Published in final edited form as:

Psychother Res. 2012 July ; 22(4): 475–488. doi:10.1080/10503307.2012.674790.

How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence?

INGA DENNHAG¹, MARY BETH CONNOLLY GIBBONS², JACQUES P. BARBER², ROBERT GALLOP³, and PAUL CRITS-CHRISTOPH²

¹Psychology, Umeå University, Umeå, 90187 Sweden

²Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania, Philadelphia, USA

³Statistics and Applied Mathematics, West Chester University, West Chester, USA

Abstract

The study utilized a generalizability theory analysis of adherence and competence ratings to evaluate the number of sessions and patients needed to yield dependable scores at the patient and therapist levels. Independent judges' ratings of supportive expressive therapy ($n = 94$), cognitive therapy ($n = 103$), and individual drug counseling ($n = 98$) were obtained on tapes of sessions from the NIDA Collaborative Cocaine Treatment Study. Generalizability coefficients revealed that, for all three treatments, ratings made on approximately five to 10 sessions per patient are needed to achieve sufficient dependability at the patient level. At the therapist level, four to 14 patients need to be evaluated (depending on the modality), to yield dependable scores. Many studies today use fewer numbers.

Keywords

statistical methodology; generalizability theory; adherence; competence

Therapist adherence and competence evaluations have been an important subject for psychotherapy outcome studies since treatment manuals were first developed. The term therapist competence has been variably used, denoting constructs as disparate as therapists' self-reported increase in self-confidence or assessor rating of structured interviews or clinicians' ability to demonstrate techniques in role play (e.g., Ekers, Lovell, & Playle, 2006; Mannix et al., 2006; Sholomskas et al., 2005). General therapist competence would also include relationship skills and other non-specific skills. Intervention competence (Kaslow, 2004) is a term more limited in its scope and more easily assessed within the framework of a specific treatment approach. This term refers to competence utilized by the therapist, in a specific type of intervention, in a specific treatment (Sharpless & Barber, 2009). In the current paper, we use the term "competence" to refer to intervention competence, not overall therapist competence.

It is also important to clarify the difference between adherence and competence, because these terms are commonly used together. Adherence refers to the extent to which a therapist delivers the prescribed techniques or interventions, and also involves not delivering

components that are proscribed by the therapy. Competence refers not only to the technical accuracy with which interventions are delivered but also to the appropriateness or adequacy with which these are delivered, taking the therapeutic context into account (e.g., Waltz, Addis, Koerner, & Jacobson, 1993). Relevant aspects of the context include, but are not limited to, client variables such as degree of impairment, the particular problems manifested by a given client, the client's life situation and stress, and factors such as stage in therapy.

From a research and a training perspective, it is important to know whether a specific therapist is competent in delivering a specific treatment. Using ratings of treatment sessions, researchers may want to certify therapists as competent in the specific treatment modality prior to initiating a clinical trial. Once a clinical trial has begun, researchers also need to evaluate whether treatments are delivered as intended and with adequate skill, and can be differentiated from comparison treatments. These applications of adherence/competence scales in research studies are important even though such ratings have often been found to not predict treatment outcome (Webb, DeRubeis, & Barber, 2010). However, the literature is inconsistent on this relationship. In cognitive therapy, adherence (DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999) and competence (controlling for adherence) (Shaw et al., 1999) have been found to be related to outcome. In addition, one recent study found that cognitive therapy competence predicted next-session outcomes (Strunk, Brotman, DeRubeis, & Hollon, 2010), and another study found that competence in the delivery of supportive-expressive therapy was related to outcome (Barber, Crits-Christoph, & Luborsky, 1996). Furthermore, such adherence and competence scales continue to be used for clinical training purposes even though associations with outcome have been inconsistent (e.g., Keen & Freeston, 2008).

Despite these important potential uses of adherence and competence evaluations, it is surprising that there has been little attention to how many treatment sessions are needed to provide an adequate assessment of adherence/competence for a given therapy dyad, or how many patients are needed to provide an adequate assessment of adherence/competence for a therapist's typical level of performance in delivering a specific psychotherapy. We are aware of only one study that has evaluated the number of sessions needed to provide an adequate assessment of competence (in this case, for cognitive behavior therapy) (Keen & Freeston, 2008). However, the evaluations in this study were not anchored to a specific treatment manual or clinical work with a specific patient population. Although the number of sessions needed for adequate assessment of adherence/competence was not provided, a recent study by Imel, Baer, Martino, Ball and Carroll (2011) indicated that the variability in adherence/competence scores within a therapist's caseload can be large, suggesting that many patients might be needed to adequately determine a given therapist's typical level of adherence/competence.

To answer the question of how many treatment sessions or patients are needed to create a stable (dependable) adherence and competence score, generalizability theory (Cronbach, Rajaratnam, & Gleser, 1963; Shavelson & Webb, 1991; Wasserman, Levy, & Loken, 2009) can be used. Generalizability theory tells us the adequacy with which one can generalize from the study sample to the sample's population. Using generalizability theory, one can calculate generalizability coefficients that index the precision, or dependability, of a measure in the context of variability due to various conditions of measurement. In brief, generalizability theory extends the concept of reliability by looking at multiple sources of variability rather than only one. Generalizability coefficients can provide information, for example, on whether scores obtained from only session are adequately stable (i.e., reliably estimate the typical level of the variable across multiple sessions) or whether more sessions are needed. Similarly, generalizability coefficients can be calculated to estimate the dependability of therapist level scores, taking into account variability from patient to patient

within a therapist. In applications where decisions are made about therapists, such as whether or not a particular therapist is competent enough in a specific treatment to participate in a randomized clinical trial evaluating the efficacy of the treatment, it is the dependability of competence scores at the therapist level of analysis that would be relevant.

Like intraclass correlations coefficients (ICCs), generalizability coefficients provide an index of adequacy of measurement ranging from 0 to 1.00. However, the difference between GCs and ICCs is that the generalizability theory makes it possible to assess multiple sources of measurement error simultaneously in order to characterize the variance components due to each source of measurement error (Cronbach et al., 1963; Shavelson & Webb, 1991; Webb, Shavelson, & Haertel, 2006). By identifying sources of error, steps can be taken to minimize error through modification of a study design (Wasserman et al., 2009) (e.g., inclusion of more sessions or patients). Different sources of error or facets (as called in generalizability theory) might include sessions, patients, therapists, and raters.

Generalizability theory is used to conduct two types of studies: (1) generalizability (G) studies that examine the characteristics of measurement within the context of the facets sampled, and (2) decision (D) studies that use the results obtained in a G study to explore and estimate how precision of a measure would change if the measurement procedures are changed (e.g., more raters, more sessions, or more therapists sampled). One important specification in any G or D study is whether the facets are fixed (only the levels of the facet included in the study are of interest) or random (levels of the facet in the study are sampled from a population or universe of interest). Further information on generalizability theory and its applications to psychotherapy research can be found in Wasserman et al. (2009) and Crits-Christoph et al. (2011) (the latter providing code for conducting such analyses).

Generalizability coefficients $\geq .80$ have been described as acceptable (Cardinet, Johnson, & Pini, 2010; Wass, van der Vleuten, Shatzer, & Jones, 2001). Descriptive interpretations of various levels of ICCs have been also proposed, with ICCs in the range of .40 to .59 described as “fair,” and in the range of .60 to .74 described as “good” (Cicchetti & Sparrow, 1981). In reality, no one cut-off value for ICCs or generalizability coefficients is appropriate for all purposes. These differing interpretations relate to the purpose of a research study or clinical application of an assessment instrument. For “high stakes” decisions about individuals in which false positives would be a concern, such as professional licensing, a coefficient of at least .90 has been proposed as the standard (Downing, 2004).

To investigate how many sessions per patient and patients per therapist are used in typical psychotherapy studies that used measures of adherence/competence, a literature search was conducted. Studies in Webb, DeRubeis and Barber’s (2010) meta-analysis were inspected and supplemented with newer studies (2010 to May of 2011) using combinations of the search terms “therapy,” “outcome,” “effectiveness,” “adherence,” “competence” in PsycINFO. All included studies used a quantifiable measure of adherence and/or competence rated by independent judges on tape-recorded treatment sessions, and used treatment manuals. A total of 24 studies were included in the review (see Table I). In these 24 studies, a mean of two and a half sessions per patient (median = 2.0, $SD = 2.2$) and a mean of eight patients per therapist (median = 5.8, $SD = 5.8$) were evaluated for adherence/competence.

The aim of the present study was to conduct a generalizability theory analysis of intervention adherence and competence ratings in three different treatments for cocaine dependence: supportive expressive therapy, cognitive therapy and individual drug counseling. Generalizability coefficients were calculated on both patient level and therapist level. To calculate data on the patient level is to aggregate data across sessions so that each

patient only has one average score. To calculate data on the therapist level is to aggregate data across patients so that each therapist only has one average score. After conducting a G study assessing variability in adherence/competence scores due to raters, patients, sessions, and therapists, we conducted a D study to explore the impact of increasing the number of sessions and patients on the precision of measurement at both the patient and therapist level.

Method

Overview

Therapist adherence and competence data of supportive expressive therapy, cognitive therapy and individual drug counseling, drawn from the National Institute of Drug Abuse Cocaine Collaborative Treatment Study (NIDA CCTS; Crits-Christoph et al., 1999), were used to do the generalizability coefficients analysis. The NIDA CCTS study was a randomized multicenter clinical trial that compared different cocaine treatments. The study was approved by an Institutional Review Board and has been reported in several studies (e.g. Crits-Christoph et al., 1997, 1999).

Participants

In the NIDA CCTS 364 patients were randomized to therapy but only data from patients whose sessions had been rated by independent judge ($n = 295$) were used in the current analyses. The patients were recruited through newspaper advertisement, from substance abuse treatment centers, by a friend or an acquaintance, mental health centers, and from private mental health providers. The inclusion criteria included a current or in early remission cocaine dependence diagnosis of DSM-IV (First, Spitzer, Gibbon, & Williams, 1994) and age between 18 and 60 years old (for more details of inclusion and exclusion criteria see Crits-Christoph et al., 1999). Patients in supportive expressive therapy ($n = 94$) were 33.7 years old ($SD = 6.2$), male ($n = 75$ [80%]), primarily Caucasian ($n = 56$ [60%]), employed ($n = 59$ [63%]), living alone ($n = 33$ [35%]), and had used cocaine in the past 30 days ($M = 9.7$, $SD = 6.9$). Patients in cognitive therapy ($n = 103$) were 35.0 years old ($SD = 6.0$), male ($n = 86$ [83%]), primarily Caucasian ($n = 59$ [57%]), employed ($n = 63$ [61%]), living alone ($n = 44$ [43%]), and had used cocaine in the past 30 days ($M = 9.8$, $SD = 8.2$). Patients in individual drug counseling ($n = 98$) were 33.3 years old ($SD = 6.7$), male ($n = 71$ [72%]), primarily Caucasian ($n = 58$ [59%]), employed ($n = 61$ [62%]), living alone ($n = 43$ [44%]), and had used cocaine in the past 30 days ($M = 10.4$, $SD = 7.6$).

The 39 therapists were volunteers recruited from newspaper announcements or from the staff of the study sites. All therapists were trained in their respective treatment modality during an intensive training phase. The intensive training consisted of reading the treatment manual, participating in a 2-day workshop, and treating four training cases under supervision. Those therapists who demonstrated their ability in delivering treatment as intended and in a competent manner during training were invited to participate in the main trial. Therapists in supportive expressive therapy ($n = 12$) were 38.9 years old ($SD = 5.1$), male ($n = 8$ [67%]), primarily Caucasian ($n = 11$ [92%]), had an MD, PhD, /PsyD or /EdD ($n = 9$ [75%]) or an MSW, MA, BA, AA or RN ($n = 3$ [25%]), had on average 11.7 years of clinical experience ($SD = 6.1$), and had 8.7 years of substance use clinical experience ($SD = 7.4$). Therapists in cognitive therapy ($n = 15$) were 40.0 years old ($SD = 6.3$), male ($n = 12$ [80%]), primarily Caucasian ($n = 14$ [93%]), had an MD, PhD, PsyD or EdD ($n = 12$ [80%]) or an MSW, MA, BA, AA or RN ($n = 3$ [20%]), had on average 13.4 years of clinical experience ($SD = 8.4$), and had 6.6 years of substance use clinical experience ($SD = 6.2$). Therapists in individual drug counseling ($n = 12$) were 40.1 years old ($SD = 6.4$), male ($n = 4$ [33%]), primarily Caucasian ($n = 9$ [75%]), had an MSW, MA, BA, AA or RN ($n = 12$

[100%]), had on average 13.8 years of clinical experience ($SD = 8.5$), and had 10.4 years of substance use clinical experience ($SD = 6.5$).

Treatments

The three individual therapies were supportive-expressive therapy (SE) based on Luborsky's psychodynamic treatment manual (Luborsky, 1984) and a psychodynamic manual for cocaine abuse (Mark & Luborsky, 1992); cognitive therapy (CT) based on Beck's cognitive model for substance abuse (Beck et al., 1993); and individual drug counseling (IDC) (Mercer & Woody, 1992) based on the 12-step model of addiction. The active phase of the treatments lasted 6 months and consisted of two sessions every week for the first 3 months, and of one session every week for the last 3 months. A more thorough description of the treatments is provided in Crits-Christoph et al. (1997, 1999).

Raters and Audiotapes

Independent judges were used to evaluate whether each specific treatment was delivered in a competent manner. The judges only rated those who were in their own expert field e.g. SE experts rated SE therapists' skills. The independent judges (SE: $n = 3$; CT: $n = 2$; IDC: $n = 2$) were not otherwise related to the research study, but were experts in their respective treatments. All judges in each treatment modality rated the same sessions from each patient. Two audiotapes were rated for each patient. They were randomly selected, one tape from sessions 2 to 11 and one tape from session 12 to the end of the therapy.

Overall, the independent judges rated 148 tapes of SE therapy, 192 tapes of CT, and 181 tapes of IDC. The average number of sessions rated per patient by the independent judges was 1.6 ($SD = .6$) for SE, 1.9 ($SD = 0.4$) for CT, and 1.8 ($SD = 0.5$) for IDC. The average number of patients rated per therapist was 7.8 ($SD = 3.5$) for SE therapy, 6.9 ($SD = 2.3$) for CT, and 8.2 ($SD = 3.9$) for IDC.

Measures

Adherence/Competence Scale for SE Cocaine Dependence (ACS-SEC; Barber, Krakauer, Calvo, Badgio, & Faude, 1997)—This 44-item scale was used to measure adherence and competence in the delivery of SE for cocaine dependence. The ratings for each item ranged from 1 (not at all) to 7 (very much). In addition to ratings of frequency of use of techniques (adherence) and quality of use of techniques (competence), ratings of the appropriateness of using difference interventions were also made. An intervention was judged as very much appropriate if the "ideal" frequency of that intervention was used in the context of a specific session. If adherence was rated as not at all for a given item, the competence rating for that item was assigned the same number as the appropriateness rating (thus competence could be high even if frequency was low). Because of high correlations between appropriateness and quality ($r = .94$), the scales were summarized into one overall competence scale. The scale contains items that focus on the therapist's ability to establish and maintain a supporting relationship (13 items) and items that reflect the therapist's focus on interventions to explore the maladaptive relationship patterns (31 items). In the present study, the internal consistency was excellent for both adherence and competence when the patient was the unit of analysis ($\alpha = .93$ and $\alpha = .98$), and when the therapist was the unit for the analysis ($\alpha = .95$ and $\alpha = .99$). The interjudge reliability for the current study sample was shown to be good for the adherence scale (ICC [2,2] = .84), but weaker for the competence scale (ICC [2,2] = .68) (Barber et al., 2008). The adherence ACS-SEC scale has also been found to discriminate SE from CT and IDC, but that was not obtained for the competence scale (Barber, Foltz, Crits-Christoph, & Chittams, 2004).

Cognitive Therapy Adherence and Competence Scale (CTACS, Barber, Liese, & Abrams, 2003)—This 21-item scale was developed for the NIDA CCTS and measures adherence and competence in the delivery of CT for cocaine dependence. Like the SE scale, the CTACS includes ratings of frequency, quality, and appropriateness of interventions. The scales' items are rated separately on a 0 (low) to 6 (high) Likert-type scale. When an intervention does not occur but should have, adherence was rated as zero, as were appropriateness and quality. However, if the absence of an intervention was appropriate, the adherence scale was rated as zero but the appropriateness scale was rated higher. The appropriateness and quality scales were highly correlated ($r > .90$; Barber et al., 2003) and summarized into one overall competence scale. The content of the scale included cognitive therapy structure, development of a collaborative relationship, case conceptualization, and cognitive and behavioral techniques. In the present study, the internal consistency was good to excellent for both adherence and competence when the patient was the unit of analysis (judges: $\alpha = .89$ vs. $\alpha = .96$) and when the therapist was the unit for the analysis (judges: $\alpha = .96$ vs. $\alpha = .98$). The interrater reliability for the independent judges in the present study's CT sample ($n = 92$) was shown to be acceptable (ICC = .67 for adherence, ICC = .73 for competence) (Barber et al., 2003). The CTACS scale as rated by independent judges also discriminated CT from SE and IDC in the NIDA CCTS study (Barber et al., 2004).

Adherence/Competence Scale for IDC for Cocaine Dependence (ACS-IDCCD, Barber, Mercer, Krakauer, & Calvo, 1996)—This scale measured adherence and competence of IDC in the current study. Each item was rated separately on a scale from 1 (low) to 7 (high). Originally, the scale contained 38 items, but for the current study 34 items of main techniques were used to create an overall score (see Barber et al., 2006). For the current study, the internal consistency was good to excellent for both adherence and competence when the patient was the unit of analysis (judges: $\alpha = .84$ and $\alpha = .92$), and when the therapist was the unit for the analysis (judges: $\alpha = .97$ and $\alpha = .97$). The interjudge ICC was acceptable for both adherence and competence (Barber, Mercer et al., 1996). The scale discriminated IDC from CT and SE therapy in the NIDA CCTS study (Barber et al., 2004; Barber, Mercer et al., 1996).

Statistical procedures

Dependability of measurement can be separately examined for each of the different facets in a study. Here, we calculated generalizability coefficients for both patient and therapist levels. Estimations of generalizability coefficients were based on a random effects model including terms for sessions, patients, therapists and raters. The mixed models were calculated in the software PASW Statistics 18 (SPSS). These mixed models included additional terms to accommodate the nesting of the repeated measures across sessions within patients, patients nested within therapist, and interactions among these terms. We calculated the generalizability coefficient (GC) according to formulas given on pages 28 and 29 in Webb et al. (2006), but with the addition of the variance components due to rater. On the patient-level we used the formula:

$$GC_p = \frac{\sigma^2_t + \sigma^2_{p:t}}{\sigma^2_t + \sigma^2_{p:t} + \frac{\sigma^2_{ts}}{n_s} + \frac{\sigma^2_{p:ts}}{n_s} + \frac{\sigma^2_{tr}}{n_r} + \frac{\sigma^2_{pr}}{n_r} + \frac{\sigma^2_{(p:t)sr,e}}{n_s n_r}}$$

where GC_p is the generalizability coefficient on the patient level, σ^2_t is the variance attributable to therapists, $\sigma^2_{p:t}$ is the variance attributable to patients nested in therapists, n_s is the number of 2 sessions, n_r is the number of raters, σ^2_{ts} is the variance attributable to the therapist by session interaction, $\sigma^2_{p:ts}$ is the variance attributable to patient nested in therapist by session, σ^2_{tr} is the variance attributable to the therapist by rater interaction, σ^2_{pr}

is the variance attributable to the patient by rater interaction, $\sigma^2_{(p:t)sr,e}$ is the variance attributable to error variance (i.e., not attributed to other factors in the design, the residual).

The therapist-level generalizability coefficients (GC) were calculated as shown in the following formula:

$$GC_t = \frac{\sigma^2_t}{\sigma^2_t + \frac{\sigma^2_{ts}}{n_s} + \frac{\sigma^2_{p:t}}{n_p} + \frac{\sigma^2_{p:ts}}{n_p n_s} + \frac{\sigma^2_{tr}}{n_r} + \frac{\sigma^2_{pr}}{n_p n_r} + \frac{\sigma^2_{(p:t)sr,e}}{n_p n_s n_r}}$$

where GC_t is the generalizability coefficient on the therapist level, σ^2_t is the variance attributable to therapists, σ^2_{ts} is the variance attributable to the therapist by session interaction, $\sigma^2_{p:t}$ is the variance attributable to patients nested in therapists, n_s is the number of sessions, n_p is the number of patients, n_r is the number of raters, $\sigma^2_{p:ts}$ is the variance attributable to patient nested in therapist by session, σ^2_{tr} is the variance attributable to the therapist by rater interaction, σ^2_{pr} is the variance attributable to the patient by rater interaction, $\sigma^2_{(p:t)sr,e}$ is the variance attributable to error variance (i.e., not attributed to other factors in the design, the residual).

Results

Generalizability Coefficients Measuring Dependability in Present Sample

Generalizability coefficients indexing patient-level dependability of scores were calculated using the actual numbers of sessions rated in the present study. With this number of sessions (ranging from 1.6 to 1.9, on average, for the three treatments), all conditions had inadequate GCs (SE: adherence GC = .44, competence GC = .60; CT: adherence GC = .64, competence GC = .66; IDC: adherence GC = .48, competence GC = .59).

Generalizability coefficients were also calculated to index therapist-level dependability using the actual number of patients included in the current study. All coefficients were marginally acceptable for CT and IDC condition (CT judges: adherence GC = .77 and competence GC = .75; IDC judges: adherence GC = .71 and competence GC = .77). For SE therapy, the coefficients were lower (SE judges: adherence GC = .58, competence GC = .65).

Variance components (expressed as percent of total variance) for each of the terms in the statistical models are presented in Table II. As can be seen, beyond the residual variation, the largest source of variation that impacts both patient and therapist GCs was consistently the Patient by Session interaction term, reflecting the fact that adherence and competence ratings varied more over sessions for some patients than for others.

Estimated Minimum Number of Sessions Needed to Yield Adequate Dependability of Adherence and Competence Assessments at the Patient Level

Having completed a “G study” that examined sources of variability and precision of measurement using the conditions inherent in the data collection (i.e., actual numbers of sessions and therapists), we then proceeded to conduct a “D study” that estimated generalizability coefficients using hypothetical numbers of sessions and therapists. Generalizability coefficients were calculated for various numbers of sessions (ranging from one to 14) per patient on the patient level. The numbers of the patients per therapist in these calculations were the actual numbers used in the study (the number of patients per therapist did not influence the GC on the patient level). Results indicated that generalizability coefficients increased steadily with increasing numbers of sessions per patient until reaching

an asymptote. For adherence, the largest increment happened in the first several sessions, and after five to 10 sessions the curve flattens out (Figure 1). The same tendency but even stronger was observed for competence ratings, and after five sessions the curve flattens out (Figure 2).

The number of sessions per patients needed to achieve a GC $\geq .80$ for adherence ratings at the patient level varied depending on the treatment modality. For IDC, 10 sessions were estimated to be needed, while only five sessions were estimated to be needed for CT; for SE, nine sessions were needed. Similar numbers of sessions were needed across the three treatments to achieve a patient-level GC $\geq .80$ on competence ratings (SE: six sessions; CT: five sessions; IDC: seven sessions).

Estimated Minimum Number of Sessions and Patients Needed to Yield Adequate Dependability of Adherence and Competence Assessments at the Therapist Level

Generalizability coefficients were calculated for various numbers of sessions (1–6 sessions) together with various numbers of patients (1–14 patients) on the therapist level. Results showed that most of the increase of the GC occurs in the first numbers of sessions or patients (Tables III and IV). To achieve a $\geq .80$ GC at the therapist level, the numbers of patients per therapist needed for dependable adherence and competence ratings, respectively, were SE, 13 and > 14 ; CT, 4 and 5; IDC, 8 and 7. If fewer than six sessions per patient are rated, even more patients per therapist are needed to achieve a therapist-level GC $\geq .80$.

Discussion

In order to answer the question how many sessions per patient, and patients per therapist, are needed to achieve dependable adherence and competence ratings in the treatment of cocaine dependence, a generalizability theory analysis was performed in a large sample of cocaine-dependent patients randomized to receive supportive expressive therapy, cognitive therapy, or individual drug counseling. The most important conclusion was that more sessions (for adherence: five in CT; nine in SE, 10 in IDC; for competence: five to seven across the treatments) than is typically sampled (two and a half) in most studies of other psychotherapies are needed to achieve dependable generalizability coefficients on the patient level when measuring therapists' adherence and competence. On the therapist level, the number of patients per therapist sufficient to achieve good ($\geq .80$) dependability of measurement varied by treatment modality, with CT needing the fewest (four or five) and SE the most (13 or 14).

Using a more liberal $\geq .70$ or greater cut-off for adequate dependability of measurement at the patient level, five sessions for SE and IDC, and three sessions for CT, would be needed for adherence ratings, and three or four sessions for competence ratings. Similarly, at the therapist level, fewer patients per therapist (three for CT; nine or 10 for SE; five for IDC) are needed to achieve a $\geq .70$ level of dependability for adherence/competence ratings (assuming six sessions are rated).

The literature review presented in the introduction of this article showed that, across 24 studies, there were typically two and a half sessions per patient and eight patients per therapist rated for adherence/competence. However, only three of these studies evaluated five or more sessions per patient (Huppert, Barlow, Gorman, Shear, & Woods, 2006; Ogrodniczuk & Piper, 1999; Shaw et al., 1999), and only 11 studies used seven or more patients per therapist (Barber, Crits-Christoph, & Luborsky, 1996; Barber et al., 2006, 2008; Carroll, Nich, & Rounsaville, 1997; Feeley et al., 1999; Gaston, Thompson, Gallagher, Cournoyer, & Gagnon, 1998; Hogue et al., 2008; Loeb et al., 2005; Luborsky, McLellan,

Woody, O'Brien, & Auerbach, 1985; Newman et al., 2011; Ogrodniczuk et al., 1999). Thus, assuming the results found here are generalizable to other scales, treatments, and patient populations, many studies in the literature probably evaluated an inadequate number of sessions and/or patients per therapist to create highly dependable adherence/competence scores at the patient and therapist levels, respectively. Inadequate dependability of adherence/competence scores may in part explain why many studies have failed to find associations between adherence/competence ratings and treatment outcome (Webb et al., 2010).

In the only previous study that reports the number of sessions needed to evaluate therapist competence, it was estimated that 17 videotaped sessions are needed to achieve a .80 GC in regard to cognitive therapy competence (Keen & Freeston, 2008). However, in addition to not focusing on a specific treatment manual or patient population, this study allowed therapists to choose two sessions to be evaluated, which may have created a positive bias in competence ratings compared to our study, in which therapists did not choose the sessions to be rated. Moreover, the design of the Keen and Freeston (2008) study did not allow for examination of patient-level and therapist-level GCs. These differences aside, it is relevant to note that Keen and Freeston (2008) also concluded that there needs to be a marked increase in the number of samples of clinical work assessed to be able to make reliable judgments of therapist competence in delivering cognitive behavior therapy.

For clinical dissemination of an evidence-based treatment approach, financial constraints make it unlikely that therapists would be evaluated on, for example, as many as five sessions for each of seven patients (35 sessions rated in total) to arrive at a decision as to whether or not a therapist is adequately skilled in a treatment approach. The two dissemination studies we found with variables of interests investigated two and three sessions per patient, respectively (Glisson et al., 2010; Weisz et al., 2009). Furthermore, our results were in line with the Imel et al. (2011) study, which suggested that variability in adherence/competence scores within the therapists' caseload can be large. The larger the variation of adherence and competence within the therapist, the more measurements are needed to obtain reliable measurements. The current results therefore suggest caution in interpreting previous efforts at measuring treatment adherence/competence where the number of sessions or patients is low, and provide some suggestions for future studies, or clinical applications, in which therapist-level adherence/competence evaluations are planned. Our results also suggest that for both clinical trials and dissemination studies, investigators should evaluate the dependability of their adherence/competence ratings to ensure that these assessments are adequate to draw conclusions about adherence/competence at the level of the dyad or therapist. In the certification of therapists during dissemination of an evidence-based treatment, if higher levels of dependability (e.g., >80) cannot be achieved due to financial constraints in rating sessions/patients, then it should be acknowledged that false positives (certifying a therapist as competent when they are not) and false negatives (not certifying a therapist as competent when they are) may be high.

Although it is important to have dependable measurements of adherence/competence, other factors may come into play when choosing which sessions are selected in a given study. For testing certain hypotheses, it may be more important to sample early-in-treatment sessions rather than later-in-treatment sessions. Adherence/competence in late sessions, for example, might be influenced by the degree of progress made by patients thus far in therapy, and therefore if the goal of a study is to predict outcome from adherence/competence ratings, late sessions might need to be avoided (so as not to introduce a confound due to patient progress influencing adherence/competence). However, regardless of which phase of therapy is targeted in research, dependability of measurement at the patient level, or therapist level, needs to be adequately addressed (e.g., by sampling multiple early sessions).

The issue of dependability of measurement is of course a broader issue than the focus on adherence/competence in the current manuscript. Other psychotherapy process variables probably suffer from the same problem of lack of adequate dependability at the patient level because of patient by session interaction effects, and lack of adequate dependability at the therapist level because of inadequate numbers of patients per therapist included in a study. Outcome measures also may be less than ideally dependable. For example, investigators typically measure symptoms on a given day. Such symptoms, however, are likely to fluctuate over the course of different days within a week. Assessment of outcome over multiple days and calculation of an aggregate outcome across days may yield greater precision of outcome measurement. Improvement in the precision of measurement will allow for greater accuracy in the estimation of effect sizes (e.g., correlations of process with outcome) and greater ability to detect small to moderate effects that would be rendered non-significant because of attenuation due to measurement error in the independent and/or dependent measures.

Several limitations of the current study should be mentioned. Other variables besides the number of sessions and patients might have impacted the generalizability coefficients. For example, the type of treatment, or the specific scale/construct for each treatment, may have influenced the level of generalizability coefficients found here. In the current study, scales for supportive expressive therapy had lower dependability than the other modalities. Although our study does not directly address whether the problem is the therapy or the particular scale, it should be noted that extensive development work was conducted on several of the scales (Barber & Crits-Christoph, 1996; Barber et al., 1997, 2003). Moreover, interjudge reliability and internal consistency reliability for the SE adherence and competence scales was generally very good, with the exception that interjudge reliability on the SE competence scale was marginal (.68). Nevertheless, it is possible that different scales for measuring adherence/competence for SE, CT, and IDC therapies for cocaine dependence might yield more dependable scores using fewer treatment sessions and patients than we estimate are needed based on the current data. Assuming, however, that the problem with low dependability for SE therapy is not a function of inferior adherence/competence scales, it may be that more sessions and patients are needed to create dependable scores for this treatment because of the more inferential nature of the key clinical concepts in this form of therapy. An adequate assessment of a therapist's performance in a psychodynamic therapy may require evaluation of the therapist's handling of a range of patient clinical material (e.g., defenses, resistances, repetitive interpersonal themes, transference phenomena) that only becomes evident over many sessions and different types of patients. Thus, researchers and clinical trainers should be aware of the possibility that adequate judgments of therapist adherence/competence in psychodynamic therapy may require sampling many sessions and patients, or may require further delineation of the construct and further scale development.

Another shortcoming with the current study was that the judges were not randomized to their different treatment modalities, and therefore not free from selection bias. The raters' ability to be objective can be influenced by the raters' characteristics, perceptions, and attitudes (e.g., Fiske, 1977). The present study limited this impact through the use of treatment manuals, training of the raters, the use of a protocol for rater procedures, and ongoing calibration between the raters. Another limitation of the current study is that adherence and intervention competence may be particularly difficult to evaluate in a substance-dependent population because of a high frequency of disruptive life events that interfere with implementation of specific interventions. The generalizability of the findings to other types of treatments, other patient populations, and other measures of adherence and intervention competence is unknown.

In summary, the present results suggest that researchers and clinical trainers need to give greater attention to the dependability of judgments about therapist adherence/competence. Dependability may vary depending on the specific treatment, and there may be a need to evaluate a larger number of sessions and patients than typically is done.

Acknowledgments

This study used data from the National Institute on Drug Abuse Collaborative Cocaine Treatment Study (NIDA CCTS), Philadelphia, USA. The study was supported by the Psychotherapy Research Center, Philadelphia, USA, the Department of Psychology, Umeå, Sweden, and the Swedish Council for Working life and Social Research, Stockholm, Sweden.

References

- Barber JP, Crits-Christoph P. Development of a therapist adherence / competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research*. 1996; 6:81–94.10.1080/10503309612331331608 [PubMed: 22242608]
- Barber JP, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*. 1996; 64:619–622.10.1037/0022-006X.64.3.619 [PubMed: 8698958]
- Barber JP, Foltz C, Crits-Christoph P, Chittams J. Therapist's adherence and competence and treatment discrimination in the NIDA Collaborative Cocaine Treatment Study. *Journal of Clinical Psychology*. 2004; 60:29–41.10.1002/jclp.10186 [PubMed: 14692007]
- Barber JP, Gallop R, Crits-Christoph P, Barrett MS, Klostermann S, McCarthy KS, Sharpless BA. The role of the alliance and techniques in predicting outcome of supportive-expressive dynamic therapy for cocaine dependence. *Psychoanalytic Psychology*. 2008; 25:461–482.10.1037/0736-9735.25.3.461
- Barber JP, Gallop R, Crits-Christoph P, Frank A, Thase ME, Weiss RD, Gibbons C. The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*. 2006; 16:229–240.10.1080/10503300500288951
- Barber JP, Krakauer I, Calvo N, Badgio PC, Faude J. Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. *Journal of Psychotherapy, Practice and Research*. 1997; 6:12–14. Retrieved from: <http://jppr.psychiatryonline.org/cgi/reprint/6/1/12>. [PubMed: 9058557]
- Barber JP, Liese B, Abrams M. Development of the cognitive therapy adherence and competence Scale. *Psychotherapy Research*. 2003; 13:205–221.10.1093/ptr/kpg019
- Barber JP, Mercer D, Krakauer I, Calvo N. Development of an adherence/ competence rating scale for individual drug counseling. *Drug and Alcohol Dependence*. 1996; 43:125–132.10.1016/S0376-8716(96)01305-1 [PubMed: 9023068]
- Beck, AT.; Wright, FD.; Newman, CF.; Liese, BS. *Cognitive therapy of substance abuse*. New York, NY: Guilford Press; 1993.
- Cardinet, J.; Johnson, J.; Pini, G. *Applying generalizability theory using EduG*. New York, NY: Routledge/Taylor & Francis; 2010.
- Carroll KM, Nich C, Rounsaville BJ. Contribution of the therapeutic alliance to outcome in active versus control psychotherapies. *Journal of Consulting and Clinical Psychology*. 1997; 65:510–514.10.1037/0022-006X.65.3.510 [PubMed: 9170775]
- Cicchetti DV, Sparrow SS. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*. 1981; 86:127–137. [PubMed: 7315877]
- Crits-Christoph P, Johnson J, Gallop R, Connolly Gibbons MB, Hamilton J, Tu X. A generalizability theory analysis of group process ratings in the treatment of cocaine dependence. *Psychotherapy Research*. 2011; 21:252–266.10.1080/10503307.2010.551429 [PubMed: 21409739]
- Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken LS, Moras K. The NIDA Collaborative Cocaine Treatment Study: rationale and methods. *Archives of General Psychiatry*.

- 1997; 54:721–726. Retrieved from <http://archpsyc.ama-assn.org/content/vol54/issue8/index.dtl>. [PubMed: 9283507]
- Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken LS, Beck AT. Psychosocial treatments for cocaine dependence. *Archives of General Psychiatry*. 1999; 56:493–502. Retrieved from <http://archpsyc.ama-assn.org/cgi/content/full/56/6/493>. [PubMed: 10359461]
- Cronbach L, Rajaratnam N, Gleser GC. Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*. 1963; 16:137–163.
- DeRubeis RJ, Feeley M. Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research*. 1990; 14:469–482.
- Downing S. Reliability: on the reproducibility of assessment data. *Medical Education*. 2004; 38:1006–1012.10.1111/j.1365-2929.2004.01932.x [PubMed: 15327684]
- Ekers DM, Lovell K, Playle JF. The use of CBT based, brief, facilitated selfhelp interventions in primary care mental health service provision: Evaluation of a 10-day training programme. *Clinical Effectiveness in Nursing*. 2006; 9:88–96.10.1016/j.cein.2006.06.001
- Feeley M, DeRubeis RJ, Gelfand L. The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*. 1999; 67:578–582.10.1037/0022-006X.67.4.578 [PubMed: 10450629]
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. *Structured Clinical Interview for Axis I DSM–IV Disorders—Patient Edition*. Washington DC: American Psychiatric Press; 1994.
- Fiske, DW. Methodological issues in research on the psychotherapist. In: Gurman, AS.; Razin, AM., editors. *Effective psychotherapy*. New York, NY: Pergamon Press; 1977. p. 23–43.
- Gaston L, Ring JM. Preliminary results on the inventory of therapeutic strategies. *Journal of Psychotherapy Practice and Research*. 1992; 1:135–146. Retrieved from <http://jppr.psychiatryonline.org/cgi/content/abstract/1/2/135>. [PubMed: 22700091]
- Gaston L, Thompson L, Gallagher D, Cournoyer L, Gagnon R. Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. *Psychotherapy Research*. 1998; 8:190–209.10.1080/10503309812331332307
- Glisson C, Schoenwald SK, Hemmelgarn A, Green P, Dukes D, Armstrong KS, Chapman JE. Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *Journal of Consulting and Clinical Psychology*. 2010; 78:537–550.10.1037/a0019160 [PubMed: 20658810]
- Hogue A, Henderson CE, Dauber S, Barajas PS, Fried A, Liddle HA. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*. 2008; 76:544–555.10.1037/0022-006X.76.4.544 [PubMed: 18665684]
- Horowitz MJ, Marmar CR, Weiss DS, DeWitt KN, Rosenbaum R. Brief psychotherapy of bereavement reactions: The relationship of process to outcome. *Archives of General Psychiatry*. 1984; 41:438–448. Retrieved from <http://archpsyc.ama-assn.org/cgi/reprint/41/5/438>. [PubMed: 6721669]
- Huppert JD, Barlow DH, Gorman JM, Shear MK, Woods SW. The interaction of motivation and therapist adherence predict outcome in cognitive behavioral therapy for panic disorder: Preliminary findings. *Cognitive and Behavioral Practice*. 2006; 13:198–204.10.1016/j.cbpra.2005.10.001
- Imel ZE, Baer JS, Martino S, Ball SA, Carroll KM. Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence*. 2011; 115:229–236.10.1016/j.drugalcdep.2010.11.010 [PubMed: 21227601]
- Kaslow NJ. Competencies in professional psychology. *American Psychologist*. 2004; 59:774–781.10.1037/0003-066X.59.8.774 [PubMed: 15554846]
- Keen AJA, Freeston MH. Assessing competence in cognitive-behavioural therapy. *The British Journal of Psychiatry*. 2008; 193:60–64.10.1192/bjp.bp.107.038588 [PubMed: 18700221]
- Loeb KL, Wilson GT, Labouvie E, Pratt EM, Hayaki J, Walsh BT, Fairburn CG. Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: A study of process and outcome. *Journal of Consulting and Clinical Psychology*. 2005; 73:1097–1107.10.1037/0022-006X.73.6.1097 [PubMed: 16392983]

- Luborsky, L. Principles of psychoanalytic psychotherapy: A manual for Supportive-Expressive treatment. New York, NY: Basic Books; 1984.
- Luborsky L, McLellan AT, Woody GE, O'Brien CP, Auerbach A. Therapist success and its determinants. *Archives of General Psychiatry*. 1985; 42:602–661. Retrieved from <http://archpsyc.ama-assn.org/cgi/reprint/42/6/602>. [PubMed: 4004503]
- Manassis K, Mendlowitz SL, Scapillato D, Avery D, Fiksenbaum L, Freire M, Owens MD. Group and individual cognitive-behavioral therapy for childhood anxiety disorders: a randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2002; 41:1423–1430.10.1097/00004583-200212000-00013 [PubMed: 12447028]
- Mannix KA, Blackburn IM, Garland A, Gracie J, Moorey S, Reid B, et al. Effectiveness of brief training in cognitive behaviour therapy techniques for palliative care practitioners. *Palliative Medicine*. 2006; 20:579–584.10.1177/0269216306071058 [PubMed: 17060250]
- Mark, D.; Luborsky, L. A manual for the use of Supportive-Expressive psychotherapy in the treatment of cocaine abuse. Philadelphia, PA: Department of Psychiatry, University of Pennsylvania; 1992.
- Mercer, D.; Woody, G. Unpublished manuscript. Veterans Affairs Medical Center, University of Pennsylvania; PA: 1992. Addiction counseling.
- Mohr DC, Boudewyn AC, Goodkin DE, Bostrom A, Epstein L. Comparative outcomes for individual cognitive-behavior therapy, supportive-expressive group psychotherapy, and sertraline for the treatment of depression in multiple sclerosis. *Journal of Consulting and Clinical Psychology*. 2001; 69:942–949.10.1037/0022-006X.69.6.942 [PubMed: 11777121]
- Newman MG, Castonguay LG, Borkovec TD, Fisher AJ, Boswell JF, Szkodny LE, Nordberg SS. A randomized controlled trial of cognitive-behavioral therapy for generalized anxiety disorder with integrated techniques from emotion-focused and interpersonal therapies. *Journal of Consulting and Clinical Psychology*. 2011; 79:171–181.10.1037/a0022489 [PubMed: 21443321]
- Ogrodniczuk JS, Piper WE. Measuring therapist technique in psychodynamic psychotherapies, development, and use of a new scale. *Journal of Psychotherapy Practice and Research*. 1999; 8:142–154. Retrieved from <http://jppr.psychiatryonline.org/cgi/content/abstract/8/2/142>. [PubMed: 10079461]
- O'Malley SS, Foley SH, Rounsaville BJ, Watkins JT, Sotsky SM, Imber SD, Elkin I. Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology*. 1988; 56:496–501.10.1037/0022-006X.56.4.496 [PubMed: 3198805]
- Paivio SC, Holowaty KAM, Hall IE. The influence of therapist adherence and competence on client reprocessing of child abuse memories. *Psychotherapy*. 2004; 41:56–68.10.1037/0033-3204.41.1.56
- Sachs JS. Negative factors in brief psychotherapy: An empirical assessment. *Journal of Consulting and Clinical Psychology*. 1983; 51:557–564.10.1037/0022-006X.51.4.557 [PubMed: 6619363]
- Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, Carroll KM. We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*. 2005; 73:106–115.10.1037/0022-006X.73.1.106 [PubMed: 15709837]
- Sharpless BA, Barber JP. A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review*. 2009; 29:47–56.10.1016/j.cpr.2008.09.008 [PubMed: 18952334]
- Shavelson, R.J.; Webb, N.M. Measurement methods for the social sciences series. Vol. 1. Thousand Oaks, CA: Sage; 1991. Generalizability theory: A primer.
- Shaw BF, Elkin I, Yamaguchi J, Olmsted M, Vallis TM, Lowery A, Imber SD. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*. 1999; 67:837–846. Retrieved from <http://spider.apa.org/ftdocs/ccp/1999/december/ccp676837.html>. [PubMed: 10596506]
- Strauman TJ, Vieth AZ, Merrill Ka, Kolden GG, Woods TE, Klein MH, Kwapil L. Self-system therapy as an intervention for self-regulatory dysfunction in depression: a randomized comparison with cognitive therapy. *Journal of Consulting and Clinical psychology*. 2006; 74:367–376.10.1037/0022-006X.74.2.367 [PubMed: 16649881]

- Strunk DR, Brotman MA, DeRubeis RJ, Hollon SD. Therapist competence in cognitive therapy for depression: predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*. 2010; 78:429–437.10.1037/a0019631 [PubMed: 20515218]
- Svartberg M, Stiles TC. Predicting patient change from therapist competence and patient–therapist complementarity in shortterm anxiety-provoking psychotherapy: A pilot study. *Journal of Consulting and Clinical Psychology*. 1992; 60:304–307.10.1037/0022-006X.60.2.304 [PubMed: 1592962]
- Trepka C, Rees A, Shapiro DA, Hardy GE, Barkham M. Therapist competence and outcome of cognitive therapy for depression. *Cognitive Therapy and Research*. 2004; 28:143–157.10.1023/B:COTR.0000021536.39173.66
- Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*. 1993; 61:620–630.10.1037/0022-006X.61.4.620 [PubMed: 8370857]
- Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001; 357:945–949. [PubMed: 11289364]
- Wasserman RH, Levy K, Loken E. Generalizability theory in psychotherapy research: The impact of multiple sources of variance on the dependability of psychotherapy process ratings. *Psychotherapy Research*. 2009; 19:397–408.10.1080/10503300802579156 [PubMed: 19235094]
- Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical psychology*. 2010; 78:200–211.10.1037/a0018912 [PubMed: 20350031]
- Webb, NM.; Shavelson, RJ.; Haertel, EH. Reliability coefficients and generalizability theory. In: Rao, CR.; Sinharay, S., editors. *Handbook of statistics*. Vol. 26. Amsterdam: Elsevier; 2006. p. 1-124.
- Weisz JR, Southam-Gerow MA, Gordis EB, Connor-Smith JK, Chu BC, Langer DA, Weiss B. Cognitive-behavioral therapy versus usual clinical care for youth depression: an initial test of transportability to community clinics and clinicians. *Journal of Consulting & Clinical Psychology*. 2009; 77:383–396.10.1037/a0013877 [PubMed: 19485581]
- Wood JJ, Piacentini JC, Southam-Gerow M, Chu BC, Sigman M. Family cognitive behavioral therapy for child anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45:314–321.10.1097/01.chi.0000196425.88341.b0 [PubMed: 16540816]

Figure 1. Patient-level generalizability coefficients for various numbers of sessions at which the adherence is assessed.

Figure 2. Patient-level generalizability coefficients for various numbers of sessions at which the competence is assessed.

Table 1

Number of sessions per patient and number of patients per therapist for 24 studies of adherence/competence

Study	n	Individual, family or group therapy	Diagnosis	Treatments	Number of sessions per patient	Number of patients per therapist
Barber et al., 1996	29	Ind.	Major depressive disorder	SE	1	7.2
Barber et al., 2006	95	Ind.	Cocaine dependence	IDC	1	8
Barber et al., 2008	108	Ind.	Cocaine dependence	SE	1.2	8.3
Carroll et al., 1997	121	Ind.	Cocaine dependence	CBT, desipramine and CM	2	11
Feeley et al., 1999	32	Ind.	Major depressive disorder	CT vs. CT and imipramine	3	8
Gaston & Ring, 1992	16	Ind.	Major depressive disorder	BDT vs. CT	3	4
Gaston et al., 1998	91	Ind.	Major depressive disorder	BT vs. CT vs. BDT	1.5	9.1
Hogue et al., 2008	136	Ind. vs. family	Adolescent substance use	CBT vs. MDFT	3.2	15.5
Horowitz et al., 1984	52	Ind.	Grief reactions	BDT	4	5.8
Huppert et al., 2006	205	Ind.	Panic disorder	CBT	5.8	4.0
Loeb et al., 2005	81	Ind.	Bulimia nervosa	CBT vs. IPT	3	10.1
Luborsky et al., 1985	110	Ind.	Drug use disorder	CBT, IDC, vs. SE	3	12.3
Manassis et al., 2002	78	Ind. vs. group	Childhood anxiety disorder	CBT	1	unknown
Mohr et al., 2001	63	Ind. vs. group	Major depressive disorder and multiple sclerosis	CBT vs. SEG	1	4.7
Newman et al., 2011	83	Ind.	General anxiety disorder	CBT vs. CBT and interpersonal	2.5	28
Ogrodniczuk & Piper 1999	50	Ind.	Mixed	BDT vs. ST	1	2.8
Ogrodniczuk & Piper 1999	50	Ind.	Mixed	BDT vs. ST	1	6.2
Ogrodniczuk & Piper 1999	144	Ind.	Mixed	BDT vs. ST	9	18
O'Malley et al., 1988	35	Ind.	Major depressive disorder	IPT	1	3.2
Paivio et al., 2004	37	Ind.	Experienced childhood abuse	EFT	2.5	3.7
Sachs, 1983					3	2.6
Shaw et al., 1999	40	Ind.	Major depressive disorder		8.4	5
Strauman et al., 2006	45	Ind.	Major depressive disorder or dysthymic disorder	CT vs. SST	1.4	5
Svartberg & Stiles, 1992	15	Ind.	Mainly anxiety diagnosis	STAPP	1	1.9
Trepka et al., 2004	30	Ind.	Major depressive disorder	CT	1	5
Wood et al., 2006	40	Ind. vs. family	Childhood anxiety disorder	CBT	2	4.4
Glisson et al., 2010	615	Family	Mixed	MST	3	Unknown

Study	<i>n</i>	Individual, family or group therapy	Diagnosis	Treatments	Number of sessions per patient	Number of patients per therapist
Weisz et al., 2009	57	Mainly ind.	Youth depression	CBT vs UC	1.6	1

Note. BT, behavioral therapy; BDT, brief dynamic therapy; CBT, cognitive behavior therapy; CM, clinical management; a therapy control condition; CT, cognitive therapy; EFT, emotion-focused therapy; IDC, individual drug counseling; IPT, interpersonal therapy; MDFT, multidimensional family therapy; MST, multi-systemic therapy; SE, supportive expressive therapy; SEG, supportive expressive group therapy; SST, self system therapy; ST, supportive therapy; STAPP, short-term anxiety-provoking psychotherapy; US, usual care.

Table II

Proportion of variance explained by model terms

Model term									
Scale	Patient	Therapist	Session	Rater	Therapist by session	Patient by session	Patient by rater	Therapist by rater	Residual
Supportive-expressive therapy									
Adherence	.17	.09	.00	.01	.00	.44	.00	.01	.27
Competence	.16	.12	.10	.04	.00	.14	.00	.06	.38
Cognitive therapy									
Adherence	.10	.26	.00	.22	.06	.21	.02	.02	.12
Competence	.14	.29	.01	.01	.05	.20	.04	.05	.21
Individual drug counseling									
Adherence	.12	.17	.00	.01	.01	.36	.00	.03	.31
Competence	.10	.13	.00	.36	.00	.13	.02	.02	.24

Note. Terms in bold are sources of variation that can reduce the size of patient and therapist level generalizability coefficients.

Table III

Generalizability coefficients at the therapist level for adherence ratings by independent judges across various numbers of sessions and patients.

	Number of sessions					
	1	2	3	4	5	6
Number of patients per therapist						
Supportive expressive therapy						
1	0.12	0.17	0.21	0.23	0.25	0.26
2	0.21	0.29	0.34	0.37	0.40	0.41
3	0.28	0.38	0.44	0.47	0.49	0.51
4	0.34	0.45	0.51	0.54	0.56	0.58
5	0.39	0.51	0.56	0.59	0.61	0.63
6	0.43	0.55	0.60	0.63	0.65	0.67
7	0.47	0.59	0.64	0.67	0.68	0.70
8	0.50	0.62	0.67	0.69	0.71	0.72
9	0.53	0.64	0.69	0.72	0.73	0.74
10	0.56	0.66	0.71	0.73	0.75	0.76
11	0.58	0.68	0.73	0.75	0.77	0.78
12	0.60	0.70	0.74	0.77	0.78	0.79
13	0.62	0.72	0.76	0.78	0.79	0.80
14	0.63	0.73	0.77	0.79	0.80	0.81
Cognitive therapy						
1	0.36	0.48	0.53	0.56	0.59	0.60
2	0.50	0.62	0.67	0.70	0.72	0.73
3	0.57	0.68	0.73	0.76	0.78	0.79
4	0.61	0.72	0.77	0.79	0.81	0.82
5	0.64	0.75	0.79	0.82	0.83	0.84
6	0.66	0.77	0.81	0.83	0.85	0.86
7	0.68	0.78	0.82	0.84	0.86	0.87
8	0.69	0.79	0.83	0.85	0.87	0.87
9	0.70	0.80	0.84	0.86	0.87	0.88
10	0.71	0.80	0.84	0.86	0.88	0.89

	Number of sessions					
	1	2	3	4	5	6
Number of patients per therapist						
11	0.71	0.81	0.85	0.87	0.88	0.89
12	0.72	0.81	0.85	0.87	0.89	0.89
13	0.72	0.82	0.86	0.88	0.89	0.90
14	0.73	0.82	0.86	0.88	0.89	0.90
Individual drug counseling						
1	0.21	0.30	0.36	0.39	0.42	0.44
2	0.33	0.45	0.51	0.55	0.57	0.59
3	0.42	0.54	0.60	0.63	0.65	0.67
4	0.48	0.60	0.65	0.68	0.70	0.71
5	0.53	0.64	0.69	0.72	0.73	0.75
6	0.57	0.67	0.72	0.74	0.76	0.77
7	0.60	0.70	0.74	0.76	0.78	0.78
8	0.62	0.72	0.75	0.78	0.79	0.80
9	0.64	0.73	0.77	0.79	0.80	0.81
10	0.66	0.74	0.78	0.80	0.81	0.82
11	0.67	0.76	0.79	0.81	0.82	0.82
12	0.69	0.77	0.80	0.81	0.82	0.83
13	0.70	0.77	0.80	0.82	0.83	0.84
14	0.71	0.78	0.81	0.82	0.83	0.84

Note. Generalizability coefficients .80 are in bold.

Table IV

Generalizability coefficients at the therapist level for competence ratings by independent judges across various numbers of sessions and patients

		Number of sessions					
		1	2	3	4	5	6
Number of patients per therapist							
Supportive expressive therapy							
1	0.21	0.27	0.30	0.32	0.33	0.34	
2	0.33	0.41	0.45	0.47	0.48	0.49	
3	0.42	0.50	0.53	0.55	0.56	0.57	
4	0.48	0.55	0.58	0.60	0.61	0.62	
5	0.52	0.60	0.62	0.64	0.65	0.65	
6	0.56	0.63	0.65	0.67	0.67	0.68	
7	0.59	0.65	0.67	0.69	0.69	0.70	
8	0.61	0.67	0.69	0.70	0.71	0.72	
9	0.63	0.69	0.71	0.72	0.72	0.73	
10	0.65	0.70	0.72	0.73	0.73	0.74	
11	0.66	0.71	0.73	0.74	0.74	0.75	
12	0.68	0.72	0.74	0.75	0.75	0.76	
13	0.69	0.73	0.75	0.75	0.76	0.76	
14	0.70	0.74	0.75	0.76	0.76	0.77	
Cognitive therapy							
1	0.35	0.45	0.49	0.52	0.53	0.54	
2	0.49	0.59	0.63	0.65	0.67	0.68	
3	0.57	0.66	0.70	0.72	0.73	0.74	
4	0.61	0.70	0.73	0.75	0.77	0.77	
5	0.64	0.73	0.76	0.78	0.79	0.80	
6	0.66	0.75	0.78	0.79	0.80	0.81	
7	0.68	0.76	0.79	0.81	0.82	0.82	
8	0.70	0.77	0.80	0.82	0.83	0.83	
9	0.71	0.78	0.81	0.82	0.83	0.84	
10	0.72	0.79	0.81	0.83	0.84	0.84	

	Number of sessions					
	1	2	3	4	5	6
Number of patients per therapist						
11	0.72	0.79	0.82	0.83	0.84	0.85
12	0.73	0.80	0.83	0.84	0.85	0.85
13	0.73	0.80	0.83	0.84	0.85	0.86
14	0.74	0.81	0.83	0.85	0.85	0.86
Individual drug counseling						
1	0.26	0.35	0.39	0.42	0.44	0.45
2	0.41	0.51	0.55	0.58	0.60	0.61
3	0.50	0.60	0.64	0.66	0.68	0.69
4	0.57	0.66	0.69	0.71	0.73	0.73
5	0.62	0.70	0.73	0.75	0.76	0.77
6	0.65	0.73	0.76	0.77	0.78	0.79
7	0.68	0.75	0.78	0.79	0.80	0.81
8	0.71	0.77	0.80	0.81	0.82	0.82
9	0.73	0.79	0.81	0.82	0.83	0.83
10	0.74	0.80	0.82	0.83	0.84	0.84
11	0.76	0.81	0.83	0.84	0.84	0.85
12	0.77	0.82	0.84	0.84	0.85	0.85
13	0.78	0.83	0.84	0.85	0.86	0.86
14	0.79	0.83	0.85	0.86	0.86	0.86

Note. Generalizability coefficients .80 are in bold.