Creating a Climate for Therapist Improvement: A Case Study of an Agency Focused on Outcomes and Deliberate Practice

Simon B. Goldberg
University of Wisconsin-Madison

Robbie Babins-Wagner
Calgary Counselling Center, Alberta, Canada

Tony Rousmaniere
University of Washington-Seattle

Sandy Berzins
University of Calgary and Calgary Counselling Center, Alberta, Canada

William T. Hoyt
University of Wisconsin-Madison

Jason L. Whipple
University of Alaska Fairbanks

Scott D. Miller
International Center for Clinical Excellence, Chicago, Illinois

Bruce E. Wampold
University of Wisconsin-Madison and Modum Bad Psychiatric Center, Vikersund, Norway

Recent evidence suggests that psychotherapists may not increase in effectiveness over accrued experience in naturalistic settings, even settings that provide access to patients' outcomes. The current study examined changes in psychotherapists' effectiveness within an agency making a concerted effort to improve outcomes through the use of routine outcome monitoring coupled with ongoing consultation and the planful application of feedback including the use of deliberate practice. Data were available for 7 years of implementation from 5,128 patients seen by 153 psychotherapists. Results indicate that outcomes indeed improved across time within the agency, with increases of $d = 0.035$ ($p = .003$) per year. In contrast with previous reports, psychotherapists in the current sample showed improvements within their own caseloads across time ($d = 0.034$, $p = .042$). It did not appear that the observed agency-level improvement was due to the agency simply hiring higher-performing psychotherapists or losing lower-performing psychotherapists. Implications of these findings are discussed in relation to routine outcome monitoring, expertise in psychotherapy, and quality improvement within mental health care.

Keywords: expertise, quality improvement, therapist effects, psychotherapy training, routine outcomes monitoring

The effectiveness of psychotherapy has long been recognized and has been demonstrated in both randomized controlled trials as well as naturalistic settings (Minami et al., 2008; Smith & Glass, 1977; Stiles, Barkham, Twigg, Mellor-Clark, & Cooper, 2006;

Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin-Madison; Robbie Babins-Wagner, Calgary Counselling Center, Alberta, Canada; Tony Rousmaniere, Student Health and Counselling, University of Washington-Seattle; Sandy Berzins, Department of Psychiatry, University of Calgary, and Calgary Counselling Centre; William T. Hoyt, Department of Counseling Psychology, University of Wisconsin-Madison; Jason L. Whipple, Department of Psychology, University of Alaska Fairbanks; Scott D. Miller, International Center for Clinical Excellence, Chicago, Illinois; Bruce E. Wampold, Department of Counseling Psychology, University of Wisconsin-Madison, and Modum Bad Psychiatric Center, Vikersund, Norway.

Correspondence concerning this article should be addressed to Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin-Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI 53706. E-mail: sbgoldberg@wisc.edu

Wampold & Imel, 2015). The average psychologically distressed person who receives psychotherapy will be better off than 80% of the distressed people who do not (Hubble, Duncan, & Miller, 1999; Wampold & Imel, 2015). To date, dozens of studies show the effects of psychotherapy to be at least as large as the effects of psychotropic medications, with lower costs, fewer troubling side effects, and longer-lasting results (Forand, DeRubeis, & Amsterdam, 2013; Gøtzsche, Young, & Crace, 2015).

Despite evidence for the benefits of psychotherapy, mental health care, along with other facets of health care, has recently come under scrutiny. There have been substantial decreases in state funding for mental health services in the United States in recent years (National Alliance on Mental Illness, 2012). While tighter budgets are understandably difficult for agencies and clinicians seeking to provide effective treatment, attention to the quality and effectiveness of mental health treatment is a worthwhile concern, and indeed a longstanding interest in the field of health care. The quality improvement movement has existed in medicine for several decades and is intended to improve the efficiency and effectiveness of medical care (Lynn et al., 2007). Batalden and Davidoff (2007) define quality improvement in medicine as "the combined and unceasing efforts of everyone—health-

care professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development (learning)" (p. 2).

One method of quality improvement that has been promoted in medicine is the use of audit and feedback (Ivers et al., 2012). This method involves measuring a practitioner's professional practice or performance, comparing this practice to professional standards or targets, and feeding back results to the individual practitioner. In a recent meta-analytic review of 140 studies, audit and feedback methods were shown to improve practitioner compliance with desired practices as well as clinical outcomes (Ivers et al., 2012).

The audit and feedback method may not apply directly to mental health care (although the spirit of quality improvement certainly does; Parry, 1992); at present, there does not exist a single, agreed upon standard for professional practice or performance that has been linked compellingly with clinical outcomes (Wampold & Imel, 2015). It is reasonable to assume that psychotherapy, unlike certain medical procedures, may be accomplished effectively in a variety of ways (Baardseth et al., 2013; Laska, Gurman, & Wampold, 2014). However, an analogue to the audit and feedback method in medicine—routine outcome monitoring (ROM)—has been developed for mental health care, and a growing body of research evidence supports its use in the context of psychotherapy (Shimokawa, Lambert, & Smart, 2010; Wampold, 2015). ROM involves the ongoing monitoring of patient progress over the course of therapy, with outcome data available to the psychotherapist, patient, or both (Wampold, 2015).

The use of ROM as a method for improving clinical outcomes is attributed to Howard and colleagues (Howard, Moras, Brill, Martinovich, & Lutz, 1996) and has been further developed in recent years (Lambert, Hansen, & Finch, 2001; Miller, Duncan, & Hubble, 2005). In general, these systems involve collecting self-reports of patients' psychological functioning that are then compared to normative data drawn from other psychotherapy patients. Psychotherapists are then provided with feedback regarding a specific patient's progress relative to an expected change trajectory (Lambert, Harmon, Slade, Whipple, & Hawkins, 2005; Miller et al., 2005). Evidence from randomized controlled trials comparing feedback versus no feedback conditions suggests that feedback may improve outcomes primarily by reducing the rates of patients who are not improving as expected (Shimokawa et al., 2010). A variety of such systems have been developed with promising psychometric and clinical properties including the Outcome Questionnaire (OQ) System (Lambert, 2015), Partners for Change Outcome Management System (PCOMS; International Center for Clinical Excellence, 2012), Treatment Outcome Package (TOP; Boswell, Kraus, Castonguay, & Youn, 2015), Clinical Outcomes in Routine Evaluation (CORE) System (Barkham, Mellor-Clark, & Stiles, 2015), Counseling Center Assessment of Psychological Symptoms (CCAPS; Youn et al., 2015), Behavioral Health Measure (BHM; Kopta, Owen, & Budge, 2015), and a collaborative outcomes research network (ACORN; Brown, Simon, Cameron, & Minami, 2015).

Although promising evidence exists that feedback systems can improve patient outcomes particularly for individuals predicted to be treatment nonresponders (Shimokawa et al., 2010), it is not clear that mere access to feedback improves psychotherapists' overall efficacy across time. A recent longitudinal study conducted

over the course of 18 years on a sample of 6,591 patients seen by 170 psychotherapists who had access to patient outcomes through ROM suggests that psychotherapists do not appear to improve over time, and may even show very slight decrements in outcomes as experience accrues (Goldberg et al., 2016).

Tracey, Wampold, Lichtenberg, and Goodyear (2014) highlight some of the challenges facing psychotherapists in their attempts to improve clinical outcomes. Tracey et al. asserted that, unlike other professions (e.g., astronomers, accountants), psychotherapists do not improve their outcomes merely by gaining experience alone, in part due to the unpredictability of outcomes and the limited utility of feedback about past actions. Instead, they propose systematic feedback on patients' progress within psychotherapy coupled with feedback relative to therapeutic actions and practice improving those skills as the key to improving psychotherapists' outcomes over time.

How might the conditions Tracey et al. (2014) propose as necessary for psychotherapists' improvement be cultivated in the context of mental health care? One possibility involves creating change at the agency, or organizational level. The quality improvement literature in medicine has identified multiple levels at which efforts to improve health care may take place. Ferlie and Shortell (2001) identified four primary levels: the individual, group/team, organization, and larger system/environment levels. Much of the literature on ROM systems has focused on the individual provider level through examining the ways in which psychotherapists' access to particular forms of feedback does or does not improve patient outcomes. The current study, however, aimed to assess the impact of ROM over time within a mental health agency actively working to improve outcomes. This agency, as discussed below, included ROM as a central component of professional development at the agency-level. These efforts are in keeping with the characteristics identified by Ferlie and Shortell (2001) for the organizational level of change in quality improvement, notably creating "an overall climate and culture for change" (p. 287).

The agency in the current case study sought to provide conditions necessary for quality improvement (Tracey et al., 2014), including the use of deliberate practice (i.e., measuring baseline performance, obtaining specific and ongoing feedback, engaging in deliberate practice through rehearsing and evaluating a plan for improvement; Miller, Hubble, Chow, & Seidel, 2013) and in line with the spirit of quality improvement (Batalden & Davidoff, 2007). In contrast to other settings using feedback in less active and less structured ways, the agency in this case study used outcome data in ongoing agency-wide trainings. These regular meetings focused on discussing the progress of specific psychotherapy cases, with an emphasis on cases that were not progressing, and intentional practice to develop therapeutic skills.

The effects of these efforts over the course of 7 years were examined in the current study. Outcomes were examined at the agency level as well as the within-psychotherapist (i.e., individual) level to assess the impact of these agency-level efforts to improve outcomes. Three research questions guided this work. The broadest question was whether patient outcomes in the agency improved over time. If such improvements were seen, two possible explanations exist: psychotherapists showed improved outcomes over time (i.e., within-psychotherapist improvement, in contrast to previous reports; Goldberg et al., 2016) or psychotherapists who are hired later in time were more effective than the psychotherapists

who were hired earlier (i.e., the agency was hiring better psychotherapists).

Method

Participants

Patients. Data were drawn from a community mental health agency in a large Canadian city. Data were collected over the course of 7 years. All patients in the current sample were seen in individual psychotherapy. The focal sample included 5,128 patients who were in the clinical range at baseline (i.e., Outcome Questionnaire-45 [OQ-45] total scores 63 or above; Lambert et al., 2004), seen by 153 psychotherapists. The sample included 2,948 female patients (57.7%), 2,107 male patients (41.3%), 16 transgender patients (0.3%), and 35 patients (0.7%) who chose not to indicate a gender identification. The average age was 33.69 (SD = 11.32, Mdn = 31.00, range = 17 to 86). Age data were unavailable for 101 participants. The sample represented a range of socioeconomic backgrounds with a median income of C\$34,000 (SD = 66308, $Mean$ = C\$52,134, range = C\$0 to C\$1,600,000). Racial/ethnic composition of the sample was as follows: 49 African or Caribbean (1.0%), 10 Arab or Middle Eastern (0.2%), 266 Asian (5.2%), 79 Latino (1.5%), 69 Native or First Nations (1.3%), 3,321 Caucasian (64.8%), 299 Other (5.8%). Racial/ethnic data were unavailable for 1,035 participants (20.2%) due to participants declining the opportunity to provide these data.

Several data processing steps were necessary in order to arrive at this final sample and meet recommendations for employing multilevel modeling with psychotherapy data (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009; Baldwin & Imel, 2013). First, the sample was reduced to include only those patients who received at least three sessions of individual psychotherapy with the same clinician. This was based on the rationale that fewer than three sessions would not adequately reflect a meaningful dose of treatment (Baldwin et al., 2009; Howard, Kopta, Krause, & Orlinsky, 1986). Patients who saw multiple clinicians were excluded to avoid cross-classification between psychotherapists. Further, we included only the first episode of care, defining a new episode when a patient either saw a new clinician or a period of 45 days or longer elapsed between sessions. There was a definite positive skew in length of treatment in the current data, as has been noted in prior naturalistic data (e.g., Stiles, Barkham, Connell, & Mellor-Clark, 2008), with the majority of cases representing lengths of treatment at or below the mean number of sessions. The mean number of sessions per patient in the sample was 6.53 (SD = 5.02, Mdn = 5, range = 3 to 99).

Psychotherapists. Psychotherapy was provided by 153 psychotherapists. The therapists had on average 4.42 years of data available in the current data set (SD = 1.89, Mdn = 4.28, range = 0.52 to 6.84). Psychotherapists in the data set began at the center either as licensed professionals, provisional professionals (accruing hours for licensure), or practicum students (completing hours as part of a graduate training program). The distribution within these three categories was as follows: 22 licensed professionals (14.4%), 54 provisional professionals (35.3%), and 77 practicum students (50.3%). Psychotherapists who were not yet licensed (i.e., provisional professionals and practicum students) received one hour of individual and two hours of group supervision per week.

As the focus of this work was on psychotherapist effects, it was vital to assure that as reliable estimates of psychotherapist effects as possible were obtained. To this end, the sample was reduced to include only patients whose psychotherapists saw 10 or more patients within the clinical range. This data reduction step increases the reliability of psychotherapist-level estimates (Baldwin, Imel, & Atkins, 2012; Crits-Christoph, Connolly Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011). Psychotherapists saw on average 33.52 patients (SD = 26.24, Mdn = 28, range = 10 to 167).

The primary means to assign patients to psychotherapists was based on available slots in the psychotherapist schedules, although occasionally patients requested a psychotherapist who was either a male or female and such requests were honored. Assignment was not based on patient severity, chronicity, or prognosis. Although assignment to psychotherapist was not completely random, it could be described as quasi-random.

Agency. Data were drawn from a nonprofit mental health agency located in a large Canadian city. The agency provides a full range of individual, couples, and family counseling services. Agency staff is primarily composed of licensed professionals with Master's degrees or PhDs in social work, psychology, marriage and family therapy, or pastoral counseling. Master's level trainees in these disciplines also provide services at the agency.

The agency first became interested in collecting outcome data in 2000 and began routine collection of the OQ-45 in 2004. Between 2004 and 2008, counselors were invited to collect OQ data prior to every session, with summaries of outcome data produced at the end of each calendar year. It was clarified to agency staff from the outset of data collection that these data would not be used for performance management and would not appear on any yearly performance plans or reports. At this point, no patient-specific outcome data were available.

In 2008, the agency reviewed rates of adoption of data collection among agency staff, noting that the majority of staff (60%) were not consistently collecting outcome data. The agency policy supporting the use of measures was changed in Fall of 2008 requiring staff to provide the measures to all patients. Patient completion of measures was included as a performance measure, although patients' responses on these measures remained not tied to performance evaluation.

Considerable staff turnover occurred following the adoption of this policy, with ~40% of licensed professionals on staff resigning within four months. Of note, resistance to the adoption of ROM has been documented in the literature previously (e.g., Duncan & Murray, 2012; Miller et al., 2013; Unsworth, Cowie, & Green, 2012). Additional clinicians were recruited at that time, with the intention of selecting staff who were interested in collecting and using outcomes in routine practice. The agency was again fully staffed within four months.

In addition to requiring clinicians to provide measures to all patients (patients still could opt out of completing these measures), the agency began monthly clinical consultations with an external consultant to discuss cases and build a culture of feedback. Consultations were two hours and took place over videoconference. Clinicians were asked to provide a graph of a particular patient's progress in psychotherapy for discussion. Clinicians were explicitly encouraged to bring cases that were not progressing, in order to build a culture of feedback (termed "error-centric" consultations within the agency). During these consultations, staff would iden-

tify, review, and make plans for better meeting the needs of cases that were not progressing as expected. Consultation discussions followed the feedback-informed treatment model of case review (Miller et al., 2013), beginning with a presentation of results from the outcome measure. Background information, diagnostic language, and conversations about treatment approaches were kept at a minimum. Instead, discussion was organized around the therapeutic alliance; in particular, discussion focused on clarifying client goals and preferences and identifying means most likely to facilitate engagement.

The culture of feedback developed over time at the agency. Monthly case consultations were often focused on specific topics (e.g., depression, anxiety, therapeutic alliance). In addition, clinicians became more engaged in this practice as they began using feedback from consultation in their psychotherapy sessions and reporting positive effects. The current data set includes data drawn from the full implementation phase, beginning in fall of 2008.

Measures

The outcome measure used in this dataset was the OQ-45 (Lambert et al., 2004). This 45-item self-report measure was designed specifically to capture change that occurs during the course of psychotherapy. The measure has been widely used and shown to possess desirable psychometric properties, including high internal consistency reliability ($\alpha = .94$ for the total score in the current sample) and adequate test–retest reliability over a 3-week range (from .78 to .84; Snell, Mallinckrodt, Hill, & Lambert, 2001). Three subscales have been defined on the OQ-45: Symptom Distress (e.g., "I feel no interest in things," "I feel nervous"), Interpersonal Relations (e.g., "I am concerned about family troubles," "I have trouble getting along with friends and close acquaintances"), and Social Role Performance (e.g., "I feel that I am not doing well at work/school," "I feel stressed at work/school"). The use of the total score has been common practice and is supported by factor analytic work (Bludworth, Tracey, & Glidden-Tracey, 2010).

Statistical Methodology

Estimating treatment effects. Standardized mean difference scores (i.e., Cohen's *d* [1988]) were computed at the patient level using the difference between patients' pre- and posttreatment OQ-45 total scores divided by the sample's pretreatment standard deviation (as in Goldberg et al., 2016). Patient-level *d*s were included as the outcome in two-level models (patients nested within therapists) described below. Patients' *d*s were computed as pre- minus-post-treatment, with a more positive effect size reflecting a larger drop in symptoms during treatment.

Estimating changes in outcome across time.

Improvement at the agency level. The primary aim of this study was to assess the possibility of improvement in patient outcomes across time in an agency that is focusing intensively on improving patient care. As such, it was first important to assess whether the agency was showing improved outcomes across time. An initial two-level model (patient prepost *d*s nested within therapists) was constructed predicting patient outcomes from patient treatment date (Model 1).

$$Y_{ij} = b_{00}(Intercept) + b_{10}(Patient\ Treatment\ Date) + [U_{0j} + e_{ij}],$$
$$\text{(Model 1)}$$

where Y_{ij} reflects the prepost *d* for given patient *i* seen by therapist *j*, b_{00} is the fixed intercept, b_{10} is the fixed slope (reflecting change in *d*s as a function of treatment date). U_{0j} represents a random intercept (such that therapists' average outcome could vary around the grand mean) and e_{ij} is the error or residual term.[1] Time in this model (patient treatment date) was scaled so that time = 0 represented the midpoint of treatment for the earliest patient seen in the data set[2] (see Figure 1). This model assesses whether there is improvement in outcomes within the agency across time.

If indeed improvement was noted at the agency level, it was next important to assess if this improvement was due to the fact that psychotherapists were improving over time or to the fact that the agency was hiring better psychotherapists over time. Two additional models were used to examine these possibilities.

Within-therapist improvement across time. In order to assess whether outcomes were improving within a given psychotherapist's caseload across time, two-level models were fit (patients nested within therapist) following procedures used previously (Goldberg et al., 2016). Time in this model was operationalized relative to the therapist (i.e., time = 0 relates to the therapist's first patient in the dataset).

$$Y_{ij} = b_{00}(Intercept) + b_{01}(Within - Therapist\ Time) + [U_{0j} + e_{ij}],$$
$$\text{(Model 2)}$$

where Y_{ij} reflects the prepost *d* for given patient *i* seen by therapist *j*, b_{00} is the fixed intercept, b_{01} is the fixed slope (reflecting change in *d*s across time within-therapist), U_{0j} represents the random intercept, and e_{ij} the error or residual term. Within-therapist time was computed as the difference between the patient's treatment date and the treatment date of the first patient for that patient's therapist, so that time = 0 represented the treatment date for the earliest patient seen by a particular therapist (see Figure 1). As in Model 1, the midpoint of each patient's course of therapy was used to index that patient's time of treatment. The fixed slope coefficient (b_{01}) in this model assesses whether there is improvement in outcomes for patients seen later by the same therapist.

Improvements in overall outcome across start dates (i.e., hiring better psychotherapists). The third model assessed the possibility that the agency was improving due simply to psychotherapists who were hired later having better overall outcomes. In this model, patient outcomes were predicted by the date on which a patient's therapist entered the dataset (i.e., first session date for that therapist's first patient). Time was scaled so that time = 0 reflects the first patient seen by the first therapist (i.e., first session of treatment for earliest patient in dataset).

---

[1] Random slope models were also fit for Models 1, 2, and 3, which allowed the various metrics of time (calendar year, within-therapist time, therapist start date) to vary within therapist in their relationship with patient-level outcome (*d*). The random slopes did not improve model fit ($p > .100$) in any of the three models. Thus, the simpler random intercept models were used and reported here.

[2] Of note, the midpoint of each patient's course of therapy was used as the date of their treatment (in order to place a given patient's therapy within a time course). If there is linear change in effectiveness (at either agency or therapist levels), this midpoint represents the average level of effectiveness experienced by each patient over the course of treatment.
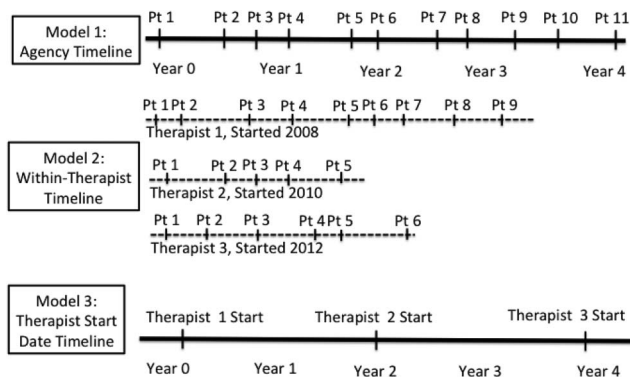
Figure 1. Timelines used for coding time. Pt = patient.

$$Y_{ij} = b_{00}(Intercept) + b_{01}(Therpaist\ Start\ Date) + [U_{0j} + e_{ij}],$$
(Model 3)

where Y_{ij} reflects the prepost d for given patient i seen by therapist j, b_{00} is the fixed intercept, b_{01} is the fixed slope (predicting patient outcome from therapist start date), U_{0j} represents the random intercept, and e_{ij} the error or residual term. Time was relative to when other therapists were hired in reference to actual, chronological calendar year (see Figure 1). This model assesses whether therapists who began work at the agency later had better overall outcomes.

Assessing for potential confounds. A final set of models was constructed to assess for potential third variables that may impact changes in treatments effects. An initial model assessed the possibility that level of baseline severity (i.e., baseline OQ total scores) changed over time (as higher levels of baseline severity could provide greater room for patients to improve and thus increase observed treatment effects; Cohen, Cohen, West, & Aiken, 2003). In this model, patients' baseline OQ scores were predicted from calendar year (as in Model 1). Subsequently, Models 1, 2, and 3 were reestimated including baseline OQ scores as a covariate to control for any temporal changes in initial severity.

The second potential confound that was explored was therapists' initial level of training. A recent report suggested that trainees may show improved outcomes over time (Owen, Wampold, Kopta, Rousmaniere, & Miller, 2016), which contrasts with results from a more heterogeneous sample of therapists (Goldberg et al., 2016). In order to assess for the potential impact of therapists' initial experience, we reestimated the model focusing on within-therapist change (i.e., Model 2) including therapists' initial training level as a main effect along with the interaction between therapists' initial training level and time. This allowed assessment of the possibility that therapists' outcomes showed differing trajectories across time based on their initial experience level.

Results

Descriptive Data

The sample overall showed a significant drop in distress as rated on the OQ over the course of treatment (pretreatment OQ: $M = 86.82$, $SD = 16.85$; posttreatment OQ: $M = 67.18$, $SD = 23.95$).

The average drop on the OQ was 19.64 points ($SD = 21.76$), with a corresponding prepost d of 1.17 ($SD = 1.29$). An unconditional model was fit initially (i.e., patients nested within therapists, with no additional predictors). The intraclass correlation coefficient (*ICC*) indicated that ~1.71% of variance in patients' prepost change was explained at the therapist level (between-therapist variance = 0.029, within-therapist variance = 1.64). The *ICC* differed significantly from zero, $\chi^2[152] = 242.43$, $p < .001$.

Examining Improvement Within the Agency Across Time (Treatment Date)

The first model examined whether patient outcomes improved across time at an agency intensively attempting to improve their outcomes. A significant effect was noted ($b_{10} = 0.035$, $p = .003$, Table 1), indicating that outcomes (patient-level prepost ds) were becoming 0.035 standardized units larger each year (see Figure 2).[3] The intercept of this model ($b_{00} = 1.02$) reflects the predicted outcome for a patient at time = 0. At time = 7 (i.e., 7 years later), a patients' predicted outcome would be $d = 1.27$, reflecting agency-level improvements across time. Having established that outcomes were indeed improving across time at the agency level, it was next important to determine whether this improvement was accounted for by (a) therapists improving over time, (b) the agency hiring higher-performing therapists over time, or (c) both.

Examining Improvement Within-Therapist Across Time

The second model examined changes in patient outcomes across therapist experience (i.e., within-therapist time). A significant effect was noted $b_{01} = 0.034$, $p = .042$; Table 1), indicating that therapists' outcomes improved 0.034 standardized units per year, a value approximately equal to the improvement of the agency over time (viz., 0.035).

Examining Improvements in Overall Outcomes Across Start Dates

The third model examined the possibility that psychotherapists hired to the agency later were simply higher performing, so observed agency-level improvements were due to hiring these more effective clinicians. The effect for therapist start date predicting patient outcomes was not significant ($b_{01} = 0.019$, $p = .129$; Table 1), indicating that therapists hired later did not have significantly better overall outcomes than therapists hired earlier.

Examining Changes in Baseline Severity and Initial Training Level as Potential Confounds

An initial model was constructed to assess the possibility that patients' baseline severity changed over the course of chronolog-

---

[3] Although our models included linear effects for time, it is theoretically unlikely that effects would actually continue to improve indefinitely. In order to assess the possibility that changes in outcome were nonlinear, a scatterplot of patient ds across treatment date was examined with both the model-derived linear regression line as well as a Loess curve (i.e., best fit line). These two lines were nearly identical. Further, a subsequent model that included both a linear and a quadratic term for time (treatment date) did not show a significant quadratic effect ($p > .10$). A significant quadratic effect for time was likewise not detected when added to the within-therapist analysis (i.e., Model 2).

Table 1

Predicting Patient Outcomes From Various Metrics of Time

Model	Predictor	b	SE	df	t	p
Model 1	Patient treatment date	.035	.012	4,974	2.97	.003
Model 2	Within-therapist time	.034	.016	4,974	2.04	.042
Model 3	Therapist start date	.019	.012	151	1.52	.129

Note. b = unstandardized multilevel model regression coefficient (which reflects change in standardized units, as the outcome [d] is in standardized units); SE = standard error; df = degrees of freedom; t = t-value; p = p-value. n = 5,128 patients seen by n = 153 psychotherapists.

ical time. A significant fixed effect for chronological time was observed ($b_{10} = 0.33$, $p = .034$), indicating that baseline OQ total scores increased by 0.33 points per year. Given this observed systematic increase in baseline severity across time, Models 1, 2, and 3 were reestimated with baseline OQ total scores included as a covariate. The significance tests for the three fixed effects previously reported did not change. Patient start date (Model 1a) and within-therapist time (Model 2a) remained significant predictors of patient-level change ($b_{10} = 0.029$, $p = .014$ and $b_{01} = 0.050$, $p = .002$, for calendar year and within-therapist time, respectively) while therapist start date (Model 3a) remained nonsignificant ($b_{01} = 0.0029$, $p = .810$) when controlling for baseline OQ scores.

The final model tested for a possible interaction between therapists' initial experience level (i.e., trainee, provisionally licensed, licensed professional) and changes across within-therapist time (i.e., Model 2). The interaction term was not significant in this model ($p > .100$), indicating that the within-therapist improvement was generally uniform across the three levels of initial experience.

Discussion

The current study examined changes in patient outcomes within a mental health agency making a concerted effort to improve clinicians' outcomes. This agency engaged in ROM coupled with the planful application of feedback through deliberate practice. Analysis of 7 years of outcome data drawn from 5,128 patients seen by 153 psychotherapists indicated that this agency's outcomes were indeed improving. A small but statistically significant increase in effectiveness was detected, reflecting an improvement of $d = 0.035$ each year. Thus, the expected outcome for patients receiving treatment at this agency increased over the course of time. It is worth noting that although this is a small effect, it is consistent with both the deliberate practice and therapist effects literature, highlighting the potentially large cumulative effect of small changes accrued over time (Ericsson, Krampe, & Tesch-Römer, 1993; Imel, Sheng, Baldwin, & Atkins, 2015).

Subsequent analyses sought to assess whether the observed agency-level improvement was due to the psychotherapists themselves improving over time (i.e., within-therapist improvement) or due to the agency hiring higher-performing psychotherapists over time. In contrast to previous reports (i.e., Goldberg et al., 2016), the psychotherapists in the current sample showed improved outcomes over time. A given psychotherapist's outcomes on average improved by $d = 0.034$ each year, with patients seen later by a given therapist showing larger reductions in psychological distress over the course of therapy. The effect size reflecting this improve-

ment in within-therapist outcomes over time is notably three times larger than the effect size reflecting decreases in effectiveness reported previously by Goldberg et al. (2016). The third model explored the possibility that the agency-level improvement in outcomes was due to the agency over time hiring better therapists (and perhaps also losing lower performing therapists). It did not appear that therapists who were hired later had better overall outcomes than therapists hired earlier, although the effect size was in this direction ($d = 0.019$). It is worth noting that as this coefficient corresponds to a therapist-level (i.e., level 2) predictor that may have been underpowered, given the sample included only 153 psychotherapists.

Additional models sought to assess the possibility that results were confounded by baseline severity or therapists' initial level of experience. Although levels of baseline severity appeared to increase over the course of chronological time, changes in baseline severity did not appear to explain the observed improvements in agency-level or within-therapist level outcomes. Similarly, the within-therapist improvement in outcomes across time was not moderated by therapists' initial level of experience.

What can be gleaned from this single agency case study in regards to the potential of improving outcomes in mental health care? For one, these results provide an example of how the quality of mental health care can be improved despite the difficulties discussed by Tracey et al. (2014) and the report from Goldberg et al. (2016) suggesting that psychotherapists in routine practice do not appear to improve with mere access to feedback alone. The current findings suggest that psychotherapists *can* improve over time. Further, this case study provides some preliminary indications as to possible conditions that may facilitate this improvement.

There were several specific actions taken by the agency under study that may have contributed to the psychotherapist improvements observed. Psychotherapists not only had access to their patients' outcomes, but they also had an ongoing forum for discussing cases, in particular cases that were not showing the expected improvements over time. In collaboration with fellow clinicians and under the guidance of an external consultant well-
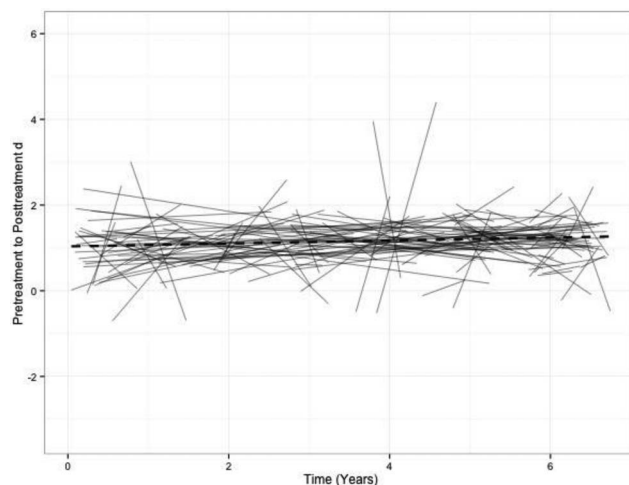

Figure 2. Changes in patient-level prepost d across time (patient treatment date). Solid lines reflect within-therapist changes in prepost d and dashed line reflects agency-level changes.

versed in routine outcome monitoring and deliberate practice, psychotherapists at this agency could gather concrete suggestions for ways of working with difficult cases. In many ways, the systems put in place by this agency reflect the kinds of agency-level cultural changes that may be helpful in creating the "overall climate and culture for change" (Ferlie & Shortell, 2001, p. 287) or a "culture of excellence" (Miller & Hubble, 2011, p. 25). Of course, lacking formal measurement of agency-level culture, it is impossible for us to draw firm conclusions regarding whether these changes did in fact occur.

Perhaps the clearest implication for psychotherapists in practice is to remain engaged in critical evaluation of one's practice throughout the course of one's career and to make efforts to improve. Mere access to outcomes does not appear to be sufficient for improvement. Rather, psychotherapists are encouraged to monitor their outcomes with an eye toward those cases that are not improving—the cases by which a psychotherapist may be discouraged, or confused, or frustrated. The goal is not simply to help the particularly difficult patient, but to develop skills that will improve performance in the future. Ideally in consultation with other professions, psychotherapists can reveal areas for growth and the opportunity to increase skill level.

Despite the importance of the results in this particular agency, caution is also needed. There are many barriers to implementing ROM in agencies (Boswell, Kraus, Miller, & Lambert, 2015; Kraus et al., 2016; Miller, Hubble, Chow, & Seidel, 2015; Wampold, 2015). Clinicians may feel threatened or micromanaged by the implementation of outcomes monitoring. In the current sample, several therapists left this agency due to incompatibility with the quality improvement efforts. And even for the individual clinician, care is warranted when attributing importance to the trajectories of individual patients. Patient characteristics account for the lion's share of variability in psychotherapy outcomes, with psychotherapists by contrast, accounting for a relatively small proportion of variance (3% to 7%; Baldwin & Imel, 2013). Thus, the trajectories of individual cases should be reviewed when not progressing based on the possibility that something important can be learned for improving the psychotherapist's practice (as was done in the current case study). And, at once, clinicians and agencies should be cautious about not using outcomes from a single (or even several) cases to form firm conclusions regarding the effectiveness of that particular clinician. The vast majority of variability in patients' outcomes is not due to the therapists, after all (Baldwin & Imel, 2013).

These results suggest several directions for future research. It may be worthwhile examining at a more fine-grained level, precisely how psychotherapists apply feedback to their clinical work in ways that improves outcomes. This work could involve examination of psychotherapy process variables (e.g., therapeutic alliance; Goldberg, Davis, & Hoyt, 2013; Horvath, Del Re, Fluckiger, & Symonds, 2011) alongside qualitative research exploring the ways in which psychotherapists use feedback from ROM and outcome-directed consultation. Additional work could examine which of the variables suggested to improve outcomes in the training literature in psychotherapy and medicine were present in this agency (e.g., rehearsing difficult conversations, using simulated case vignettes, reflecting and planning ahead; Bjork & Bjork, 2011; Chow et al., 2015; Issenberg et al., 2002; Miller & Hubble, 2011). Quantitative and qualitative work could help elucidate the extent to which this agency and agencies using similar

quality improvement methods are indeed creating a culture of change within the agency. It would likewise be important to replicate the current findings in an independent sample drawn from another agency (perhaps even a very different setting such as a private group practice) using similar feedback and consultation methods. Further refinement of the current study could examine therapist factors (e.g., therapist personality, therapist professional self-doubt; Nissen-Lie, Monsen, Ulleberg, & Rønnestad, 2013) that influence to what extent the kinds of training opportunities provided at this agency are helpful in improving outcomes.

Several limitations are worth considering. The most obvious limitation is that this study is a case study and therefore subject to the limitations of this design (e.g., limitations to generalizability). It will thus be important to replicate these findings in another agency employing similar quality improvement efforts. Another significant drawback is the lack of a control group with which to compare the observed improvements. The lack of a nonintervention comparison group, or data from the current agency prior to their attempts at improving outcomes leaves ambiguous the cause of the observed changes. The agency-level improvements, for example, could have been due to changes in patient or therapist characteristics that were not examined (i.e., characteristics beyond baseline severity or therapists initial training level). Further, although Goldberg et al.'s (2016) results suggest that psychotherapists do not improve with mere access to feedback alone, it is possible that other samples of psychotherapists may show normative increases in outcomes even in the absence of the training efforts made by the agency in the current study. While logistically difficult, more definitive evidence in support of the interventions used at this agency would be provided through a randomized controlled design in which some agencies are provided with ongoing feedback monitoring and consultation and others are not. As discussed, a lack of formal measurement of agency-level cultural changes limits our ability to conclude that this type of change did in fact occur.

Other limitations of the current work relate to the sample itself. While the sample included a large number of patients and a fairly large number of psychotherapists, analyses including therapist-level (i.e., level 2) predictors were likely underpowered and may have failed to detect significant effects (e.g., improved outcomes at baseline for psychotherapists hired later). Relatively limited data were available on the psychotherapists themselves (e.g., specialized trainings they may have received before or during this study). It would have been informative to examine whether therapist characteristics could predict response to feedback and consultation.

In conclusion, the current study suggests promising directions for future quality improvement efforts in the context of mental health care. As increasing attention is placed on the effectiveness and efficiency of treatment, the efforts made by the agency under study may provide a template from which to build. At the very least, this case study provides evidence that improvement at both the agency- and therapist-levels are possible.

References

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., . . . Wampold, B. E. (2013). Cognitive-behavioral therapy

versus other therapies: Redux. *Clinical Psychology Review, 33,* 395–405. http://dx.doi.org/10.1016/j.cpr.2013.01.004

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77,* 203–211. http://dx.doi.org/10.1037/a0015235

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). Hoboken, NJ: Wiley.

Baldwin, S. A., Imel, Z. E., & Atkins, D. C. (2012). The influence of therapist variance on the dependability of therapists' alliance scores: A brief comment on "The dependability of alliance assessments: The alliance-outcome correlation is larger than you think" (Crits-Christoph et al., 2011). *Journal of Consulting and Clinical Psychology, 80,* 947–951. http://dx.doi.org/10.1037/a0027935

Barkham, M., Mellor-Clark, J., & Stiles, W. B. (2015). A CORE approach to progress monitoring and feedback: Enhancing evidence and improving practice. *Psychotherapy, 52,* 402–411. http://dx.doi.org/10.1037/pst0000030

Batalden, P. B., & Davidoff, F. (2007). What is "quality improvement" and how can it transform healthcare? *Quality and Safety in Health Care, 16,* 2–3. http://dx.doi.org/10.1136/qshc.2006.022046

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

Bludworth, J. L., Tracey, T. J. G., & Glidden-Tracey, C. (2010). The bilevel structure of the Outcome Questionnaire-45. *Psychological Assessment, 22,* 350–355. http://dx.doi.org/10.1037/a0019187

Boswell, J. F., Kraus, D. R., Castonguay, L. G., & Youn, S. J. (2015). Treatment outcome package: Measuring and facilitating multidimensional change. *Psychotherapy, 52,* 422–431. http://dx.doi.org/10.1037/pst0000028

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25,* 6–19. http://dx.doi.org/10.1080/10503307.2013.817696

Brown, G. S., Simon, A., Cameron, J., & Minami, T. (2015). A collaborative outcome resource network (ACORN): Tools for increasing the value of psychotherapy. *Psychotherapy, 52,* 412–421. http://dx.doi.org/10.1037/pst0000033

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52,* 337–345. http://dx.doi.org/10.1037/pst0000015

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum, Inc.

Crits-Christoph, P., Connolly Gibbons, M. B., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance-outcome correlation is larger than you think. *Journal of Consulting and Clinical Psychology, 79,* 267–278. http://dx.doi.org/10.1037/a0023668

Duncan, E. A. S., & Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. *BMC Health Services Research, 12,* 96. http://dx.doi.org/10.1186/1472-6963-12-96

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406. http://dx.doi.org/10.1037/0033-295X.100.3.363

Ferlie, E. B., & Shortell, S. M. (2001). Improving the quality of health care in the United Kingdom and the United States: A framework for change. *The Milbank Quarterly, 79,* 281–315. http://dx.doi.org/10.1111/1468-0009.00206

Forand, N. R., DeRubeis, R. J., & Amsterdam, J. D. (2013). Combining medication and psychotherapy in the treatment of major mental disorders. *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 735–774). Hoboken, NJ: Wiley.

Goldberg, S. B., Davis, J. M., & Hoyt, W. T. (2013). The role of therapeutic alliance in mindfulness interventions: Therapeutic alliance in mindfulness training for smokers. *Journal of Clinical Psychology, 69,* 936–950. http://dx.doi.org/10.1002/jclp.21973

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63,* 1–11. http://dx.doi.org/10.1037/cou0000131

Gøtzsche, P. C., Young, A. H., & Crace, J. (2015). Does long term use of psychiatric drugs cause more harm than good? *British Medical Journal, 350,* h2435. http://dx.doi.org/10.1136/bmj.h2435

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48,* 9–16. http://dx.doi.org/10.1037/a0022186

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41,* 159–164. http://dx.doi.org/10.1037/0003-066X.41.2.159

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. http://dx.doi.org/10.1037/0003-066X.51.10.1059

Hubble, M. A., Duncan, B. L., & Miller, S. D. (1999). *The heart and soul of change: What works in therapy*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11132-000

Imel, Z. E., Sheng, E., Baldwin, S. A., & Atkins, D. C. (2015). Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy. *Psychotherapy, 52,* 329–336. http://dx.doi.org/10.1037/pst0000023

International Center for Clinical Excellence. (2012). *Partners for Change Outcome Management System (PCOMS): International Center for Clinical Excellence*. Retrieved from http://legacy.nreppadmin.net/ViewIntervention.aspx?id=249.

Issenberg, S. B., McGaghie, W. C., Gordon, D. L., Symes, S., Petrusa, E. R., Hart, I. R., & Harden, R. M. (2002). Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching and Learning in Medicine, 14,* 223–228. http://dx.doi.org/10.1207/S15328015TLM1404_4

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., . . . Oxman, A. D. (2012). Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews, 6,* CD000259.

Kopta, M., Owen, J., & Budge, S. (2015). Measuring psychotherapy outcomes with the Behavioral Health Measure-20: Efficient and comprehensive. *Psychotherapy: Theory, Research, and Practice, 52,* 442–448. http://dx.doi.org/10.1037/pst0000035

Kraus, D. R., Bentley, J. H., Alexander, P. C., Boswell, J. F., Constantino, M. J., Baxter, E. E., & Castonguay, L. G. (2016). Predicting therapist effectiveness from their own practice-based evidence. *Journal of Consulting and Clinical Psychology, 84,* 473–483. http://dx.doi.org/10.1037/ccp0000083

Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy: Theory, Research, and Practice, 52,* 381–390. http://dx.doi.org/10.1037/pst0000027

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. http://dx.doi.org/10.1037/0022-006X.69.2.159

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61,* 165–174. http://dx.doi.org/10.1002/jclp.20113

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., . . . Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire-45.* Orem, UT: American Professional Credentialing Services.

Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy: Theory, Research, and Practice, 51,* 467–481. http://dx.doi.org/10.1037/a0034332

Lynn, J., Baily, M. A., Bottrell, M., Jennings, B., Levine, R. J., Davidoff, F., . . . James, B. (2007). The ethics of using quality improvement methods in health care. *Annals of Internal Medicine, 146,* 666–673. http://dx.doi.org/10.7326/0003-4819-146-9-200705010-00155

Miller, S. D., Duncan, B. L., & Hubble, M. A. (2005). Outcome-informed clinical work. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (pp. 84–102). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/med:psych/9780195165791.003.0004

Miller, S. D., & Hubble, M. (2011). The road to mastery. *Psychotherapy Networker, 35,* 22–60.

Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow. *Psychotherapy: Theory, Research, and Practice, 50,* 88–97. http://dx.doi.org/10.1037/a0031097

Miller, S. D., Hubble, M. A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy: Theory, Research, and Practice, 52,* 449–457. http://dx.doi.org/10.1037/pst0000031

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76,* 116–124. http://dx.doi.org/10.1037/0022-006X.76.1.116

National Alliance on Mental Illness. (2011). *State mental health cuts: The continuing crisis.* Arlington, VA: Author.

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research, 23,* 86–104. http://dx.doi.org/10.1080/10503307.2012.735775

Owen, J., Wampold, B. E., Kopta, M., Rousmaniere, T., & Miller, S. D. (2016). As good as it gets? Therapy outcomes of trainees over time. *Journal of Counseling Psychology, 63,* 12–19. http://dx.doi.org/10.1037/cou0000112

Parry, G. (1992). Improving psychotherapy services: Applications of research, audit and evaluation. *The British Journal of Clinical Psychology, 31,* 3–19. http://dx.doi.org/10.1111/j.2044-8260.1992.tb00964.x

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78,* 298–311. http://dx.doi.org/10.1037/a0019247

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760. http://dx.doi.org/10.1037/0003-066X.32.9.752

Snell, M. N., Mallinckrodt, B., Hill, R. D., & Lambert, M. J. (2001). Predicting counseling center clients' response to counseling: A 1-year follow-up. *Journal of Counseling Psychology, 48,* 463–473. http://dx.doi.org/10.1037/0022-0167.48.4.463

Stiles, W. B., Barkham, M., Connell, J., & Mellor-Clark, J. (2008). Responsive regulation of treatment duration in routine practice in United Kingdom primary care settings: Replication in a larger sample. *Journal of Consulting and Clinical Psychology, 76,* 298–305. http://dx.doi.org/10.1037/0022-006X.76.2.298

Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological Medicine, 36,* 555–566. http://dx.doi.org/10.1017/S0033291706007136

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69,* 218–229. http://dx.doi.org/10.1037/a0035099

Unsworth, G., Cowie, H., & Green, A. (2012). Therapists' and clients' perceptions of routine outcome measurement in the NHS: A qualitative study. *Counselling and Psychotherapy Research, 12,* 71–80. http://dx.doi.org/10.1080/14733145.2011.565125

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges. *Psychotherapy: Theory, Research, and Practice, 52,* 458–462. http://dx.doi.org/10.1037/pst0000037

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

Youn, S. J., Castonguay, L. G., Xiao, H., Janis, R., McAleavey, A. A., Lockard, A. J., . . . Hayes, J. A. (2015). The Counseling Center Assessment of Psychological Symptoms (CCAPS): Merging clinical practice, training, and research. *Psychotherapy: Theory, Research, and Practice, 52,* 432–441. http://dx.doi.org/10.1037/pst0000029