The enduring effects of psychodynamic treatments vis-à-vis alternative treatments: A multilevel longitudinal meta-analysis

D. Martin Kivlighan III [a,*], Simon B. Goldberg [a], Maleeha Abbas [a], Brian T. Pace [b], Noah E. Yulish [a], Joel G. Thomas [c], Megan M. Cullen [a], Christoph Flückiger [d], Bruce E. Wampold [a,e]

[a] University of Wisconsin-Madison, United States
[b] University of Utah, United States
[c] Universtiy of Illinois, United States
[d] University of Zürich, Zürich, Switzerland
[e] Modum Bad Psychiatric Center, Vikersund, Norway

HIGHLIGHTS

- Examined the enduring impact of dynamic treatments versus non-dynamic treatments.
- Calculated four ESs; targeted, non-targeted, personality, and combined measures.
- Treatments were not significantly different at post-treatment for all four ESs.
- Post-treatment slopes for all four ESs were non-significant.
- Dynamic and non-dynamic treatments were equivalent at post-treatment and beyond.

ARTICLE INFO

ABSTRACT

Although evidence suggests that the benefits of psychodynamic treatments are sustained over time, presently it is unclear whether these sustained benefits are superior to non-psychodynamic treatments. Additionally, the extant literature comparing the sustained benefits of psychodynamic treatments compared to alternative treatments is limited with methodological shortcomings. The purpose of the current study was to conduct a rigorous test of the growth of the benefits of psychodynamic treatments relative to alternative treatments across distinct domains of change (i.e., all outcome measures, targeted outcome measures, non-targeted outcome measures, and personality outcome measures). To do so, the study employed strict inclusion criteria to identify randomized clinical trials that directly compared at least one bona fide psychodynamic treatment and one bona fide non-psychodynamic treatment. Hierarchical linear modeling (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) was used to longitudinally model the impact of psychodynamic treatments compared to non-psychodynamic treatments at post-treatment and to compare the growth (i.e., slope) of effects beyond treatment completion. Findings from the present meta-analysis indicated that psychodynamic treatments and non-psychodynamic treatments were equally efficacious at post-treatment and at follow-up for combined outcomes ($k = 20$), targeted outcomes ($k = 19$), non-targeted outcomes ($k = 17$), and personality outcomes ($k = 6$). Clinical implications, directions for future research, and limitations are discussed.

Published by Elsevier Ltd.

Contents

* Corresponding author at: Department of Counseling Psychology, University of Wisconsin-Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI 53706-1326, United States. Tel.: +1 202 422 3664.
   E-mail address: kivlighan@wisc.edu (D.M. Kivlighan).

1. Introduction

Contentious debates exist regarding the superiority of competing psychotherapy treatments. As an example, proponents of cognitive-behavioral therapy (CBT) have claimed superiority to alternative treatments for the past several decades (Eysenck, 1994; Hofmann & Lohr, 2010; Siev & Chambless, 2007; Siev, Huppert, & Chambless, 2009; Tolin, 2010). These claims are supported by various meta-analyses over the years. Specifically, Shapiro and Shapiro (1982) meta-analyzed 143 comparative studies and found that behavioral treatments were superior to psychodynamic and humanistic treatments. More recently, Tolin (2010) meta-analyzed 26 studies examining the efficacy of CBT vis-à-vis alternative treatments and concluded that CBT was superior to psychodynamic treatments for depressive and anxiety disorders. In a review of meta-analyses, Hofmann and Lohr (2010) claimed that seven meta-analyses found higher response rates for CBT compared to alternative treatments and only one found higher rates for the comparison treatment.

On the other hand, a substantial body of research continues to indicate uniform efficacy of treatments intended to be therapeutic (Baardseth et al., 2013; Benish, Imel, & Wampold, 2008; Cuijpers et al., 2013; Imel, Wampold, Miller, & Fleming, 2008; Luborsky et al., 2002; Wampold & Imel, in press; Wampold et al., 1997). For example, in a reanalysis and extension of the findings of Tolin (2010), Baardseth et al. (2013) found no evidence of the superiority of CBT compared to alternative treatments intended to be therapeutic for anxiety disorders.

The debate regarding superiority, at least from a meta-analytic perspective, has focused primarily on outcome measured at one point in time (typically at termination) and has also focused on disorder specific symptom change (Wampold & Imel, in press). Advocates of treatments that are focused on character change rather than on symptoms, such as psychodynamic therapies, suggest that the benefits of such treatments are broader based and longer lasting. For example, Shedler (2010) theorized that:

> The goals of psychodynamic therapy include, but extend beyond, alleviation of acute symptoms. Psychological health is not merely the absence of symptoms; it is the positive presence of inner capacities and resources that allow people to live life with a greater sense of freedom and possibility…. Such intrapsychic changes may account for long-term treatment benefits [of psychodynamic treatments].
>
> [pp 102, 105]

Seeking to produce evidence of the sustained benefits of psychodynamic treatments beyond treatment completion, an increasing number of meta-analyses have indicated that the benefits of psychodynamic treatments at post-treatment are maintained at follow-up, and in some instances increase over time (Abbass, Hancock, Henderson, & Kisely, 2006; Abbass, Kisely, & Kroenke, 2009; Leichsenring, Rabung, & Leibing, 2004; Town et al., 2012). For example, Abbass and colleagues (Abbass et al., 2006; Abbass et al., 2009) conducted a series of meta-analyses of controlled trials of short-term dynamic therapy (STDT) and found that STDT was superior to various types of no-treatment or minimal treatment controls on a variety of outcome measures and that the effects were sustained or grew over time. A number of other meta-analyses have substantiated the enduring effects of psychodynamic treatments (Abbass, Town, & Driessen, 2011; Driessen et al., 2010; Town, Abbass, & Hardy, 2011) and some have claimed that the benefits of psychodynamic treatments increase over time (e.g., Leichsenring et al., 2004; Town et al., 2012).

Based on these findings, Shedler (2010) asserted that, "Consistent trend[s] toward larger effect sizes at follow-up suggest that psychodynamic therapy sets in motion psychological processes that lead to ongoing change, even after therapy has ended…. [Whereas] the benefits of other (nonpsychodynamic) empirically supported therapies tend to decay over time for the most common disorders" (pp. 101,102). Shedler's (2010) assertion that the benefits of psychodynamic treatments are longer lasting than non-psychodynamic treatments does not appear to be universally accepted nor is it conclusively supported by empirical research. Although evidence suggests that the benefits of psychodynamic treatments are sustained over time and in some instances increase compared to control groups, it is unclear whether non-psychodynamic treatments produce equivalent sustained benefits beyond treatment completion.

There are few empirical studies that have addressed the question of whether the effects of some types of treatment, such as psychodynamic treatments, are longer lasting than alternative types of treatments. Meta-analyses of studies that do exist have produced mixed findings. For example, Anderson and Lambert (1995) examined the effectiveness of STDT compared to alternative treatments for a variety of disorders and found that STDT was equivalent to alternative treatments at post-treatment, but produced superior benefits compared to alternative treatments at follow-up. However, Keefe, McCarthy, Dinger, Zilcha-Mano, and Barber (2014) recently meta-analyzed the impact of psychodynamic treatments compared to alternative treatments for anxiety disorders and found that psychodynamic treatments did not significantly differ from alternative treatments at short-term follow-up and long-term follow-up.

These mixed findings may be a result of several methodological limitations. Specifically, many of the previous studies used no-treatment or minimal treatment control groups. Additionally, the majority of meta-analyses and clinical trials comparing two or more treatments did not directly compare treatments intended to be therapeutic. For example, in a meta-analysis examining the effect

of STDT, Anderson and Lambert (1995) failed to identify treatment comparisons that were intended to be therapeutic: "A treatment was classified as 'alternative' only when it was either the usual form of treatment for the disorder or it was expected to produce results similar to STDT." (p. 505). Operating under this definition, non-bona fide comparison treatments were included in the analysis of STDT vs. "alternative" treatments (i.e. hypnosis, dietary advice, supportive treatments, and mutual-help groups), resulting in a bias for STDT. In order to effectively test the superiority of a particular treatment, studies must implement designs that directly compare two or more treatments intended to be therapeutic (see Wampold & Imel, in press; Wampold et al., 1997).

Additionally, the inconsistent findings regarding the long-term superiority of psychodynamic treatments may be related to the historical focus on disorder specific symptom change in psychotherapy research (Wampold & Imel, in press). In a review of psychodynamic effectiveness, Shedler (2010) posited that the benefits of psychodynamic treatments are not limited to the alleviation of symptoms, but rather simultaneously increase clients' inner capacities and resources. If comparative studies solely assess and report symptom-oriented outcomes, they may be failing to capture the lasting benefits of less symptom-oriented therapies.

Lastly, previous meta-analyses have neglected to control for researcher allegiance and treatment dose, potentially contributing to inconsistent findings. Researcher allegiance refers to a researchers preference for a particular treatment and results in better outcomes for the preferred treatment (Munder, Brütsch, Leonhart, Gerger, & Barth, 2013; Wampold & Imel, in press). Similarly, allegiance may impact the effect of the non-preferred alternative treatment as a result of researchers poorly implementing the non-preferred treatment (Munder, Gerger, Trelle, & Barth, 2011). Treatment dose refers to the amount of sessions or treatment "dose" received by a patient. Treatment dose has been found to be significantly and positively related to treatment outcomes, as such it is imperative to control for differences in dosage in comparative analyses. To the best of our knowledge, the majority of meta-analyses examining the effectiveness of psychodynamic treatments fail to control for differences in treatment dose and explore the effect of researcher allegiance on treatment outcome, resulting in potentially biased results.

The inconsistent evidence pertaining to the long-term benefits of psychodynamic treatments suggests that additional meta-analyses addressing previous limitations are needed. As such, the purpose of the current study was to test the growth of the benefits of psychodynamic treatments compared to alternative treatments beyond treatment completion. Specifically, the present meta-analysis strictly included clinical trials that directly compared at least one bona fide psychodynamic treatment to at least one bona fide non-psychodynamic treatment for a variety of disorders. Non-psychodynamic treatments were not further classified into categories of treatments, as we were specifically interested in testing the lasting impact of psychodynamic treatments compared to non-psychodynamic treatments. Multilevel longitudinal analyses were run to perform a significance test of the growth of the impact of psychodynamic treatments compared to alternative treatments beyond treatment completion for four categories of outcome measures (i.e., all outcome measures, targeted outcome measures, non-targeted outcome measures, and personality outcome measures). Informed by the sizeable body of research finding uniform efficacy at post-treatment, we hypothesize that psychodynamic treatments and bona fide non-psychodynamic treatments will not significantly differ at post-treatment on all outcome measures. Additionally, we hypothesize a significant and positive growth (i.e., slope coefficient) in the impact of psychodynamic treatments compared to non-psychodynamic treatments from post-treatment to follow-up on all types of outcome measures. This second hypothesis is based on evidence that the benefits of psychodynamic

psychotherapy at post-treatment increase at follow-up compared to control groups.

2. Methods

2.1. Inclusion criteria

For studies to be included in the current meta-analysis they needed to (a) be published in an English-printed peer-reviewed journal, (b) have utilized randomized clinical designs, (c) have examined treatments of adult patients, (d) have utilized direct comparisons of at least two bona fide therapeutic treatments, one of which was psychodynamic and one which was not, (e) have reported outcome data at post-treatment and at least one follow-up assessment, (f) have reported the necessary statistics to calculate effect sizes, and (g) be published between the years 1972 and 2012.

2.2. Literature search

The following study implemented an exhaustive literature search of several major databases, including: PsychINFO, PsycARTICLES, PsycCRITIQUE, Medline, CINAHL, HealthSource: Nursing/Academic Edition, ERIC, Education Fulltext, SocIndex, Social Work Abstracts, Social Sciences Fulltext, and Academic Search Elite. A team of doctoral students with previous meta-analytic training searched the identified databases for relevant studies by pairing the primary search terms with the secondary search terms. The primary search terms were *psychodynamic*, *dynamic*, *psychoanalytic*, and *psychoanalysis* and the secondary search terms were *psychotherapy*, *therapy*, *controlled trial*, *clinical trial*, *randomized clinical trial*, *RCT*, *comparison study*, *direct comparison*, *effectiveness*, *efficacy*, *outcome*, *treatment*, *follow up*, and *study*. In addition, the reference lists of existing meta-analyses and reviews of psychodynamic effectiveness trials were examined to identify relevant studies. Lastly, psychodynamic researchers were contacted to identify existing databases of psychodynamic trials. The review of databases resulted in 78,772 search results and the review of other sources (i.e., existing meta-analyses and databases of psychodynamic trials) resulted in 136 additional search results, totaling 78,858 (See Fig. 1). Each search result was reviewed for potential inclusion, which resulted in the initial inclusion of 190 studies. A team of trained coders further screened these studies resulting in 66 randomized clinical trials to be evaluated as a bona-fide treatment.

2.3. Bona fide treatment criteria

Treatments within studies were evaluated as bona fide based on the criteria used by Wampold et al. (1997). Specifically, treatments are considered bona fide if they meet the following criteria. First, a trained therapist who holds at least a master's degree or is enrolled in a graduate program in a relevant mental health field delivers the treatment. Second, the treatment is an individualized treatment that is delivered face-to-face. Third, the treatment contains psychological elements based on at least two of the following criteria: (a) the study presents a description and accompanying reference for the treatment, (b) the study contains a treatment citation of an established psychotherapy approach, (c) the treatment is manualized and the manual was utilized in the study, or (d) the active ingredients of the treatment were identified and cited in the study (Wampold et al., 1997).

The team was trained to independently evaluate the treatments based on the bona fide criteria. If the independent raters disagreed on the bona fide status of a treatment, then the raters discussed the disagreement in order to come to a consensus. If the raters did not agree upon a consensus after discussion, the supervising author (BEW) independently evaluated the treatment. The 66 identified studies contained 165 treatments that were subsequently evaluated as
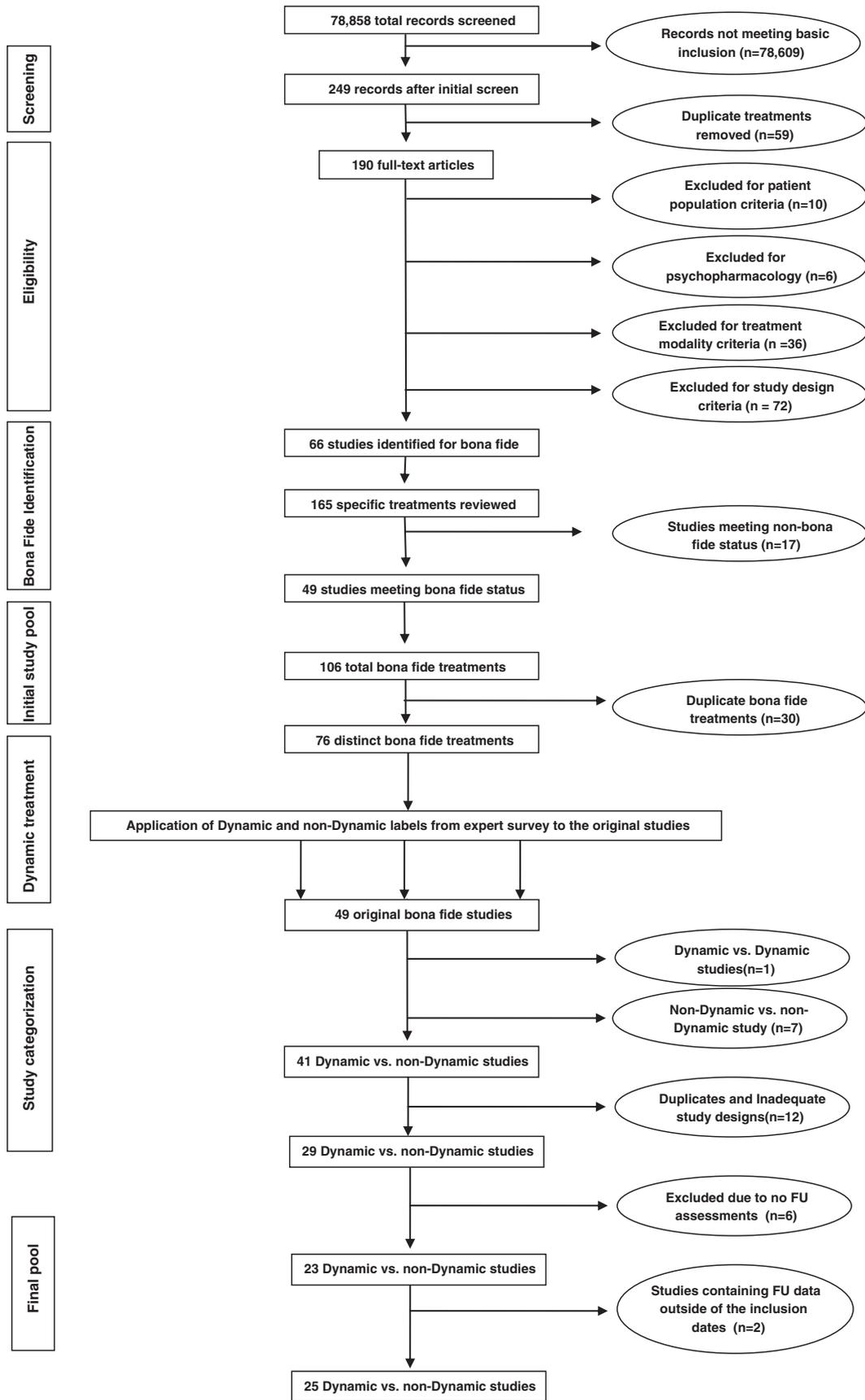
Screening

78,858 total records screened

→ Records not meeting basic inclusion (n=78,609)

249 records after initial screen

→ Duplicate treatments removed (n=59)

190 full-text articles

Eligibility

→ Excluded for patient population criteria (n=10)

→ Excluded for psychopharmacology (n=6)

→ Excluded for treatment modality criteria (n =36)

→ Excluded for study design criteria (n = 72)

Bona Fide Identification

66 studies identified for bona fide

165 specific treatments reviewed

→ Studies meeting non-bona fide status (n=17)

49 studies meeting bona fide status

Initial study pool

106 total bona fide treatments

→ Duplicate bona fide treatments (n=30)

76 distinct bona fide treatments

Dynamic treatment

Application of Dynamic and non-Dynamic labels from expert survey to the original studies

49 original bona fide studies

Study categorization

→ Dynamic vs. Dynamic studies(n=1)

→ Non-Dynamic vs. non-Dynamic study (n=7)

41 Dynamic vs. non-Dynamic studies

→ Duplicates and Inadequate study designs(n=12)

29 Dynamic vs. non-Dynamic studies

→ Excluded due to no FU assessments (n=6)

Final pool

23 Dynamic vs. non-Dynamic studies

→ Studies containing FU data outside of the inclusion dates (n=2)

25 Dynamic vs. non-Dynamic studies

Fig. 1. Flow diagram of study selection and inclusion criteria process.

bona fide. Rater agreement was acceptable (Cohen's kappa = .66). Only studies containing at least two bona fide treatments were included in the present meta-analysis. Of the 66 studies, 17 studies did not contain at least two bona fide treatments, resulting in a total of 49 studies that contained at least two bona fide treatments.

2.4. Psychodynamic and non-psychodynamic categorization

The 49 studies included 106 bona fide treatments to be classified as psychodynamic or non-psychodynamic. Of these, 30 treatments were duplicates, resulting in a total of 76 independent bona fide treatments to be classified as psychodynamic or non-psychodynamic. Shedler (2012) criticized previous psychodynamic research, stating that, "Treatment methods have been inadequately specified" (p. 106). As such, the current meta-analysis followed the psychodynamic categorization method based on ratings by psychodynamic researchers and clinicians, similar to that employed by Baardseth et al. (2013). Psychodynamic raters were identified and surveyed to categorize a treatment as psychodynamic or non-psychodynamic. Psychodynamic raters were identified and recruited through several international psychotherapy societies, including the Society for the Exploration of Psychotherapy Integration (SEPI) and the International Experiential Dynamic Therapy Association (IEDTA). As members of these societies may not exclusively ascribe to psychodynamic therapy, participants were asked to identify their primary theoretical orientation and only those that indicated psychodynamic as a primary theoretical orientation were included as psychodynamic raters. A total sample of 800 mental health practitioners and researchers was generated from the SEPI listserv ($n = 200$) and the IEDTA listserv ($n = 600$).

Survey participants were presented with the name of the treatment, as indicated in the article, as well as the corresponding citation for the treatment (e.g. Dialectical Behavior Therapy; DBT; Linehan, 1993). Participants were then instructed as follows: "Based on your background knowledge and using your own definitions of psychodynamic treatments, please indicate whether you consider the treatments listed below as a psychodynamic treatment." Identification of a treatment as psychodynamic required a majority of affirmative responses (yes).

2.5. Moderators

The following variables were coded and analyzed as moderator variables:

Researcher allegiance. Researcher allegiance is an important moderator to consider as it may effect differential implementation of preferred and non-preferred treatments as well as researcher bias during subjective evaluative processes of research (Munder et al., 2013). The team coded researcher allegiance according to the guidelines developed by Miller, Wampold, and Varhely (2008). The degree of researcher allegiance was determined by assessing study design issues, such as if the author(s) developed or advocated one of the treatments, supervised or trained the therapists for one particular treatment in the study, or if more experienced therapists were utilized for one of the treatments. Allegiance was rated using a five-point scale, where 0 represents no researcher allegiance and 4 represents evidence of strong researcher allegiance. Coding disagreements were resolved by discussion among raters to reach a consensus. The weighted kappa was 0.41, indicating moderate rater agreement for researcher allegiance.

Type of outcome measure. Outcome specificity refers to the extent to which an outcome measure assesses targeted outcomes associated with a specific disorder rather than measures of non-targeted

functioning (Minami, Wampold, Serlin, Kircher, & Brown, 2007). Furthermore, research has found that outcome specificity is related to treatment outcomes (Lambert & Bergin, 1994; Minami et al., 2007; Smith, Glass, & Miller, 1980). In an effort to directly compare equivalent outcome measures, reported outcome measures were coded as targeted outcome measures (e.g., measures of depression in a study on depression), non-targeted outcome measures (e.g., measures of life satisfaction, measures of depression in a study on anxiety), and personality outcome measures (e.g., Millon Clinical Multiaxial Inventory and Personality Disorder Belief Questionnaire). Coding the dependent variable as targeted, non-targeted, or personality outcome measures addresses the issues inherent to outcome specificity, while allowing for more nuanced examinations of the effects of psychodynamic treatments on distinct types of outcome compared to alternative treatments. Coding was conducted as described above with moderate interrater agreement (kappa = .59).

Treatment dose. Treatment dose was calculated as the number of sessions of the psychodynamic treatment minus the number of sessions of the non-psychodynamic treatment. As such zero represented equivalent number of sessions across treatments within a study. Treatment dose was entered into the multilevel models in order to control for differences in treatment length while testing the impact of the psychodynamic treatments compared to non-psychodynamic treatments.

2.6. Statistical analyses

In the current meta-analysis, the team obtained means, standard deviations, and sample sizes at every assessment point for each included study to calculate effect sizes. Effect sizes were computed as Hedges' *g*, representing the mean difference between psychodynamic treatments and non-psychodynamic treatments, where a positive *g* indicates the superiority of psychodynamic treatments. Hedges' *g* corrects for a small bias in Cohen's *d* (Borenstein, Hedges, Higgins, & Rothstein, 2009). When studies contained more than two psychodynamic or non-psychodynamic treatments compared to the alternative, we aggregated the two treatments of the same type and compared to the alternative (e.g., psychodynamic-1 & psychodynamic-2 vs. non-psychodynamic). This aggregation process was similarly applied to studies that compared one psychodynamic treatment with two non-psychodynamic treatments.

Four separate within-study effect sizes were calculated at each assessment time according to outcome type: (a) combined outcome measures, included all reported outcomes within a study, (b) targeted outcome measures, included measures that focused on the primary diagnosis of a particular study, (c) non-targeted outcome measures, included non-targeted measures of distress or well-being, and (d) personality outcome measures, included measures of personality. When a study reported multiple outcome measures, effect sizes of dependent measures were aggregated within each study according to the type of outcome measure (i.e., targeted, non-targeted, personality) and calculated under the assumption that correlations of dependent effects are .50 (see Wampold et al., 1997 for rationale). This within-study aggregation process resulted in one effect size per type of outcome at each assessment. Four separate meta-analyses were conducted according to type of outcome: combined outcome measures, targeted outcome measures, non-targeted outcome measures, and personality outcome measures.

Following the aggregation of within-study effect sizes, a multilevel longitudinal meta-analysis was conducted to account for the dependency of multiple assessment times within studies, where assessment times

were nested within studies. For these analyses, hierarchical linear modeling (HLM; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) was used to estimate a restricted maximum-likelihood random-effect model with known variances (Hedges & Olkin, 1985). We conducted a two-level longitudinal analysis with time (weeks post treatment completion) as a predictor variable of the effect size to examine the growth of the impact of psychodynamic treatments versus non-psychodynamic treatments from post-treatment to follow-up while controlling for treatment dose. Time was centered at week 0 (post-treatment). The mixed model was:

$$g(\text{outcomes}) = \gamma_{00} + \gamma_{01} * (\text{dose}) + \gamma_{10} * (\text{time}) + \gamma_{11} * (\text{dose}) \\ * (\text{time}) + u_0 + u_1 * (\text{time})$$

where γ_{00} represented the intercept, or the effect size of psychodynamic treatments versus non-psychodynamic treatments at post-treatment, γ_{01} represented the impact of treatment dose on the effect size at post-treatment, γ_{10} represented the slope or growth in the effect size of psychodynamic treatments versus non-psychodynamic treatments from post-treatment in weeks, γ_{11} represented the impact of treatment dose on the slope, u_0 represented the error term at level 1, and u_1 represented the error term at level 2 (random effect model). This model was run for each outcome type. The Q-statistic (Q) and I^2 were calculated to assess the significance and amount of between-study heterogeneity (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). Lastly, to examine potential factors impacting the effect sizes of psychodynamic versus non-psychodynamic treatments, researcher allegiance was included in the mixed model to assess for moderation.

3. Results

3.1. Psychodynamic survey

Of the total of 800 IEDTA ($n = 600$) and SEPI ($n = 200$) members who were contacted via email, approximately ten percent (9.63%, $n = 77$) responded. Of the 77 respondents, 75% ($n = 58$) identified their primary theoretical orientation as psychodynamic and therefore were included in the coding of psychodynamic treatments. The following data is reported for the 58 psychodynamic raters. The majority of survey participants (48%, $n = 28$) indicated Ph.D. as the highest level of degree completed, followed by Master's degree (22%, $n = 13$), Psy.D. (16%, $n = 9$), M.D. (7%, $n = 4$), and other (7%, $n = 4$). Participants were also asked to indicate their primary setting(s) for clinical practice. The majority of participants (72%, $n = 42$) indicated private practice as their primary setting, followed by college/university department (12%, $n = 7$), university counseling center (9%, $n = 5$), and other (7%, $n = 4$).

Of the 76 independent bona fide treatments from the 49 included studies, survey participants identified 32 treatments as psychodynamic and 44 treatments as non-psychodynamic (based on the majority decision rule described above). Following psychodynamic status coding, 7 of the 49 studies were excluded due to a direct comparison of two non-psychodynamic bona fide treatments. Another study was excluded due to direct comparison of a bona fide psychodynamic treatment with another bona fide psychodynamic treatment. Ultimately, 41 of the 49 studies met inclusion criteria of a study that directly compared a bona fide psychodynamic treatment with a bona fide non-psychodynamic treatment.

3.2. Meta-analysis

Following the identification of psychodynamic and non-psychodynamic treatments, means, standard deviations, and sample sizes were extracted from the 41 identified studies for analysis. This final review resulted in the exclusion of four studies due to an inadequate study design for the purpose of the current meta-analysis,

eight duplicate studies (whose data was represented in other included studies), and six studies that failed to report follow-up data. This process resulted in a sample of 23 comparative studies of a bona fide psychodynamic treatment and a bona fide non-psychodynamic treatment. Lastly, several identified studies did not report the needed data for analysis and principle investigators were contacted to request the needed data. This process led to the inclusion of two additional studies that fell outside of the original search dates but reported follow-up data for two of the 23 original studies, resulting in a sample of 25 studies. As multiple studies were often published on a single sample (i.e., subsequent publications of follow-up data, subsequent publications of secondary outcome measures, etc.), the present meta-analysis included 25 comparative studies of 20 independent samples. Of the 20 independent samples, three studies compared three treatments; two studies compared two psychodynamic treatments and one non-psychodynamic treatment and one study compared one psychodynamic treatment and two non-psychodynamic treatments, the remaining 17 studies compared one psychodynamic treatment and one non-psychodynamic treatment. A total of 1690 patients participated in the included 25 clinical trials resulting in a per-study mean of 84.5 and median of 53.5 patients. The 20 independent samples (25 studies) are outlined in Table 1. References for the included studies are provided in Appendix A.

3.3. Impact of psychodynamic treatments versus non-psychodynamic treatments at post-treatment

The results of the longitudinal multilevel analysis of the impact of psychodynamic versus non-psychodynamic treatments at post-treatment and beyond are reported in Table 2. At post-treatment no significant differences were found between psychodynamic treatments and non-psychodynamic treatments for combined outcome measures ($g = -0.05$; 95% CI [$-0.18, 0.07$]; $p = 0.428$; $k = 20$), targeted outcome measures ($g = -0.10$; 95% CI [$-0.23, 0.02$]; $p = 0.134$; $k = 19$), non-targeted outcome measures ($g = 0.10$; 95% CI [$-0.10, 0.29$]; $p = 0.346$; $k = 17$), and personality outcome measures ($g = -0.10$; 95% CI [$-0.52, 0.33$]; $p = 0.685$; $k = 6$) after controlling for treatment length. The analyses of between-study heterogeneity indicated that the effect sizes were not homogeneously distributed. Specifically, between-study heterogeneity was significant at post-treatment for non-targeted outcome measures ($Q = 45.31$; $p < .001$; $I^2 = 66.89\%$) and personality outcome measures ($Q = 14.38$; $p = .006$; $I^2 = 72.18\%$), but not for targeted outcome measures ($Q = 20.58$; $p = .245$; $I^2 = 17.39\%$) or combined outcome measures ($Q = 21.25$; $p = .266$; $I^2 = 15.29\%$).

Moderator analyses were run to examine the impact of researcher allegiance on treatment differences at post-treatment. The results of the moderator analyses are presented in Table 3. Allegiance was not a significant moderator for combined outcome measures ($\gamma_{02} = 0.11$; $p = .224$; $k = 20$), targeted outcome measures ($\gamma_{02} = 0.11$; $p = .209$; $k = 19$), non-targeted outcome measures ($\gamma_{02} = 0.07$; $p = .427$; $k = 17$) or personality outcome measures ($\gamma_{01} = 0.08$; $p < .850$; $k = 6$) at post-treatment. Of note, the systematic between-study variability for non-targeted and personality outcome measures (indicated by the significant Q-statistic for these outcomes) was not explained by allegiance, suggesting the presence of undetermined factors related to the heterogeneity among effects.

3.4. Sustained impact of psychodynamic treatments versus non-psychodynamic treatments beyond treatment completion

Results from the longitudinal analyses are reported in Table 2. Results indicated no significant differences in the growth (i.e., slope) of the effect size of psychodynamic treatments versus non-psychodynamic treatments as a function of weeks after

termination for combined outcome measures ($\gamma_{10} = 0.001$; 95% CI [-0.002, 0.004]; $p = .441$; $k = 20$), targeted outcome measures ($\gamma_{10} = 0.001$; 95% CI [-0.002, 0.004]; $p = .523$; $k = 19$), non-targeted outcome measures ($\gamma_{10} = 0.001$; 95% CI [-0.002, 0.004]; $p = .557$; $k = 17$), and personality outcome measures ($\gamma_{10} = -0.001$; 95% CI [-0.007, 0.005]; $p = .722$; $k = 6$) after controlling for treatment length. The analyses of between-study heterogeneity indicated that the individual slopes for combined measures, targeted measures, non-targeted measures, and personality measures did not significantly vary between studies (all Q-statistic $ps > .05$, see Table 2). Moderator analyses were run to examine the impact of researcher allegiance on the growth (i.e., slope) of effects in weeks post-treatment. The results of the moderator analyses are presented in Table 3. Researcher allegiance was a significant moderator of the slope of the impact of psychodynamic treatments compared to non-psychodynamic treatments for targeted outcome measures ($\gamma_{12} = -0.01$; $p = .042$; $k = 19$).. As psychodynamic allegiance increased by one unit, the growth in the effect size of psychodynamic treatments decreased by 0.01 for targeted outcome measures. This finding is surprising as the relationship is in the unexpected direction. Researcher allegiance was not a significant moderator (Q-statistic $ps > .10$, see Table 3) of the slopes for combined outcome measures, non-targeted outcome measures, and personality outcome measures.[1]

4. Discussion

The present meta-analysis examined the impact of bona fide psychodynamic treatments compared to bona fide non-psychodynamic treatments both at post-treatment and, most interestingly, beyond post-treatment. In particular, the growth of the impact of bona fide psychodynamic treatments compared to bona fide non-psychodynamic treatments was examined beyond the end of therapy. As hypothesized, psychodynamic and non-psychodynamic treatments were equally efficacious at post-treatment on combined outcome measures, targeted measures, non-targeted measures, and personality measures. This finding is consistent with the findings of a recent meta-analysis (Keefe et al., 2014) that found no significant differences at post-treatment between psychodynamic and alternative treatments. Specifically, Keefe et al. (2014) examined the effect of psychodynamic treatments compared to alternative treatments for anxiety disorders and found a small and non-significant effect, indicating no differences at post-treatment. Additionally, the present findings replicate several previous meta-analyses that found uniform efficacy between psychodynamic treatments and alternative treatments (Abbass et al., 2011; Anderson & Lambert, 1995; Crits-Christoph, 1992; Leichsenring, 2001; Leichsenring & Leibing, 2003; Leichsenring et al., 2004). Lastly, the finding of non-significant treatment differences at post-treatment is consistent with and expands the growing body of research that suggests uniform efficacy among psychotherapies intended to be therapeutic (Benish et al., 2008; Imel et al., 2008; Luborsky et al., 2002; Wampold & Imel, in press).

Although these analyses revealed no significant differences between psychodynamic and non-psychodynamic treatments at post-treatment, between-study heterogeneity among the effect sizes of non-targeted

and personality outcome measures was significantly greater than that expected due to sampling error alone. This degree of heterogeneity suggests that, at least theoretically, important between study variations may exist, which are due to unknown sources. Contrary to other meta-analyses (see Munder et al., 2013), researcher allegiance did not significantly account for the between-study variability in the effect sizes of non-targeted outcome measures and personality outcome measures nor was it a significant moderator of the post-treatment effect for combined outcome measures or targeted outcome measures. Therefore these moderator analyses did not identify variables that could explain the systematic variability that remained, again suggesting that important unexamined between-study differences may explain the heterogeneity among effects for both non-targeted outcome measures and personality outcome measures.

Contrary to our hypothesis, the post-treatment slope of the effect size of psychodynamic treatments versus non-psychodynamic treatments was not significant for any outcome measure types (combined outcome measures, targeted measures, non-targeted measures, and personality measures). Although previous findings suggested that the benefits of psychodynamic treatments are sustained beyond treatment completion compared to control groups (Abbass et al., 2006; Abbass et al., 2009; Leichsenring et al., 2004; Town et al., 2012), the current findings suggest that this phenomenon may not be unique to psychodynamic treatments. In other words, when bona fide psychodynamic treatments are directly compared to bona fide non-psychodynamic treatments, as identified by independent psychodynamic raters, it appears that the growth in treatment effects beyond treatment completion is equivalent. The current meta-analysis advances previous empirical efforts to examine the sustained benefits of psychodynamic treatments in several meaningful ways. Specifically, the present meta-analysis employed an exhaustive literature search, strictly included studies that directly compared at least two bona fide treatments, and used independent psychodynamic raters to identify and categorize treatments as psychodynamic. After addressing the previous limitations of meta-analyses examining the long-term benefit of psychodynamic treatments in comparison to alternative treatments, the evidence produced to date suggests that the enduring benefits of psychodynamic treatments are equivalent to non-psychodynamic treatments. Of note, the findings of the present meta-analysis are consistent with a recent meta-analysis examining the long-term impact of psychodynamic treatments (see Keefe et al., 2014).

In contrast to the analysis of between-study heterogeneity at post-treatment, between-study heterogeneity of the slopes was not significant for any type of outcome (i.e., combined outcome measures, targeted outcome measures, non-targeted outcome measures, and personality outcome measures). In addition, researcher allegiance only significantly moderated the slope for personality outcome measures. This lack of slope heterogeneity and allegiance moderation supports the strength of the primary finding that the post-treatment trajectory of change did not differ between psychodynamic and non-psychodynamic therapies.

4.1. Strengths and limitations

The current meta-analysis addressed several limitations of previous studies examining the relative efficacy of psychodynamic treatments; the inclusion of studies that directly compare two or more treatments, the identification and inclusion of only bona-fide treatments, and the use of psychodynamic raters to identify psychodynamic treatments are noteworthy strengths of the current study. Additionally, the number of studies that reported both post-treatment and follow-up assessments included in the analysis of combined outcome measures ($k = 20$) in comparison to the relatively smaller sample sizes of previous meta-analyses of psychodynamic treatments, such as Keefe et al. (2014; $k = 14$), Abbass et al. (2011; $k = 8$), Crits-Christoph (1992; $k = 11$), and Leichsenring (2001; $k = 6$), is a strength of the current meta-

---

[1] Knekt et al. (2008; 2010; 2013) reported nine FU assessments compared to other studies included in the meta-analysis that typically reported one or two FU assessments. Consequently this study may have greatly influenced the longitudinal results. To examine this possibility, we removed Knekt et al. (2008; 2010; 2013) and reran the longitudinal analyses. The analysis of the growth of the effect of psychodynamic treatments versus non-psychodynamic treatments excluding Knekt et al. (2008; 2010; 2014) indicated no significant differences on combined outcome measures ($g = 0.001$; 95% CI [-0.002, 0.004]; $p = .469$; $k = 19$), targeted outcome measures ($g = 0.001$; 95% CI [-0.002, 0.004]; $p = .556$; $k = 18$), and non-targeted outcome measures ($g = 0.001$; 95% CI [-0.002, 0.004]; $p = .679$; $k = 16$). Knekt et al. (2008; 2010; 2014) did not report any measures of personality, therefore the original analyses of personality measures were not reanalyzed.

Table 1
Included studies in the current meta-analysis.

Study	Disorder	Psychodynamic treatment	Non-psychodynamic treatment	Follow-up assessments (months)	Outcome measures
Bachar, Latzer, Kreitler, and Berry (1999)	Anorexia and bulimia	52 sessions SPT (n = 14 at T, 8 at FU)	52 sessions COT (n = 12 at T, 5 at FU)	12	DSM-SS, EAT 26, GSI, Selves Questionnaire
Barkham et al. (1996)	Depression	12 sessions PI (n = 18)	12 sessions CBT (n = 18)	3, 12	BDI, SCL-90-R, IIP, Self-Esteem Scale
Barkham, Shapiro, Hardy, and Rees (1999)	Depression	3 sessions PI (n = 54)	3 sessions CBT (n = 62)	12	Aggregated ES
Brom, Kleber, and Defares (1989)	PTSD	Ave. 18.8 sessions PT (n = 21)	Ave. 15 sessions TD (n = 23)	3	SCL-90, STAI, State-Trait Anger Inventory, Impact of Event Scale, Dutch Personality Questionnaire, Amsterdam Biographical Questionnaire, Locus of Control
Driessen et al. (2007); Driessen et al. (2013)	Depression	16 sessions SPSP (n = 177)	16 sessions CBT (n = 164)	12	HDRS-17, IDS-SR, BSI Interpersonal Sensitivity, OQ Interpersonal Relationships, OQ Social Role
Emmelkamp et al. (2006)	Avoidant personality disorder	Ave. 18.8 sessions BDT (n = 22)	Ave. 18.5 sessions CBT (n = 18)	6	LWASQ, PDBQ, SPAI, Avoidance Scale
Gallagher-Thompson and Steffen (1994)	Depression	20 sessions BPT (n = 21 at T, 20 at FU)	20 sessions CBT (n = 31 at T, 28 at FU)	3	HAM-D, BDI, GDS, Diagnostic Status
Giesen-Bloo et al. (2006)	BPD	312 sessions TF (n = 42)	312 sessions SF (n = 44)	12	BPDSI-IV, EuroQol Thermometer, WHOQOL, Psycho- and Personality Factor Score
Hardy et al. (1995)	Depression	Ave. 12 sessions PI (n = 56)	Ave. 12 sessions CBT (n = 56)	3, 12	BDI, SCL-90-R, self-esteem, IIP, PSE
Hellerstein et al. (1998)	Cluster C personality disorders	Ave. 28.5 sessions STDP (n = 14)	Ave. 31 sessions BSP (n = 12 at T, 10 at FU)	6	PTC, SCL-90-R, IIP
Knekt et al. (2008); Knekt, Laaksonen, Raitasalo, Haaramo, and Lindfors (2010); Knekt, Lindfors, Sares-Jaske, Virtala, and Harkanen (2013)	Mood and anxiety disorders	Ave. 232 sessions LTPP (n = 128); Ave. 18.5 sessions STPP (n = 101)	Ave. 9.8 sessions SFT (n = 97)	3, 7, 9, 12, 18, 24, 36, 48, 60	BDI, HAM-D, SCL-90-Anx, HAM-A, SCL-90-GSI, WAI, SAS-Work, PPF, number of sick leave days, alcohol consumption, BMI, total serum cholesterol, serum HDL cholesterol
Leichsenring et al. (2009); Salzer, Winkelbach, Leweke, Leibing, and Leichsenring (2011)	GAD	Ave. 29.1 sessions STPP (n = 28)	Ave. 28.8 sessions CBT (n = 29)	6, 12	HAM-A, PSWQ, STAI, BAI, Hospital Anxiety and Depression Scale, BDI, IIP

Maina, Forner, and Bogetto (2005)	MDD	Ave. 19.6 sessions BDT (n = 10)	Ave. 18.6 sessions BSP (n = 10)	6	HAM-D, HAM-A, CGI-S
Maina, Rosso, Crespi, and Bogetto (2007)	MDD	Ave. 15.5 sessions BDT (n = 16)	Ave. 15.35 sessions BSP (n = 16)	6	HAM-D, HAM-A, CGI-S, CGI-I
Muran, Safran, Samstag, and Winston (2005)	PD	30 sessions STPT (n = 22); 30 sessions BRT (n = 33)	30 sessions CBT (n = 29)	6	Client reported target complaints, SCL-90-R, IIP, WISPI, therapist reported target complaints, GAS
Pierloot and Vinck (1978)	Anxiety disorders	Ave. 19.67 sessions STDP (n = 9)	Ave. 19.85 sessions SD (n = 13)	3	PSS, TMAS, STAI, symptom checklist, Subjective Appraisal Scale
Svartberg, Stiles, and Seltzer (2004)	Custer C personality disorders	40 sessions STDP (n = 25)	40 sessions CT (n = 25)	6, 12, 24	SCL-90-R, IIP, Millon Clinical Multiaxial Inventory
Thompson, Gallagher, and Breckenridge (1987); Gallagher-Thompson, Hanley-Peterson, and Thompson (1990)	MDD	Ave. 18 sessions BPT (n = 30)	Ave. 18 sessions CT (n = 31); Ave. 18 sessions BT (n = 30)	12, 24	BDI, HAM-D, GDS, BSI, GAS, BPRS, Hopelessness Scale, ATQ, YLI, Social Adjustment Scale, Health and Daily Living Questionnaire, Older Person's Pleasant Events Schedule, diagnostic status
Winston et al. (1994)	PD	Ave. 40.3 sessions STDP (n = 25 at T, 19 at FU)	Ave. 40.3 Sessions BA (n = 30 at T, 19 at FU)	18	Target complaint rating, SCL-90-R, Social Adjustment Scale
Woody, McLellan, Luborsky, and O'Brien (1987)	Opiate dependence	24 sessions SE (n = 25)	24 sessions CBT (n = 31)	6	BDI, Maudsley Personality Inventory, SCL-90, Shipley Institute of Living Scale, SADS-L, SADS-C, Addiction Severity Index Interview

Note. ATQ = Automatic Thoughts Questionnaire, BA = Brief Adaptive Therapy, BAI = Beck Anxiety Inventory, BDI = Beck Depression Inventory, BDT = Brief Dynamic Therapy, BMI = Body Mass Index, BPD = Borderline Personality Disorder, BPDSI-IV = Borderline Personality Disorder Severity Index, BPRS = Brief Psychiatric Rating Scale, BPT = Brief Psychodynamic Therapy, BRT = Brief Relational Therapy, BSI = Brief Symptom Inventory, BSP = Brief Supportive Psychotherapy, BT = Behavioral Therapy, CBT = cognitive-behavioral therapy, CGI-I = Clinical Global Impression — Improvement, CGI-S = Clinical Global Impression for Severity, COT = Cognitive Orientation Treatment, CT = Cognitive Therapy, DSM-SS = Symptomatology Scale for Anorexia and Bulimia, EAT-26 = Eating Attitudes Test, ES = Effect Size, GAD = Generalized Anxiety Disorder, GAS = Global Assessment Scale, GDS = Geriatric Depression Scale, GSI = Global Severity Index, HAM-A = Hamilton Rating Scale for Anxiety, HAM-D = Hamilton Rating Scale for Depression, HDRS-17 = Hamilton Depression Rating Scale-17, IDS-SR = Inventory of Depressive Symptomatology-Self Report, IIP = Inventory of Interpersonal Problems, LTPP = Long-Term Psychodynamic Psychotherapy, LWASQ = Lehrer Woolfolk Anxiety Symptoms Questionnaire, MDD = Major Depressive Disorder, OQ = Outcome Questionnaire, PD = Personality Disorder, PDBQ = Personality Disorder Belief Questionnaire, PI = Psychodynamic–Interpersonal, PPF = Perceived Psychological Functioning Scale, PSE = Present State Examination, PSS = Psychiatric Status Schedule, PSWQ = Penn State Worry Questionnaire, PT = psychodynamic therapy, PTC = Patient-rated Target Complaints, PTSD = Post-Traumatic Stress Disorder, SADS-C = Schedule for Affective Disorders − Change version, SADS-L = Schedule for Affective Disorders − Life-time version, SAS-Work = Social Adjustment Scale-Work Subscale, SCL-90 = Symptom Checklist-90, SCL-90-Anx = Symptom Checklist-90 Anxiety Scale, SCL-90-GSI = Symptom Checklist-90 Global Severity Index, SCL-90-R = Symptom Checklist-90 Revised, SD = Systematic Desensitization, SE = Supportive Expressive Therapy, SF = Schema-Focused Therapy, SFT = Solution Focused Therapy, SPAI = Social Phobia Anxiety Inventory, SPSP = Short Psychodynamic Supportive Psychotherapy, SPT = Self Psychological Therapy, STAI = State-Trait Anxiety Inventory, STDP = Short-Term Dynamic Psychotherapy, STPP = Short-Term Psychodynamic Psychotherapy, STPT = Short-Term Psychodynamic Therapy, TD = Trauma Desensitization, TMAS = Taylor Manifest Anxiety Scale, TF = Transference-focused Therapy, WAI = Work Ability Index, WHOQOL = World Health Organization Quality of Life Assessment, WISPI = Wisconsin Personality Inventory, YLI = Young Loneliness Inventory.

Table 2

Longitudinal model of the effect size of psychodynamic treatments versus non-psychodynamic treatments controlling for treatment length.

	k	Coefficient	95% CI	Q	p (Q)	I^2
Combined outcomes	20					
Intercept γ_{00}		−0.05	(−0.18, 0.07)	21.25	.266	15.29%
Tx dose γ_{01}		0.01	(−0.01, 0.02)			
Slope γ_{10}		0.001	(−0.002, 0.004)	26.84	.082	32.94%
Tx dose γ_{11}		−0.00	(−0.0003, 0.0002)			
Targeted outcomes	19					
Intercept γ_{00}		−0.10	(−0.23, 0.02)	20.58	.245	17.39%
Tx dose γ_{01}		0.01	(−0.005, 0.02)			
Slope γ_{10}		0.001	(−0.002, 0.004)	21.96	.186	22.59%
Tx dose γ_{11}		−0.00	(−0.0002, 0.000)			
Non-targeted outcomes	17					
Intercept γ_{00}		0.10	(−0.10, 0.29)	45.31[***]	<.001	66.89%
Tx dose γ_{01}		0.002	(−0.02, 0.02)			
Slope γ		0.001	(−0.002, 0.004)	24.39	.058	38.50%
Tx dose γ_{11}		−0.00	(−0.0002, 0.0001)			
Personality outcomes	6					
Intercept γ_{00}		−0.10	(−0.52, 0.33)	14.38[**]	.006	72.18%
Tx dose γ_{01}		0.07	(−0.20, 0.34)			
Slope γ_{10}		−0.001	(−0.007, 0.005)	0.59	>.50	0.0%
Tx dose γ_{11}		0.001	(−0.012, 0.014)			

Note. Intercept γ_{00} = post-treatment effect size (g) of psychodynamic treatments compared to non-psychodynamic treatments, slope γ_{10} = growth in the effect size (g) of psychodynamic treatments compared to non-psychodynamic treatments in weeks beyond post-treatment.

[**] $p < .01$.
[***] $p < .001$.

analysis and allows for a more precise estimate of the enduring impact of psychodynamic treatments compared to non-psychodynamic treatments. Additionally, whereas previous research examining the efficacy of psychodynamic treatments failed to analyze outcome measure types separately (Abbass et al., 2011; Driessen et al., 2010; Keefe et al., 2014; Svartberg & Stiles, 1991), the present study analyzed three distinct types of outcome in addition to the overall omnibus effect size (i.e., combined outcomes), providing a more nuanced examination of treatment effects for different domains of change.

Lastly, the present study is one of few meta-analyses to utilize a multilevel longitudinal model to examine comparative treatment effects in psychotherapy research. A contribution of multilevel longitudinal meta-analyses is the ability to examine repeated assessments, while correctly accounting for the nested nature and dependency of the data. Moreover, longitudinal designs allow for the calculation of the growth (slope) of the repeated assessments, rather than performing cross-sectional analyses at arbitrarily derived follow-up periods, ultimately providing a more accurate picture of the change over time as well as a statistical test of this trajectory of change. While a few noteworthy multilevel longitudinal meta-analyses have been conducted (Flückiger, Del Re, Munder, Heer, & Wampold, 2014; Flückiger, Del Re, & Wampold, 2015; Flückiger, Del Re, Wampold, Symonds, & Horvath, 2012); this is the first multilevel longitudinal meta-analysis to examine the growth in effects sizes of two bona fide psychotherapies (i.e., psychodynamic and non-psychodynamic) beyond treatment completion. While Keefe et al. (2014) calculated and reported effect sizes at pre-determined ranges of assessment times (i.e., post-treatment, follow-up to one year, and follow-up past one year), this method only provides information of the effect size at these specific assessments and does not allow for a statistical test of the size of the effect over time.

There are several limitations that need to be considered in interpreting the findings of the present study. First, there were a small number of studies that reported follow-up data for the growth model of personality outcome measures ($k = 6$). In other words, the finding of no significant differences in the slope of psychodynamic treatments relative to non-psychodynamic treatments for personality outcome measures may be an accurate estimation of the population or it may be a result of the limited number of studies reporting follow-up assessments, limiting our ability to confidently interpret the findings of this

analysis in particular. Related, it is important to note that the number of studies varied across outcome categories at post-treatment and follow-up due to studies failing to collect data for a given outcome category, similarly resulting in reduced sample sizes for the analyses of targeted, non-targeted, and personality outcome measures.

Second, it was unknown if patients received booster sessions or additional mental health services post-treatment due to the lack of reporting in the identified studies. Third, inconsistent reporting in the included studies contributed to a lack of reported data needed to calculate effect sizes. In these cases the authors were left to calculate effect sizes using non-conventional methods and transformations of reported data in order to represent the existing body of research in its totality — methods which potentially introduce more noise into the data. Fourth, the current meta-analysis only included published studies in peer-reviewed journals, which may increase the effects of publication bias. Publication bias can result in an inflated estimate of the true effect, as studies with significant findings are often published at higher rates than studies with non-significant findings (Lipsey & Wilson, 1993). However, unpublished studies are more likely to have found no treatment differences in comparison to published studies (Rotton, Foos, Van Meek, & Levitt, 1995), therefore it is unlikely that the inclusion of unpublished studies would change the findings of no significant differences between treatments. Fifth, the agreement between coders on the classification of outcome measures into four distinct categories was moderate indicating that outcome measures were not easily discernable by trained coders as targeted, non-targeted, or personality outcomes.

Sixth, researcher allegiance was coded according to the guidelines developed by Miller et al. (2008), which focus on a researcher's positive allegiance but may fail to assess negative allegiance or poor implementation of a non-preferred treatment, potentially ignoring additional aspects of researcher allegiance. While the use of a published and widely used method to code researcher allegiance is a strength of the current meta-analysis, it is important to not the inherent limitations of this particular method. Seventh, while using independent psychodynamic raters to identify psychodynamic treatments was a strength of the present study, and for the majority of treatments the raters were able to make strong and discernable decisions as to psychodynamic or non-psychodynamic, several treatments received a weak majority decision. This may raise the question of whether the treatments

Table 3
Multilevel model of the impact of researcher allegiance on the effect size of psychodynamic treatments versus non-psychodynamic treatments.

	k	Coefficient	95% CI	Q	p (Q)	I^2
Combined outcomes	20					
Intercept γ_{00}		−0.05	(−0.18, 0.08)	18.36	.366	7.43%
Tx dose γ_{01}		0.01	(−0.006, 0.02)			
Allegiance γ_{02}		0.11	(−0.06, 0.27)			
Slope γ_{10}		0.003	(−0.001, 0.007)	29.55	.03*	42.46%
Tx dose γ_{11}		−0.00	(−0.0004, 0.0003)			
Allegiance γ_{12}		−0.003	(−0.01, 0.004)			
Targeted outcomes	19					
Intercept γ_{00}		−0.10	(−0.23, 0.03)	16.25	.436	1.52%
Tx dose γ_{01}		0.01	(−0.004, 0.02)			
Allegiance γ_{02}		0.11	(−0.05, 0.26)			
Slope γ_{10}		0.002	(−0.001, 0.005)	17.03	.383	6.07%
Tx dose γ_{11}		−0.00	(−0.0002, 0.0000)			
Allegiance γ_{12}		−0.01*	(−0.01, −0.0006)			
Non-targeted outcomes	17					
Intercept γ_{00}		0.10	(−0.12, 0.31)	46.15	<.001***	69.66%
Tx dose γ_{01}		0.003	(−0.02, 0.03)			
Allegiance γ_{02}		0.07	(−0.26, 0.41)			
Slope γ_{10}		0.002	(−0.002, 0.005)	23.50	.052	40.41%
Tx dose γ_{11}		−0.00	(−0.0002, 0.0001)			
Allegiance γ_{12}		−0.002	(−0.007, 0.003)			
Personality outcomes	6					
Intercept γ_{00}		−0.04	(−0.66, 0.58)	15.16	.002**	80.21%
Tx dose γ_{01}		0.08	(−0.25, 0.40)			
Allegiance γ_{02}		0.08	(−0.68, 0.84)			
Slope γ_{10}		−0.001	(−0.007, 0.005)	0.39	<.50	0.0%
Tx dose γ_{11}		0.001	(−0.013, 0.015)			
Allegiance γ_{12}		0.003	(−0.011, 0.018)			

Note. Intercept γ_{00} = post-treatment effect size (g) of psychodynamic treatments compared to non-psychodynamic treatments, slope γ_{10} = growth in the effect size (g) of psychodynamic treatments compared to non-psychodynamic treatments in weeks beyond post-treatment.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

included or excluded based on independent ratings were indeed correctly categorized as the raters may have had limited familiarity or knowledge of the treatments. This may also highlight the potentially fuzzy line between theoretical orientations and the psychotherapies based upon them (Baardseth et al., 2013).

A related limitation was the heavy representation of cognitive-behavioral treatments within the non-psychodynamic comparison conditions. While there was considerable variability of treatments categorized as non-psychodynamic (e.g., Dialectical Behavior Therapy, Solution-focused Therapy, Systematic Desensitization), the majority of the non-psychodynamic treatments represented a form of cognitive-behavioral treatment ($k = 14$). As such, it may be that the results of the present meta-analysis are more generalizable to the comparison of psychodynamic treatments and bona fide cognitive-behavioral treatments, rather than to any bona fide alternative treatment that may be only minimally represented in the current analysis. Eighth, the current meta-analysis did not successfully explain the significant heterogeneity of non-targeted and personality effects at post-treatment, suggesting the presence of undetermined factors related to the heterogeneity among effects. Additionally, the unexplained significant between-study heterogeneity suggests that the studies included in the analyses of non-targeted and personality effects at post-treatment uniquely differ from one another.

Finally, it may well be that the sustained benefits of psychodynamic therapies necessitate longer treatment periods than that typically provided in short-term non-psychodynamic treatments (which may emphasize cognitive or behavioral skills that can be learned more quickly). It is common practice for comparative clinical trials to match the length of competing treatments in order to control for differences in dosage. However, this methodological practice may not allow psychodynamic treatments sufficient time to produce change in the way these therapies were designed, thus putting psychodynamic treatments at a disadvantage. Similarly, the limited number of studies included in

the current meta-analysis that conducted distal assessments of change (e.g. 2 + year follow-up) may limit the ability to observe unique continued change associated with psychodynamic treatments.

4.2. Implications for practice and future research

The evidence produced by this meta-analysis did not corroborate a number of conjectures in the literature. First, there was no evidence that psychodynamic treatments are more enduring than alternative treatments or that they result in benefits beyond symptom change (e.g., in personality change) in comparison to alternative treatments. Psychodynamic treatments were as effective as other treatments, many of which were variants of CBT, on all outcome domains at the end of treatment and beyond. That is, albeit possibly for different reasons, psychodynamic treatments are as enduring as other treatments, including treatments that focus on the acquisition of skills to manage or overcome symptomatic distress. Consequently, it appears that the plethora of CBT treatments in lists of evidence-based treatments or psychological treatments with research support may be due to the preponderance of well-designed research on these treatments rather on their inherent clinical superiority.

It is worth highlighting briefly that the lack of differential effects between psychodynamic and non-psychodynamic treatments does not in any way imply that either psychodynamic or non-psychodynamic treatments are ineffective. Indeed, decades of research have confirmed that psychotherapy is remarkably and consistently effective (Smith et al., 1980; Wampold & Imel, in press). Rather, these results simply suggest that psychodynamic treatments are equivalently efficacious when compared with non-psychodynamic forms of therapy, at the end of treatment and from that point forward.

Given the findings of this meta-analysis, future research may benefit from a focus on alternative factors associated with sustained treatment

success outside of treatment type. More specifically, future research could focus on client and therapist characteristics that may be associated with the enduring effects of psychotherapy. Empirically investigating pan-theoretical psychotherapy factors may yield more promising findings in regard to the long-term benefits of psychotherapy (e.g., client attributions regarding therapeutic change; Powers, Smits, Whitley, Bystritsky, & Telch, 2008). For example, in one study, Powers et al. (2008) manipulated an exposure-based treatment plus inactive pill condition, so that participants were led to believe that the inactive pill made the exposure more tolerable, or had no effect on their ability to complete the exposure treatment. Results indicated the participants who attributed the ease of completing the exposure treatment to the inactive pill had higher relapse compared to the other conditions (Powers et al., 2008). This result suggests that attention to attributions about the treatment may be as important than the actual ingredients of the treatment. Additionally, future meta-analyses of studies with repeated assessments would benefit from utilizing multilevel longitudinal designs in order to appropriately model dependent data, reveal more sensitive and nuanced examinations of longitudinal growth patterns, and conduct statistical tests of the size of the effect at follow-up versus termination.

In summary, the findings of the present meta-analysis are consistent with and extend the growing body of research that suggests uniform efficacy among psychotherapy treatments intended to be therapeutic. Whereas the conclusion of uniform efficacy of bona fide treatments is mainly based on post-treatment analyses (Benish et al., 2008; Imel et al., 2008; Luborsky et al., 2002; Wampold & Imel, in press; Wampold et al., 2011), the present study extends these findings by analyzing the growth in treatment differences beyond treatment completion. As it appears, bona fide psychodynamic treatments are equally effective as bona fide non-psychodynamic treatments at post-treatment and beyond.

Contributors

Maleeha Abbas, Brian T. Pace, Noah E. Yulish, Joel G. Thomas, and Megan M. Cullen assisted in the literature search, coding identified studies, and editing the manuscript. D. Martin Kivlighan, III and Bruce E. Wampold designed the study. D. Martin Kivlighan, III wrote the first draft of the manuscript. D. Martin Kivlighan, III, Simon B. Goldberg, and Christoph Fluckiger conducted the statistical analyses. All authors contributed to and have approved the final manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Appendix A. References for included studies in the present meta-analysis

Bachar, E., Latzer, Y., Kreitler, S., & Berry, E. M. (1999). Empirical comparison of two psychological therapies: Self psychology and cognitive orientation in the treatment of anorexia and bulimia. *Journal of Psychotherapy Practice & Research*, 8(2), 115–128.

Barkham, M., Rees, A., Shapiro, D. A., Stiles, W. B., Agnew, R. M., Halstead, J., Culverwell, A., & Harrington, V. G. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology*, 64(5), 1079–1085. http://dx.doi.org/10.1037/0022-006X.64.5.1079

Barkham, M., Shapiro, D. A., Hardy, G. E., & Rees, A. (1999). Psychotherapy in two-plus-one sessions: Outcomes of a randomized

controlled trial of cognitive-behavioral and psychodynamic–interpersonal therapy for subsyndromal depression. *Journal of Consulting and Clinical Psychology*, 67(2), 201–211. http://dx.doi.org/10.1037/0022-006X.67.2.201

Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology*, 57(5), 607–612. http://dx.doi.org/10.1037/0022-006X.57.5.607

Driessen, E., Van, H. L., Schoevers, R. A., Cuijpers, P., van Aalst, G., Don, F. J., Hendriksen, M., Kool, S., Molenaar, P. J., Peen, J., & Dekker, J. M. (2007). Cognitive Behavioral Therapy versus Short Psychodynamic Supportive Psychotherapy in the outpatient treatment of depression: A randomized controlled trial. *BMC Psychiatry*, 7. doi:http://dx.doi.org/10.1186/1471-244X-7-58

Driessen, E., Van, H. L., Don, F. J., Peen, J., Kool, S., Westra, D., Hendriksen, M., Schoevers, R. A., Cuijpers, P., Twisk, W. R., & Dekker, J. M. (2013). The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: A randomized clinical trial. *The American Journal of Psychiatry*, 170(9), 1041–1050. http://dx.doi.org/10.1176/appi.ajp.2013.12070899

Emmelkamp, P. G., Benner, A., Kuipers, A., Feiertag, G. A., Koster, H. C., & van Apeldoorn, F. J. (2006). Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *British Journal of Psychiatry*, 189(1), 60–64. http://dx.doi.org/10.1192/bjp.bp.105.012153

Gallagher-Thompson, D., Hanley-Peterson, P., & Thompson, L. W. (1990). Maintenance of gains versus relapse following brief psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 58(3), 371–374. http://dx.doi.org/10.1037/0022-006X.58.3.371

Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology*, 62(3), 543–549. http://dx.doi.org/10.1037/0022-006X.62.3.543

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., Kremers, I., Nadort, M., & Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, 63(6), 649–658.

Hardy, G. E., Barkham, M., Shapiro, D. A., Stiles, W. B., Rees, A., & Reynolds, S. (1995). Impact of Cluster C personality disorders on outcomes of contrasting brief psychotherapies for depression. *Journal of Consulting and Clinical Psychology*, 63(6), 997–1004. http://dx.doi.org/10.1037/0022-006X.63.6.997

Hellerstein, D. J., Rosenthal, R. N., Pinsker, H., Samstag, L, Muran, J., & Winston, A. (1998). A randomized prospective study comparing supportive and dynamic therapies: Outcome and alliance. *Journal of Psychotherapy Practice & Research*, 7(4), 261–271.

Knekt, P. P., Laaksonen, M. A., Raitasalo, R. R., Haaramo, P. P., & Lindfors, O. O. (2010). Changes in lifestyle for psychiatric patients three years after the start of short- and long-term psychodynamic psychotherapy and solution-focused therapy. *European Psychiatry*, 25(1), 1–7. http://dx.doi.org/10.1016/j.eurpsy.2009.03.006

Knekt, P., Lindfors, O., Laaksonen, M. A., Raitasalo, R., Haaramo, P., & Järvikoski, A. (2008). Effectiveness of short-term and long-term psychotherapy on work ability and functional capacity—A randomized clinical trial on depressive and anxiety disorders. *Journal of Affective Disorders*, 107(1–3), 95–106. http://dx.doi.org/10.1016/j.jad.2007.08.005

Knekt, P., Lindfors, O., Sares-Jäske, L., Virtala, E., & Härkänen, T. (2013). Randomized trial on the effectiveness of long- and short-term psychotherapy on psychiatric symptoms and working ability during a 5-year follow-up. *Nordic Journal of Psychiatry*, 67(1), 59–68. http://dx.doi.org/10.3109/08039488.2012.680910

Leichsenring, F., Salzer, S., Jaeger, U., Kächele, H., Kreische, R., Leweke, F., Ruger, U., Winkelbach, C., & Leibing, E. (2009). Short-term psychodynamic psychotherapy and cognitive-behavioral therapy in generalized anxiety disorder: A randomized, controlled trial. *The American Journal*

of Psychiatry, 166(8), 875–881. http://dx.doi.org/10.1176/appi.ajp.2009.09030441

Maina, G., Forner, F., & Bogetto, F. (2005). Randomized controlled trial comparing brief dynamic and supportive therapy with waiting list condition in minor depressive disorders. *Psychotherapy and Psychosomatics*, 74(1), 43–50. http://dx.doi.org/10.1159/000082026

Maina, G., Rosso, G., Crespi, C., & Bogetto, F. (2007). Combined brief dynamic therapy and pharmacotherapy in the treatment of major depressive disorder: A pilot study. *Psychotherapy and Psychosomatics*, 76(5), 298–305. http://dx.doi.org/10.1159/000104706

Muran, J., Safran, J. D., Samstag, L., & Winston, A. (2005). Evaluating an alliance-focused treatment for personality disorders. *Psychotherapy: Theory, Research, Practice, Training*, 42(4), 532–545. http://dx.doi.org/10.1037/0033-3204.42.4.532

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(1), 87–98.

Salzer, S., Winkelbach, C., Leweke, F., Leibing, E., & Leichsenring, F. (2011). Long-term effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in generalized anxiety disorder: 12-month follow-up. *The Canadian Journal of Psychiatry/La Revue Canadienne De Psychiatrie*, 56(8), 503–508.

Svartberg, M., Stiles, T. C., & Seltzer, M. H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster c personality disorders. *The American Journal of Psychiatry*, 161(5), 810–817. http://dx.doi.org/10.1176/appi.ajp.161.5.810

Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology*, 55(3), 385–390. http://dx.doi.org/10.1037/0022-006X.55.3.385

Winston, A., Laikin, M., Pollack, J., & Samstag, L. (1994). Short-term psychotherapy of personality disorders. *The American Journal of Psychiatry*, 151(2), 190–194.

Woody, G. E., McLellan, A., Luborsky, L., & O'Brien, C. P. (1987). Twelve-month follow-up of psychotherapy for opiate dependence. *The American Journal of Psychiatry*, 144(5), 590–596.

References

Abbass, A.A., Hancock, J.T., Henderson, J., & Kisely, S. (2006). Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database of Systematic Reviews*(4). http://dx.doi.org/10.1002/14651858.CD004687.pub3 (Article no. CD004687).

Abbass, A., Kisely, S., & Kroenke, K. (2009). Short-term psychodynamic psychotherapy for somatic disorders: Systematic review of meta-analysis of clinical trials. *Psychotherapy and Psychosomatics*, 78(5), 265–274. http://dx.doi.org/10.1159/000228247.

Abbass, A., Town, J., & Driessen, E. (2011). The efficacy of short-term psychodynamic psychotherapy for depressive disorders with comorbid personality disorder. *Psychiatry: Interpersonal and Biological Processes*, 74(1), 58–71. http://dx.doi.org/10.1521/psyc.2011.74.1.58.

Anderson, E.M., & Lambert, M.J. (1995). Short-term dynamically oriented psychotherapy: A review and meta-analysis. *Clinical Psychology Review*, 15, 503–514. http://dx.doi.org/10.1016/0272-7358(95)00027-M.

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(3), 395–405. http://dx.doi.org/10.1016/j.cpr.2013.01.004.

Benish, S.G., Imel, Z.E., & Wampold, B.E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review*, 28(5), 746–758. http://dx.doi.org/10.1016/j.cpr.2007.10.005.

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

Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *The American Journal of Psychiatry*, 149(2), 151–158.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K.S. (2013). A meta-analysis of cognitive–behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie*, 58(7), 376–385.

Driessen, E., Cuijpers, P., de Maat, S.M., Abbass, A.A., de Jonghe, F., & Dekker, J.M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis. *Clinical Psychology Review*, 30(1), 25–36. http://dx.doi.org/10.1016/j.cpr.2009.08.010.

Eysenck, H.J. (1994). The outcome problem in psychotherapy: What have we learned? *Behaviour Research and Therapy*, 32(5), 477–495. http://dx.doi.org/10.1016/0005-7967(94)90135-X.

Flückiger, C., Del Re, A.C., Munder, T., Heer, S., & Wampold, B.E. (2014). Enduring effects of evidence-based psychotherapies in acute depression and anxiety disorders versus treatment as usual at follow-up — A longitudinal meta-analysis. *Clinical Psychology Review*, 34(5), 367–375. http://dx.doi.org/10.1016/j.cpr.2014.05.001.

Flückiger, C., Del Re, A.C., & Wampold, B. E. (2015). The sleeper effect: Artifact or phenomenon—A brief comment on Bell et al. (2013). *Journal Of Consulting And Clinical Psychology*, 83(2), 438–442. http://dx.doi.org/10.1037/a0037220.

Flückiger, C., Del Re, A.C., Wampold, B.E., Symonds, D., & Horvath, A.O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology*, 59, 10–17.

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

Hofmann, S.G., & Lohr, J.M. (2010). To kill a dodo bird. *The Behavior Therapist*, 33(1), 14–15.

Huedo-Medina, T.B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, 11, 193–206.

Imel, Z.E., Wampold, B.E., Miller, S.D., & Fleming, R.R. (2008). Distinctions without a difference: Direct comparisons of psychotherapies for alcohol use disorders. *Psychology of Addictive Behaviors*, 22(4), 533–543. http://dx.doi.org/10.1037/a0013171.

Keefe, J.R., McCarthy, K.S., Dinger, U., Zilcha-Mano, S., & Barber, J.P. (2014). A meta-analytic review of psychodynamic therapies for anxiety disorders. *Clinical Psychology Review*, 34(4), 309–323. http://dx.doi.org/10.1016/j.cpr.2014.03.004.

Lambert, M.J., & Bergin, A.E. (1994). The effectiveness of psychotherapy. In A.E. Bergin, & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 143–189) (4th ed.). Oxford England: John Wiley & Sons.

Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review*, 21(3), 401–419. http://dx.doi.org/10.1016/S0272-7358(99)00057-4.

Leichsenring, F., & Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *The American Journal of Psychiatry*, 160(7), 1223–1232. http://dx.doi.org/10.1176/appi.ajp.160.7.1223.

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

Linehan, M.M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* New York, NY, US: Guilford Press.

Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *Journal of American Psychology*, 48(12), 1181–1209. http://dx.doi.org/10.1037//0003-066X.48.12.1181.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T.P., Berman, J.S., Levitt, J.T., et al. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice*, 9(1), 2–12. http://dx.doi.org/10.1093/clipsy/9.1.2.

Miller, S., Wampold, B., & Varhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, 18(1), 5–14. http://dx.doi.org/10.1080/10503300701472131.

Minami, T., Wampold, B.E., Serlin, R.C., Kircher, J.C., & Brown, G. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology*, 75(2), 232–243. http://dx.doi.org/10.1037/0022-006X.75.2.232.

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review*, 33(4), 501–511. http://dx.doi.org/10.1016/j.cpr.2013.02.002.

Munder, T., Gerger, H., Trelle, S., & Barth, J. (2011). Testing the allegiance bias hypothesis: A meta-analysis. *Psychotherapy Research*, 21(6), 670–684. http://dx.doi.org/10.1080/10503307.2011.602752.

Powers, M.B., Smits, J.A.J., Whitley, D., Bystritsky, A., & Telch, M.J. (2008). The effect of attributional processes concerning medication taking on return of fear. *Journal of Consulting and Clinical Psychology*, 76(3), 478–490.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., Congdon, R.T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling.* Chicago, IL: Scientific Software International.

Rotton, J., Foos, P.W., Van Meek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behavior & Personality*, 10(1), 1–13.

Shapiro, D.A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, 92, 581–604. http://dx.doi.org/10.1037//0033-2909.92.3.581.

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, 65(2), 98–109. http://dx.doi.org/10.1037/a0018378.

Shedler, J. (2012). The efficacy of psychodynamic psychotherapy. In R.A. Levy, J. Ablon, & H. Kächele (Eds.), *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (pp. 9–25). Totowa, NJ US: Humana Press.

Siev, J., & Chambless, D.L. (2007). Specificity of treatment effects: Cognitive therapy and relaxation for generalized anxiety and panic disorders. *Journal of Consulting and Clinical Psychology*, 75, 513–522.

Siev, J., Huppert, J.D., & Chambless, D.L. (2009). The Dodo Bird, treatment technique, and disseminating empirically supported treatments. *The Behavior Therapist*, 32(4), 71–76.

Smith, M.L., Glass, G.V., & Miller, T.I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press. http://dx.doi.org/10.1017/S0022216X00022835.

Svartberg, M., & Stiles, T.C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 59(5), 704–714. http://dx.doi.org/10.1037/0022-006X.59.5.704.

Tolin, D.F. (2010). Is cognitive-behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical Psychology Review*, 30(6), 710–720. http://dx.doi.org/10.1016/j.cpr.2010.05.003.

Town, J. M., Abbass, A., & Hardy, G. (2011). Short-term psychodynamic psychotherapy for personality disorders: A critical review of randomized controlled trials. *Journal Of Personality Disorders, 25*(6), 723–740. http://dx.doi.org/10.1521/pedi.2011.25.6.723.

Town, J.M., Diener, M.J., Abbass, A., Leichsenring, F., Driessen, E., & Rabung, S. (2012). A meta-analysis of psychodynamic psychotherapy outcomes: Evaluating the effects of research-specific procedures. *Psychotherapy, 49*(3), 276–290. http://dx.doi.org/10.1037/a0029564.

Wampold, B.E., Budge, S.L., Laska, K.M., Del Re, A.C., Baardseth, T.P., Flűckiger, C., et al. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review, 31*(8), 1304–1312. http://dx.doi.org/10.1016/j.cpr.2011.07.012.

Wampold, B.E., & Imel, Z.E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York, NY, US: Routledge/Taylor & Francis Group (in press).

Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, 'all must have prizes'. *Psychological Bulletin, 122*(3), 203–215. http://dx.doi.org/10.1037/0033-2909.122.3.203.