Enhancing Treatment Outcome of Patients at Risk of Treatment Failure: Meta-Analytic and Mega-Analytic Review of a Psychothera....

Article *in* Journal of Consulting and Clinical Psychology · June 2010

DOI: 10.1037/a0019247 · Source: PubMed

3 authors, including:

Kenichi Shimokawa

7 PUBLICATIONS **771** CITATIONS

SEE PROFILE

Lambert Michael

Brigham Young University - Provo Main Cam…

274 PUBLICATIONS **15,775** CITATIONS

SEE PROFILE

Journal of Consulting and Clinical Psychology
2010, Vol. 78, No. 3, 298–311

0022-006X/10/$12.00 DOI: 10.1037/a0019247

Enhancing Treatment Outcome of Patients at Risk of Treatment Failure: Meta-Analytic and Mega-Analytic Review of a Psychotherapy Quality Assurance System

Kenichi Shimokawa, Michael J. Lambert, and David W. Smart
Brigham Young University

Objective: Outcome research has documented worsening among a minority of the patient population (5% to 10%). In this study, we conducted a meta-analytic and mega-analytic review of a psychotherapy quality assurance system intended to enhance outcomes in patients at risk of treatment failure. ***Method:*** Original data from six major studies conducted at a large university counseling center and a hospital outpatient setting (N = 6,151, mean age = 23.3 years, female = 63.2%, Caucasian = 85%) were reanalyzed to examine the effects of progress feedback on patient outcome. In this quality assurance system, the Outcome Questionnaire–45 was routinely administered to patients to monitor their therapeutic progress and was utilized as part of an early alert system to identify patients at risk of treatment failure. Patient progress feedback based on this alert system was provided to clinicians so that they could intervene before treatment failure occurred. Meta-analytic and mega-analytic approaches were applied in intent-to-treat and efficacy analyses of the effects of feedback interventions. ***Results:*** Three forms of feedback interventions—integral elements of this quality assurance system—were effective in enhancing treatment outcome, especially for signal alarm patients. Two of the three feedback interventions were also effective in preventing treatment failure (clinical support tools and the provision of patient progress feedback to therapists). ***Conclusions:*** The current state of evidence appears to support the efficacy and effectiveness of feedback interventions in enhancing treatment outcome.

Keywords: treatment outcomes, treatment failure, patient deterioration, feedback, psychotherapy quality assurance

Supplemental materials: http://dx.doi.org/10.1037/a0019247.supp

In this era of accountability, health care systems, including mental health care systems, have been placed under tremendous pressure to demonstrate the effectiveness of their services in bringing about improved patient outcomes (Lambert, Bergin, & Garfield, 2004; Reed & Eisman, 2006). Despite substantial evidence that psychotherapy is generally effective (Lambert & Ogles, 2004), the outcome literature has also documented the phenomenon of patients leaving treatment worse off than when entering treatment (e.g., Lambert & Ogles, 2004; Mohr, 1995). Hansen, Lambert, and Forman (2002) found that deterioration occurred in both clinical trials and in routine care, with 5% to 10% of adult clients having a negative outcome. The situation is even worse for children, who have negative outcome rates between 10% and 20% (Warren et al., in press). Although a causal relationship cannot be easily drawn between psychotherapy and negative outcome, the sheer number of patients whose quality of life worsens despite receiving psychotherapy should be alarming to a profession that seeks to alleviate the suffering of, or at least do no harm to, its service consumers. Failing cases also have serious economic implications for patients and third-party payers, who paid for or reimbursed ineffective treatments. These serious implications make prevention of treatment failure an important goal of quality assurance systems.

Although much effort has been invested in the past few decades to demonstrate effectiveness of psychotherapy in the form of comparative treatment studies, often referred to as *empirically supported treatments* (e.g., American Psychological Association, 2006), the limitations of such research and misuse of its evidence have been extensively debated (e.g., Reed & Eisman, 2006). Alternative evidence-based methods have also been proposed. An example of such an alternative is patient-focused research, which advocates systematic evaluation of patient response to treatment throughout the course of therapy (Howard, Moras, Brill, Martinovich, & Lutz, 1996). The advocates of this approach recommend providing feedback on patients' progress to therapists. Such feedback allows therapists to make treatment decisions based on changes in patient distress, rather than merely offering fixed-

Kenichi Shimokawa and Michael J. Lambert, Department of Psychology, Brigham Young University; David W. Smart, Counseling and Career Center, Brigham Young University.

This study was based on a doctoral dissertation conducted by Kenichi Shimokawa under the guidance of Michael J. Lambert in partial fulfillment of a doctorate in clinical psychology at Brigham Young University. This research was supported in part by an endowed chair research grant to Michael J. Lambert from Brigham Young University. Michael J. Lambert is a part owner of OQ Measures, which is the copyright holder of the Outcome Questionnaire–45 and OQ–Analyst used in the studies reviewed. We acknowledge Scott A. Baldwin at the Department of Psychology, Brigham Young University, for his advice in the statistical analyses.

Correspondence concerning this article should be addressed to Michael J. Lambert, Department of Psychology, Brigham Young University, 272 TLRB, Provo, UT 84602. E-mail: michael_lambert@byu.edu

length, evidence-based treatment protocols. Therapists can then make treatment decisions based on the formally measured treatment response of individual patients, rather than trusting treatments to have certain positive effects. Patient-focused research relies heavily on feedback of patient-reported functioning to providers, along with identification of patients who are at-risk for treatment failure. The basic rationale behind the concept of providing feedback to clinicians is straightforward. Therapists can be more responsive to patient needs if they know that the patient is not succeeding as intended. As has been repeatedly demonstrated in clinical research, any prediction relying on statistical or actuarial methods tends to fare better than clinical judgment alone (Ægisdóttir et al., 2006; Grove, 2005). This notion is especially salient when clinicians are making predictions about treatment failures. For example, Hannan et al. (2005) asked 40 therapists to predict which of their (550) patients would deteriorate. The therapists identified only one of the 40 cases who eventually deteriorated. Hatfield, McCullough, Plucinski, and Krieger (2009) found that only 32% of therapists recorded patient worsening in their case notes, despite dramatic escalation in their symptoms in the week prior to meeting with their therapist.

The psychological community has increasingly recognized the importance of providing feedback to clinicians regarding their patients' progress. For instance, the American Psychological Association (2006) noted that one of the "most pressing research needs" (p. 278) in evidence-based practice in psychology includes the very type of research we present here—that is "providing clinicians with real-time patient feedback to benchmark progress in treatment and clinical support tools to adjust treatment as needed" (p. 278).

A quality assurance system based on a feedback model was developed by Lambert and colleagues (e.g., Lambert, Hansen, & Finch, 2001). This feedback model is based on the routine administration of the Outcome Questionnaire–45 (OQ-45; Lambert, Morton, et al., 2004). Given the limitations of clinical judgment and the advantages of actuarial-based prediction making, the quality assurance system was established on the following three major principles: (a) development of a reasonable estimate of expected progress of the average patient; (b) the data driven process of comparing the progress of individual patients with expected progress to identify patients who are at risk of experiencing a negative outcome (Finch, Lambert, & Schaalje, 2001; Lambert, Whipple, Bishop, et al., 2002; Spielmans, Masters, & Lambert, 2006); and (c) provision of patient progress feedback to the therapist (and case managers when applicable) to adjust treatment as necessary.

Six major studies have been conducted to evaluate the effects of providing feedback about patient progress (Harmon et al., 2007; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Whipple et al., 2003). In this meta-analysis, data from these clinical trials are combined and analyzed to better understand the effects of the various procedures examined in these studies. The interventions examined in these studies range from providing therapists with progress feedback to supplying them with problem-solving tools (clinical support tools, CST) for identifying the causes of deterioration and making suggestions for resolution of identified problems. The CST intervention relies on assessment of therapeutic alliance, patient motivation, and social support with corresponding recommendations for effective actions. It is important to note that all six studies assigned patients to treatment conditions within therapists (i.e., therapists offered treatment as usual (TAU) as well as the experimental interventions).

To summarize the six feedback studies, several acronyms are used to identify their main features. As patients entered treatment and subsequently participated in the past feedback studies, they were assigned to one of the following conditions: Therapists received OQ-45-based patient progress feedback (Fb); both therapists and patients received feedback (T/P Fb); and therapists received no feedback (TAU). As treatment continued, patients divided themselves into two groups on the basis of their treatment progress as measured by the OQ-45. Patients whose progress negatively deviated from the expected course of progress (i.e., signal alarm cases) were classified as not-on-track (NOT) cases. Patients who progressed as expected were classified as on-track (OT) cases. In three of the feedback studies, of those patients in the Fb and T/P Fb conditions and later classified as NOT, half the patients were given the additional intervention—CST. Accordingly, the groups of patients were termed CST Fb or P/T CST Fb. Because the CST Fb and P/T CST Fb groups were indistinguishable in treatment effects in the past studies, these two groups are aggregated as the CST Fb group in this study.

All but one of the feedback studies were conducted at a large university counseling center. The first feedback study (Study 1; Lambert, Whipple, et al., 2001) randomly assigned patients into the Fb group or TAU group and found statistically significant effects of the Fb intervention in keeping NOT patients in treatment longer and improving the outcome of the same patients in relation to TAU. The second feedback study (Study 2; Lambert, Whipple, Vermeersch, 2002) also demonstrated the outcome-enhancing effect of the Fb intervention for NOT patients. It should be noted that Study 2 assigned patients into treatment conditions on the basis of the semester in which they sought treatment. To further enhance the outcome of NOT patients, the third feedback study (Study 3; Whipple et al., 2003) used the CST Fb, Fb, and TAU conditions and found a superior outcome for patients in the CST Fb group when compared with those in the Fb or TAU groups. Consistent with findings from Studies 1 and 2, the Fb group had superior outcome over the TAU group. It should be pointed out that therapists decided the assignment of NOT patients into the CST Fb intervention rather than such patients being randomly assigned.

The fourth study (Hawkins et al., 2004) was conducted in a hospital outpatient setting and randomly assigned patients into one of the three conditions: P/T Fb, Fb, or TAU. Unlike the previous studies, treatment effects were not separately tested on the NOT and OT patients. Thus, the overall feedback effects were reported on the combined NOT and OT samples, showing improved outcome for those in the P/T Fb over the Fb group and for those in the Fb over the TAU group. The fifth study (Harmon et al., 2007) incorporated random assignment to feedback conditions (P/T Fb, Fb, or TAU) and random assignment of NOT patients to CST Fb or No CST Fb (near equivalent of Fb) conditions. This study replicated the outcome-enhancing effects of the CST Fb and Fb groups as reported in previous studies. This study did not replicate, however, the effects of the P/T Fb intervention found by Hawkins et al. (2004).

One methodological limitation of Study 5 and Study 6 (Slade et al., 2008) should be noted. On the basis of the outcome-enhancing effects of providing progress feedback to therapists, the counseling center in which the series of feedback studies took place adopted routine administration of OQ progress feedback as part of their standard of care at the same time Study 5 commenced. This policy change prevented implementation of the TAU condition in Studies 5 and 6, thus making direct comparisons between the TAU condition and various experimental conditions no longer available. Study 6 was the first study to implement the OQ–Analyst, a computer software that provided immediate, electronic progress feedback. Study 6 also made changes to the CST measures and essentially replicated the design of Study 5, except for the use of OQ–Analyst. The effects of CST Fb and Fb on the outcome were replicated, but the effects of P/T Fb were not. After the completion of Study 3, Lambert et al. (2003) conducted a small meta-analysis of the first three studies. Overall, the results suggested decreased rate of deterioration and increased rate of improvement among NOT patients in the Fb group when compared with TAU.

Now that three more major studies have been completed with further developments in the feedback system and the effects of the P/T Fb and CST Fb interventions have not been summarized across studies, conducting another research synthesis study appeared appropriate. As repeatedly demonstrated in the previous feedback studies, OQ feedback interventions appear to be effective in enhancing outcome for NOT patients, while having little impact on OT cases. Thus, the primary purpose of this meta-analysis was to investigate the effects of various OQ feedback interventions on the outcomes of patients whose progress was identified as NOT. Although subtle differences existed in the operationalization of feedback interventions across studies, given similarities in methodologies, all of the feedback interventions were grouped into one of the following:

- *NOT Fb:* NOT patients whose OQ progress feedback was provided to their therapists only.[1]
- *NOT P/T Fb:* NOT patients whose OQ progress feedback was provided directly to both patients and therapists.[2]
- *CST Fb:* NOT patients whose OQ progress feedback and CSTs were provided to their therapists.[3]
- *NOT TAU:* NOT patients whose therapists received no feedback intervention.[4]
- *OT Fb:* OT patients whose OQ progress feedback was provided to their therapists only.[5]
- *OT P/T Fb:* OT patients whose OQ progress feedback was provided directly to both patients and therapists.[6]
- *OT TAU:* OT patients whose therapists received no feedback intervention.[7]

Selection Criteria and Participants

All of the six OQ feedback studies published to date were included in this analysis. Each study's demographic variables, mean OQ total score at pretreatment, and *n* and percentage of patients identified as NOT cases are reported in Table 1.

The statistical methods used in the previous feedback studies reflect two distinct approaches: effectiveness analysis based on the intent-to-treat (ITT) principle and efficacy analysis (Atkins, 2009; Lachin, 2000). These approaches reflect two distinct philosophies in terms of the interpretation of their results. The former addresses the overall effect of a treatment at the population level, regardless of various treatment compliance issues that may arise in naturalistic clinical settings. This method essentially includes the data of all patients solely on the basis of the initial assignment to treatment conditions. The latter approach addresses the effect of a given treatment on a subset of patients who met certain compliance criteria to be considered completers of the treatment regimen. The studies, which examined the effects of the Fb intervention against the TAU used the effectiveness analyses. Alternatively, two of the three studies that used the CST Fb condition applied post hoc screening criteria to analyze a subset of patients who completed the prescribed feedback interventions. Given these differences in analytical approaches, we evaluated each feedback treatment under both approaches, using the original data sets of all six studies included in this study.

In the ITT analyses, all participants in the CST Fb, NOT P/T Fb, NOT Fb, NOT TAU, OT P/T Fb, OT Fb, and OT TAU groups were included. These analyses provide the most conservative estimates of the treatment effects because they even incorporate the data of individuals whose posttreatment scores are missing, including the data of those with only the intake and warning OQ scores. Patients with only one data point were grouped within the OT groups. To obtain conservative estimates of these patients' posttest scores, we carried their last observed data point (or their only data point) forward and treated it as their posttest score, utilizing the last observation carried forward method. The breakdown of the number of participants in each treatment condition across all six studies was as follows: NOT Fb (n = 427), NOT P/T Fb (n = 222), CST Fb (n = 415), NOT TAU (n = 318), OT Fb (n = 2,390), OT P/T Fb (n = 935), and OT TAU (n = 1,444).

In the efficacy analyses, we retrospectively defined the inclusion criteria that represented the least necessary condition in which the effects of the OQ feedback interventions (i.e., Fb, P/T Fb, and CST Fb) could be measured.[8] For the analyses of CST Fb interventions,

[1] Studies 1, 2, 3, and 4 included the NOT Fb condition.

[2] Studies 4, 5, and 6 included NOT P/T Fb condition.

[3] Studies 3, 5, and 6 utilized CST Fb interventions. Because of study designs, Studies 5 and 6 used variations of CST Fb groups: CST Fb group, P/T CST Fb group, a 1-week delayed CST Fb group, and a 2-weeks delayed CST Fb group. Because of statistically nonsignificant findings between the CST groups, we combined them as the CST Fb group in this meta-analysis.

[4] Studies 1, 2, 3, and 4 included the NOT TAU condition.

[5] All six studies included the OT Fb condition.

[6] Studies 4, 5, and 6 included the OT P/T Fb condition.

[7] Studies 1, 2, 3, and 4 included the OT TAU condition.

[8] The efficacy sample inclusion criteria for the NOT Fb and NOT P/T Fb groups were defined as follows: attended at least five sessions (for Studies 1, 2, 3, 4, and 5) or four sessions (for Study 6 because of electronic immediate progress feedback), completed the OQ in at least three sessions, and had the last recorded OQ score come from at least two sessions (for Studies 1, 2, 3, 4, and 5) or one session (for Study 6) after the patient was identified as a NOT case. The efficacy sample inclusion criteria for the OT Fb and OT P/T Fb groups were set more loosely than their NOT counterparts, given that a majority of OT patients left treatment before the effects of feedback treatments could be measured (i.e., nearly 70% attended four or fewer sessions). Accordingly, the OT Fb and P/T Fb criteria were defined as the following: attended at least two sessions and filled out the OQ in at least two of the sessions attended.

Table 1
Characteristics of Clients From Studies Used in the Meta-Analyses and Mega-Analyses

Study	Clients/therapists[a] (*N*)	Age		Female (%)	Caucasian (%)	Dosage		Intake OQ-45		NOT	
		M	*SD*			*M*	*SD*	*M*	*SD*	*n*	%
Lambert, Whipple, Smart, et al. (2001)	609/36	22.23	3.92	70.0	87.4	4.68	3.89	69.23	23.20	66	10.8
Lambert, Whipple, Vermeersch, et al. (2002)	1,422/56	22.37	3.74	66.7	85.0	4.49	3.39	69.87	22.58	240	16.9
Whipple et al. (2003)	1,339/49	23.01	3.56	63.5	86.0	5.14	4.80	69.27	23.37	278	20.8
Hawkins et al. (2004)	306/5	30.51	10.77	63.1	94.1	6.06	6.45	83.23	23.74	101	33.0
Harmon et al. (2007)	1,374/72	22.68	3.68	61.0	83.0	6.74	6.44	71.23	22.61	369	26.9
Slade et al. (2008)	1,101/73	24.25	3.29	57.5	82.7	5.81	5.67	71.50	22.07	328	29.8

Note. OQ-45 = Outcome Questionnaire–45; NOT = clients whose progress was identified by OQ-45 algorithms as being not on track.
[a] Numbers of clients and therapists prior to applying any exclusion criteria. Thus, the numbers reported here do not match those reported in the original articles for studies that used exclusion criteria (i.e., Lambert et al., 2002; Hawkins et al., 2004; and Whipple et al., 2003).

we used the exclusion criteria as defined in the original articles in Studies 5 and 6 (Harmon et al., 2007; Slade et al., 2008). Study 3 was the first study to implement the CST intervention; however, this study did not use exclusion criteria similar to those applied in Studies 5 and 6. Thus, we retrospectively defined and applied the minimum inclusion criteria required for a given patient to be considered a completer of the CST intervention in Study 3.[9] Through the application of the aforementioned inclusion criteria, the following number and percentage of participants were included in each treatment condition when aggregated across studies: NOT Fb, $n = 263$ (61.6%); NOT P/T Fb, $n = 177$ (79.7%); CST Fb, $n = 217$ (52.2%); OT Fb, $n = 1{,}651$ (69.0%); OT P/T Fb, $n = 777$ (83.1%).

Dependent Measures and Computation of Effect Sizes

The effects of OQ feedback interventions were compared on the following dependent measures: mean posttreatment OQ total score, the odds of patients achieving clinically significant improvement at posttreatment, and the odds of the occurrence of clinically significant worsening (or deterioration) at posttreatment. Mean number of sessions attended by patients in each condition was also compared for the ITT analyses but not for the efficacy analyses, because different numbers of sessions attended by patients were part of the inclusion criteria.

Following recommendations by Overton (1998) and Hedges and Vevea (1998), we used a random-effects model, given that the research on providing feedback for the purpose of enhancing treatment outcome is relatively new in psychotherapy outcome literature, that studies included in this meta-analysis contained slight variations in research designs, and that the purpose of the present study was to investigate the applicability of our findings to a broader clinical context. Hedges's (1981) standardized mean difference *g* was used as the unit of effect size for mean posttreatment OQ total score comparisons and mean number of sessions attended by patients between feedback groups and control groups. Formulas for obtaining Hedges's *g* are provided in Appendix A. Random weights were then assigned to individual standardized mean differences to obtain the estimated weighted mean effect size per comparison. Formulas for calculating random weights and estimated weighted mean effect sizes (or combined effect sizes) are presented in Appendix B. Because lower OQ scores indicate lower levels of distress, negative effect sizes in posttreatment OQ total scores comparisons signify superior outcome of the treatment condition in question. Although the P/T Fb and CST Fb groups were not directly compared against the TAU condition in some of the studies, we considered that such comparisons would afford a more intuitive interpretation of the effects of feedback interventions in relation to the TAU condition. Thus, we conducted mega-analyses on the pooled data set from all of the six feedback studies to calculate the effect sizes of feedback interventions (i.e., P/T Fb and CST Fb) in relation to TAU. Such an approach with large *n* provides an alternative method to traditional meta-analysis in research synthesis (e.g., DeRubeis, Gelfand, Tang, & Simons, 1999; Serretti, Cusin, Rausch, Bondy, & Smeraldi, 2006).

Possible heterogeneity of effect sizes and publication biases were tested. Given the small number of studies included in this study, mega-analytic approaches were used to test for the homogeneity of effect sizes. To test for heterogeneity of effect sizes in mean posttreatment OQ score differences, separate analyses of covariance (ANCOVAs) were performed for each pooled treatment group, with study as the factor, posttreatment OQ total score as the dependent variable, and pretreatment OQ total score as the covariate. To test for equivalence in pretreatment distress level across groups, we conducted an independent

[9] To identify a given patient as a NOT case, administer the CST intervention, and measure the effects of the CST intervention in Study 3, the patient needed to have attended at least six sessions (three of which occurred after the patient was identified as a NOT case) and to have filled out the OQ in at least three of the sessions. Application of these inclusion criteria, nonetheless, does not guarantee the inclusion of only those who completed the CST intervention. For instance, NOT patients who attended more than the required number of sessions, but did not complete the OQ after the administration of the CST intervention would still be included in the analysis despite lacking the posttreatment score.

samples *t* test for each between-group comparison. To test for heterogeneity of effect sizes in mean number of sessions attended, we conducted one-way analyses of variance (ANOVAs) for each pooled treatment group, with number of sessions attended as the dependent variable and study as the factor. Classic fail-safe *N* test (Rosenthal, 1979), Orwin's (1983) fail-safe *N* test, and Duval and Tweedie's (2000) trim and fill were performed to address possible publication biases.

Another set of treatment outcomes investigated in this study was differences in proportions and odds of patient outcome classification based on clinical significance indices. The use of clinical significance indices based on a clinical cutoff score and reliable change index methods proposed by Jacobson and Truax (1991) is one of the hallmarks of the OQ-45-based quality assurance system. As demonstrated by Beckstead et al. (2003) and Lunnen and Ogles (1998), the OQ-45-based clinical significance classification of patient outcome appears to reflect meaningful change as well as the functional/dysfunctional state of patients. In this quantitative review, the clinical significance status for each patient at termination was classified in one of the three categories: deterioration/reliable worsening, no change, or clinically significant improvement.

The results are presented in three ways. First, *n* and percentage of patients in each of the three clinical significance categories for each feedback intervention group across all six studies were aggregated and reported. Second, the odds of the occurrence of deteriorated/reliably worsened cases were compared for each feedback intervention group against its control group (i.e., TAU or Fb groups, depending on the comparisons being made) in the unit of odds ratio. Third, the odds of the occurrences of clinically significant improvement were similarly compared for each feedback intervention against its control. To expedite the statistical calculations, Comprehensive Meta-Analysis (Version 2) was utilized in the calculation of effect sizes.

Analyses of Effect Sizes

Effects of Feedback Interventions on Posttreatment OQ Total Score in NOT Patients

The combined effect size and the results of tests of publication bias for each of the comparisons presented next are summarized in Table 2.

Fb effect. The results of a one-way ANCOVA, testing for heterogeneity of effects across studies, with study as the factor, posttreatment OQ total score as the dependent variable, and pretreatment OQ total score as the covariate, indicated no significant study effect among the Fb group in ITT analysis, $F(5, 420) = 0.221$, $p = .951$, or efficacy analysis, $F(5, 250) = 1.49$, $p = .192$. However, statistically significant study effect was found for the TAU group, $F(3, 313) = 2.79$, $p = .041$. The result of the independent samples *t* test of pretreatment mean OQ scores between pooled Fb and TAU groups was not significant in ITT analysis, $t(743) = -0.28$, $p = .778$, or efficacy analysis, $t(579) = -0.48$, $p = .631$, indicating that Fb and TAU were comparable at pretest distress level. Thus, despite the heterogeneity among the TAU groups in mean posttest scores, given equivalent pretreatment OQ scores across groups, we deemed it appropriate to proceed with aggregating the TAU data in favor of ecological validity. ITT meta-analysis indicated that effect sizes of individual studies comparing the NOT Fb and NOT TAU groups ranged from $g = -0.42$, $p < .001$, 95% CI [−0.68, −0.17], to $g = 0.08$, $p = .742$, 95% CI [−0.41, 0.58] (see Table 1 in the supplementary materials for a complete list of individual effect sizes and the forest plot). The aggregate effect size was statistically significant at the .05 level, $g = -0.28$, $p = .003$, 95% CI [−0.47, −0.10]—equivalent of a 6.4 OQ total score difference on average. When the efficacy sample inclusion criteria were applied to the same comparison groups, the results showed greater treatment effect favoring the Fb intervention. As shown in Table 2 in the supplementary materials,

Table 2
Meta-Analysis and Mega-Analysis of Effects of Feedback Interventions on Mean Posttest OQ-45 Total Scale Score

Analysis type and feedback condition	*k*	Group 1/Group 2 *N*	ES [95% CI]	Classic fail-safe *N*	Orwin's fail-safe *N*	Trim-and-fill ES (studies trimmed)
Intent-to-treat analysis						
CST Fb vs. Fb	3	415/246	−0.16* [−0.33, −0.002]	0	0	−0.16 (0)
P/T Fb vs. Fb	3	222/188	−0.16 [−0.36, 0.03]	0	0	−0.16 (0)
Fb vs. TAU	4	269/318	−0.28** [−0.47, −0.10]	6	2	−0.28 (0)
CST Fb vs. TAU[a]	—	415/318	−0.44*** [−0.59, −0.30]	—	—	—
P/T Fb vs. TAU[b]	—	222/318	−0.36*** [−0.54, −0.19]	—	—	—
Efficacy analysis						
CST Fb vs. Fb	3	181/169	−0.19 [−0.43, 0.05]	0	1	−0.19 (0)
P/T Fb vs. Fb	3	177/147	−0.16 [−0.37, 0.06]	0	0	−0.11 (1)
Fb vs. TAU	4	136/318	−0.53*** [−0.78, −0.28]	20	8	−0.67 (2)
CST Fb vs. TAU[a]	—	217/318	−0.70*** [−0.88, −0.52]	—	—	—
P/T Fb vs. TAU[b]	—	177/318	−0.55*** [−0.73, −0.36]	—	—	—

Note. Negative effect sizes indicate lower distress level. Dashes indicate that values are not applicable because a given analysis was based on mega-analysis. OQ-45 = Outcome Questionnaire–45; *k* = number of studies; ES = weighted effect size (Hedges's *g*; random effect model); CI = confidence interval; classic fail-safe *N* = the number of null studies needed to bring the combined *p* value to above .05 (two-tailed); Orwin's fail-safe *N* = the number of studies (with null mean Hedges's *g*) needed to bring the combined effect size (fixed model) to above −0.2; CST = clinical support tools; Fb = feedback; P/T = patient/therapist; TAU = treatment as usual.

[a] Mega-analysis with pooled CST Fb group versus pooled TAU group. [b] Mega-analysis with pooled P/T Fb group versus pooled TAU group.

* $p < .05$. ** $p < .01$. *** $p < .001$.

effect sizes for individual studies ranged from $g = -0.78$, $p < .001$, 95% CI [−1.11, −0.45], to $g = -0.18$, $p = .523$, 95% CI [−0.73, 0.37]. The aggregate effect size was significant at the .05 level, $g = -0.53$, $p < .001$, 95% CI [−0.78, −0.28], which equates to 12.0 OQ total points difference on average.

P/T Fb effect. Although the ideal evaluation would have been to compare all of the feedback interventions against the TAU group, the last two feedback studies (Studies 5 and 6) containing the P/T Fb groups did not have TAU groups, as explained earlier. Thus, P/T Fb groups were compared against Fb groups, where Fb groups were used as the benchmark to evaluate incremental benefits of the P/T Fb intervention. The results of a one-way ANCOVA to test the heterogeneity of effects showed that study effect did not reach statistical significance for the P/T Fb groups in ITT analysis, $F(2, 218) = 1.58$, $p = .208$, or efficacy analysis, $F(2, 173) = 1.62$, $p = .201$. As presented in Table 3 in the supplementary materials, ITT analyses of NOT P/T Fb versus NOT Fb indicated none of the individual effect sizes were significant at the .05 level, with individual effect sizes ranging from $g = -0.44$, $p = .071$, 95% CI [−0.92, 0.04], to $g = -0.10$, $p = .526$, 95% CI [−0.39, 0.20]. The aggregate effect size also did not reach statistical significance at the .05 level, $g = -0.16$, $p = .099$, 95% CI [−0.36, 0.03]. As presented in Table 4 in the supplemental materials, when the efficacy criteria were applied, individual effect sizes ranged from $g = -0.39$, $p = .177$, 95% CI [−0.96, 0.18], to $g = -0.06$, $p = .734$, 95% CI [−0.40, 0.28]. The aggregated effect size was similar to that of the ITT analysis, $g = 0.16$, $p = .163$, 95% CI [−0.37, 0.06]. These results suggest that, in terms of treatment outcome at termination, providing progress feedback to both clinicians and patients adds no significant incremental benefit to providing progress feedback only to clinicians (who may or may not share it with patients).

Pretreatment mean OQ total score comparison between the pooled P/T Fb and pooled TAU groups did not reach statistical significance for either ITT analysis, $t(538) = 0.65$, $p = .518$, or efficacy analysis, $t(493) = 0.24$, $p = .810$, indicating that the two groups did not differ significantly in their initial level of disturbance. ITT posttreatment score difference was significant at the .05 level, $g = -0.36$, $p < .001$, 95% CI [−0.54, −0.19], equivalent to 7.9 points difference in mean OQ total scores. Efficacy posttreatment score difference was also significant, $g = -0.55$, $p < .001$, 95% CI [−0.73, −0.36], equivalent to mean OQ total score difference of 11.7 points. These results suggest that NOT patients in the P/T Fb condition experience greater therapeutic gain as measured by the OQ-45 at termination than those in TAU. Such therapeutic benefits are more pronounced among those who stayed in treatment long enough to experience the benefits of P/T Fb intervention.

CST Fb effect. As in the case of P/T Fb analyses, two of the three studies that tested the effects of the CST interventions (Studies 5 and 6) did not use the TAU condition. Thus, the CST Fb groups were also compared with the Fb groups to estimate their incremental clinical benefits over the Fb condition. The results of one-way ANCOVAs to test for heterogeneity of effects for the CST Fb group did not reach statistical significance in ITT analysis, $F(2, 411) = 1.20$, $p = .137$, or efficacy analysis, $F(2, 213) = 0.48$, $p = .617$. As presented in Table 5 in the supplementary materials, the ITT analysis indicated that individual effect sizes ranged from $g = -0.23$, $p = .094$, 95% CI [−0.49, 0.04], to $g = -0.11$, $p = .415$, 95% CI [−0.38, 0.16]. The combined effect size was significant at the .05 level, $g = -0.16$, $p = .048$, 95% CI [−0.33, −0.002], indicating approximately 3.6 OQ total points difference on average, favoring the CST Fb group. When the efficacy criteria were applied to both the CST Fb and Fb groups, the combined effect size was $g = -0.19$, $p = .113$, 95% CI [−0.43, 0.05], the equivalent of approximately 4.2 OQ total points difference on average. Individual effect sizes ranged from $g = -0.32$, $p = .053$, 95% CI [−0.65, 0.01], to $g = -0.11$, $p = .606$, 95% CI [−0.32, 0.55] (see Table 6 in the supplementary materials). It should be pointed out that, contrary to the outcome comparison between the CST Fb group and the Fb group reported in Study 3 (Whipple et al., 2003), which reported results favoring the CST Fb group, application of the efficacy criteria in this study yielded a result favoring the Fb group. Although the two groups appeared demographically similar at pretreatment, given that random assignment of NOT patients to CST Fb and Fb groups was not used in this study, such contradictory findings may have been due to unknown artifacts resulting from therapists' selection of patients into treatment conditions. When Study 3 was removed from the efficacy analysis, the aggregate effect size of the CST Fb group over the Fb group improved to $g = -0.29$, $p = .013$, 95% CI [−0.52, −0.06], the equivalent of approximately 6.2 OQ points difference on average. These results suggest that, on average, those NOT patients who receive the CST intervention in routine care in addition to the Fb intervention experience small additional therapeutic gains represented in about 3 to 4 OQ points reduction over those who receive only progress feedback intervention. Those who stay in treatment long enough to experience the benefit of the CST intervention experience, on average, further distress reduction over those who experience the benefit of the Fb intervention alone. More studies with random-assignment-based comparison between the CST Fb and Fb conditions may help researchers better estimate the effect of the CST Fb intervention.

Pretreatment mean OQ total score comparison between the pooled CST Fb and pooled TAU groups did not reach statistical significance for either ITT analysis, $t(731) = -0.34$, $p = .732$, or efficacy analysis, $t(533) = 0.73$, $p = .468$, indicating the two groups were comparable at pretreatment. ITT analysis indicated that posttreatment score difference between CST Fb group and NOT TAU was significant at the .05 level, $g = -0.44$, $p < .001$, 95% CI [−0.59, −0.30], the equivalent of 9.5 points difference in mean OQ total scores. Efficacy posttreatment score difference was also significant, $g = -0.70$, $p < .001$, 95% CI [−0.88, −0.52], the equivalent of a mean OQ total score difference of 14.6 points. These results suggest NOT patients who receive the CST Fb intervention experience, on average, significantly more therapeutic gain than those in the TAU condition. Such therapeutic gain is more pronounced among who stay in treatment long enough to see the benefit of the CST Fb intervention.

Effects of Feedback Interventions on Clinical Significance

The *n* and percentage of the clinical significance classification of patient outcome at termination were aggregated by each treatment condition and are presented in Table 3. The summary of combined effects for the odds of deterioration/reliable worsening and the results of tests of publication bias are presented in Table 4.

Table 3
Clinical Significance Classification of Not-on-Track Patients by Treatment Conditions

Treatment condition and clinical significance	CST Fb	NOT P/T Fb	NOT Fb	NOT TAU	OT P/T Fb	OT Fb	OT TAU
Intent-to-treat sample							
Worsened/deteriorated	47 (11.3%)	35 (15.8%)	58 (13.6%)	64 (20.1%)	20 (2.1%)	45 (1.9%)	43 (3.0%)
No change	212 (51.1%)	101 (45.5%)	237 (55.5%)	183 (57.5%)	507 (54.2%)	1485 (62.1%)	940 (65.1%)
Improved/recovered	156 (37.6%)	86 (38.7%)	132 (30.9%)	71 (22.3%)	408 (43.6%)	860 (36.0%)	461 (31.9%)
Efficacy sample							
Worsened/deteriorated	12 (5.5%)	26 (14.7%)	24 (9.1%)	64 (20.1%)	20 (2.6%)	40 (2.4%)	43 (3.0%)
No change	91 (41.9%)	71 (40.1%)	140 (53.2%)	183 (57.5%)	349 (44.9%)	794 (48.1%)	940 (65.1%)
Improved/recovered	114 (52.5%)	80 (45.2%)	99 (37.6%)	71 (22.3%)	408 (52.5%)	817 (49.5%)	461 (31.9%)

Note. CST = clinical support tools; Fb = feedback; NOT = not on track; P/T = patient/therapist; TAU = treatment as usual; OT = on track.

The summary of combined effects for the odds of clinically significant improvement and the results of tests of publication bias are presented in Table 5.

Fb effect. When the odds of patient deterioration/reliable worsening at termination of the NOT Fb group were compared against NOT TAU, the results of ITT analyses indicated that the combined effect was significant at the .05 level, $OR = 0.62$, $p = .040$, 95% CI [0.40, 0.98], with effect sizes of individual studies ranging from $OR = 0.21$, $p = .063$, 95% CI [0.04, 1.09], to $OR = 0.72$, $p = .315$, 95% CI [0.39, 1.32] (see Table 7 in the supplementary materials). When the efficacy criteria were applied to the Fb group, the combined odds of deterioration for the Fb group decreased to $OR = 0.44$, $p = .015$, 95% CI [0.23, 0.85], with odds ratios of individual studies ranging from $OR = 0.21$, $p = .041$, 95% CI [0.05, 0.94], to $OR = 0.60$, $p = .238$, 95% CI [0.25, 1.41] (see Table 8 in the supplementary materials). These results suggest that the odds of deterioration among NOT patients in TAU are approximately 1.5 times higher than the odds for those who received the Fb intervention in routine practice. The results further suggest that the odds of deterioration among TAU are about 2.3 times higher than the odds for those who had stayed in treatment long enough to receive the benefit of the Fb intervention. When the odds of patients achieving clinically significant improvement at termination were compared between the NOT Fb group and NOT TAU, the results indicated significantly increased odds at the .05 level favoring the Fb group, $OR = 1.70$, $p = .005$, 95% CI [1.17, 2.46], with individual effect sizes ranging from $OR = 1.44$, $p = .539$, 95% CI [0.45, 4.65], to $OR = 2.17$, $p = .012$, 95% CI [1.19, 3.97] (see Table 9 in the supplementary materials). When the efficacy criteria were applied, the combined odds ratio of the occurrence of clinically significant improvement among the NOT Fb group against the NOT TAU group was $OR = 2.55$, $p < .001$, 95% CI [1.64, 3.98], with odds ratios of individual studies ranging from $OR = 1.23$, $p = .766$, 95% CI [0.32, 4.67], to $OR = 2.97$, $p = .003$, 95% CI [1.44, 6.11] (see Table 10 in the supplementary materials). These results suggest clinical benefit of the Fb intervention in reducing the occurrence of

Table 4
Meta-Analysis and Mega-Analysis of Effects of Feedback Interventions on Treatment Outcome: Combined Odds Ratio of Reliable Worsening/Deterioration

Analysis type and feedback condition	*k*	*OR* [95% CI]	Classic fail-safe *N*	Orwin's fail-safe *N*	Trim-and-fill ES (studies trimmed)
Intent-to-treat analysis					
CST Fb vs. Fb	3	0.76 [0.46, 1.26]	0	0	0.76 (0)
P/T Fb vs. Fb	3	1.35 [0.76, 2.41]	0	0	1.35 (0)
Fb vs. TAU	4	0.62* [0.40, 0.98]	3	1	0.70 (2)
CST Fb vs. TAU[a]	—	0.51** [0.34, 0.76]	—	—	—
P/T Fb vs. TAU[b]	—	0.74 [0.47, 1.17]	—	—	—
Efficacy analysis					
CST Fb vs. Fb	3	0.66 [0.29, 1.52]	0	1	0.83 (2)
P/T Fb vs. Fb	3	1.89 [0.90, 3.96]	0	1	2.95 (2)
Fb vs. TAU	4	0.44* [0.23, 0.85]	3	4	0.58 (2)
CST Fb vs. TAU[a]	—	0.23*** [0.12, 0.44]	—	—	—
P/T Fb vs. TAU[b]	—	0.68 [0.42, 1.13]	—	—	—

Note. Dashes indicate that values are not applicable because a given analysis was based on mega-analysis. *k* = number of studies; *OR* = combined odds ratio (random effect model); CI = confidence interval; classic fail-safe *N* = the number of null studies needed to bring the combined *p* value to above .05 (two-tailed); Orwin's fail-safe *N* = the number of studies (with an odds ratio of 1.00) needed to bring the combined odds ratio (fixed model) to above 0.66; ES = effect size. Odds ratios smaller than 1.00 indicate lower odds of client deterioration. CST = clinical support tools; Fb = feedback; P/T = patient/therapist; TAU = treatment as usual.
[a] Mega-analysis with pooled CST Fb group versus pooled TAU group. [b] Mega-analysis with pooled P/T Fb group versus pooled TAU group.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
Meta-Analysis and Mega-Analysis Effects of Feedback Interventions on Treatment Outcome: Combined Odds Ratio of Clinically Significant Improvement

Feedback condition	*k*	*OR* [95% CI]	Classic fail-safe *N*	Orwin's fail-safe *N*	Trim-and-fill ES (studies trimmed)
Intent-to-treat analysis					
CST Fb vs. Fb	3	1.53* [1.08, 2.18]	2	1	1.40 (1)
P/T Fb vs. Fb	3	1.44 [0.95, 2.19]	0	0	1.44 (0)
Fb vs. TAU	4	1.70** [1.17, 2.46]	3	2	1.72 (1)
CST Fb vs. TAU[a]	—	2.01*** [1.51, 2.92]	—	—	—
P/T Fb vs. TAU[b]	—	2.20*** [1.51, 3.21]	—	—	—
Efficacy analysis					
CST Fb vs. Fb	3	1.83 [0.89, 3.76]	4	2	1.83 (0)
P/T Fb vs. Fb	3	1.38 [0.88, 2.18]	0	1	1.25 (2)
Fb vs. TAU	4	2.55*** [1.64, 3.98]	11	6	2.33 (1)
CST Fb vs. TAU[a]	—	3.85*** [2.65, 5.60]	—	—	—
P/T Fb vs. TAU[b]	—	2.97*** [1.93, 4.27]	—	—	—

Note. Dashes indicate that values are not applicable because a given analysis was based on mega-analysis. *k* = number of studies; *OR* = combined odds ratio (random effect model); CI = confidence interval; classic fail-safe *N* = the number of null studies needed to bring the combined *p* value to above .05 (two-tailed); Orwin's fail-safe *N* = the number of studies (with an odds ratio of 1.00) needed to bring the combined odds ratio (fixed model) to above 1.5; ES = effect size. Odds ratios greater than 1.00 indicate higher odds of client improvement. CST = clinical support tools; Fb = feedback; P/T = patient/therapist; TAU = treatment as usual.
[a] Mega-analysis with pooled CST Fb group versus pooled TAU group. [b] Mega-analysis with pooled P/T Fb group versus pooled TAU group.
* $p < .05$. ** $p < .01$. *** $p < .001$.

treatment failure while increasing the odds of patients experiencing clinically significant improvement.

P/T Fb effect. When the odds of the occurrence of deterioration/reliable worsening were compared between the NOT P/T Fb and NOT Fb groups, the results of ITT analyses indicated that effect sizes of individual studies ranged from *OR* = 1.00, *p* = 1.000, 95% CI [0.42, 2.38], to *OR* = 1.74, *p* = .184, 95% CI [0.77, 3.95] (see Table 11 in the supplementary materials). The combined effect size was not statistically significant, *OR* = 1.35, *p* = .306, 95% CI [0.758, 2.413]. When the efficacy criteria were applied, the combined odds ratio of deterioration cases increased for the P/T Fb condition, although the results did not reach the .05 significance level, *OR* = 1.89, *p* = .094, 95% CI [0.90, 3.96]. Individual effect sizes ranged from *OR* = 0.68, *p* = .788, 95% CI [0.04, 11.53], to *OR* = 2.95, *p* = .047, 95% CI [1.02, 8.54] (see Table 12 in the supplementary materials). Although statistical significance was not achieved, the results suggest a higher rate of deterioration among NOT patients in the P/T Fb condition than among those in the Fb condition.

ITT comparisons of the odds of clinically significant improvement yielded the combined effect of *OR* = 1.44, *p* = .086, 95% CI [0.95, 2.19], with individual effect sizes ranging from *OR* = 1.26, *p* = .495, 95% CI [0.65, 2.47], to *OR* = 1.94, *p* = .179, 95% CI [0.74, 5.10] (see Table 13 in the supplementary materials). The efficacy analyses indicated that the combined effect size was similar to that obtained from the ITT sample, *OR* = 1.38, *p* = .164, 95% CI [0.88, 2.18], with a similar range of individual study effect sizes, *OR* = 1.25, *p* = .521, 95% CI [0.63, 2.50], to *OR* = 1.56, *p* = 0.459, 95% CI [0.48, 5.00] (see Table 14 in the supplementary materials). Although statistical significance was not reached, the results suggest higher odds of clinically significant improvement among NOT patients in the P/T Fb condition than among those in the Fb condition. These results suggest that provision of direct progress feedback to NOT patients has potential clinical effects that may enhance outcome in some patients even beyond what can be achieved by provision of progress feedback to clinicians alone, although such feedback may have possible iatrogenic effects in some patients.

The odds of deterioration/reliable worsening between the pooled NOT P/T Fb group and pooled NOT TAU group did not reach statistical significance in ITT analysis, *OR* = 0.74, *p* = .199, 95% CI [0.47, 1.17], or efficacy analysis, *OR* = 0.68, *p* = .134, 95% CI [0.42, 1.13]. The odds of clinically significant improvement between the pooled NOT P/T Fb group and pooled NOT TAU were significant in both ITT analysis, *OR* = 2.20, *p* < .001, 95% CI [1.51, 3.21], and efficacy analysis, *OR* = 2.87, *p* < .001, 95% CI [1.93, 4.27]. These results suggest that the P/T Fb intervention, in comparison with TAU, does not decrease the odds of deterioration but increases the odds of improvement among NOT patients.

CST Fb effect. ITT comparisons between the CST Fb groups and the NOT Fb groups indicated that individual effect sizes of deterioration/reliable worsening ranged from *OR* = 0.59, *p* = .342, 95% CI [0.20, 1.76], to *OR* = 1.043, *p* = .916, 95% CI [0.48, 2.29] (see Table 15 in the supplementary materials) with the combined effect size of *OR* = 0.76, *p* = .288, 95% CI [0.46, 1.26]. The results of efficacy analyses indicated the combined effect was *OR* = 0.66, *p* = .329, 95% CI [0.29, 1.52], with individual effect sizes ranging from *OR* = 0.54, *p* = .356, 95% CI [0.15, 2.0], to *OR* = 0.83, *p* = .756, 95% CI [0.25, 2.74] (see Table 16 in the supplementary materials). When the odds of patients achieving clinically significant improvement were compared in the ITT analyses, the combined effect size was *OR* = 1.53, *p* = .016, 95% CI [1.08, 2.18], favoring the CST Fb, with individual study effect sizes ranging from *OR* = 1.22, *p* = .467, 95% CI [0.69, 2.184], to *OR* = 1.97, *p* = .050, 95% CI [1.00, 3.87] (see Table 17 in the supplementary materials). The results of the efficacy analyses of comparing the odds of patients achieving clinically significant improvement yielded the combined effect size of *OR* = 1.83, *p* = .098, 95% CI [0.89, 3.76], with individual effect sizes ranging from *OR* = 1.167, *p* = .729, 95% CI [0.487, 2.729], to *OR* =

3.610, $p < .001$, 95% CI [1.847, 7.057] (see Table 18 in the supplementary materials). These results suggest that the CST Fb, compared with the NOT Fb, increases the odds of patients achieving clinically significant improvement, but the odds of deterioration/reliable worsening do not seem to decrease, at least at a statistically significant level.

The odds of deterioration/reliable worsening between the pooled CST Fb group and the pooled NOT TAU group reached statistical significance in both ITT analysis, $OR = 0.51$, $p = .001$, 95% CI [0.34, 0.76], and efficacy analysis, $OR = 0.23$, $p < .001$, 95% CI [0.12, 0.44]. The odds of clinically significant improvement between the pooled CST Fb group and the pooled TAU were significant in both ITT analysis, $OR = 2.20$, $p < .001$, 95% CI [1.51, 3.21], and efficacy analysis, $OR = 2.87$, $p < .001$, 95% CI [1.93, 4.27]. The odds of clinically significant improvement between the same groups reached statistical significance in ITT analysis, $OR = 2.01$, $p < .001$, 95% CI [1.51, 2.92], and efficacy analysis, $OR = 3.85$, $p < .001$, 95% CI [2.65, 5.60]. These results indicate that the odds of patients in TAU experiencing deterioration are approximately 2.0 times higher than those receiving the CST Fb in routine care settings (ITT). When comparing against those who complete the CST intervention, the odds of deterioration/reliable worsening among the TAU patients are approximately 4.3 times higher than the odds for those in the CST group. The results further indicate that the odds of patients in the CST Fb group achieving clinically significant improvement in routine care settings (ITT) are approximately 2.0 times higher than the odds for those in TAU. The odds of clinically significant improvement among those who complete the CST Fb intervention are about 3.9 times higher than the odds for those in the TAU.

Effects of Feedback Interventions on Session Attendance

The number of therapy sessions utilized by patients was thought of as an effect of feedback interventions in previous studies. Because the number of sessions attended by patients was part of the efficacy criteria, between-group comparisons of the mean number of sessions attended were appropriate only for the ITT analyses. The summary of effect sizes and the results of tests of publication bias are presented in Table 6.

Fb effect. A one-way ANOVA was conducted on the NOT Fb groups to test for heterogeneity of effect sizes, which resulted in a significant study effect, $F(5, 421) = 2.78$, $p = .017$. Although speculations could be made about the presence of possible moderators (e.g., treatment settings in which original studies took place), the data were pooled across studies, given that the difference between the highest mean attendance ($M = 10.8$) and lowest mean attendance ($M = 8.44$), as tested by an independent samples t test, was not significant, $t(116) = 1.93$, $p = .056$. The one-way ANOVA on the NOT TAU group yielded a significant study effect, $F(3, 314) = 6.55$, $p < .001$, with a significant difference between the highest mean attendance ($M = 11.22$; Study 4) and the lowest mean attendance ($M = 6.03$; Study 1), $t(61) = -2.563$, $p = .013$. In this case, given such a large discrepancy in mean session attendance, the presence of a moderator or moderators might have contributed to the heterogeneity. Study 1, in particular, was the only study resulting in a very large effect size, whereas the other studies yielded small effect sizes. The causes for such wide dispersion are not known at this time. Thus, although the data were pooled in favor of ecological validity in this study, future investigation of moderators appears warranted. The combined effect of the differences in mean session attendance between the NOT Fb and NOT TAU groups was $g = 0.27$, $p = .217$, 95% CI [−0.16, 0.69], with individual effect sizes ranging from –0.10, $p = .459$, 95% CI [−0.37, 0.17], to $g = 1.09$, $p < .001$, 95% CI [0.58, 1.60] (see Table 19 in the supplementary materials). The results did not show a statistically significant difference in the mean number of sessions utilized between the Fb and TAU groups.

P/T Fb effect. The results of a one-way ANOVA on the NOT P/T Fb groups did not support the presence of heterogeneity among mean session attendance across studies, $F(2, 219) = 2.67$, $p = .071$. The combined effect size of differences in mean session attendance between the NOT P/T Fb and NOT Fb groups was $g = 0.12$, $p = .311$, 95% CI [0.11, 0.35], with individual effect sizes ranging from $g = -0.22$, $p = .356$, 95% CI [−0.69, 0.25], to $g = 0.23$, $p = .145$, 95% CI [−0.08, 0.54] (see Table 20 in the

Table 6
Meta-Analysis of Effects of Feedback Interventions on Mean Number of Session Attendance

		Intent-to-treat analysis				
Feedback condition	*k*	Group1/Group 2 *N*	ES [95% CI]	Classic fail-safe *N*	Orwin's fail-safe *N*	Trim-and-fill ES (studies trimmed)
CST Fb vs. Fb	3	415/246	0.41* [0.05, 0.76]	14	3	0.41 (0)
P/T Fb vs. Fb	3	222/188	0.12 [−0.11, 0.35]	0	0	0.12 (0)
Fb vs. TAU	4	269/318	0.27 [−0.16, 0.70]	4	0	0.42 (1)
CST Fb vs. TAU[a]	—	415/318	0.48*** [0.33, 0.63]	—	—	—
P/T Fb vs. TAU[b]	—	222/318	0.40*** [0.23, 0.58]	—	—	—

Note. Dashes indicate that values are not applicable because a given analysis was based on mega-analysis. k = number of studies; ES = weighted effect size (Hedges's g; random effect model); CI = confidence interval; classic fail-safe N = the number of null studies needed to bring the combined p value to above .05 (two-tailed; based on fixed model); Orwin's fail-safe N = the number of studies (with null mean Hedges's g) needed to bring the combined effect size (fixed model) to below 0.2. Positive effect sizes indicate a greater number of sessions attended. CST = clinical support tools; Fb = feedback; P/T = patient/therapist; TAU = treatment as usual.

[a] Mega-analysis with pooled CST Fb group versus pooled TAU group. [b] Mega-analysis with pooled P/T Fb group versus pooled TAU group.

* $p < .05$. *** $p < .001$.

supplementary materials). The results did not support the presence of an increase in mean number of sessions in P/T Fb groups. The effect size of the pooled NOT P/T Fb group in relation to the pooled NOT TAU group was $g = 0.40$, $p < .0001$, 95% CI [0.23, 0.58], indicating the attendance of 2.6 more sessions by those in the P/T Fb group (2.5 sessions of which occurred after the signal alarm event).

CST Fb effect. A one-way ANOVA on the CST Fb groups resulted in a significant study effect, $F(2, 412) = 4.50$, $p = .012$. The combined effect of the difference in mean session attendance between the CST Fb and NOT Fb groups was $g = 0.41$, $p = .024$, 95% CI [0.05, 0.76], with individual effect sizes ranging from $g = 0.22$, $p = .106$, 95% CI [−0.05, 0.48], to $g = 0.816$, $p < .001$, 95% CI [0.48, 1.16] (see Table 21 in the supplementary materials). Study 3 resulted in a significantly larger number of sessions attended (mean difference of 4.6 sessions). Given the possible bias reflected in the assignment process in Study 3, another weighted mean effect size was calculated after removing the data from Study 3. The result was significant at the .05 level, $g = 0.22$, $p = .020$, 95% CI [0.035, 0.410], suggesting significantly more sessions attended on average (1.8 more sessions) by NOT patients in the CST Fb groups than by those in the Fb groups on average in routine care. The effect size of the pooled CST group in relation to the pooled TAU group was $g = 0.48$, $p < .001$, 95% CI [0.33, 0.63], indicating the attendance of 3.4 more sessions by those in the CST group (2.7 sessions of which occurred after the signal alarm event).[10]

Fb Effects on OT Patients

We used mega-analytic approaches to test the effects of P/T Fb and Fb interventions on OT patients. Prior to comparing feedback intervention groups against TAU, we tested heterogeneity of effects in the same manner as for the NOT samples. Statistically significant heterogeneity was detected at the .05 level in both (a) ANCOVAs of mean posttest OQ scores by study with pretest OQ score as a covariate and (b) ANOVAs of mean pretest scores by study. The primary reason for this heterogeneity was the significantly higher mean posttest and pretest scores of patients in Study 4 (Hawkins et al., 2004). When the data of patients from Study 4 were removed from the analyses, study effects no longer reached statistical significance. When the data were pooled by treatment conditions across studies and tested for equivalence in mean pretest scores by independent samples t tests, no significant differences were found. Considering that patients from Study 4 were found in all three of the OT treatment conditions (i.e., OT Fb, OT P/T Fb, and OT TAU groups) and on the basis of the equivalent mean pretest scores by treatment conditions, the data were pooled by OT treatment conditions in favor of ecological validity.

Fb effects. Contrary to the findings in a previous meta-analysis (Lambert et al., 2003), when the mean numbers of sessions attended by patients in OT Fb and OT TAU were compared in ITT analysis, no statistically significant difference was found, $g = 0.01$, $p = .792$, 95% CI [−0.06, 0.07]. Although an overall decrease in session attendance was not observed, patients in the OT Fb group, on average, experienced greater therapeutic gains. In terms of mean posttest OQ score difference, ITT analysis was significant at the .05 level, $g = -0.12$, $p < .001$, 95% CI [−0.19, −0.06], the equivalent of approximately a 2.8 OQ point reduction, whereas efficacy analysis was also significant at the .05 level, $g = -0.30$, $p < .001$, 95% CI [−0.37, −0.23], the equivalent of a 6.5 OQ point reduction.

When the odds of patient deterioration/reliable worsening at termination of the OT Fb group were compared against OT TAU, the results of ITT mega-analysis indicated that the effect size was significant at the .05 level, $OR = 0.63$, $p = .030$, 95% CI [0.41, 0.95]. Efficacy analysis was not significant at the .05 level, $OR = 0.81$, $p = .341$, 95% CI [0.52, 1.25]. When the odds of the occurrence of patient reliable/clinically significant improvement at termination were compared, both ITT and efficacy analyses were significant at the .05 level with respective odds ratios of $OR = 1.20$, $p = .010$, 95% CI [1.04, 1.38], and $OR = 2.09$, $p < .001$, 95% CI [1.80, 2.42]. These results suggest that, although the number of sessions utilized essentially remains the same, OT patients in the Fb condition experience superior treatment outcome on average and have decreased odds of experiencing deterioration than those in TAU.

P/T Fb effects. When the mean number of sessions attended was compared between the OT P/T Fb and OT TAU groups in ITT analysis, the results were significant at the .05 level, $g = 0.10$, $p < .02$, 95% CI [0.01, 0.18], the equivalent of approximately 0.4 more sessions attended by the OT P/T Fb. In terms of mean posttest OQ score differences, both ITT and efficacy analyses, respectively, yielded significant results at the .05 level, $g = -0.18$, $p < .001$, 95% CI [−0.26, −0.96], the equivalent of approximately a 4.1 OQ point reduction on average, and $g = -0.32$, $p < .001$, 95% CI [−0.40, −0.23], the equivalent of approximately a 7.1 OQ score point reduction on average.

When the odds of the occurrence of patient reliable worsening/deterioration at termination were compared, neither ITT nor efficacy analyses were significant at the .05 level with respective odds ratios of $OR = 0.71$, $p = .215$, 95% CI [0.42, 1.22], and $OR = 0.86$, $p = 0.585$, 95% CI [0.55, 1.47]. When the odds of the occurrence of reliable/clinically significant improvement were compared, ITT and efficacy analyses yielded significant results at the .05 level, $OR = 1.65$, $p < .001$, 95% CI [1.39, 1.96], and $OR = 2.36$, $p < .001$, 95% CI [1.97, 2.82], respectively. These results suggest that, in comparison to TAU, patients who receive P/T Fb intervention on average experience superior treatment outcome in terms of distress reduction and improved odds of achieving reliably positive change, whereas the odds of reliable worsening/deterioration remain the same as those for patients in TAU.

Discussion

This meta- and mega-analytic study evaluated the effects of three types of patient progress feedback interventions used in the OQ-based quality assurance system: progress feedback to therapists, progress feedback to both patients and therapists, and CST in addition to progress feedback. These interventions were aimed at monitoring individual patient progress in treatment, identifying patients at risk of treatment failure, and intervening before termination occurred. The effects of these interventions were evaluated

[10] The results were equivalent when the data of patients in Study 3 were removed from the analysis, $g = 0.45$, $p < .001$, 95% CI [0.29, 0.60], or 3.2 more sessions attended by the CST Fb group.

with patients whose progress in treatment was identified as NOT (i.e., patients at risk of leaving treatment worse off than when entering treatment as well as those identified as OT). Two sets of analyses were conducted to estimate the effects of feedback interventions that can be expected in routine practice (ITT analyses) and among patients who stay in treatment until the effects of feedback interventions could be measured (efficacy analyses). The effects of the feedback interventions were evaluated on the basis of between-group differences in mean OQ total scores at termination of treatment, rate and odds of clinically significant change status at termination, and mean number of sessions attended.

Overall, the effects of feedback interventions on patients who were identified as being at risk of treatment failure (NOT) were more substantial than the effects for OT patients. When compared with the NOT TAU in ITT analyses, the combined effects (Hedges's g) of mean posttreatment OQ total scores for the NOT Fb, NOT P/T Fb, and the CST Fb groups were −0.28, −0.36, and −0.44, respectively. The overall percentages of reliable worsening/deterioration (clinically significant improvement) among the NOT TAU, NOT Fb, NOT P/T Fb, and CST Fb groups were 20.1% (22.3%), 13.6% (30.9%), 15.8% (38.7%), and 11.3% (37.6%), respectively. The odds ratio of reliable worsening/deterioration (clinically significant improvement) for the NOT Fb, NOT P/T Fb, and CST Fb groups in relation to NOT TAU were 0.62 (1.70), 0.74 (2.20), and 0.51 (2.01), respectively. These results indicate that all forms of feedback interventions were effective in enhancing outcome while reducing treatment failures among NOT patients, with the exception of the P/T Fb intervention in its effects in preventing treatment failure. These results also show that, when the treatment impact is evaluated on the level of routine care (ITT analysis), the three types of feedback interventions are similar in their effects on treatment outcomes.

The effects of feedback interventions on those who satisfied the least necessary conditions to likely have been the actual recipients of the feedback interventions were also estimated (efficacy analyses). Such criteria comprised attending a minimum number of sessions (i.e., at least four to six sessions, depending on feedback conditions) and completing the OQ a minimum number of times (i.e., at least three to four administrations, depending on conditions). The effect sizes (Hedges's g) for the mean posttreatment OQ total score differences among the NOT Fb, NOT P/T Fb, and CST Fb groups in comparison with NOT TAU were −0.53, −0.55, and −0.70, respectively. Furthermore, the percentages of patients experiencing reliable worsening/deterioration (clinically significant improvement) for the NOT TAU, NOT Fb, NOT P/T Fb, and CST Fb groups were 20.1% (22.3%), 9.1% (37.6%), 14.7% (45.2%), and 5.5% (52.5%), respectively. The combined odds ratios of reliable worsening/deterioration (clinically significant improvement) for the NOT Fb, NOT P/T Fb, and CST Fb groups were 0.44 (2.55), 0.68 (2.97), and 0.23 (3.85), respectively. These results indicate greater treatment effects in all of the outcome criteria evaluated in this study, except for the NOT P/T Fb condition in its effect to reduce reliable worsening/deterioration.

Contrary to the previous meta-analysis (Lambert et al., 2003), this study highlighted effects of feedback interventions on OT patients. Although not to the magnitude experienced by the NOT counterparts, patients in OT P/T Fb and OT Fb appear to have experienced more distress reduction and increased odds of experiencing reliable/clinically significant improvements than those in OT TAU.

It is interesting to note the pattern of outcomes seen in the P/T Fb patients. Specifically, this intervention yielded increased treatment-enhancing effects while yielding a similar rate of reliable worsening/deterioration when compared with that of the TAU group. These results suggest the possibility of a mechanism that interacts with provision of direct progress feedback to patients in a way that enhances outcomes for some while inhibiting outcome enhancement for others.

Another aim of this meta-analysis was to investigate the incremental benefit of the NOT P/T Fb and CST Fb interventions compared with the NOT Fb intervention. Because previous reports (Harmon et al., 2007; Slade et al., 2008) provided the results of efficacy analyses for only the CST Fb intervention, we compared this intervention with the NOT Fb intervention under equivalent inclusion criteria. Although the comparative magnitude of the effects of the CST Fb over the Fb were smaller than previously reported (primarily because of comparing the efficacy CST Fb samples against the ITT NOT Fb samples in previous studies), results of ITT analyses produced statistically significant effects in terms of superior distress reduction at posttreatment ($g = -0.16$, $p < .05$) and increased odds of clinically significant improvement ($OR = 1.53$, $p < .05$). Although statistical significance was observed in comparisons between the CST Fb and NOT Fb groups, there were overlaps in 95% confidence intervals between the CST Fb versus NOT TAU comparisons and the NOT Fb versus NOT TAU comparisons. Efficacy analyses did not yield statistically significant greater additive treatment effects for the CST Fb group over the NOT Fb group. These statistically nonsignificant results, however, should not be automatically assumed to be indications of no additive effect, as reflected in greater effect sizes yielded in the efficacy analyses. Statistically nonsignificant results may be due to lack of statistical power because sample sizes were reduced by exclusion criteria. Future trials featuring the CST Fb and P/T Fb interventions may further researchers' understanding of the magnitude of these interventions.

Comparison of ITT analyses and efficacy analyses opens questions about possible mechanisms of change. Because the primary element of the exclusion criteria used in efficacy analyses was the number of sessions attended by patients, a question arises as to how much of the improvements in the results of efficacy analyses were a function of the dose–response effect (Hansen, Lambert, & Forman, 2002). Improvements in treatment outcome among the efficacy samples also suggest a higher proportion of poorer outcomes among patients who left treatment before the feedback interventions could have taken the effect. If so, it appears important that proactive effort be given to retain at-risk patients in treatment, even more so for those experiencing worsening at early stages of therapy.

As a supplemental analysis, to test the possibility of disproportional occurrences of reliable worsening/deterioration and clinically significant improvement based on the length of treatment, we calculated the percentages and odds of such outcomes on the pooled data set of all NOT cases ($N = 1{,}382$). Of those NOT patients who left treatment after five or fewer sessions (early terminators; $n = 381$), 22.8% deteriorated, whereas 12.1% made significant clinical improvement. In contrast, of those NOT patients who stayed in treatment for six sessions or more (late

terminators; $n = 1{,}001$), 11.7% later reliably worsened/deteriorated, whereas 39.9% made significant clinical improvement. When early terminators and late terminators were compared, the odds ratio of reliable worsening/deterioration for the early terminators was 2.24, whereas the odds ratio of clinically significant improvement was 0.21. These findings underline the need to retain NOT patients in treatment longer. Future research effort to uncover the therapeutic and countertherapeutic processes of engaging NOT patients in treatment is recommended. Future research concerning the process of deterioration also appears to be an important area to be explored further.

Although research synthesis, such as this study, provides various statistical advantages in data analyses, this study has limitations, many of which were inherent in the original studies. Reliability of treatment implementation may have been an issue in individual studies because the use of feedback interventions by therapists was not closely controlled or monitored. Although statistical power increased as a result of data synthesis/pooling, the magnitude of true effects may have been underestimated. Because random assignment to conditions was not incorporated in two of the studies (Lambert, Whipple, Vermeersch, et al., 2002; Whipple et al., 2003), selection bias may have occurred, resulting in heterogeneous samples of patients. Although this issue was not detected in original studies, application of uniform inclusion/exclusion criteria in this study revealed some heterogeneity. Similarly, an argument can be made against causal statements based on the data of studies not directly compared in the same randomized trials, such as in the case of comparing the CST Fb group and the pooled TAU group in the studies of Harmon et al. (2007) and Slade et al. (2008). However, pooled TAU data from multiple studies may provide the most reliable benchmark for comparing alternative treatment strategies.

Another criticism may be made regarding the possibility of mono-method bias because this line of research used the OQ-45 both as the outcome measure and as the method for identifying NOT cases. We recognize that arguments can be made for using multimethod, multiperspective outcome assessment to capture more breadth of information related to patient treatment progress and outcome. Such methods may be valuable in enhancing comprehensive understanding of the impact of feedback interventions. However, routine assessment and monitoring of outcome requires an instrument that is time and cost efficient. In routine care where treatment termination is determined largely by patients and treatment length is unknown at the outset, the use of multiple outcome measures is not feasible. Given the established reliability and validity of the OQ-45 as a sensitive measure of treatment outcome (Vermeersch et al., 2004) and given that its classification of patient change is concordant with other frequently used measures, we considered the use of the OQ-45 as the sole assessment tool well suited for the purpose of quality assurance in routine clinical practice.

Another limitation of this line of research is the exclusive use of the OQ total score in outcome monitoring and feedback provision. Exploration of the OQ subscale scores may enhance researchers' understanding of the mechanisms of the feedback interventions as well as the processes of change in psychotherapy.

Although we do not view the following as a limitation to this line of research, it is important to point out that the feedback procedures advocated here are deemed to be more appropriate for cases that are predicted to deteriorate and not for all patients. To better understand the effects of feedback interventions in a broad context of routine clinical practice, we investigated the overall effects of feedback interventions on all patients included in original studies (both OT and NOT patients) by pooling the full data sets of the six studies ($N = 6{,}151$; ITT analyses). Of those who received any form of feedback intervention, 4.7% of patients experienced reliable worsening or deterioration, whereas 37.4% of patients experienced clinically significant improvement. Of those patients in routine care (i.e., no feedback; TAU), 6.1% reliably worsened/deteriorated, whereas 30.2% achieved clinically significant improvement. Accordingly, overall odds of deterioration among the pooled feedback interventions group in relation to patients receiving TAU were statistically significant ($OR = 0.76$, $p = .024$). The overall odds of clinically significant improvement among those in the pooled feedback group were also statistically significant ($OR = 1.38$, $p < .001$). The overall effects in terms of posttreatment mean OQ total scores showed significantly less disturbance ($g = -0.12$, $SE = 0.03$, $p < .001$). This effect size translates to 2.9 OQ total points reduction on average. The overall reduction of deteriorated cases, the increase in clinically significant improvement, and the decrease in distress level at termination occurred within a context of utilizing an average of 0.9 more of a session of care.

Despite the limitations discussed here, the accumulating evidence is substantial in favor of the routine use of progress feedback and clinical problem-solving tools. When considering clinicians' difficulty with identifying patients at risk of treatment failure (Hannan et al., 2005), the current state of evidence seems sufficient to warrant routine use of these feedback interventions. Nonetheless, further replications across different patient populations by different research groups are needed before the boundary conditions of effectiveness will be known.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist, 34,* 341–382.

American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285.

Atkins, D. C. (2009). Clinical trials methodology: Randomization, intent-to-treat, and random-effects regression. *Depression and Anxiety, 26,* 697–700.

Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goats, M. K., & Vermeersch, D. A. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *Behavior Analyst Today, 4,* 79–90.

Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

DeRubeis, R. J., Gelfand, L. A., Tang, T. Z., & Simons, A. D. (1999). Medications versus cognitive behavior therapy for severely depressed outpatients: Mega-analysis of four randomized comparisons. *American Journal of Psychiatry, 156,* 1007–1013.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 255–263.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy

quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy, 8,* 231–242.

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology, 61,* 1233–1243.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61,* 155–163.

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9,* 329–343.

*Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist–client feedback and clinical support tools. *Psychotherapy Research, 17,* 379–392.

Hatfield, D., McCullough, L., Plucinski, A., & Krieger, K. (2009). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy.* Advance online publication. doi:10.1002/cpp.656

*Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 14,* 308–327.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

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

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51,* 1059–1064.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Lachin, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials, 21,* 167–189.

Lambert, M. J., Bergin, A. E., & Garfield, S. L. (2004). Introduction and historical overview. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed.). New York, NY: Wiley.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172.

Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., Reid, R. C., . . . Burlingame, G. M. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire).* Orem, UT: American Professional Credentialing Services.

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

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology & Psychotherapy, 9,* 149–164.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10,* 288–301.

*Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11,* 49–68.

*Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. K. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology & Psychotherapy, 9,* 91–103.

Lunnen, C., & Ogles, B. M. (1998). A multi-perspective, multi-variable evaluation of reliable change. *Journal of Consulting and Clinical Psychology, 66,* 400–410.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice, 2,* 1–27.

OQ–Analyst [Computer software]. Salt Lake City, UT: OQ Measures.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

Reed, G. M., & Eisman, E. J. (2006). Uses and misuses of evidence: Managed care, treatment guidelines, and outcomes measurement in professional practice. In C. D. Goodheart, A. E. Kazdin, & R. J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet* (pp. 13–35). Washington, DC: American Psychological Association.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 85,* 638–664.

Serretti, A., Cusin, C., Rausch, J. J., Bondy, B., & Smeraldi, E. (2006). Pooling pharmacogenetic studies on the serotonin transporter: A mega-analysis. *Psychiatry Research, 145,* 61–65.

*Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy, 15,* 287–303.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy, 13,* 202–214.

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology, 51,* 38–49.

Warren, J. (in press). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology.*

*Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50,* 59–68.

Appendix A

Calculation of Effect Sizes and Standard Errors

Effect Sizes

Hedges's standardized mean difference (g) for mean posttest OQ scores and mean session attendance comparisons is calculated as the following (Comprehensive Meta-Analysis [Version 2]; Hedges, 1981):

1. Calculate standardized difference in means (d) by dividing the raw score difference in means by pooled standard deviation of two samples ($m_1 - m_2$) in comparison: $d = (m_1 - m_2)/s_{pooled}$, where s_{pooled} is calculated by using the following formula:

$$s_{pooled} = \sqrt{\frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{(n_1 + n_2 - 2)}}, \quad \text{(A1)}$$

where n_1 and n_2 represent the sample sizes of Samples 1 and 2, and s_1 and s_2 represent the standard deviations of Samples 1 and 2.

2. Compute correction factor J for correcting bias: $J = 1 - [3/(4df - 1)]$, where df is given by

$$df = n_1 + n_2 - 2. \quad \text{(A2)}$$

3. Compute Hedges's standardized mean difference (g) by multiplying d by a correction factor (J):

$$g = d \times J. \quad \text{(A3)}$$

Standard Errors

1. Obtain standard error for standard difference in means (d):

$$SE(d) = \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{d^2}{2(n_1 + n_2)}}. \quad \text{(A4)}$$

2. Correct for bias by multiplying standard error of standardized mean difference by a correction factor J:

$$SE(g) = SE(d) \times J. \quad \text{(A5)}$$

Appendix B

Assignment of Random Weight

Random weight and the main effect are calculated as the following (Comprehensive Meta-Analysis (Version 2); Hedges & Vevea, 1998). A random weight (w) assigned to each individual study (i) is defined as follows:

$$w_i = \frac{1}{v_i^*}, \quad \text{(A6)}$$

where v_i^* represents the sum of within-study variance (i) and the between-studies variance (τ^2):

$$v_i^* = v_i + \tau^2. \quad \text{(A7)}$$

The mean effect size ($\bar{g}$) is calculated as follows:

$$\bar{g} = \frac{\sum_{i=1}^{k} w_i g_i}{\sum_{i=1}^{k} w_i}. \quad \text{(A8)}$$

The variance of the mean effect is defined as the reciprocal of the sum of the individual study weights. Thus, the standard error (SE) of the mean effect is the square root of the sampling variance:

$$SE(\bar{g}) = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}. \quad \text{(A9)}$$

Received November 6, 2009
Revision received February 1, 2010
Accepted February 15, 2010 ■