Psychotherapy
2012, Vol. 49, No. 3, 276–290

0033-3204/12/$12.00 DOI: 10.1037/a0029564

A Meta-Analysis of Psychodynamic Psychotherapy Outcomes: Evaluating the Effects of Research-Specific Procedures

Joel M. Town
Dalhousie University

Marc J. Diener
Long Island University-CW Post

Allan Abbass
Dalhousie University

Falk Leichsenring
University of Giessen

Ellen Driessen
VU University of Amsterdam

Sven Rabung
University Medical Center Hamburg-Eppendorf, and Alpen-Adria-Universität

The aim of this research was to examine the extent to which the use of research-specific procedures in psychodynamic psychotherapy impacts upon treatment effectiveness and which variables moderate this potential relationship. Effects of audio/video recording of sessions, use of treatment manuals, and checks of treatment fidelity were examined. A meta-analysis was conducted on randomized controlled trials of psychodynamic psychotherapy. Forty-six independent treatment samples totaling 1615 patients were included. The magnitude of change between pretreatment and posttreatment aggregated across all studies (45 treatment samples) for overall outcome was large ($\bar{d}$ = 1.01), and further improvement was observed between posttreatment and an average 12.8-month follow-up ($\bar{d}$ = 0.18). Subgroup analyses comparing studies that used research-specific procedures and those that did not revealed that for posttreatment data no differences in treatment effects were found. However, the use of treatment manuals and fidelity checks were significantly associated with improvement between the end of treatment and follow-up assessment. Within the limitations of analyses, this data offered preliminary evidence that use of research-specific procedures does not contribute in a negative manner to posttreatment outcomes in psychodynamic psychotherapy, and their use contributes to positive differences that emerge with time. These findings, although observational in nature, make a case for reconsidering how dimensions of clinical utility and experimental control may be integrated in psychodynamic psychotherapy to enable further elucidation of principles that evidently work.

Keywords: psychodynamic, psychotherapy, meta-analysis, research, moderator

Supplemental materials: http://dx.doi.org/10.1037/a0029564.supp

Psychotherapy outcome research has evolved to regard the need for adequate specificity and standardization of psychotherapy treatments as essential, despite limitations in the generalizability of the controlled trial methodology (e.g., Blatt & Zuroff, 2005; Roth & Fonagy, 2005). Nevertheless, having a true and accurate picture of the nature of a treatment delivered is relevant for controlled research, naturalistic studies, and practice-based evidence more generally to enable a valid assessment of effectiveness. Psychotherapy research and to a lesser extent clinical and training facilities, therefore, now commonly use manual-based treatments, audio/video recording of sessions, and formal checks on treatment fidelity. From here on, audio/video recording, treatment manuals, and fidelity checks, as a collective are commonly described as *research-specific procedures*.

Changes in clinical practice and psychotherapy training (Crits-Christoph, Frank, Chambless, Brody, & Karp, 1995) can be traced to efforts to incorporate these research methods more systematically, moving away from the historical position of less structured methods and theoretical texts lent from psychoanalysis (Matarazzo & Garner, 1992). However, psychodynamic clinicians and teachers, past and present, continue to vary widely in both their attitudes toward these

Joel M. Town and Allan Abbass, Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada; Marc J. Diener, Clinical Psychology Doctoral Program, Long Island University—CW Post Brookville, NY, USA; Falk Leichsenring, Clinic of Psychosomatics and Psychotherapy, University of Giessen, Giessen, Germany; Ellen Driessen, Department of Clinical Psychology, Faculty of Psychology and Education, VU University of Amsterdam, Amsterdam, The Netherlands; Sven Rabung, Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, and University of Klagenfurt, Klagenfurt, Austria.

The authors would like to dedicate this article to the memory of Dr. Leigh McCullough: With grateful appreciation for Dr. McCullough's passionate support, collaboration, and a lifetime's achievements in advancing the field of psychotherapy research and clinical theory, practice, and training.

Correspondence concerning this article should be addressed to Joel M. Town, Abbie J. Lane Building, 7th Floor, Room 7516, 5909 Veterans' Memorial Lane, Halifax, Nova Scotia B3H 2N1, Canada. E-mail: joel.town@dal.ca

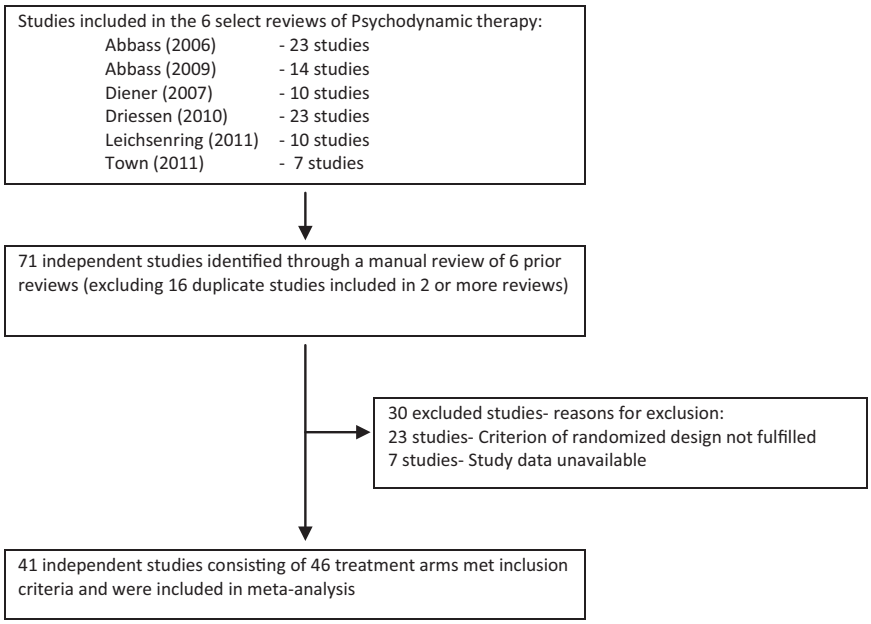

Figure 1. Process of selection of trials.

methods and their use of them. Further, the empirical question of whether the requisites of experimental research diverge from the requisites of good treatment remains unclear. This article examines the impact on clinical outcome, if any, of using audio/video recording, treatment manuals, and treatment fidelity checks in psychodynamic psychotherapy.

Background

It has been well-documented that psychodynamic therapists have historically tended to be antipathetic toward scientific investigation (Parry, Roth, & Fonagy, 2005). One view is that what can be learnt from current paradigms of research is of limited significance because of over simplistic attempts to quantify the complex mental activity presumed within psychoanalytic theory (Green, 1996). Further objections lie in theoretically based assumptions that research methods confound treatment objectives. For example, traditional psychoanalytic theory might assume that audio/video recording compromises the neutrality of the therapist, arguably by contributing to the existing therapist–patient power imbalance: the sanctity of the therapeutic relationship is sullied, therefore distorting the transference–countertransference matrix. Theoretically, this could harm the therapeutic process, enhance patient resistance, and impede treatment progress. Assuming iatrogenic effects such as this can and do regularly occur also with use of treatment manuals and fidelity checks (which empirical data do not currently substantiate), more widespread use of psychodynamic treatment protocols incorporating these research-specific components is unlikely.

Audio/Video Recording in Psychodynamic Psychotherapy

A number of recent articles from psychodynamically trained clinicians have highlighted the utility of, and applications for, this technology (Abbass et al., 2011; McCullough, Bhatia, Ulvenes, Berggraf, & Osborn, 2011; Manring, Greenberg, Gregory, & Gallinger, 2011), prompting the statement that such innovations, "will take psychotherapy training, research, supervision, and treatment forward toward increased effectiveness" (Barnett, 2011, p. 103). However, variation in the use of audio/video equipment by psychodynamic therapists exists, and it would be overly simplistic to suggest that this can be wholly attributed to theories, like that described earlier, that have since been updated and revised. Other possible contributory factors relevant to the wider psychotherapy field include (a) therapists' and/or supervisors' anxiety, (b) strain and pressures that come with set-up and time needed to review tapes, (c) resource limitations preventing access to technology, (d) difficulties guaranteeing confidentiality and/or security of tapes, and (e) training deficits/limited knowledge about the technology.

Despite the aforementioned possible limitations or "barriers" to audio/video recording, psychodynamically orientated researchers have described the advantages of using video recording for psychotherapy training purposes (Abbass, 2004; Binder, 1993, 1999; Hilsenroth, Defife, Blagys, & Ackerman, 2006; Levenson & Strupp, 1999), including clinical supervision within psychoanalytic programs (Haggerty & Hilsenroth, 2011). As an educational experience, video playback enables microanalysis of key in-session events (Aveline, 1992; Binder, 1993; Levenson, 2006), and the opportunity to expose less experienced therapists to treatment nuances. Using recordings, therefore, may help trainees overcome rigid adherence to technique, which can occur at the expense of the therapeutic alliance (Strupp, Butler, & Rosser, 1988). Video recording in particular allows the detection and analysis of subtle nonverbal cues communicated between patient and therapist that would otherwise go unnoticed. Furthermore, the use of video playback can enable greater specificity in guiding the timing and application of therapeutic technique, providing anchored instruction (Binder, 2004). This is consistent with the finding that use of

video recording within a psychotherapy training program enabled acquisition of psychodynamic techniques (Hilsenroth et al., 2006). Haggerty and Hilsenroth (2011) argued that recording therapy sessions can also minimize the effect of memory limitations (e.g., forgetting, misattribution, absent mindedness, bias) that restrict how accurately session content can be examined using more indirect methods.

Beyond providing a resource that can supplement and facilitate the supervisory process, session recording also facilitates the ability of therapists to conduct self-analysis (Alpert, 1996; Aveline, 1992; Huhra, Yamokoski-Maynhart, & Prieto, 2008; Wolberg, 1954). More generally, therapists' exposure to videotapes of their own sessions and those of others promotes increased self-awareness and improved anxiety tolerance (Abbass, 2004). Although few studies have examined the impact of using recordings, a recent meta-analysis on the utility of affect-focused techniques in psychodynamic therapies found that the most robust moderator variable was the presence of use of audio/videotapes within supervision ($r = .29$; Diener, Hilsenroth, & Weinberger, 2007). Despite this accumulation of evidence, to our knowledge, the empirical question of whether audio/video technology contributes positively or negatively to clinical outcomes has not been formally answered (Brown, 1990; Friedmann, Yamamoto, Wolkon, & Davis, 1978).

Treatment Manuals in Psychodynamic Psychotherapy

Psychotherapy treatment manuals provide information on the nature of the disorder being treated and the specific treatment strategies and technique, duration, and format to be delivered (Clarkin, 1998). Early psychodynamic psychotherapy manuals reported time-limited, focal forms of treatment (e.g., Supportive-Expressive Psychotherapy, Luborsky, 1984; Time-Limited Dynamic Psychotherapy, Strupp & Binder, 1984), and more recently, contemporary longer-term treatments have been manualized (e.g., Transference focused psychotherapy, Clarkin, Yeomans, & Kernberg, 2006; Mentalization-based treatment, Bateman & Fonagy, 2004). The introduction of manuals, primarily as a research tool for operationalizing interventions under empirical investigation, appeared to also offer the potential for improving outcomes through delivering adherent empirically based treatments (Wilson, 1995, 1996), superior to reliance on the clinical judgment of therapists (Drozd & Goldfried, 1996; Wilson, 1998). However, based on national practitioner surveys (e.g., Addis & Krasnow, 2000), a reluctance to incorporate manualized methods is not unique to psychodynamically informed therapists. Although the merit in common practitioner concerns about manualized approaches (Addis, Wade, & Hatgis, 1999) may be questioned (Fonagy, 1999), the empirical literature indicates that treatment manuals do not ensure effective delivery of therapy (Binder, 1993; 1999; Butler & Strupp, 1993). However, based on a small number of studies that produced conflicting results, and given the limitations of study methodology, it is less clear whether that which is gained in adherence to specific techniques through manuals (Crits-Christoph et al., 1998; Henry, Strupp, Butler, Schacht, & Binder, 1993; Hilsenroth, Defife, Blagys, & Ackerman, 2006; Multon et al., 1996) may be at the expense of other therapeutic factors and therapeutic progress (Strupp & Anderson, 1997).

Conclusions about the value of treatment manuals note their limitations but overall support their clinical and training utility as a source of "conceptual support" for defining patient problems and guiding the content of interventions (Binder et al., 1993; Strupp & Anderson, 1997). Furthermore, manuals have aided psychotherapy research efforts to better understand mechanisms of therapeutic change. Treatment manuals have almost become a necessary element in psychotherapy outcome studies (Kazdin, 1994; Kiesler, 1994; Lambert & Bergin, 1994); thus, they offer an accepted medium for operationalizing the independent variable and as such, a framework for assessing fidelity.

Treatment Fidelity in Psychodynamic Psychotherapy

Treatment fidelity (also commonly referred to as *treatment integrity*) refers to the extent to which therapy is delivered as intended (Kazdin, 1994). *Adherence* considers whether the core components or techniques, typically described in a treatment manual, are implemented, and *treatment competence* is the skill or accuracy with which these interventions are delivered: both components define treatment fidelity but may be measured separately. Although treatment manuals offer direction for standardizing how therapy can be delivered, formal assessment of treatment implementation usually involves objective review of audio/video recordings of sessions: methods may include use of a reliable scale of measurement, independent trained raters, and interrater reliability calculations.

Empirical research measuring treatment adherence and competence has overall demonstrated no consistent relationship with outcome, irrespective of the adequacy of the methods used. The lack of effect may be due to the methods used having been inadequate to detect an association, or that other treatment characteristics, therapist, or patient factors, confound a possible relationship with outcome (Perepletchikova & Kazdin, 2005). Leichsenring et al. (2011, p. 319) concluded that the lack of association could be because fidelity plays a more complex role in the outcome of psychotherapy.

According to psychological theory, psychotherapy is intended to relate to therapeutic change. Thus, independent of the result, treatment fidelity is relevant to any discussion on outcome because it can point to the likely validity of statements about effectiveness/efficacy (Kazdin, 1994, p. 38). For example, treatment fidelity is relevant for determining whether (a) a treatment delivered is representative of the theoretical constructs and mechanisms presumed to underpin its purpose, (b) the extent to which treatment effects are causally attributed to the treatment applied, and (c) whether these methods are generalizable to clinical practice (Leichsenring et al., 2011).

Preliminary results from a meta-analysis, albeit with a small sample ($k = 9$) of studies comparing psychodynamic psychotherapy and CBT, found a positive relationship between psychodynamic studies reporting treatment fidelity procedures and outcome (Leichsenring et al., 2011). This was true, however, for only a few variables (Leichsenring et al., 2011). Although nonsignificant, some of the correlations were substantial, corresponding to medium-large effect sizes. This was interpreted as possible evidence that the implementation of fidelity checks may offer a small effect on the outcome for psychodynamic treatment studies. Re-

search was recommended to confirm a result that could potentially have practice implications for psychodynamic psychotherapy.

Current Study

Psychodynamic psychotherapy clinicians, researchers, and stakeholders alike continue to look for methods, techniques, and strategies for titrating and improving the size, longevity, and transmissibility of treatment effects. Progress is necessary to increase the clinical utility of empirical findings and reduce the researcher–clinician gap. Given the aforementioned wide-ranging relevance of research-specific procedures, we propose that dismissing their potential application requires demonstrable evidence that they are associated with significantly worse outcomes. Proponents committed to psychodynamic principles regularly use audio/video recording, treatment manuals, and fidelity checks in clinical trials, though, to our knowledge, no well-controlled clinical trials exist that used a dismantling procedure to examine the possible effects of their use within any psychotherapy modality, psychodynamic, or otherwise. This gap in the literature may contribute to variation in attitudes toward their use and also limits conclusions on the effects of such use.

This review will therefore examine empirically the extent to which use of research-specific procedures in psychodynamic psychotherapy effect outcomes. As a secondary research question, we sought to identify which variables moderate the potential relationship between the use of these research methods in psychodynamic psychotherapy and outcome.

Method

Selection of Studies

Using extensive searches previously conducted, we selected articles using the list of studies included in the six most recently published reviews on psychodynamic psychotherapy conducted by the coauthors (Abbass, Hancock, Henderson, & Kisely, 2006; Abbass, Kisely, Kroenke, 2009; Diener et al., 2007; Driessen et al., 2010; Leichsenring & Rabung, 2011; Town, Abbass, & Hardy, 2011) (Table 1 Characteristics of reviews). Criteria used for selecting studies differed between reviews: One review selected studies with a parallel group design consisting of a nonactive control arm (Abbass et al., 2006), whereas others included active treatment comparisons (Driessen et al., 2010; Leichsenring & Rabung, 2011; Town et al., 2011). Four of these reviews selected only Short-Term Psychodynamic Psychotherapy (STPP), excluding studies where the average treatment length was greater than 40 therapy sessions (Abbass et al., 2006; Abbass, Kisely, & Kroenke, 2009; Driessen et al., 2010; Town et al., 2011), the fifth selected only Long-Term Psychodynamic Psychotherapies (LTPP) (Leichsenring & Rabung, 2011), and the sixth described studies of psychodynamic psychotherapy that presented data relevant to therapist facilitation of patient affect experiencing/expression (Diener et al., 2007). We also noted that the original searches were conducted at different times, so additional studies may have since been published. However, the literature from which data were selected is representative of a large body of studies ($k = 71$) commonly cited on the overall effectiveness of short- and long-term psychodynamic psychotherapy. We felt this provided a valid rationale for using studies from this list of previously published and recent meta-analyses.

Inclusion Criteria

Having identified studies of interest based on their inclusion in one of the six published reviews on psychodynamic psychotherapy, a second screening process was used to confer eligibility. We included studies in which the psychotherapy treatment (a) was described by the authors as psychodynamic or psychoanalytic in nature, (b) was provided in an individual or group format (e.g., not Internet delivered or self-help), and (c) applied verbal techniques (e.g., treatments using art as a form of expression were excluded). All participants had to be at least 18 years old and considered to have a common mental disorder. The latter included anxiety disorders, depressive disorders, stress-related physical conditions, and interpersonal or personality problems mixed with symptom disorders. Psychotic disorders were excluded. We only included randomized controlled trials (RCTs) in which psychodynamic psychotherapy was implemented as an active treatment arm and

Table 1
Characteristics of Psychodynamic Therapy Reviews From Which Studies Were Selected

Review	Psychodynamic therapy format	Study design	Population	Databases searched	Date search completed	Studies included
Abbass et al., 2006	STPP; ≤40 sessions; individual tx	RCT	Common mental disorders	PsycINFO, CENTRAL, MEDLINE, CINAHL, EMBASE	April 2005	23
Abbass et al., 2009	STPP; ≤40 sessions; individual tx	RCT, non-RCT	Somatic disorders	PsycINFO, MEDLINE, Cochrane Library	2007	14
Diener et al., 2007	No restrictions	RCT, non-RCT	Mixed	PsycINFO	February 2006	10
Driessen et al., 2010	STPP; ≤40 sessions; individual/group	RCT, non-RCT	Depressive disorders	PsycINFO, PubMed, Cochrane Library, EMBASE, Web of Science	November 2007	23
Leichsenring & Rabung, 2011	LTPP; >1 year or 50 sessions; individual/group	RCT, non-RCT	Common mental disorders	PsycINFO, MEDLINE, Current Contents	April 2010	10
Town et al., 2011	STPP; ≤40 sessions; individual	RCT	Personality disorders	PsycINFO, MEDLINE, CINAHL, EMBASE	January 2008	7

Note. STPP = Short-Term Psychodynamic Psychotherapy; LTPP = Long-Term Psychodynamic Psychotherapy; RCT = Randomized Controlled Trial.

participants were randomly assigned to treatment. Studies were excluded if there was insufficient data to enable calculation of within-group effect sizes.

Data Extraction

Using a structured and standardized coding scheme, written to specify study characteristics of interest, we (JMT, MJD, AA, ED, SR) independently extracted the necessary data for calculation of effect sizes alongside the following study information: full study citation, short- versus long-term treatment length, mean number of treatment sessions, primary patient diagnostic group, treatment modality, use of treatment manuals, treatment fidelity checked, outcome informant, use of audio/video recording, and use of pharmacotherapy. For primary diagnostic group, we included depressive disorders, anxiety disorders, somatoform disorders, and personality disorders consistent with *Diagnostic and Statistical Manual* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) categories. A fifth group, "mixed-other disorders," was added to ensure that other patient groups from eligible studies were captured within the dataset.

When necessary, coding ambiguities were discussed and the coding manual was revised/clarified as needed. Given the potential ambiguity in coding decisions for the "Outcome type" moderator variable (e.g., "General psychiatric symptoms" vs. "depressive symptoms," etc.), the first (JMT) and third author (AA) recoded all data for this variable. Kappa coefficients calculated to check interrater reliability between raters were excellent (Cohen's $\kappa = 0.95$). Any discrepancies between these ratings and the original ones were resolved via consensus discussion.

In the case of nine papers (Cooper, Murray, Wilson, & Romaniuk, 2003; de Jonghe et al., 2004; Hamilton et al., 2000; Linnet and Jemec, 2001; Maina et al., 2005; Monsen & Monsen, 2000; Shapiro et al., 1994/Hardy et al., 1995; Svartberg, Stiles, & Seltzer, 2004; Winston et al., 1994), ratings were completed by two coders because of overlap of studies in the authors' respective databases. Interrater reliability was computed for categorical data with the kappa coefficient, and all values (range = 0.84 to 1.0) were in the excellent range (>0.74; Fleiss, 1981). Intraclass correlation coefficients (ICC) for ratings of mean treatment length also demonstrated excellent agreement between raters (ICC[2,1] = 1.0; Shrout & Fleiss, 1979). Discrepancies in ratings were resolved via consensus discussion.

Effect size calculation consisted of within-group standardized mean difference scores using Equations 4.18–4.19 from Borenstein, Hedges, Higgins, and Rothstein (2009; Abbass, Town, & Driessen, 2012; cf., Dunlap, Cortina, Vaslow, & Burke, 1996), and the standard error was calculated using Equations 4.20–4.21 from Borenstein et al. (2009).

Effect sizes were calculated for both pre–post outcomes and posttreatment-follow up outcomes (when available). In cases in which multiple follow-ups were reported, the longest time period for follow-up was selected, with one exception: The follow-up period selected for Bateman and Fonagy (2001) was the 18-month follow-up. Although there is an additional 8-year follow-up available for this study (Bateman & Fonagy, 2008), the 18-month follow-up was selected because this interval is more commonly used to evaluate treatment effects than the longer, 8-year alternative. Effect sizes were coded as positive if the data indicated improvement in the treatment group or negative if the data indicated deterioration in the treatment group. When studies reported multiple effect sizes that were relevant (e.g., multiple outcome measures were used), these effect sizes were averaged, following standard meta-analytic convention (Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000).

The original coding manual distinguished between eight types of outcome measures. To limit the number of analyses, however, only the following outcomes were examined in the analyses: (a) Overall outcome (aggregate of all eight original outcome types), (b) General psychiatric symptoms, (c) Anxiety symptoms, (d) Depressive symptoms, and (e) Personality, that is, a condensed outcome variable consisting of the original "Personality functioning/traits" outcomes as well as outcome data based on the Inventory of Interpersonal Problems (Horowitz, Alden, Wiggins, & Pincus, 2000).

Quantitative Data Synthesis

Effect sizes were aggregated *across* studies using the random-effects method of Hedges and colleagues (Hedges & Vevea, 1998). Random-effects methods are considered to be more representative of real-world data (National Research Council, 1992) and yield results that are more generalizable than their fixed-effect counterparts (Hedges & Vevea, 1998). These calculations were performed using version 2 of the *Comprehensive Meta-Analysis* (*CMA*; *Borenstein* et al., 2005) software. Cohen's (1988) benchmarks for the magnitude of effect sizes were selected to aid in interpreting the results. Homogeneity tests and related analyses (e.g., calculation of I^2) were conducted to examine the degree of variation between effect sizes (Borenstein et al., 2009). All reported p values in the present study are two-tailed unless otherwise noted.

Moderator Analyses

Because the primary analyses of the present study involved outcome comparisons between studies that did versus did not use a particular methodological variable (e.g., recordings), all moderator analyses were conducted by comparing outcomes for studies that did and did not use that methodological variable separately for each of the relevant moderator variable levels. Thus, for example, in examining the moderator impact of use of treatment manuals on the use of recordings, outcomes for studies that used recordings were compared with outcomes for studies that did not use recordings, first for studies that did use treatment manuals and then for studies that did not use such manuals.

Continuous moderator analyses were conducted using mixed-effects (method of moments) meta-regression analyses for effect size data aggregated across all outcomes only and for pre-post data only, with the average effect size for each study serving as the dependent variable and each continuous moderator variable serving as a covariate. Because *CMA* software will not conduct a *multiple* meta-regression analysis, each covariate was examined using a separate meta-regression. Meta-regression analyses were performed using the following moderators as predictor variables of effect size: (a) Publication year, and (b) Mean number of treatment sessions. When this latter information was not explicitly reported or when data were not provided to allow calculation of mean, the

planned/estimated number of treatment sessions based on treatment completion was used.

For all categorical moderator analyses, *Q* tests, analogous to analysis of variance in primary research (Borenstein et al., 2009; Lipsey & Wilson, 2001), were calculated to determine whether the various levels of the moderator variable differed significantly from each other. When all subgroups in a particular analysis had at least six studies, estimates of the variance of true effect sizes were not pooled; when at least one subgroup, however, had fewer than six studies, estimates of the variance of true effect sizes were pooled because the accuracy yielded by pooling is likely to be greater than any real differences between subgroups (Borenstein et al., 2009).

When examining the impact of *audio/video recordings*, categorical analyses were conducted for each of the following moderator variables: (a) Use of pharmacotherapy (e.g., antidepressant medication): yes (medication use was allowed during treatment and/or follow-up), no (medication use was not permitted during treatment or follow-up), or unclear (it was unclear whether or not medication use was permitted during treatment and/or follow-up); (b) Use of a treatment manual: yes or no (regarding use of treatment manual reported in methodology); (c) Treatment fidelity: None (i.e., fidelity checks of treatment were reported, and psychodynamic treatment was shown to not be delivered as required by the study; note that no studies in the present meta-analysis were given this rating), Unknown (i.e., fidelity checks were not reported on the psychodynamic treatment), Confirmed (i.e., fidelity checks of treatment were reported, and psychodynamic treatment was shown to be delivered as required by the study); (d) Treatment length: Short-term (≤40 sessions) versus long-term (>40 sessions). When examining the impact of use of a *treatment manual*, categorical analyses were conducted for the following moderators: (a) Use of audio/video recording; (b) Use of pharmacotherapy; (c) Treatment fidelity; (d) Treatment length. Finally, when examining the impact of *treatment fidelity*, categorical analyses were conducted for the following moderators: (a) Use of audio/video recording; (b) Use of pharmacotherapy; (c) Use of a treatment manual; (d) Treatment length.

Publication Bias

Potential publication bias of the overall meta-analysis was assessed in multiple ways including (a) Begg and Mazumdar's (1994) rank correlation, (b) Egger's regression intercept (Egger, Davey Smith, Schneider, & Minder, 1997), and (c) Duval and Tweedie's (2000a, 2000b) trim and fill procedure.

Results

Inclusion of Studies

The initial search, screening studies included in six previously conducted reviews of the psychodynamic psychotherapy literature, resulted in the identification of 71 independent trials (this did not include 16 duplicate studies included in one or more of the 6 original reviews). In the second phase, the full-text of each trial was screened by one of the authors for the presence of this review's inclusion and exclusion criteria. Twenty-three studies were excluded because the criterion of randomized design was not fulfilled. Seven studies were excluded because the necessary data to enable effect size calculation were not available (Baldoni et al., 1995; Creed et al., 2003; Guthrie, Creed, Dawson, & Tomenson, 1993; Morris, 1975; Piper, Azim, McCallum, & Joyce, 1990; Sjodin, 1983; Svedlund, Sjodin, Ottosson, & Dotevall, 1983). Hardy et al. (1995) was excluded because data were reported on a subsample of Shapiro et al. (1994). This process yielded 41 independent RCTs of psychodynamic psychotherapy suitable for inclusion and meta-analysis. Within this literature, five studies (Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Huber and Klug, 2006; Knekt et al., 2008; Vinnars, Barber, Noren, Gallop, & Weinryb, 2005; Winston et al., 1994) reported two psychotherapy arms described as psychodynamic or psychoanalytic in origin and thus eligible for inclusion as forms of psychodynamic psychotherapy. In total, 46 psychodynamic psychotherapy treatment groups were available for analysis. (A complete reference list for all meta-analysed studies is available in a data supplement that accompanies the online version of this article.)

Study characteristics. The 46 independent psychodynamic psychotherapy treatment samples consisted of 1,615 subjects. The mean number of subjects per treatment sample was 35 ($SD = 28$), and the range was 8 to 128.[1] The included studies reported on subjects displaying a range of common mental health presentations. The primary diagnostic groups treated were categorized as Depressive disorders ($k = 15$), Anxiety disorders ($k = 4$), Somatoform disorders ($k = 5$), Personality disorders ($k = 13$), and mixed/other disorder ($k = 9$).

The majority of studies ($k = 38$) reported psychodynamic psychotherapy with an average treatment length of ≤40 sessions (short-term), whereas eight studies averaged >40 sessions (long-term) in treatment duration. One study described a group-based psychodynamic intervention (Liberman and Eckman, 1981), and all other treatments were delivered in an individual format. Audio or video recording of psychodynamic treatment sessions was implemented in 24 studies, and not used or not reported in 22 studies. Manualized psychodynamic treatments were used in 33 studies, whereas 13 did not report the use of a treatment manual. Treatment fidelity checks were conducted in more than half of the included studies ($k = 25$) and not checked or reported in others ($k = 21$).

Although a standardized method for evaluating study quality was not used here, several points on the methodology of the studies can be highlighted. Because only RCTs were included, all studies used random assignment to allocate subjects to condition. The assessment of treatment outcome at follow-up was present in 31 studies (average follow-up on overall outcome = 12.8 months, $SD = 9.2$). Pharmacotherapy (e.g., use of antidepressants) was

[1] Not every effect size or every study included in the present meta-analytic review was included in each analysis since, for example, separate analyses were conducted for pre–post versus posttreatment/follow-up effect sizes and Cooper et al. (2003) provided only posttreatment/follow-up data. The descriptive data reported in the main body of the article provide the numbers aggregated across every study and every effect size, and so they are not indicative of the actual sample sizes for particular analyses in the present study. The analysis with the largest sample size was the overall meta-analysis for psychodynamic psychotherapy using pre–post effect sizes. This analysis included 45 independent treatment samples from 42 peer-reviewed journal articles as well as one book chapter (plus unpublished data for that book chapter) made up of 40 RCTs. The total number of participants for this specific analysis was 1,613, with a range of sample sizes from 13 to 116, and a mean sample size of 36 ($SD = 29$).

permitted in 23 studies, not permitted in 13, and it was unclear whether medication was monitored in the remaining 10 studies.

Quantitative Data Synthesis

Effect sizes of psychodynamic psychotherapy. Including all of the 45 conditions of psychodynamic psychotherapy, results indicated a large magnitude of change in outcome from pretreatment to posttreatment when all outcome measures were aggregated, and that this change was statistically significant ($\bar{d}$ = 1.01, 95% confidence interval [CI] = 0.86–1.16, $p < .001$). Results also indicated demonstrable heterogeneity across effect sizes (Q[44] = 157.51, p[one-tailed] < .001), and the degree of variation that could be attributed to true differences between effect sizes fell in the medium to large range (I^2 = 72.07, τ^2 = 0.17, τ = 0.42; Higgins, Thompson, Deeks, & Altman, 2003). This heterogeneity is examined further by the attempt described later to identify moderator variables that could account for observed differences between effect sizes.

For outcomes measured between posttreatment and follow-up, the effect size was 0.18 (95% CI = 0.07–0.29), indicating a statistically significant improvement in overall outcome (p = .002). There was no demonstrable heterogeneity across effect sizes (Q[29] = 37.40, p[one-tailed] = .14), and the degree of variation that could be attributed to true differences between effect sizes fell slightly below Higgins et al.'s (2003) benchmark for small (I^2 = 22.46, τ^2 = 0.02, τ = 0.15).

The treatment effects in psychodynamic psychotherapy seen between pre- and posttreatment (see Table 2) were also calculated specifically for different outcome domains, with results ranging from 0.75 to 1.20, all ps < .001. These analyses found statistically significant change in outcome scores that were in the large range, and approaching the conventional 0.80 cutoff for personality functioning. At pretreatment to follow-up, change in personality functioning was large ($\bar{d}$ = 0.96, 95% CI = 0.60–1.27, $p < .001$). We calculated changes in outcomes from posttreatment to follow-up. Effect sizes ranged from 0.22 to 0.50, all ps < .05, indicating small, but significant, improvement across all outcome domains.

Publication bias for overall effect sizes. Results indicated no demonstrable evidence of publication bias for the overall meta-analysis of pre–post outcome using either Begg and Mazumdar's (1994) rank correlation method (Kendall's tau [with continuity correction, Borenstein, Hedges, Higgins, & Rothstein, 2005] = 0.13, p[one-tailed] = .10) or Egger's (Egger et al., 1997) regression intercept method (intercept = 0.79, p[one-tailed] = .16). In addition, results from Duval and Tweedie's (2000a, 2000b) trim and fill procedure suggested that the impact of potential publication bias would likely be minimal even if it did exist (zero studies trimmed; adjusted and observed estimates of effect size were identical).

For posttreatment to follow-up change, publication bias results suggested some potential for publication bias using Egger's (Egger et al., 1997) intercept method (intercept = 1.11, p[one-tailed] = .05), and only a trend toward significance for Begg and Mazumdar's (1994) approach (Kendall's tau [with continuity correction] = 0.17, p[one-tailed] = .099). Results from Duval and Tweedie's (2000a, 2000b) trim and fill procedure indicated that the impact of any potential publication bias would likely be minimal even if it did exist (zero studies trimmed; adjusted and observed estimates of effect size were identical).

Utility of audio/video recordings by outcome type. Table 3 summarizes all effect sizes and test statistics for each outcome category and type of contrast for pre–post change. The pooled effect size indicating the size of the change in overall outcome in studies with audio/video recording was 0.92 (95% CI = 0.71–1.14, $p < .001$) and 1.11 (95% CI = 0.90–1.32, $p < .001$) when recording was absent. Across each of the four individual outcome categories, the effect size for pre–post change was also large ($\bar{d}'s$ ranging from 0.85 to 1.32) and statistically significant both in studies that utilized audio/video recordings as well as studies that did not; the one exception was personality functioning, for which studies that used audio/video recording demonstrated a statistically significant effect in the medium–large range ($\bar{d}$ = 0.63, 95% CI = 0.38–0.89, $p < .001$).

The comparison for effect size between studies that did use recordings and those that did not was not statistically significant for any outcome type, though a small effect indicating potentially better outcomes was found in studies that did not use audio/video recording for depressive symptoms ($\Delta\bar{d}$ = 0.22) and personality functioning ($\Delta\bar{d}$ = 0.30).

Table 4 summarizes all effect sizes and test statistics based on posttreatment to follow-up contrasts for individual outcome types. A comparison for anxiety symptoms was not conducted because only a single study was coded as using audio/video recording. Comparisons of within-group contrasts at posttreatment to follow-up revealed statistically significant improvement with and without audio/video recording when all outcomes were aggregated together ($\bar{d}$ = 0.21, 95% CI = 0.01–0.41, p = .042; and $\bar{d}$ = 0.16, 95% CI = 0.02–0.31, p = .030; respectively). For individual outcome domains analyzed, results ranged from 0.09 to 0.42; see Table 4 for more details).

Table 2
Random Effects Meta-Analysis of Psychodynamic Psychotherapy: Pre- to Posttreatment and Posttreatment to Follow-Up Change

	Pre–post treatment					Post/follow-up				
	k	$\bar{d}$	95% CI	Z value	p	k	$\bar{d}$	95% CI	Z value	p
Overall	45	1.01	0.86–1.16	13.12	<.001*	30	0.18	0.07–0.29	3.11	.002*
Depression	28	1.20	1.0–1.40	11.76	<.001*	14	0.24	0.04–0.44	2.39	.017**
Anxiety	18	0.87	0.67–1.07	8.41	<.001*	11	0.50	0.04–1.0	2.11	.035**
Personality functioning	20	0.75	0.50–0.99	6.02	<.001*	12	0.25	0.01–0.48	2.04	.041**
General psychiatric	31	1.07	0.87–1.27	10.4	<.001*	19	0.22	0.07–0.36	2.96	.003*

* $p < .01$. ** $p < .05$.

Table 3
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Use of Audio/Video Recordings: Within-Group Change and Between-Group Differences for Pre- to Posttreatment Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
A/V	24	0.92	0.71–1.14	8.38	<.001*	−0.19	1.45	.229
No A/V	21	1.11	0.90–1.32	10.32	<.001*			
Depression								
A/V	15	1.10	0.82–1.37	7.79	<.001*	−0.22	1.19	.275
No A/V	13	1.32	1.03–1.61	8.80	<.001*			
Anxiety								
A/V	6	0.93	0.56–1.29	5.01	<.001*	0.08	0.12	.732
No A/V	12	0.85	0.59–1.10	6.51	<.001*			
Personality functioning								
A/V	13	0.63	0.38–0.89	4.85	<.001*	−0.30	1.20	.274
No A/V	7	0.93	0.46–1.39	3.91	<.001*			
General psychiatric								
A/V	16	0.98	0.64–1.31	5.64	<.001*	−0.17	0.70	.402
No A/V	15	1.15	0.92–1.38	9.93	<.001*			

Note. Δd = the magnitude of difference in outcome between studies that did use audio-visual recording and studies that did not do so.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure.
* $p < .001$.

Based on between-groups comparisons at posttreatment to follow-up, a small effect size was found for personality functioning favoring use of audio/video recording ($\Delta\bar{d} = 0.33$). All between-groups statistical comparisons were, however, nonsignificant.

Utility of treatment manuals by outcome type. Table 5 summarizes all effect sizes and test statistics for each outcome category and type of contrast for pre–post change. The pooled effect size for overall outcome in studies with treatment manuals was 0.98 (95% CI = 0.79–1.17, $p < .001$) and 1.08 (95% CI = 0.82–1.35, $p < .001$) when treatment manuals were absent. Across each of the four individual outcome categories, the effect size for pre–post change was also large ($\bar{d}$'s ranging from 0.82 to 1.28) and statistically significant both in studies that used treatment manuals as well as studies that did not; the one exception was personality functioning, for which studies that used treatment manuals ($\bar{d} = 0.74$, 95% CI = 0.41–1.06, $p < .001$) as well as studies that did not use treatment manuals ($\bar{d} = 0.78$, 95% CI = 0.48–1.09, $p < .001$) both demonstrated a statistically significant effect in the medium–large range.

Table 4
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Use of Audio/Video Recordings: Within-Group Change and Between-Group Differences for Posttreatment to Follow-Up Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
A/V	14	0.21	0.01 to 0.41	2.04	.042*	0.05	0.14	.707
No A/V	16	0.16	0.02 to 0.31	2.17	.030*			
Depression								
A/V	6	0.27	−0.13 to 0.67	1.33	.185	0.06	0.06	.808
No A/V	8	0.21	0.01 to 0.41	2.08	.037*			
Anxiety[b]								
A/V	—	—	—	—	—	—	—	—
No A/V	—	—	—	—	—	—	—	—
Personality functioning								
A/V	6	0.42	−0.04 to 0.88	1.78	.076	0.33	1.61	.205
No A/V	6	0.09	−0.12 to 0.30	0.84	.403			
General psychiatric								
A/V	8	0.25	−0.06 to 0.56	1.58	.115	0.03	0.02	.878
No A/V	11	0.22	0.05 to 0.39	2.56	.010*			

Note. Δd = the magnitude of difference in outcome between studies that did use audio-visual recording and studies that did not do so.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure. [b] The comparison for anxiety symptoms was not conducted because only a single study was coded as using audio/video recording.
* $p < .05$.

Table 5
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Use of Treatment Manual: Within-Group Change and Between-Group Differences for Pre- to Posttreatment Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
Manual	32	0.98	0.79–1.17	10.28	<.001*	−0.10	0.39	.533
No manual	13	1.08	0.82–1.35	8.03	<.001*			
Depression								
Manual	18	1.16	0.90–1.41	8.87	<.001*	−0.12	0.33	.565
No manual	10	1.28	0.94–1.62	7.41	<.001*			
Anxiety								
Manual	11	0.98	0.64–1.31	5.74	<.001*	0.16	0.63	.426
No manual	7	0.82	0.61–1.02	7.75	<.001*			
Personality functioning								
Manual	14	0.74	0.41–1.06	4.47	<.001*	−0.04	0.04	.838
No manual	6	0.78	0.48–1.09	4.97	<.001*			
General psychiatric								
Manual	24	1.04	0.78–1.31	7.72	<.001*	−0.09	0.20	.656
No manual	7	1.13	0.86–1.39	8.42	<.001*			

Note. Δd = the magnitude of difference in outcome between studies that did use treatment manuals and studies that did not do so.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure.
* $p < .001$.

All comparisons for effect size between studies that did use treatment manuals and those that did not were not statistically significant for any outcome type, and effects fell short of the conventional benchmark for a small effect (i.e., all Δds < 0.20).

Table 6 summarizes all effect sizes and test statistics based on posttreatment to follow-up contrasts for individual outcome types. Within-group contrasts revealed statistically significant improvement but only for studies that used treatment manuals ($\bar{d}$ = 0.22, 95% CI = 0.06–0.37, p = .007) for overall outcome. Similarly, effect sizes for the individual outcome domains in studies that used treatment manuals ranged from 0.25 to 0.71, all ps < .05 (see Table 6 for details). All between-groups statistical comparisons were nonsignificant, although a medium effect size for anxiety symptoms was obtained favoring use of treatment manuals ($\Delta\bar{d}$ = 0.56).

Utility of fidelity checks by outcome type. Table 7 summarizes all effect sizes and test statistics for each outcome category and type of contrast for pre–post change. The pooled effect size indicating the size of the change in overall outcome in studies with confirmed fidelity was 1.03 (95% CI = 0.81–

Table 6
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Use of Treatment Manuals: Within-Group Change and Between-Group Differences for Posttreatment to Follow-Up Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
Manual	21	0.22	0.06 to 0.37	2.71	.007*	0.10	0.59	.441
No manual	9	0.12	−0.08 to 0.31	1.20	.230			
Depression								
Manual	10	0.26	0.02 to 0.50	2.16	.031**	0.07	0.11	.744
No manual	4	0.19	−0.21 to 0.58	0.91	.363			
Anxiety								
Manual	7	0.71	0.11 to 1.31	2.33	.020**	0.56	1.26	.261
No manual	4	0.15	−0.64 to 0.93	0.36	.717			
Personality functioning								
Manual	7	0.33	0.01 to 0.66	2.02	.043**	0.19	0.60	.440
No manual	5	0.14	−0.25 to 0.52	0.70	.484			
General psychiatric								
Manual	14	0.25	0.07 to 0.43	2.74	.006*	0.09	0.29	.593
No manual	5	0.16	−0.12 to 0.44	1.11	.027			

Note. Δd = the magnitude of difference in outcome between studies that did use treatment manuals and studies that did not do so.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure.
* $p < .001$. ** $p < .05$.

Table 7
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Fidelity: Within-Group Change and Between-Group Differences for Pre- to Posttreatment Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
Fidelity[b]	24	1.03	0.81–1.25	9.05	<.001*	0.04	0.07	.796
No fidelity[c]	21	0.99	0.79–1.20	9.48	<.001*			
Depression								
Fidelity[b]	13	1.14	0.83–1.44	7.29	<.001*	−0.12	0.33	.567
No fidelity[c]	15	1.26	0.99–1.52	9.21	<.001*			
Anxiety								
Fidelity[b]	5	1.14	0.74–1.54	5.63	<.001*	0.35	2.36	.124
No fidelity[c]	13	0.79	0.54–1.01	6.51	<.001*			
Personality functioning								
Fidelity[b]	11	0.68	0.40–0.96	4.76	<.001*	−0.13	0.27	.607
No fidelity[c]	9	0.81	0.39–1.22	3.81	<.001*			
General psychiatric								
Fidelity[b]	16	1.02	0.67–1.37	5.75	<.001*	−0.09	0.18	.675
No fidelity[c]	15	1.11	0.89–1.33	10.04	<.001*			

Note. Δd = the magnitude of difference in outcome between studies with confirmed treatment fidelity and studies without reported treatment fidelity checks.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure. [b] Indicates studies that conducted treatment fidelity checks. [c] Indicates studies that did not report treatment fidelity checks.
* $p < .001$.

1.25, $p < .001$) and 0.99 (95% CI = 0.79–1.20, $p < .001$) when treatment fidelity checks were not reported. Effect sizes for individual outcome domains ranged from 0.68 to 1.26, all of which were statistically significant (all *p*s < .001; see Table 7 for details).

The comparison for effect size between studies with confirmed fidelity versus studies that did not report fidelity checks was not statistically significant for any outcome type. Effect size calculation of the magnitude of between-groups comparisons found a small effect for anxiety symptoms ($\Delta d = 0.35$) favoring studies with confirmed fidelity.

Table 8 summarizes all effect sizes and test statistics based on posttreatment to follow-up contrasts for individual outcome types. Comparison of within-group contrasts at posttreatment to follow-up revealed statistically significant improvement in four of five analyses for studies with confirmed fidelity (see table for specifics). Statistically significant within-group contrasts at posttreatment to follow-up in studies where fidelity checks were not conducted were only present for general psychiatric symptoms ($\bar{d} = 0.20$, 95% CI = 0.02–0.38, $p = .031$) and depressive symptoms ($\bar{d} = 0.23$, 95% CI = 0.01–0.46, $p = .045$).

Based on between-groups comparisons at posttreatment to follow-up, a statistically significant difference was found for anxiety symptoms with a magnitude of a large effect size favoring studies with confirmed fidelity ($\Delta\bar{d} = 1.40$, $p = .004$). All remaining between-groups statistical comparisons were, however, nonsignificant.

Moderator analyses of the impact of research-specific procedures. We conducted moderator analyses (for each respective research procedure, we considered the effect of (a) *audio-video recording/treatment manual/treatment fidelity checks* [where applicable]; (b) *pharmacotherapy use*; (c) *treatment length*; (d) *mean number of treatment sessions*), including all studies, with outcome defined as change in pretreatment to posttreatment scores for overall outcome. Only a single study was coded as not using a treatment manual that had confirmed fidelity. Similarly, only a single study was coded for an absence of treatment manual but presence of audio/video recording. These sets of analyses were therefore not conducted. Moderator results showed that no categorical analyses yielded a medium effect size (i.e., all $\Delta\bar{d}$'s < 0.50; no *p* values could be computed for these analyses because the meta-analytic software cannot calculate moderator analyses for subgroup data; the $\Delta\bar{d}$'s were calculated by subtracting the relevant difference scores).[2] In addition, all meta-regressions using the mean number of treatment sessions or the year of publication did not significantly predict outcomes for studies that did have any of the research-specific variables (all *p*s > .05); the same was true for all meta-regressions for studies that did *not* have any of the research-specific variables.

Discussion

There is mounting evidence from recent meta-analyses that demonstrate the effectiveness of psychodynamic psychotherapy (Abbass et al., 2006; Abbass et al., 2009; Driessen et al., 2010; Leichsenring &

[2] For example, to examine the moderating role of treatment length on the impact of audio/video recording, studies that did not use audio/video recording ($\bar{d} = 1.00$) were compared with studies that did use audio/video recording ($\bar{d} = 0.92$) for the group of studies that had short-term treatment. The difference associated with use of recording is therefore 1.00 − 0.92 = 0.08 for studies with short-term treatments. Next, studies that did not use audio/video recordings ($\bar{d} = 1.38$) were compared with studies that did use audio/video recordings ($\bar{d} = 0.94$) for the long-term treatments only. These yields a difference score of 1.38 − 0.94 = 0.44. When the previously mentioned difference score of 0.08 is subtracted from this difference score of 0.44, the resulting effect size is 0.36.

Table 8
Random Effects Meta-Analysis of Psychodynamic Psychotherapy by Fidelity: Within-Group Change and Between-Group Differences for Posttreatment to Follow-Up Outcomes

	Within-group					Between-group		
Subgroup	*k*	$\bar{d}$	95% CI	*Z*-value	*p*	Δd[a]	*Q*-value	*p*
Overall								
Fidelity[b]	17	0.22	0.04 to 0.40	2.41	.016**	0.08	0.39	.535
No fidelity[c]	13	0.14	−0.02 to 0.31	1.70	.089			
Depression								
Fidelity[b]	8	0.27	−0.06 to 0.59	1.58	.113	0.04	0.02	.880
No fidelity[c]	6	0.23	0.01 to 0.46	2.00	.045**			
Anxiety								
Fidelity[b]	3	1.54	0.72 to 2.35	3.69	<.001*	1.40	8.50	.004*
No fidelity[c]	8	0.14	−0.34 to 0.61	0.56	.574			
Personality functioning								
Fidelity[b]	7	0.33	0.01 to 0.66	2.02	.043**	0.19	0.60	.440
No fidelity[c]	5	0.14	−0.25 to 0.52	0.70	.484			
General psychiatric								
Fidelity[b]	10	0.27	0.00 to 0.54	1.98	.047**	0.07	0.18	.669
No fidelity[c]	9	0.20	0.02 to 0.38	2.15	.031**			

Note. Δd = the magnitude of difference in outcome between studies with confirmed treatment fidelity and studies without reported treatment fidelity checks.
[a] Positive effect sizes indicate differences in favor of use of research-specific procedure. [b] Indicates studies that conducted treatment fidelity checks. [c] Indicates studies that did not report treatment fidelity checks.
* $p < .01$. ** $p < .05$.

Rabung, 2011; Town et al., 2011). The reoccurring finding that therapeutic gains may not only be maintained after psychodynamic treatment but they continue to improve with time is especially noteworthy (Shedler, 2010). Based on the results from the current meta-analysis of only randomized controlled trials including 46 psychodynamic psychotherapy treatment samples ($\bar{d}$ = 1.01 for pre–post overall outcome; $\bar{d}$ = .18 for overall outcome between posttreatment and an average of 12.8 months of follow-up), we now believe there is unambiguous empirical support for these findings. Evidence of continued gains posttreatment is of particular importance because it suggests that under the right circumstances, psychotherapy may facilitate changes in the underlying psychological structures and intrapsychic processes presumed to mediate psychopathology, thus enabling long-lasting benefits that translate to the real world, far beyond and after therapy is complete. But, crucially much remains unknown about the precise mechanisms of psychotherapeutic change (Kazdin, 2007). Toward this goal, treatment protocols may use procedures such as audio/video recording, treatment manuals, and fidelity checks that allow specification of what is actually delivered in psychotherapy. To examine the utility and impact of these methods on treatment effectiveness in psychodynamic psychotherapy, in this meta-analysis, we conducted individual subgroup analyses to compare treatment effects in studies using recording equipment, treatment manuals, and fidelity checks versus psychodynamic psychotherapy in which each respective procedure was absent. Within the limitations of these analyses, which require results to be considered observational in nature, all comparisons between groups across different outcome domains indicated no statistically significant differences in treatment effects at posttreatment. The presence/absence of significant within-group differences in treatment effects between posttreatment and follow-up do, however, provide preliminary evidence that using research-specific procedures may in fact bring an advantage, particularly on measures of anxiety (Δds ranging from .56 to 1.40) and, to a lesser extent, personality functioning (Δds ranging from .19 to .33), which appear after treatment is completed.

The use of audio/video recording, treatment manuals, and fidelity checks in psychodynamic psychotherapy can be controversial. Based on some proponents of psychoanalytic theory, everything else being equal, one might suppose negative effects. The limitations of this conceptual argument may be questioned, given other proponents clearly committed to psychodynamic principles regularly incorporate audio/video recording, manualized treatments, and fidelity checks in clinical trials. Given the absence of statistically significant differences in posttreatment effects, overall, the current findings suggest the use of these methods, at this time, is not demonstrably associated with either better or worse outcomes at the end of therapy in efficacy trials. If you consider the magnitude of the difference in treatment effects between studies rather than rely solely on the criteria of whether they reached statistical significance, this revealed some limited (nonsignificant) differences based on the use of audio/video equipment. Of the five pre- to posttreatment between-group comparisons, two fell in the range of a small effect indicating better outcomes for the studies that did not use audio/video recording (Δd = .22 for depressive symptoms; Δd = .30 for personality functioning). At follow-up, however, this trend was not observed; by contrast, studies that used audio/video recording demonstrated *better* outcomes for personality functioning (Δd = .33).

The potential benefits of recording psychotherapy sessions for purposes of retrospective review (Abbass, 2004; Alpert, 1996; Binder, 1993a, 1999) and supervision (Aveline, 1992; Haggerty & Hilsenroth, 2011) find only limited support from the current findings. This is perhaps unsurprising given that we might expect not just the occurrence of recordings but also the type, quality, and perceived helpfulness of the review/supervision process received to impact on treatment outcome. Research efforts should pay attention to the longer-term effects of audio/video review both for

treatment outcome as well as therapist training and development of treatment competencies.

Furthermore, moderator analysis examining the magnitude of difference in pre- to posttreatment outcome between those studies that used recordings and those that did not, revealed a small effect ($\bar{d}$ = 0.36), favoring *long-term* treatment (>40 sessions). Closer inspection of the pooled mean effect size within studies of *long-* or *short-term* treatment with or without *audio/video recording* showed that the larger treatment effects seen in studies not using recording was in fact driven by five studies that provided treatment, with an average duration of approximately 3 years (145 sessions); $\bar{d}$ = 1.40. The three studies of *long-term* psychodynamic psychotherapy that used *audio/video recording*, in contrast, provided on average only 59 sessions with a pooled posttreatment effect of $\bar{d}$ = .95, comparable with *short-term* psychodynamic psychotherapy (≤40 sessions) outcomes (k = 38), with or without recording ($\bar{d}$'s ranging from .92–.99). Thus, the observation of larger effects in psychodynamic psychotherapy without audio/video recording may be better accounted for by treatment dose rather than the impact of recording equipment.

The lack of a significant association between outcome posttreatment and the use of either treatment manuals or fidelity checks is consistent with findings from the common factors literature that outcome variance due to treatment differences is minor (Lambert, 1992; Wampold et al., 1997). A recent meta-analysis (k = 9) that included five overlapping studies analyzed here, also found no significant correlations between effects sizes and treatment integrity procedures comparing psychodynamic psychotherapy with cognitive–behavioral therapy. Although effect sizes at follow-up were not examined, the magnitude of correlations at posttreatment did offer tentative evidence that the attention given to implementing procedural interventions for monitoring fidelity may particularly benefit outcome in psychodynamic psychotherapy (Leichsenring et al., 2011). The failure to replicate some of these posttreatment finding in the current study could be accounted for by a methodological difference between studies: here study differences were assessed based on a categorical distinction of the presence/absence of research procedures, which ignores possible variation in the nature and quality of the procedures used. However, a potentially new emerging finding from this study is evidence that continued therapeutic improvement occurring after psychodynamic psychotherapy may be associated to the use of research-specific procedural factors. We were stuck to find that nine of a possible 10 within-group analyses (90%), examining posttreatment to follow-up outcome change in studies according to the presence of treatment manuals or fidelity checks, were statistically significant when the same was true in only two of 10 analyses (20%) for studies excluding these procedures. Between-groups statistical comparisons were nonsignificant in all but one case; however, of the 10 analyses, two fell narrowly below the benchmark for a small effect (Δd = .19) and two were in the small/medium to large range (Δd's = .33–1.40). As such, this offers further evidence that research procedural interventions connected to assessment of how treatment is delivered, may moderate outcome. These research procedures may also just be a proxy variable for other influences that account for therapeutic change (Perepletchikova & Kazdin, 2005). We do, however, recognize that the between-group analyses showed largely small effects and included a small number of studies, meaning the addition of new studies may change the results. We advise caution when interpreting these findings, and further studies are required to confirm the results.

An alternative interpretation of these results is that treatment effects present in psychodynamic psychotherapy studies that used research-specific procedures translate largely only to improvements on measures of anxiety and personality functioning. By statistical convention, small but consistently positive effects (Δd = .19 for use of manual; Δd = .19 for use of fidelity checks; Δd = .33 for use of audio/video recording) were present on personality functioning scores in follow-up. Results at 18-month and 8-year follow-up, in one included study of manualized long-term psychodynamic psychotherapy, specifically illustrate that continued improvement and wide-ranging changes in functioning can occur at long-term outcome in a personality disordered sample (Bateman & Fonagy, 2001, 2008). More attention should be paid to the importance of efforts to pursue a "goal-orientated structure to treatment" (Critchfield & Benjamin, 2006, p. 62) and therapist consistency (Livesley, 2007) for treating personality problems. The magnitude of improvement in anxiety symptoms, however, was more pronounced, both between pre- and posttreatment and at follow-up assessment. Most notably, a very large statistically significant treatment effect ($\bar{d}$ = 1.40, p = .004) was found indicating further improvement after therapy was completed in studies that checked fidelity. This finding, when taken alongside the generally good evidence for the efficacy of manualized behavioral and cognitive–behavioral approaches with anxiety disorders, may indicate that prioritizing the delivery of specific, concrete techniques is important. This is consistent with symptoms of anxiety being potentially more tractable to strategies because they are less likely to represent a disorder of the self (Roth & Fonagy, 2005).

It remains to be seen whether these observations reflect chance findings. Having not accounted for the possible moderating effects of therapist experience and level of training, this may represent a confound within the present study: we might predict larger effects and less variation between studies regardless of the use of research procedures in more experienced and well-trained therapists. Future research using larger sample sizes and greater power is necessary to address these issues more definitively.

Critical Analysis

The results of this meta-analysis should be considered alongside its limitations. First, this studies' reliance on subgroup analyses to detect differences associated to different research-specific procedures cannot be used to prove causality: it is possible that differences in effect sizes may be attributed to differences between the sets of studies. Second, although a large number of studies were included in analyses, some controlled studies that may have met the inclusion were missed because they were not included in past publications from which eligible studies were selected. Furthermore, the original searches were conducted at different times (between 2005 and 2010) using different criteria so additional studies may have since been published. Third, the possibility of publication bias stemming from a reliance on reviews of the published outcome literature, which may fail to detect trials with null findings, should be considered: publication bias analyses conducted suggest this is, however, unlikely. Fourth, the quality of included studies was variable (e.g., medication was allowed alongside psychotherapy in 23 studies). Although the results of subanalyses to examine the moderating effect of use of treatment

manual, treatment fidelity, and medication use showed that these variables (*pharmacotherapy use, treatment length, mean number of sessions, year of publication*) did not account for a significant proportion of variance in treatment effects, power for testing moderators is often low (Borenstien et al., 2009; Hedges & Pigott, 2004). Additional points for consideration on methodological study quality such as blinding of observer-rated outcomes, reporting of intention to treat analyses, and the reliability of random allocation procedure may, however, have had an influence on treatment effects. The study can point to the inclusion of a sizable body of literature, including only controlled trials, using random allocation to group, predefined selection criteria for patient participation, and the use of reliable outcome measures, to state the case for the validity of the reported findings.

Conclusion

The current findings confirm that psychodynamic psychotherapy produces large overall treatment effects that continue to improve in long-term follow-up. Based on subgroup analysis, we can say that there is provisional evidence indicating that using research-specific treatment procedures is not associated with specifically better or worse outcomes posttreatment. However, significant treatment gains occurring after therapy appear more likely when these research-specific procedures are used. Although well-controlled research is ultimately necessary to infer causality and the mechanisms of change contributing to outcome are unclear, the current findings makes a case for reconsidering how the benefits of integrating research-specific procedures in psychodynamic psychotherapy can be maximized to outweigh perceived limitations.

References

Abbass, A., Arthey, S., Elliott, J., Fedak, T., Nowoweiski, D., Markovski, J., & Nowoweiski, S. (2011). Web-conference supervision for advanced psychotherapy training: A practical guide. *Psychotherapy, 48,* 109–118. doi:10.1037/a0022427

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

Abbass, A., Town, J. M., & Driessen, E. (2012). Intensive short-term dynamic psychotherapy: A systematic review and meta-analysis of outcome research. *Harvard Review of Psychiatry, 20,* 97–108. doi:10.3109/10673229.2012.677347

Abbass, A. A., Hancock, J. T., Henderson, J., & Kisely, S. (2006). Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database of Systematic Reviews,* CD004687. doi:10.1002/14651858.CD004687.pub3

Abbass, A. A. (2004). Small-group videotape training for psychotherapy skills development. *Academic Psychiatry, 28,* 151–155. doi:10.1093/brief-treatment/mhn001

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes towards psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68,* 331–339. doi:10.1037//0022-006X.68.2

Addis, M. E., Wade, W. A., & Hatgis, C. (1999). Barriers to dissemination of evidence based practices: Addressing practitioners concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice, 6,* 430–441.

Alpert, M. C. (1996). Videotaping psychotherapy. *Journal of Psychotherapy Practice and Research, 5,* 93–104.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Aveline, M. (1992). The use of audio and videotape recordings of therapy sessions in the supervision and practice of dynamic psychotherapy. *British Journal of Psychotherapy, 8,* 347–358. doi:10.1111/j.1752-0118.1992.tb01198.x

Baldoni, F., Baldaro, B., & Trombini, G. (1995). Psychotherapeutic perspectives in urethral syndrome. *Stress Medicine, 11,* 79–84. doi:10.1002/smi.2460110115

Barnett, J. E. (2011). Utilizing technological innovations to enhance psychotherapy supervision, training, and outcomes. *Psychotherapy, 48,* 103–108. doi:10.1037/a0023381

Bateman, A., & Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: An 18-month follow-up. *American Journal of Psychiatry, 158,* 36–42.

Bateman, A., & Fonagy, P. (2008). Eight-year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *The American Journal of Psychiatry, 165,* 631–638. doi:10.1176/appi.ajp.2007.07040636

Bateman, A. F., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization-based treatment.* New York: Oxford University Press.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. Retrieved from http://www.biometrics.tibs.org/

Binder, J. L., Strupp, H. H., Bongar, B., Lee, S. S., Messer, S., & Peake, T. H. (1993). Recommendations for improving psychotherapy training based on experiences with manual guided training and research: Epilogue. *Psychotherapy, 30,* 599–600.

Binder, J. L. (1993). Is it time to improve psychotherapy training? *Clinical Psychology Review, 12,* 301–318. doi.org/10.1016/0272-7358(93)90015-E

Binder, J. L. (1999). Issues in teaching and learning time-limited psychodynamic psychotherapy. *Clinical Psychology Review, 19,* 705–719. doi.org/10.1016/S0272-7358(98)00078-6

Blatt, S. J., & Zuroff, D. C. (2005). Empirical evaluations of the assumptions in identifying evidence based treatments in mental health. *Clinical Psychology Review, 25,* 459–486.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta- analysis version 2.* Englewood, NJ: Biostat.

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

Brown, E. (1990). Problems of the use of video recordings in training for psychotherapy. *Psychotherapy and Psychosomatics, 53,* 139–141. doi:10.1159/000288356

Butler, S. F., & Strupp, H. H. (1993). Effects of training experienced dynamic therapists to use a psychotherapy manual. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 191–210). New York: Basic Books.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry, 164,* 922–928.

Clarkin, J. F., Yeomans, F., & Kernberg, O. F. (2006). *Psychotherapy of borderline personality: Focusing on object relations.* Washington, DC: American Psychiatric Publishing, Inc.

Clarkin, J. F. (1998). Intervention research: Development and manualization. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (Vol. 3). New York: Pergamon Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, P. J., Murray, L., Wilson, A., & Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of

post-partum depression: Part I. Impact on maternal mood. *British Journal of Psychiatry, 182,* 412–419.

Creed, F., Fernandes, L., Guthrie, E., Palmer, S., Ratcliffe, J., Read, N., . . . Tomenson, B. (2003). The cost-effectiveness of psychotherapy and paroxetine for severe irritable bowel syndrome. *Gastroenterology, 124,* 303–317. doi:10.1053/gast.2003.50055

Critchfield, K. L., & Benjamin, L. S. (2006). Integration of therapeutic factors in treating personality disorders. In L. E. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 253–271). New York: Oxford University Press.

Crits-Christoph, P., Frank, E., Chambless, D. L., Brody, C., & Karp, J. F. (1995). Training in empirically validated treatments: What are clinical psychology students learning. *Professional Psychology: Research and Practice, 26,* 514–522.

Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Liese, B., . . . Woody, G. (1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology, 66,* 484–492.

de Jonghe, F., Hendriksen, M., van Aalst, G., Kool, S., Peen, J., Van, R., . . . Dekker, J. (2004). Psychotherapy alone and combined with pharmacotherapy in the treatment of depression. *British Journal of Psychiatry, 185,* 37–45. doi:10.1192/bjp.185.1.37

Diener, M. J., Hilsenroth, M. J., & Weinberger, J. (2007). Therapist affect focus and patient outcomes in psychodynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry, 164,* 936–941. doi:10.1176/appi.ajp.164.6.936

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

Drozd, J. F., & Goldfried, M. R. (1996). A critical evaluation on the state-of-the-art in psychotherapy outcome research. *Psychotherapy, 33,* 171–180.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. doi:10.1037/1082-989X.1.2.170

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98. Retrieved from http://www.amstat.org/publications/jasa.cfm

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

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal, 315,* 629–634. Retrieved from http://www.British Medical Journal.com/

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* New York: Wiley.

Fonagy, P. (1999). Achieving evidence-based psychotherapy practice: A psychodynamic perspective on the general acceptance of treatment manuals. *Clinical Psychology: Science and Practice, 6,* 442–444.

Friedmann, C., Yamamoto, J., Wolkon, G., & Davis, L. (1978). Videotaped recordings of dynamic psychotherapy: Supervisory tool or hindrance. *American Journal of Psychiatry, 135,* 1388–1391.

Guthrie, E., Creed, F., Dawson, D., & Tomenson, B. (1993). A randomized controlled trial of psychotherapy in patients with refractory irritable bowel syndrome. *British Journal of Psychiatry, 163,* 315–321.

Haggerty, G., & Hilsenroth, M. J. (2011). The use of video in psychotherapy supervision. *British Journal of Psychotherapy, 27,* 193–210. doi:10.1111/j.1752-0118.2011.01232.x

Hamilton, J., Guthrie, E., Creed, F., Thompson, D., Tomenson, B., Bennett, R., . . . Liston, R. (2000). A randomized controlled trial of psychotherapy in patients with chronic functional dyspepsia. *Gastroenterology, 119,* 661–669.

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,* 997–1004.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426–445. doi:10.1037/1082-989X.9.4.426

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. doi:10.1037/1082-989X.3.4.486

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology, 61,* 434–440.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. doi:10.1136/bmj.327.7414.557

Hilsenroth, M. J., Defife, J., Blagys, M., & Ackerman, S. (2006). Effects of training in short-term psychodynamic psychotherapy: Changes in graduate clinician technique. *Psychotherapy Research, 16,* 292–303.

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP: Inventory of Interpersonal Problems Manual.* The Psychological Corporation.

Huber, D., & Klug, G. (2006). Munich psychotherapy study (mps): The effectiveness of psychoanalytic long-term psychotherapy for depression. In Society for Psychotherapy Research (Ed.). *Book of abstracts. From research to practice* (p. 154). Ulmer Textbank.

Huhra, R. L., Yamokoski-Maynhart, C. A., & Prieto, L. R. (2008). Reviewing videotape in supervision: A developmental approach. *Journal of Counseling and Development, 86,* 412–418. doi:10.1002/j.1556-6678.2008.tb00529.x

Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 19–71). New York: Wiley.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3,* 1–27. doi:10.1146/annurev.clinpsy.3.022806.091432

Kiesler, D. J. (1994). Standardization of intervention: The tie that binds psychotherapy research and practice. In P. F. Talley, H. H. Strupp, S. F. Butler (Eds.), *Psychotherapy research and practice: Bridging the gap* (pp. 143–153). New York: Basic Books.

Knekt, P., Lindfors, O., Harkanen, T., Valikoski, M., Virtala, E., Laaksonen, M. A., . . . Renlund, C. (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. http://dx.doi.org/10.1017/S003329170700164X;/Border doi:http://dx.doi.org/10.1017/S003329170700164X

Lambert, M. J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (pp. 94–129). New York: Basic Books.

Leichsenring, F., & Rabung, S. (2011). Long-term psychodynamic psychotherapy in complex mental disorders: Update of a meta-analysis. *British Journal of Psychiatry, 199,* 15–22. doi:10.1192/bjp.bp.110.08277

Leichsenring, F., Salzer, S., Hilsenroth, M., Leibing, E., Leweke, F., & Rabung, S. (2011). Treatment integrity: An unresolved issue in psychotherapy research. *Current Psychiatry Reviews, 3,* 313–321.

Levenson, H., & Strupp, H. H. (1999). Recommendations for the future of

training in brief dynamic psychotherapy. *Journal of Clinical Psychology, 55,* 385–391.

Levenson, H. (2006). Time-limited dynamic psychotherapy. In A. B. Rochlen (Ed.), *Applying counseling theories: An outline case-based approach* (pp. 75–90). Upper Saddle River, NJ: Pearson.

Liberman, R. P., & Eckman, T. (1981). Behavior therapy versus insight-oriented therapy for repeated suicide attempters. *Archives of General Psychiatry, 38,* 1126–1130.

Linnet, J., & Jemec, G. (2001). Anxiety level and severity of skin condition predicts outcome of psychotherapy in atopic dermatitis patients. *International Journal of Dermatology, 40,* 632–636.

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

Livesley, W. J. (2007). Integrated therapy for complex cases of personality disorder. *Journal of Clinical Psychology: In Session, 64,* 207–221.

Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive/expressive treatment.* New York: Basic Books.

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,* 43–50. doi:10.1159/000082026

Manring, J., Greenberg, R. P., Gregory, R., & Gallinger, L. (2011). Learning psychotherapy in the digital age. *Psychotherapy, 48,* 119–126. doi:10.1037/a0022185

Martarazzo, R. G., & Garner, A. M. (1992). Research on training for psychotherapy. In D. K. Freedheim, H. J. Freudenberger, J. W. Kessler, S. Messer, D. R. Petevsm, H. H. Strupp, P. L. Watchel (Eds.), *History of psychotherapy: A century of change* (pp. 850–877). Washington, DC: American Psychological Association.

McCullough, L., Bhatia, M., Ulvenes, P., Berggraf, L., & Osborn, K. (2011). Learning how to rate video-recorded therapy sessions: A practical guide for trainees and advances clinicians. *Psychotherapy, 48,* 127–137. doi:10.1037/a0023131

Monsen, K., & Monsen, J. T. (2000). Chronic pain and psychodynamic body therapy: A controlled outcome study. *Psychotherapy, 37,* 257–269. doi:10.1037/h0087658

Morris, N. E. (1975). *A group self-instruction method for the treatment of depressed outpatients (Doctoral dissertation).* Toronto, Canada: University of Toronto.

Multon, K., Kivlighan, D., & Gold, P. (1996). Changes in counselor adherence over the course of training. *Journal of Counseling Psychology, 43,* 356–363.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy Press.

Parry, G., Roth, T., & Fonagy, P. (2005). Psychotherapy research, health policy, and service provision. In A. Roth & P. Fonagy (Eds.), *What works for whom: A critical review of psychotherapy research* (2nd ed., pp. 43–65). London: The Guilford Press.

Piper, W. E., Azim, H. F., McCallum, M., & Joyce, A. S. (1990). Patient suitability and outcome in short-term individual psychotherapy. *Journal of Consulting and Clinical Psychology, 58,* 475–481.

Roth, A., & Fonagy, P. (2005). *What works for whom: A critical review of psychotherapy research* (2nd ed.). New York: Guilford Press.

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522–534. doi:10.1037/0022-006X.62.3.522

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. doi:10.1037/0033-2909.86.2.420

Sjodin, I. (1983). Psychotherapy in peptic ulcer disease. A controlled outcome study. *Acta Psychiatrica Scandinavica, 67*(Suppl 307), 1–90.

Strupp, H. H., & Anderson, T. (1997). On the limitations of therapy manuals. *Clinical Psychology: Science and Practice, 4,* 76–82.

Strupp, H. H., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy.* New York: Basic Books.

Strupp, H. H., Butler, S. F., & Rosser, C. L. (1988). Training in psychodynamic psychotherapy. *Journal of Consulting and Clinical Psychology, 56,* 689–695.

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. *American Journal of Psychiatry, 5,* 810–817. doi:10.1176/appi.ajp.161.5.810

Svedlund, J., Sjodin, I., Ottosson, J. O., & Dotevall, G. (1983). Controlled study of psychotherapy in irritable bowel syndrome. *Lancet, 2,* 589–592.

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,* 723–740. doi:10.1521/pedi.2011.25.6.723

Vinnars, B., Barber, J. P., Noren, K., Gallop, R., & Weinryb, R. M. (2005). Manualized supportive-expressive psychotherapy versus nonmanualized community-delivered psychodynamic therapy for patients with personality disorders: Bridging efficacy and effectiveness. *American Journal of Psychiatry, 10,* 1933–1940. doi:10.1176/appi.ajp.162.10.1933

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,* 203–215.

Wilson, G. T. (1995). Empirically validated treatments as a basis for clinical practice: Problems and prospects. In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 163–196). Reno, NV: Context Press.

Wilson, G. T. (1996). Manual-based treatments: The clinical application of research findings. *Behavior Research and Therapy, 34,* 295–314.

Wilson, G. T. (1998). Manual-based treatment and clinical practice. *Clinical Psychology: Science and Practice, 5,* 363–375.

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

Wolberg, L. R. (1954). *The technique of psychotherapy* (4th ed.). New York: Grune & Stratton.

Received June 21, 2012
Accepted June 25, 2012 ■