | 7/6 |
| Psychodynamic therapy Luborsky's SE | p = .341 | −0.244 | 3/10 |
| Psychodynamic therapy adherence check | p = .984 | −0.005 | 7/6 |
| % Dropout in psychodynamic therapy | p = .038 | −0.018 | df = 11 |
| Log relative risk of dropout | p = .000 | −0.701 (RR favoring PDT increases ES) | df = 11 |
| Data analysis: intention to treat? | p = .661 | 0.106 | 8/5 |
| Reliable diagnosis | p = .702 | −0.100 | 9/4 |
| Primary disorder: GAD | p = .166 | −0.343 | 4/9 |
| Primary disorder: panic | p = .337 | 0.307 | 2/11 |
| Primary disorder: social phobias | p = .802 | 0.062 | 4/9 |
| Primary disorder: social anxiety disorder | p = .604 | −0.149 | 2/11 |
| Exposure component in psychodynamic therapy | p = .866 | −0.040 | 5/8 |
| Comparison treatment: CBTs | p = .521 | 0.178 | 10/3 |
| RCT-PQRS total quality score (items 1–24) | p = .887 | 0.002 | df = 11 |
| RCT-PQRS overall quality score (item 25) | p = .675 | −0.037 | df = 11 |
| RCT-PQRS dichotomous ≥ 24 total quality score | p = .702 | −0.100 | 9/4 |

*Moderators of effect size.* Again, there was no significant relation between short-term follow-up effect size and Total RCT-PQRS quality score ($p = 0.556$), Overall/Item 25 score ($p = 0.465$), or being over or under "adequate" study quality ($p = 0.167$). There was no association between either sample size ($p = 0.771$) or between-therapy dose discrepancy ($p = 0.782$) and between-groups effect.

PDT performed significantly worse with a medium effect size against comparison treatments in the 3 studies treating GAD symptomatology versus other studies ($p = 0.032$, $β = −.677$; GAD studies: $g = −0.607$ [$−1.121$ to $−0.093$], $p = 0.026$; non-GAD studies: $g = 0.070$ [$−0.245$ to $0.384$], $p = 0.624$). The difference in effect between GAD trials relative to remaining non-GAD trials explained a proportion of heterogeneity in the follow-up analysis ($I^2$ dropped to 63.37%), though

after adjustment there was still significantly more heterogeneity among all studies included in the follow up than would be expected given within-study variances (after adjustment Q-test $p = 0.018$). However, using the moderation effect of studies involving GAD vs. non GAD explained less variability of effect among studies than simply eliminating the particular Durham GAD trial (32.6% of heterogeneity eliminated for removal of Durham vs. 14.7% of heterogeneity for moderation). This may be because the Durham trial has a substantially more negative effect size ($g = −1.11$) even among the remaining GAD trials ($g = −0.17$ & $−0.35$). Keeping in mind that removing one of 3 trials at follow-up reduces power, removing the Durham trial resulted in non-significant differences between PDT and other active treatments at short-term follow-up ($p = 0.249$, $β = −0.337$). These findings again



| Author(s) and Year | | Hedge's g [95% CI] |
|---|---|---|
| Alstrom et al. (1984a) | | 0.72 [ 0.08 , 1.36 ] |
| Alstrom et al. (1984b) | | −0.18 [ −0.92 , 0.57 ] |
| Beutel et al. (2013) | | −0.10 [ −0.67 , 0.46 ] |
| Bögels et al. (in press) | | 0.13 [ −0.43 , 0.68 ] |
| Brom et al. (1989) | | 0.23 [ −0.23 , 0.70 ] |
| Durham et al. (1994) | | −1.11 [ −1.59 , −0.62 ] |
| Knekt et al. (2008) | | 0.13 [ −0.31 , 0.58 ] |
| Leichsenring et al. (2009) | | −0.35 [ −0.86 , 0.17 ] |
| Leichsenring et al. (2013) | | −0.26 [ −0.45 , −0.06 ] |
| Pierloot & Vinck (1978) | | −0.17 [ −0.99 , 0.65 ] |
| RE Model / p=0.484 | | −0.11 [ −0.45 , 0.23 ] |

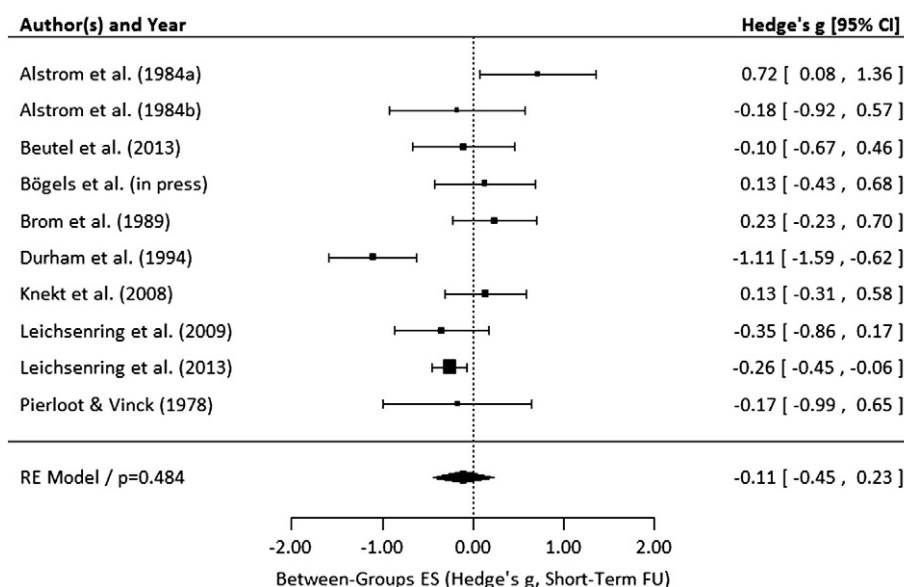Between-Groups ES (Hedge's g, Short-Term FU)

**Fig. 3.** Forest plot of between-group effect sizes of anxiety symptom outcomes of PDT as compared to active treatments at short-term follow-up (<1 year post-termination). *Note.* The squares represent the effect sizes for each study, the size of the square the relative weighing of the study in the analysis, and the errors bars indicate a 95% confidence interval. Positive *g* represents advantage for PDT relative to other active treatments.

**Table 3**
Moderator findings for active treatment comparisons (short-term follow-up).

| Moderator | Significance of moderator | β of effect | df or Subgroup sizes (yes/no) |
|---|---|---|---|
| Year of publication | p = .700 | −0.005 | df = 8 |
| Average age (years) | p = .351 | −0.033 | df = 8 |
| Percent female | p = .308 | 0.010 | df = 8 |
| Total sample size | p = .771 | −.001 | df = 8 |
| Number of psychodynamic sessions | p = .953 | −0.001 | df = 8 |
| Dose difference between therapies | p = .782 | 0.011 per session advantage of PDT | df = 8 |
| Avg. # years psychodynamic therapy experience | p = .664 | 0.019 | df = 8 |
| # of Psychodynamic therapists | p = .756 | −0.004 | df = 6 |
| Psychodynamic therapy manualization | p = .929 | −0.029 | 4/6 |
| Psychodynamic therapy disorder-specific model | p = .851 | 0.070 | 5/5 |
| Psychodynamic therapy Luborsky's SE | p = .516 | −0.246 | 2/8 |
| Psychodynamic therapy adherence check | p = .796 | −0.085 | 4/6 |
| % Dropout in psychodynamic therapy | p = .441 | −0.009 | df = 8 |
| Log relative risk of dropout | p = .002 | −1.474 | df = 8 |
| Data analysis: intention to treat? | p = .943 | 0.023 | 5/5 |
| Reliable diagnosis | p = .167 | −0.448 | 6/4 |
| Primary disorder: GAD | p = .032 | −0.677 (from intercept 0.070) | 3/7 |
| Primary disorder: social phobias | p = .365 | 0.294 | 4/6 |
| Primary disorder: social anxiety disorder | p = .962 | 0.018 | 2/8 |
| Exposure component in psychodynamic therapy | p = .768 | 0.098 | 4/6 |
| RCT-PQRS total quality score (items 1–24) | p = .556 | −0.010 | df = 8 |
| RCT-PQRS overall quality score (item 25) | p = .465 | −0.086 | df = 8 |
| RCT-PQRS dichotomous ≥ 24 total quality score | p = .167 | −0.448 | 6/4 |

suggest that the Durham trial may be an outlier rather than merely presenting a representative effect of PDT for GAD.

Conversely, we found no significant differences in effect size when comparing studies treating a social phobia versus remaining studies ($p = 0.365$), nor specifically social anxiety disorder ($p = 0.962$). Only trials employing CBT comparisons included follow-up—with the exception of Knekt et al. (2008)—so investigating comparison treatment modality as a moderating variable was not possible.

In addition, there was again a significant relation between the log RR of dropout and effect size such that lower relative dropout in the psychodynamic condition predicted more favorable effect sizes for PDT ($p = 0.003$, $\beta = -1.462$). The inclusion of this term explained the previously large amounts of residual heterogeneity in the effect size estimate. No other moderators were significant (see Table 3).

*3.5.3. Long-term follow-up*

Finally, we examined the effect size of PDT versus active treatment conditions at follow-up at least one year after termination.

*Anxiety outcome analyses.* Only 5 studies reported on follow-up data tracking patients a year or more after therapy termination. The longest follow-up available for each study was used (see Table 1), and no significant difference was found between PDT and other active treatments ($g = -0.257$ [−0.899 to 0.385], $p = 0.329$, subject $n = 678$, study $K = 5$).[6] The result was insensitive to a leave-one-out check and Henmi & Copas' publication bias sensitivity check. A large amount of heterogeneity was detected ($Q$-test $p = 0.005$, $I^2 = 83.70$) that was entirely driven by the inclusion of the Durham et al. (1994) trial (when removed, $Q$-test $p = 0.359$, $I^2 = 32.57$; $g = -0.101$, $p = 0.375$). This again suggested that Durham et al. (1994) was a negative outlier at follow-up.

*Secondary outcome measure analyses.* Using the remission rates from these studies produced a similar null finding (log OR = −0.232 [−1.291 to 0.826], $p = 0.536$, subject $n = 624$, study $K = 4$) with medium-high heterogeneity ($Q$-test $p = 0.212$, but $I^2 = 63.03$) again entirely driven by the Durham trial (when removed $I^2 = 15.63$, log OR = −0.107, $p = 0.584$).

---

[6] See Online Supplemental Fig. 3 for the forest plot showing these results.

*Moderators of effect size.* At long-term follow-up all studies were "adequate" or above quality as per Gerber et al. (2011) norms, and all but one (Knekt et al., 2008) used a CBT active treatment comparison. Three major categorical differences between the studies existed: two of the five studies were manualized with a variant of supportive–expressive PDT, and two studies each dealt with social anxiety or GAD patients exclusively. No moderator was significant. Due to the small number of studies no further moderators were examined.

## 4. Discussion

Despite commonly held views (cf. Tolin, 2010), PDT did not differ in effect from other active treatments in the treatment of anxiety disorders. These results were consistent across primary and secondary outcome measures and were maintained at follow-up up to and over a year. We also found evidence of a medium-sized controlled effect size for PDT in comparison with control conditions. However, one needs to keep in mind that the small number of comparisons used—primarily examining social phobias—and the different intensities of control conditions (e.g., wait-list, minimum treatment control, "placebo" therapy) restrict our meta-analytic conclusion concerning control conditions.

Both the weighted average dropout (17.1%) and the insignificant difference in relative risk of dropout between psychodynamic and comparison active treatments suggest that PDT as conducted in RCTs is relatively well-tolerated at a rate comparable to CBTs (cf. Hofmann & Smits, 2008). We further found that PDTs and other active treatments did not differ from each other in remission rates at termination or follow-up, though with substantial between-study differences in remission rates. Depression and interpersonal problem outcomes at termination were also insignificantly different between psychodynamic and other active treatments, though again with high heterogeneity.

Most analyses undertaken exhibited at least a moderate level of between-study effect size heterogeneity, indicating that while the overall sample of PDTs may be close in effect size to other active treatments, "true" differences may exist between subgroups of studies. However, a significant proportion of observed heterogeneity appeared to be due to two studies with uniquely large effects (Durham et al., 1994; Milrod, Leon, Busch, et al., 2007). We addressed the lack of homogeneity between studies by including the likely presence of significant heterogeneity in the a priori power analyses. The power analyses suggested adequate power (≥80%) to detect medium and large effects (but not

small effects) even in the case of high heterogeneity. It is thus possible that a small ($d \leq .20$), undetected effect size difference may "truly" exist, but it is unlikely that medium-to-large effects were missed. Indeed, for our between-groups comparisons the nonsignificant effect estimates were generally in the range of $\pm g \leq 0.20$ (in favor of PDT at termination, and the reverse at follow-up), though were closer to 0 when heterogeneity-influencing outliers (Durham et al., 1994 at follow-ups, Milrod, Leon, Busch, et al., 2007 at termination) were removed.

There were few significant moderators of effect size that helped explain between-study heterogeneity of effect, perhaps due to the study sample size limiting power and the distribution of potential study characteristics (e.g., only three trials of Luborsky's supportive–expressive therapy). Most reliably, both the dropout rate in PDT and relative risk of dropout compared to other active treatments negatively predicted both uncontrolled and between-group effect sizes for PDT. One interpretation could be that this association is a form of "double dipping" into effect sizes (i.e., if a patient is improving, they are more likely to stay in therapy). Another interpretation is that outcomes were better when patients in a study were able to tolerate the psychodynamic treatment (e.g., found it credible) and thus received a proper "dose" of therapy.

Therapist experience in PDT positively predicted uncontrolled effect size and between-group effect at termination, though not at follow-up. This finding should be interpreted cautiously as therapist experience was reported in different ways in different studies: some as floors for inclusion, some the actual level of experience. More experienced therapists may be better able to recognize and maintain competent therapeutic focus on typical dynamics underlying specific anxiety disorders. However, past research on the role of therapist experience and training on therapy efficacy has been mixed (Beutler et al., 2004). Differences in therapist experience may not be best conceptualized by years in the field but rather by specific time and training concerning particular psychopathology and techniques (Beutler, 1999). This distinction could be especially relevant to PDT, as Leichsenring et al. (2013) found in their social anxiety trial that substantially fewer psychodynamic therapists compared to CBT therapists had ever used any manualized therapy, and that about a third of CBT therapists had used the specific trial manual compared to none of the psychodynamic therapists. Perhaps as a consequence, in this trial PDT therapists' first study cases fared significantly worse than subsequent cases, which were equivalent to CBT cases, possibly because PDT therapists were still adjusting their technique both to manualization generally and the trial PDT specifically (Leichsenring, 2011).

While across the different anxiety disorders PDT was as efficacious as other active treatments, PDT seemed to fare significantly worse for GAD patients at follow-up (but not at termination, though see Footnote 5) compared to other active treatments. Given the relatively poorer quality of the therapy in some of the GAD trials,[7] it is possible that the size of the GAD effect may be mis-estimated. The Durham et al. (1994) RCT of GAD flagged in our analyses as a negative outlier contributing a third or more of all heterogeneity at short-term and long-term follow-up (more than heterogeneity explained by using GAD as a moderator), and even among GAD trials at follow-up had an unusually negative effect size. This suggests that at follow-up the Durham trial is an outlier within the overall effect distribution, and thus may not represent a standard effect of PDT. In addition, follow-up analyses can be problematic as they are often confounded when patients do not follow protocol and seek out additional treatment post-termination (e.g., Knekt et al., 2008 wherein the effect sizes at follow-up were biased by the short-term therapy groups seeking additional treatment). Taken together, the moderator analysis herein cannot provide definitive

answers regarding the efficacy of PDT versus other active treatments for GAD, and future study using high fidelity, well-conceptualized treatments may be necessary (e.g., Leichsenring et al., 2013).

On the other hand, PDT studies treating social phobias, social anxiety disorder, or panic disorder did not show a significantly different effect from the overall distribution of effect sizes, though the latter two comparisons had small study samples (2 each) that would prevent detecting smaller effect size differences. We also did not find evidence that PDT performed better or worse when compared against generally-construed cognitive-behavioral treatments than against other modalities, though this does not preclude superiority or inferiority for more specific subgroups (see Footnote 5).

Finally, we did not find any indication that PDTs containing a secondary, but explicit exposure component were superior in effect size compared to PDTs that did not have these elements. This may be because exposure was usually self-guided and possibly too minimal to have a noticeable effect as per the moderation power. A recent meta-analysis of additive and dismantling component psychotherapy trials suggested that adding a new technique was found to be significantly beneficial for primary symptomatic outcomes, but with only a small effect size ($d = 0.14$ at termination, $d = 0.28$ at follow-up) that would be difficult to detect in our moderator analyses (Bell, Marcus, & Goodlad, 2013). Alternatively, it is entirely possible that elements of exposure or CBT are already present in most PDTs even without the explicit addition of an exposure component (e.g., that repeated recounting of emotional events may entail exposure; cf. Lambert & Ogles, 2004). Moreover, it may also be true that CBTs do sometimes intervene on ostensibly psychodynamic therapeutic concepts (e.g., interpersonal focus in social anxiety disorder; see also Ablon & Jones, 1998; McCarthy & Barber, 2009; cf. Lambert & Ogles, 2004).

### 4.1. Limitations and future directions

A minority of studies included in the meta-analyses fell below "adequate" quality as per Gerber et al.'s (2011) provisional quality cutoff (e.g., Pierloot & Vinck, 1978). In lieu of heavy quality filtering, we decided to quantitatively explore markers of quality in relation to effect size, as more extensive filtering may raise questions as to whether authors excluded studies in a biased manner. Our exploratory moderator analyses on general and specific aspects of study quality found no significant relationships between the RCT-PQRS quality scores of an RCT and uncontrolled or between-groups effect size (though with one possible exception of a positive relation between uncontrolled effect and total quality score, see section 3.3.3), nor between a selection of particular quality variables and effect size (e.g., adherence checks). Furthermore, the two largest effect outliers (in opposite directions) were both of higher rather than lower quality (Durham et al., 1994; Milrod, Leon, Busch, et al., 2007). This suggested that RCT quality as was measurable did not strongly influence our results. This finding is concordant with results from the Gerber et al. (2011) quality-based review of published PDT trials, which across trials found no association between study quality and the outcome of PDT compared to other treatments. Overall, more high-quality studies of PDT for anxiety disorders are still needed (e.g., Leichsenring et al., 2013).

Several studies were of higher quality, but consisted of small subject samples (e.g., Crits-Christoph et al., 2005). Studying small samples leads to concerns about publication bias (Kraemer et al., 1998; cf. Bhar et al., 2010). Essentially, small studies with positive findings (as defined by the investigators) are more likely to be published than small studies with negative findings. In addition, smaller trials may be more likely to show unusual or imprecisely estimated effects compared to larger trials. Unfortunately, most controlled studies of psychotherapy tend to be relatively small. This is true for all types of psychotherapy, including well-validated treatments such as CBTs (Leichsenring & Rabung, 2011a). Interestingly, there is recent simulation evidence suggesting that single very high-powered trials mis-estimate true effects under

---

[7] Pierloot & Vinck, 1978 being a failed trial for both the behavioral and psychodynamic groups and Durham et al., 1994 using a psychodynamic approach with no cited general or disorder-specific theory of treatment and using a long-term model for short-term therapy.

conditions of true heterogeneity and/or publication bias, relative to a series of more modestly powered studies meta-analyzed in tandem, which were more robust against heterogeneity and/or publication bias (IntHout, Ioannidis, & Borm, in press). Thus, well-conducted meta-analyses including smaller studies provide valuable evidence to consider alongside single large trials, especially under conditions that may harm the precision of larger studies (e.g., true effect heterogeneity as in clinical trials of psychotherapy).

We took several analytic precautions against publication bias in this meta-analysis. Our meta-analytic estimator (the Sidik–Jonkman estimator) has been shown to estimate more accurate effect sizes in the presence of publication bias compared to standard random effects meta-analysis (e.g., Henmi & Copas, 2010; Sidik & Jonkman, 2005b). In addition, we did not find evidence of publication bias with Egger's regression test, with leave-one-out sensitivity checks (except for noted contributors to heterogeneity), with meta-regressing study size on effect size, nor by using the Henmi and Copas (2010) publication bias sensitivity check. As several consulted reviews of PDTs have scoured the published and unpublished literature for PDT trials in addition to our own search, we have some degree of confidence that most unpublished trials that could be found have likely been found, and that we have minimized our risk of omitting published and eligible trials (Abbass et al., 2006; Gerber et al., 2011; Leichsenring & Rabung, 2011b; Leichsenring et al., 2004; Slavin-Mulford & Hilsenroth, 2012). It is further unlikely that any unidentified small study would change our primary estimate, given the null findings of our sensitivity analyses that included removal of single small and large studies.

As our primary analyses assessed the effectiveness of PDT for anxiety disorders as a class, findings cannot necessarily be fully extended to individual disorders. Per our moderator analyses, PDT may be less effective on average for GAD. On the other hand, in future trials PDT may be likely to show only small effect size differences at most compared to other therapies when treating the general category of social phobias. From two studies each, there is also evidence that PDT may effectively treat social anxiety disorder and panic disorder. However, not all anxiety disorders were represented equally in the published literature (e.g., only one PTSD trial) and no primary trial of simple phobia or OCD met criteria for inclusion, the latter possibly because in the experience of many clinicians, primary OCD may not respond to PDT (cf. Rice, 2004). More controlled trials investigating PDT for different anxiety disorders would help evaluate for which specific disorders PDT is effective.

Moderator analyses were limited due to the lack of consistent reporting of patient demographics (e.g., Axis-II disorders, comorbidity) and process data (e.g., relation of adherence to outcome) in conjunction with the small study sample. This restricted our ability to further examine the classic question of differential treatment efficacy or "what works for whom" (Paul, 1967). Preliminary data indicate that PDT may be particularly beneficial for personality disordered panic patients (Milrod, Leon, Barber, et al., 2007), and differences in the distribution of personality functioning between patient samples may explain heterogeneity between studies. There is some indication from primary trials of depression and personality disorders that PDT may be comparably less efficacious for patients with avoidant personality features than it is for patients with more obsessive–compulsive personality features, while CBT may show the opposite pattern (Barber, Morse, Krakauer, Chittams, & Crits-Christoph, 1997; Barber & Muenz, 1996; Emmelkamp et al., 2006). If there are indeed populations of patients who preferentially respond to psychodynamic treatment frames, it would be useful for future trials to not only try to "match" current evidence-based treatments in efficacy but to proffer direct evidence to recommend PDT to particular patients based on their characteristics (cf. Beutler & Harwood, 2000).

Finally, like nearly any psychotherapy meta-analysis, our results cannot disentangle what specific components of the therapies—intended or unintended—contributed to their efficacy. Effect size heterogeneity may be especially common to clinical meta-analyses, reflecting differences in clinical practice between even well-monitored trials (e.g., Webb et al., 2013 describe significant differences in cognitive therapy technique using the same manual between high-quality trials and even within sites of the same trial; see also Malik, Beutler, Alimohamed, Gallagher-Thompson, & Thompson, 2003). We observed such significant heterogeneity in this meta-analysis, which is typical for meta-analyses of psychotherapy RCTs (e.g., Cuijpers et al., 2013 for CBT of depression; Budge et al., 2013 for treatments of personality disorders). Regardless, adherence ratings have been shown to have an inconsistent relationship to change across types of therapies and disorders treated (Webb, DeRubeis, & Barber, 2010), indicating more focused research into technique use is necessary, and that ostensible differences in clinical effect between trials may not be well-indexed by general adherence to a particular manual. PDTs may share some common change mechanisms with CBTs (e.g., acquisition of compensatory skills; Gibbons et al., 2009), some particular to PDTs but not CBTs (e.g., insight or self understanding; Gibbons et al., 2009), and some that work to opposite effect in PDTs versus CBTs (e.g., deepening versus avoidance of affect; Ulvenes et al., 2012). Future trials of PDT for anxiety disorders should examine processes of change common among effective therapies and specific to PDTs (e.g., Pitman, Slavin-Mulford, & Hilsenroth, 2014), in terms of specific techniques applied by therapists rather than overall adherence, the downstream mediators of technique use that promote change, and complex interactions between techniques, therapeutic factors (e.g., the therapeutic alliance), and patient variables.

## 5. Conclusion

Overall, our meta-analysis suggests that PDTs as studied in RCTs are as effective at treating anxiety disorders as other active treatments, and more effective than control groups. Medium-high heterogeneity of effect sizes and our moderator analyses suggest that relevant "true" differences may exist between studies (e.g., treating GAD, therapist characteristics). However, removing heterogeneity outliers did not lead to changes in our conclusions. These findings set the stage to recommend conducting further high-quality, controlled trials of anxiety-specific formulations of psychodynamic therapy, especially for understudied areas (e.g., panic disorder, social anxiety disorder, PTSD). We endorse the view that no treatment is going to work for all patients, and that therefore future research should address and refine the subgroups of anxiety patients for whom PDT may be particularly efficacious.

### Acknowledgments

### Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cpr.2014.03.004.

### References

*Alstrom, J. E., Norlund, C. L., Persson, G., Harding, M., & Ljungqvist, C. (1984a). Effects of four treatment methods on agoraphobic women not suitable for insight-oriented psychotherapy. *Acta Psychiatrica Scandinavica, 70*, 1–17.

*Alstrom, J. E., Norlund, C. L., Persson, G., Harding, M., & Ljungqvist, C. (1984b). Effects of four treatment methods on social phobic patients not suitable for insight-oriented psychotherapy. *Acta Psychiatrica Scandinavica, 70*, 97–110.

*Beutel, M. E., Scheurich, V., Knebel, A., Michal, M., Wltink, J., Tschan, R., et al. (2013). Implementing panic-focused psychodynamic therapy into clinical practice. *Canadian Journal of Psychiatry, 58*, 326–334.

*Bögels, S., Wijts, P., Oort, F., & Sallaerts, S. (2014). Psychodynamic psychotherapy versus cognitive behaviour therapy for social anxiety disorder: An efficacy and partial effectiveness trial. *Depression and Anxiety* (in press).

*Bressi, C., Porcellana, M., Marinaccio, P.M., Nocito, E. P., & Magri, L. (2010). Short-term psychodynamic psychotherapy versus treatment as usual for depressive and anxiety disorders: A randomized clinical trial of efficacy. *The Journal of Nervous and Mental Disease, 198*, 647–652.

*Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology, 57*, 607–612.

*Crits-Christoph, P., Gibbons, M. B. C., Narducci, J., Schamberger, M., & Gallop, R. (2005). Interpersonal problems and the outcome of interpersonally oriented psychodynamic treatment of GAD. *Psychotherapy: Theory, Research, Practice, Training, 42*, 211–224.

*Durham, R. C., Fisher, P. L., Treliving, L. R., Hau, C. M., Richard, K., & Stewart, J. B. (1999). One year follow-up of cognitive therapy, analytic psychotherapy and anxiety management training for generalized anxiety disorder: Symptom change, medication usage and attitudes to treatment. *Behavioural and Cognitive Psychotherapy, 27*, 19–35.

*Durham, R. C., Murphy, T., Allan, T., Richard, K., Treliving, L. R., & Fenton, G. W. (1994). Cognitive therapy, analytic psychotherapy and anxiety management training for generalised anxiety disorder. *The British Journal of Psychiatry, 165*, 315–323.

*Knekt, P., Lindfors, O., Härkänen, T., Välikoski, M., Vertala, E., Laaksonen, M.A., et al. (2008). Randomized trial on the effectiveness of long- and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychological Medicine, 38*, 689–703.

*Knekt, P., Lindfors, O., Renlund, C., Kaipainen, M., Mäkelä, P., Järvikoski, A., et al. (2004). A randomized trial of the effects of four forms of psychotherapy on depressive and anxiety disorders. In P. Knekt, & O. Lindfors (Eds.), *Studies in social security and health., Vol. 77*. Helsinki, Finland: Kela: The Social Insurance Institution.

*Knijnik, D. Z., Kapczinski, F., Chachamovich, E., Margis, R., & Eizirik, C. L. (2004). Psicoterapia psicodinâmica em grupo para fobia social generalizada [Group psychodynamic therapy for generalized social phobia]. *Revista Brasileira de Psiquiatria, 26*, 77–81.

*Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., et al. (2013). Psychodynamic therapy and cognitive therapy in social anxiety disorder: A multi-center randomized controlled trial. *American Journal of Psychiatry, 170*, 759–767.

*Leichsenring, F., Salzer, S., Jaeger, U., Kächele, H., Kreische, R., Leweke, F., et al. (2009). Short-term psychodynamic psychotherapy and cognitive-behavioral therapy in generalized anxiety disorder: A randomized, controlled trial. *American Journal of Psychiatry, 166*, 875–881.

*Milrod, B., Leon, A.C., Busch, F., Rudden, M., Schwalberg, M., Clarkin, J., et al. (2007). A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *American Journal of Psychiatry, 164*, 265–272.

*Pierloot, R., & Vinck, J. (1978). Differential outcome of short-term dynamic psychotherapy and systematic desensitization in the treatment of anxious out-patients: A preliminary report. *Psychologica Belgica, 18*, 87–98.

*Salzer, S., Winkelbach, C., Leweke, F., Leibing, E., & Leichsenring, F. (2011). Long-term effects of short-term psychodynamic psychotherapy and cognitive-behavioural therapy in generalized anxiety disorder: 12-month follow-up. *Canadian Journal of Psychiatry, 56*, 503–508.

Abbass, A. A., Hancock, J. T., Henderson, J., & Kisley, S. R. (2006). Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database of Systematic Reviews, 18*, CD004687.

Ablon, J., & Jones, E. (1998). How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychotherapy Research, 8*, 71–83.

Andrews, G., Henderson, S., & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation: Overview of the Australian National Mental Health Survey. *British Journal of Psychiatry, 178*, 145–153.

Andrews, G., Slade, T., & Issakidis, C. (2002). Deconstructing current comorbidity: data from the Australian National Survey of Mental Health and Well-Being. *British Journal of Psychiatry, 181*, 306–314.

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., et al. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review, 33*, 395–405.

Barber, J. P., Barrett, M. S., Gallop, R., Rynn, M.A., & Rickels, K. (2012). Dr. Barber and colleagues reply [letter to the editor]. *Journal of Clinical Psychiatry, 73*, 718–719.

Barber, J.P., & Luborsky, L. (1991). A psychodynamic view of simple phobias and prescriptive matching: A commentary. *Psychotherapy: Theory, Research, Practice, Training, 28*, 469–472.

Barber, J. P., Morse, J. Q., Krakauer, I. D., Chittams, J., & Crits-Christoph, K. (1997). Change in obsessive–compulsive and avoidant personality disorders following time-limited supportive–expressive therapy. *Psychotherapy: Theory, Research, Practice, Training, 34*, 133–143.

Barber, J. P., & Muenz, L. R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: Empirical findings from the treatment for depression collaborative research program. *Journal of Consulting and Clinical Psychology, 64*, 951–958.

Barber, J. P., Muran, J. C., McCarthy, K. S., & Keefe, J. R. (2013). Research on dynamic therapies. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). Hoboken, NJ: John Wiley & Sons.

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., et al. (2011). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*  (1st ed.). New York, NY: Oxford University Press.

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

Beutler, L. E. (1999). The psychotherapist as a neglected variable in psychotherapy: An illustration by reference to the role of therapist experience and training. *Clinical Psychology: Science & Practice, 4*, 44–52.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy: A practical guide to systematic treatment selection.* New York, NY: Oxford University Press.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., et al. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed.). New York, NY: Wiley.

Bhar, S. S., Thombs, B.D., Pignotti, M., Bassel, M., Jewett, L., Coyne, J. C., et al. (2010). Is longer term psychodynamic therapy more effective than shorter term therapies? Review and critique of the evidence. *Psychotherapy and Psychosomatics, 79*, 208–216.

Blagys, M.D., & Hilsenroth, M. J. (2000). Distinctive features of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process Literature. *Clinical Psychology: Science & Practice, 7*, 167–188.

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

Boswell, J. F., Farchione, T. J., Sauer-Zavala, S., Murray, H. W., Fortune, M. R., & Barlow, D. H. (2013). Anxiety sensitivity and interoceptive exposure: A transdiagnostic construct and change strategy. *Behavior Therapy, 44*, 417–431.

Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., et al. (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-year prospective study. *American Journal of Psychiatry, 162*, 1179–1187.

Budge, S. L., Moore, J. T., Del Re, A.C., Wampold, B. E., Baardseth, T. P., & Nienhuis, J. B. (2013). The effectiveness of evidence-based treatments for personality disorders when comparing treatment-as-usual and bona fide treatments. *Clinical Psychology Review, 33*, 1057–1066.

Busch, F. N., Milrod, B.L., Singer, M. B., & Aron, A.C. (2011). *Manual of panic focused psychodynamic psychotherapy—Extended range, Vol. 36*,  New York: Routledge.

Cafri, G., Kromrey, J.D., & Brannick, M. T. (2009). A SAS macro for statistical power calculations in meta-analysis. *Behavioral Research Methods, 41*, 35–46.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Crits-Christoph, P., Wolf-Palacio, D., Ficher, M., & Rudick, D. (1995). Brief supportive–expressive psychodynamic therapy for generalized anxiety disorder. In J. P. Barber, & P. Crits-Christoph (Eds.), *Dynamic therapies for psychiatric disorders (Axis I)* (pp. 43–83). New York: Basic Books.

Cuijpers, P., Berking, M., Andersson, G., Qugiley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Canadian Journal of Psychiatry, 58*, 376–385.

Driessen, E., Cuijpers, P., de Maat, S.C. M., Abbass, A. A., de Jonghe, F., & Dekker, J. J. M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis. *Clinical Psychology Review, 30*, 25–36.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*(34), 917–928.

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

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

Emmelkamp, P.M. G., Benner, A., Kuipers, A., Feiertag, G. A., Koster, H. C., & van Appeldoorn, F. J. (2006). Comparison of brief dynamic and cognitive-behavioral therapies in avoidant personality disorder. *British Journal of Psychiatry, 189*, 60–64.

Fonagy, P., Roth, A., & Higgitt, A. (2005). Psychodynamic psychotherapies: Evidence-based practice and clinical wisdom. *Bulletin of the Menninger Clinic, 69*, 1–58.

Freud, S. (1990). In J. Strachey (Ed.), *Inhibitions, symptoms, and anxiety. Standard edition of the complete psychological works of Sigmund Freud. Trans., Vol. 20*. New York, NY: Norton (Original work published 1926).

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

Gerson, B. P. R., & Carlier, I. V. E. (1994). Treatment of work related trauma in police officers: Posttraumatic stress disorder and posttraumatic decline. In M. B. Williams, & J. F. Sommer (Eds.), *Handbook of posttraumatic therapy* (pp. 325–333). Westport, CT: Greenwood.

Gibbons, M. B. C., Crits-Christoph, P., Barber, J. P., Stirman, S. W., Gallop, R., Goldstein, L. A., et al. (2009). Unique and common mechanisms of change across cognitive and dynamic psychotherapies. *Journal of Consulting and Clinical Psychology, 77*, 801–813.

Goisman, R. M., Warshaw, M. G., & Keller, M. B. (1999). Psychosocial treatment prescriptions for generalized anxiety disorder, panic disorder, and social phobia, 1991-1996. *American Journal of Psychiatry, 156*, 1819–1821.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology, 32*, 50–55.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*, 203–217.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*, 426–445.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Henmi, M., & Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine, 29*, 2969–2983.

Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry, 62*, 182–189.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 15*, 1539–1558.

Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine, 23*, 1663–1682.

Hoffman, E. J., & Mathew, S. J. (2008). Anxiety disorders: A comprehensive review of pharmacotherapies. *Mount Sinai Journal of Medicine, 75*, 248–262.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*(69), 621–632.

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

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2012). Obtaining evidence by a single well-powered trial or several modestly powered trials. *Statistical Methods in Medical Research* (in press).

Inthout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*, e25.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*, 640–648.

Johansson, R., Björklund, M., Hornborg, C., Karlsson, S., Hesser, H., Ljótsson, B., et al. (2013). Affect-focused psychodynamic psychotherapy for depression and anxiety through the internet: A randomized controlled trial. *PeerJ, 1*, e102.

Kashdan, T. B., Barrios, V., Forsyth, J. P., & Steger, M. F. (2006). Experiential avoidance as a generalized psychological vulnerability: Comparisons with coping and emotion regulation strategies. *Behaviour Research and Therapy, 44*, 1301–1320.

Klein, D. F., Zitrin, C. M., Woerner, M. G., & Ross, D. C. (1983). Treatment of phobias. II. Behavior therapy and supportive psychotherapy: Are there any specific ingredients? *Archives of General Psychiatry, 40*, 139–145.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*, 2693–2710.

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

Koen, N., & Stein, D. J. (2011). Pharmacotherapy of anxiety disorders: A critical review. *Dialogues in Clinical Neuroscience, 13*, 423–437.

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

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*, 317–325.

Lambert, M. J., & Ogles, B.M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 139–193) (5th ed.). New York: Wiley.

Leichsenring, F. (2011). Angst vor Menschen: Soziale Phobien [Fear of people: Social phobias]. *Paper presented at the 62. Jahrestagung de Deutsche Gesellschaft für Psychoanalyse, Psychotherapie, Psychosomatik und Tiefenpsychologie62nd Yearly Conference of the German Society for Psychoanalysis, Psychotherapy, Psychosomatics and Depth Psychology.*

Leichsenring, F., Beutel, M., & Leibing, E. (2007). Psychodynamic psychotherapy for social phobia: A treatment manual based on supportive–expressive therapy. *Bulletin of the Menninger Clinic, 71*, 56–83.

Leichsenring, F., & Rabung, S. (2011a). Double standards in psychotherapy research. *Psychotherapy and Psychosomatics, 80*, 48–51.

Leichsenring, F., & Rabung, S. (2011b). Long-term psychodynamic psychotherapy in complex mental disorders: Update of a meta-analysis. *British Journal of Psychiatry, 199*, 15–22.

Leichsenring, F., Rabung, S., & Leibing, E. (2004). The efficacy of short-term psychodynamic therapy in specific psychiatric disorders: A meta-analysis. *Archives of General Psychiatry, 61*, 1208–1216.

Leichsenring, F., & Salzer, S. (2014s). A unified protocol for the transdiagnostic psychodynamic treatment of anxiety disorders: An evidence-based approach. *Psychotherapy* (in press).

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

Malik, M. L., Beutler, L. E., Alimohamed, S., Gallagher-Thompson, D., & Thompson, L. (2003). Are all cognitive therapies alike? A comparison of cognitive and noncognitive therapy processes and implications for the application of empirically supported treatments. *Journal of Consulting and Clinical Psychology, 71*, 150–158.

Martini, B., Rosso, G., Chiodelli, D. F., Cori, D. D., & Maina, G. (2011). Brief dynamic therapy combined with pharmacotherapy in the treatment of panic disorder with concurrent depressive symptoms. *Clinical Neuropsychiatry, 8*, 204–211.

McAleavey, A. A., Castonguay, L. G., & Goldfried, M. R. (2014). Clinical experiences in conducting cognitive-behavior therapy for social phobia. *Behavior Therapy, 45*, 21–35.

McCarthy, K. S., & Barber, J. P. (2009). The multitheoretical list of therapeutic interventions (MULTI): Initial report. *Psychotherapy Research, 19*, 96–113.

McTeague, L. M., & Lang, P. J. (2012). The anxiety spectrum and the reflex physiology of defense: From circumscribed fear to broad distress. *Depression and Anxiety, 29*, 264–281.

Milrod, B., Leon, A., Barber, J. P., Markowtiz, J., & Graf, E. (2007). Do comorbid personality disorders moderate panic-focused psychotherapy? An exploratory examination of the American Psychiatric Association practice guideline. *Journal of Clinical Psychiatry, 68*, 885–891.

Monsen, K. (1989). *Psykodynamisk kroppsterapi [Psychodynamic body therapy].* Oslo, Norway: Tano.

Olatunji, B. O., Cisler, J. M., & Deacon, B. J. (2010). Efficacy of cognitive behavioral therapy for anxiety disorders: A review of meta-analytic findings. *Psychiatric Clinics of North America, 33*, 557–577.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*, 109–118.

Pitman, S., Slavin-Mulford, J., & Hilsenroth, M. (2014). Psychodynamic techniques related to outcome for anxiety disorder patients at different points in treatment. *Journal of Nervous and Mental Disease, 202*, 391–396.

R Core Team (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing (Retrieved from http://www.R-project.org).

Rice, E. (2004). Reflections on the obsessive–compulsive disorders: A psychodynamic and therapeutic perspective. *Psychoanalytic Review, 91*, 23–44.

Shear, M. K., Houck, P., Greeno, C., & Masters, S. (2001). Emotion-focused psychotherapy for patients with panic disorder. *American Journal of Psychiatry, 158*, 1993–1998.

Shrout, P. E. (1995). Reliability. In M. T. Tsuang, M. Tohen, & G. E. P. Zahner (Eds.), *Textbook in psychiatric epidemiology* (pp. 213–227). New York: Wiley-Liss.

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

Sidik, K., & Jonkman, J. N. (2005a). A note on variance estimation in random effects meta-regression. *Journal of Biopharmaceutical Statistics, 15*, 823–838.

Sidik, K., & Jonkman, J. N. (2005b). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society, Series C, 54*, 367–384.

Slavin-Mulford, J., & Hilsenroth, M. (2012). Evidence-based psychodynamic treatments for anxiety disorders: A review. In R. A. Levy, J. S. Ablon, & H. Kaechele (Eds.), *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (pp. 117–137). New York: Springer Press.

Somers, J. M., Goldner, E. M., Waraich, P., & Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: A systematic review of the literature. *Canadian Journal of Psychiatry, 51*, 100–113.

Summers, R. J., & Barber, J. P. (2009). *Dynamic psychotherapy: A guide to evidence-based practice.* New York, NY: Guilford Press.

Szkodny, L. E., Newman, M. G., & Goldfried, M. R. (2014). Clinical experiences in conducting empirically supported treatments for generalized anxiety disorder. *Behavior Therapy, 45*, 7–20.

Tambs, K., Czajkowsky, N., Røysamb, E., Neale, M. C., Reichborn-Kjennerud, T., Aggen, S. H., et al. (2009). Structure of genetic and environmental risk factors for dimensional representations of DSM-IV anxiety disorders. *British Journal of Psychiatry, 195*, 301–307.

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical Psychology Review, 30*, 710–720.

Ulvenes, P. G., Berggraf, L., Hoffart, A., Stiles, T. C., Svartberg, M., McCullough, L., et al. (2012). Different process for different therapies: Therapist actions, therapeutic bond, and outcome. *Psychotherapy, 49*, 291–302.

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

Wachtel, P. L. (1977). *Psychoanalysis and behavior therapy: Toward an integration.* New York, NY: Basic.

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

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

Webb, C. A., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J.D., & Shelton, R. C. (2013). Convergence and divergence in the delivery of cognitive therapy in two randomized clinical trials. *Behaviour Research and Therapy, 51*, 493–498.

Wolf, A. W., & Goldfried, M. R. (2014). Clinical experiences in using cognitive-behavioral therapy to treat panic disorder. *Behavior Therapy, 45*, 36–46.