Psychotherapy
2015, Vol. 52, No. 4, 422–431

0033-3204/15/$12.00 http://dx.doi.org/10.1037/pst0000028

Treatment Outcome Package: Measuring and Facilitating Multidimensional Change

James F. Boswell
University at Albany, State University of New York

David R. Kraus
Outcome Referrals, Inc., Framingham, Massachusetts

Louis G. Castonguay and Soo Jeong Youn
Pennsylvania State University

The Treatment Outcome Package (TOP; D. R. Kraus, Seligman, & Jordan, 2005) is a multidimensional routine progress and outcome measure developed for use in diverse naturalistic practice settings. In this article, we (a) provide a brief review and summary of the extant psychometric and research support for the TOP, (b) provide examples of the TOP's use in clinical training and practice, and (c) discuss the implications of the TOP for future psychotherapy training, research, and practice. In particular, we focus on the implications of risk-adjusted progress monitoring for systems of care and mental health care decision making.

Keywords: outcome monitoring, feedback, measurement-based care, treatment outcome package

It has been almost 20 years since Howard and colleagues (Howard, Moras, Brill, Martinovich, & Lutz, 1996) published their seminal paper on the evaluation of progress in psychotherapy. Approximately 2 years prior to this, the Society for Psychotherapy Research and the American Psychological Association convened a Core Battery Conference on the topic of routine progress and outcome monitoring, which culminated in a book that outlined recommendations for measuring patient change (Strupp, Horowitz, & Lamber, 1997). Recommendations included characteristics such as the ability to measure subjective distress, specific symptomatic states, and social and interpersonal functioning, as well as feasible integration into routine clinical practice, the ability to track multiple administrations, and sensitivity to change. These recommendations heavily influenced the development of the Treatment Outcome Package (TOP; D. R. Kraus, Seligman, & Jordan, 2005).

The TOP is a multidimensional outcome measure designed for use in naturalistic treatment settings to track behavioral health treatment progress and outcomes (D. R. Kraus & Castonguay, 2010; Kraus et al., 2005). It is utilized by clinicians and systems of care in a variety of treatment and other real-world settings to promote both patient-centered care and program evaluation. The TOP has three age-specific versions (child, adolescent, and adult). Across all versions, key characteristics include (a) assessment of multiple symptom and functioning domains, including quality of life; (b) real-time graphical feedback report that consists of standardized scores relative to the nonclinical population and actuarially derived alerts for patients who are at risk of hospitalization and deterioration/treatment failure; (c) risk adjustment based on case-mix variables that include medical morbidities and life stressors; and (d) benchmarking that allows for more direct comparisons to be made among providers and systems of care.

In this article, we (a) provide a brief review and summary of the extant psychometric and research support for the TOP, (b) provide examples of the TOP's use in clinical training and practice, and (c) discuss the implications of the TOP for future psychotherapy training, research, and practice. In particular, we focus on the implications of risk-adjusted progress monitoring for systems of care and decision making.

Psychometric Research and Support

The TOP has been developed and revised over the past 20 years with extensive input and feedback from mental health consumers and providers (Youn, Kraus, & Castonguay, 2012). The adult TOP Clinical Scales consist of 58 items assessing 12 symptom and functional domains (risk-adjusted on the basis of the case-mix assessment): work functioning, sexual functioning, social conflict, depression, panic (somatic anxiety), psychosis, suicidal ideation, violence, mania, sleep, substance abuse, and quality of life. Case-mix variables are patient characteristics that are likely related to treatment outcome, such as severity, medical morbidity, and life stress. Domain-specific symptom severity is quantified as the risk-adjusted individual z scores for each clinical scale (lower scores indicating healthier functioning). Global symptom severity is also assessed by summing all items or by averaging the z scores (i.e., standard deviation units relative to the general population mean) across each of the clinical scales. When calculated using raw (nonstandardized) scores, higher TOP total scores indicate healthier functioning.

James F. Boswell, Department of Psychology, University at Albany, State University of New York (SUNY), Albany, New York; David R. Kraus, Outcome Referrals, Inc., Framingham, Massachusetts; Louis G. Castonguay and Soo Jeong Youn, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania.

Correspondence concerning this article should be addressed to James F. Boswell, Social Science 307, Department of Psychology, University at Albany, SUNY, 1400 Washington Avenue, Albany, NY 12222. E-mail: jboswell@albany.edu

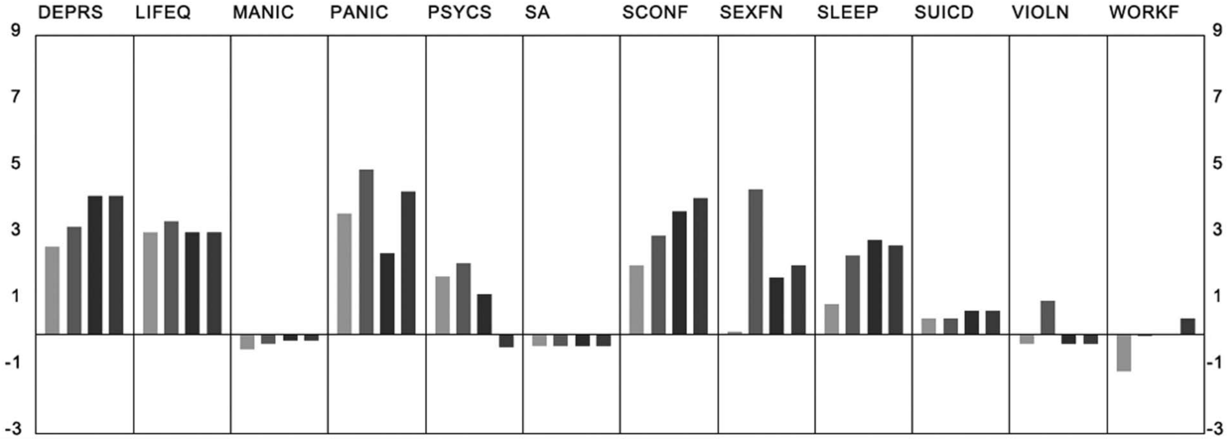
Figure 1 provides a sample TOP client report that includes four repeated outcome assessments between 2007 and 2008. All scores are standardized to the general population, therefore, easing the interpretation of the findings: Zero represents the general population average, and positive scores on the *Y* axis represent the number of standard deviations more severe than the general population average. If the *Y* axis score is negative, it represents a potential patient strength.

Presenting TOP scores as standardized scores obviates the need for users to remember or look up normal or clinical cutoffs. The feedback report highlights clinical significance in a color coding/shades of gray schematic that draws the clinician's attention to domains that warrant special attention. For example, in Figure 1, this patient's highest score occurred at the second administration on May 8, 2008, on the Panic domain, with a score nearly 5 standard deviations above the norm. Most TOP scores are log-normally distributed with little risk of ceiling effects (D. R. Kraus et al., 2005), and as such, the *z* score can represent highly elevated scores in the pathological range. The feedback chart and graph represented in Figure 1 can display up to 20 assessment points, which allows a clinician to monitor treatment progress even across multiple episodes of care. Although a total score can be calculated and has previously been used in research (e.g., Zack et al., 2015), the TOP total score is not routinely presented to clinicians because pilot testing of its display appeared to take clinicians' focus off of the unique changes that were occurring at the domain level.

Factor Structure and Reliability

The TOP was initially developed on the basis of a large pool of items that spanned common symptom and functional domains. Items were subsequently edited or eliminated incorporating clinician and patient feedback, based on an iterative process that helped ensure face validity. Following preliminary factor analyses, the penultimate iteration included 93 mental health symptom, functional, and quality-of-life items. These were administered to a large sample (N = 19,801) of newly admitted psychiatric patients presenting for care in nearly 400 different treatment services across the United States (D. R. Kraus et al., 2005). This sample was divided into five random subsamples to support cross-validation studies. A series of confirmatory factor models across the subsamples were tested and compared using multiple fit indices, chi-square difference tests, and interpretability. Strong support was provided (over multiple subsamples) for the 12 factor solution via diverse goodness-of-fit statistics: comparative fit index (.95), Tucker-Lewis index (.94), and root-mean-square error of approximation (.04).

The TOP clinical scales possess good to excellent internal consistency reliability (alphas = .69–.93), as well as test–retest reliability (ICCs = .87–.94), with the exception of the Mania subscale. The Mania subscale has a test–retest ICC of .76 and an internal consistency that has ranged between .55 in outpatient settings (N = 472,969) and .70 in long-term inpatient settings, where the incidence of mania is more likely (N = 5,107). The range in reliability measures for the Mania scale is likely related to the relatively low incidence of manic symptoms in outpatient settings and the apparent bimodal distribution of items such as "Feeling on top of the world." Such items do not have a linear relationship to health, because feeling on top of the world all of the time might be an indication of mania, while never feeling on top of the world might be an indication of depression. The TOP total score has demonstrated excellent internal consistency reliability (e.g., alpha = .94) in multiple studies (Boswell, Kraus, Nordberg,



	DEPRS	LIFEQ	MANIC	PANIC	PSYCS	SA	SCONF	SEXFN	SLEEP	SUICD	VIOLN	WORKF
09/05/2008	4.14	3.06	-0.18	4.28	-0.37	-0.34	4.07	2.04	2.65	0.70	-0.28	0.47
07/10/2008	4.14	3.06	-0.18	2.42	1.20	-0.34	3.66	1.68	2.82	0.70	-0.28	-0.02
05/08/2008	3.20	3.38	-0.29	4.95	2.12	-0.34	2.93	4.33	2.36	0.47	0.97	-0.03
12/27/2007	2.61	3.06	-0.43	3.61	1.73	-0.34	2.04	0.05	0.88	0.47	-0.28	-1.12
	***	***		***			***	**	**			

Figure 1. Sample Treatment Outcome Package client feedback report with four administrations. Scores represent the number of standard deviations above the normal population. DEPRS = Depression; LIFEQ = Quality of Life; PSYCS = Psychosis; SA = Substance Abuse; SCONF = Social Conflict; SEXFN = Sexual Functioning; SUICD = Suicide; VIOLN = Violence; WORKF = Work Functioning.

& Castonguay, 2009; D. R. Kraus et al., 2015; Zack et al., 2015), yet a formal examination of test–retest reliability has not been conducted.

Validity

The TOP subscales and total score have demonstrated good to excellent convergent validity based on correlations with existing measures of similar constructs (Boswell et al., 2009; D. R. Kraus et al., 2005; see Table 1). In addition, criterion validity has been demonstrated on the basis of the TOP's ability to discriminate between members of the general population and clinical samples (D. R. Kraus et al., 2005). In a series of logistic regression analyses, TOP scores were able to correctly classify 80%–89% of participants as being from either a clinical or the general population (average = 84%). Interestingly, Quality of Life scores were the most consistent predictor of clinical versus nonclinical status.

Table 1
Correlations Between TOP Scores and Alternative Self-Report Measures of Similar Constructs

TOP subscale	Comparison measure	*r*
Depression	BDI-II	.92
	BSI Depression	.90
	MMPI-2 Depression	.73
Violence	BSI Hostility	.77
Social Conflict	BSI Phobic	.52
	BSI Interpersonal	.42
	BASIS-32 Relationships	.60
Quality of Life	MMPI-2 Depression	.60
	BASIS-32 Daily Role	.65
	SF-36 Vitality	.57
Sleep	SF-36 Physical Functioning	.29
	SF-36 Role Physical	.39
Sexual Functioning	SF-36 Role Physical	.41
	SF-36 Vitality	.47
Work Functioning	BASIS-32 Daily Role	.51
Psychosis	BSI Psychoticism	.72
	BASIS-32 Psychosis	.80
	MMPI-2 PA	.36
	MMPI-2 SC	.28
Panic	BSI Somatization	.72
	BSI Phobic	.82
	BSI Anxiety	.41
	BASIS-32 Depression/Anxiety	.73
Mania	MMPI-2 Hypomania	.43
	SF-36 Mental Health	.32
Suicidality	BDI-II	.60
	BSI Depression	.69
	BASIS-32 Depression/Anxiety	.72
	SF-36 Mental Health	.69
TOP total	BSI	.91
	BASIS-32	.89

Note. TOP = Treatment Outcome Package; BDI = Beck Depression Inventory (Beck, Steer, & Brown, 1996); BSI = Brief Symptom Inventory (Derogatis, 1975); MMPI-2 = Minnesota Multiphasic Personality Inventory (Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001); BASIS-32 = Behavior and Symptom Identification Scale; SF-32 = Short-Form Health Survey (McHorney, Ware, & Raczek, 1993); PA = Paranoia; SC = Schizophrenia. See D. R. Kraus et al. (2005) for all correlations.

Floor and Ceiling Effects

An implicit goal of most measurement tools is the ability to assess the full range of a construct. In addition, clinicians and researchers are interested in the level of health (or lack thereof) at baseline and some follow-up time point, as well as the magnitude of change between these two. Therefore, floor and ceiling effects should be addressed as much as possible. A total of $N = 216{,}642$ TOP administrations were analyzed for both floor and ceiling effects (D. R. Kraus et al., 2005). Data were analyzed for frequency counts of patients who scored at the theoretical maximum or minimum score of each TOP subscale. Results indicated minimal ceiling effects ($M = 1\%$), ranging between 0.1% (Work Functioning) to 4.0% (Quality of Life); therefore, demonstrating that clinically severe extremes are capable of being captured. The percentages of patients at the theoretical minimum (or floor) ranged between 2.8% (Quality of Life) and 59.1% (Violence; $M = 18\%$). It is important to remember that a standardized score of 1.0 represents the clinical cutoff, and a score of 0.0 is the general population average. The theoretical minimum for the TOP subscales can extend beyond 2 standard deviations below zero (−2.34); thus, the "healthy tails" are still well captured.

Sensitivity to Change

Several studies have provided support for the TOP's sensitivity to change (e.g., D. R. Kraus et al., 2005; Nordberg, Castonguay, Fisher, Boswell, & Kraus, 2014; Zack et al., 2015). In one study (D. R. Kraus et al., 2005), $N = 20{,}098$ adult patients were administered the TOP at the start of treatment and a median of 49 days later. The average initial *z* score for the sample was 1.04 standard deviations above the normative population (range = −.31 for Mania to 2.38 for Suicidality). Cohen's *d* effect sizes (baseline-to-follow-up change) ranged between $d = .16$ (Mania) and $d = .53$ (Depression). On the basis of reliable change index (RCI) criteria established by Jacobson, Roberts, Berns, and McGlinchey (1999), 91% of patients demonstrated reliable improvement on at least one TOP subscale. Because of differences in reliability estimates among subscales, each subscale and the total score require a different RCI. For example, Zack et al. (2015) calculated the RCI for the total score based on alpha = .94, RCI = $1.96 * SD(1 - r)^{1/2}$, resulting in an RCI of 26.7 points for the total score (baseline total score $M = 255.82$, $SD = 39.32$; post total score $M = 306.59$, $SD = 19.06$). Table 2 provides RCI scores for each TOP domain, calculated as reported by D. R. Kraus, Castonguay, Boswell, Nordberg, and Hayes (2011), RCI = $1.96 * SD(1 - r)^{1/2}$. RCIs for each domain are calculated by using the TOP standardized scores and the reliability estimates reported by D. R. Kraus et al. (2005).

Clinical Training and Practice Examples

The TOP is first and foremost a routine outcome-monitoring and feedback tool designed to enhance outcomes and reduce the risk of negative treatment effects. The TOP system is used in a number of settings, from large health service networks such as State Child Welfare jurisdictions and Massachusetts Medicaid, to large primary care practices and solo-practice clinicians in multiple countries. Depending on the setting, it can be administered at intake to aid in case conceptualization and treatment planning, at intake and posttreatment as an overall indicator of treatment effectiveness, or

Table 2
Reliable Change Index (RCI) for Each TOP Domain

TOP domain	RCI[a]
Sexual Functioning	.44
Work Functioning	.44
Violence	.46
Social Functioning	.39
Panic/Anxiety	.52
Substance Abuse	.44
Psychosis	.52
Quality of Life	.39
Sleep	.34
Suicidality	.46
Depression	.39
Mania	.98

Note. RCIs calculated as reported by D. R. Kraus et al. (2011); RCI = 1.96 * $SD(1 - r)^{1/2}$. TOP = Treatment Outcome Package.
[a] RCIs calculated using the TOP standardized scores and the reliability estimates reported in the work of D. R. Kraus et al. (2005).

at regular intervals (e.g., session to session) to more closely monitor treatment progress to make treatment adjustments. The TOP can be completed in various electronic formats, including Web-based assessments via computer, tablet, and smartphone (https://www.wellnesscheck.net), as well as via paper and pencil if that is preferred, with tools submitted by fax for scoring. Reports can be returned by e-mail, fax, or Web access. TOP scoring is provided as a service to clinicians and includes real-time client report scoring as well as monthly aggregate reports that include change over time and benchmarks to similar programs, services, or clinicians from within the state or across the country. There are myriad ways that the TOP has been used in clinical training and practice settings, and we highlight a few examples below.

Practice Research Networks

Due in part to it being designed for use in real-world settings (based on extensive stakeholder feedback), the TOP has been a natural anchoring point for practice research networks (PRNs) in both broad community and training clinic settings. For example, the TOP has served as the primary outcome measure in multiple Pennsylvania Psychological Association (PPA) PRN investigations (e.g., Castonguay, Boswell, Zack et al., 2010; Castonguay et al., 2014). In addition to providing valuable information on community treatment outcomes, once a routine outcome assessment tool such as the TOP has been adopted, it becomes easier (relatively speaking) to integrate more rigorous process assessment and research designs (e.g., experimental manipulations). For example, the PPA-PRN recently completed a study in which community clinicians were randomly assigned to either a condition that involved receiving routine feedback on patient outcomes and their technique use or a condition with no such feedback (see Koerner & Castonguay, 2015).

As further evidence of the capacity of routine progress and outcome monitoring tools to facilitate science–practice integration, the TOP has played a crucial role in training clinic PRNs. For example, the Pennsylvania State University Psychological Clinic functions as both a community mental health center (CMHC) and a graduate training clinic. As part of routine practice, patients are asked to complete the TOP at intake and prior to each psychotherapy appointment using Internet-connected devices. Patient responses are automatically processed and scored, and a feedback report is electronically returned to the therapist for review typically within a minute of completion. The therapist is directly alerted by e-mail that a report is ready and this report highlights the need for urgent care when it is recommended. This is an invaluable tool for several reasons. Research has convincingly demonstrated that outcome monitoring and feedback enhances patient outcomes and reduces the risk of deterioration (Lambert & Shimokawa, 2011). Furthermore, by using the TOP with their training cases, trainee-therapists have a standardized method for monitoring their own progress not only with individual patients over time, but over the course of a specific practicum or years within the training program. Early routine utilization of standardized assessment and the integration of empirically derived feedback into the treatment process itself, in our opinion, also foster one's professional identity as a scientist-practitioner (Boswell, Kraus, Miller, & Lambert, 2015; Boswell & McHugh, 2015).

As a graduate trainee, James F. Boswell (hereinafter referred to as *the therapist*) worked with a woman who presented with complex comorbidity; she met criteria for posttraumatic stress disorder (PTSD), major depressive disorder (recurrent), and panic disorder. Consistent with the diagnostic assessment, this patient demonstrated clinically elevated scores on several TOP domains, including Depression, Panic, Suicidality, Sleep, Social Conflict, Sexual Functioning, and Quality of Life. The patient and therapist agreed that the PTSD was primary, and relatively early in treatment began working toward the use of prolonged exposure (PE; Foa & Rothbaum, 1998). The patient completed the TOP biweekly and initially demonstrated symptom improvement, which was attributed both to remoralization (Howard, Lueger, Maling, & Martinovich, 1993) and the learning of basic coping skills. Prior to the implementation of PE, the patient was informed that she might notice an acute exacerbation in symptoms in the coming weeks. Consistent with direct report and observation, nearly all of the patient's TOP scores increased, and then varied dramatically over several subsequent administrations. The beginning of each session was spent reviewing the most recent TOP graphical feedback report, which appeared to validate the patient's experience. The following would be a typical introduction from the therapist: "I made a copy of your latest TOP progress report, and unless you would prefer to begin with something else, I would like us to review this together. I am wondering about your initial impressions." If this was unclear from the patient's response, a typical follow-up question was "Is this feedback consistent with your subjective experience, and if not, how do you understand the discrepancy?" Depending on the patient's response, the therapist would express his shared understanding or comment on other aspects deemed to be potentially important; for example, "it appears that several domains, such as your anxiety, have stabilized, but your mood continues to be quite low; perhaps we should discuss that further before reviewing your homework."

This particular patient's scores eventually began to stabilize and then decrease. Toward the end of the treatment, she commented that the regular review of her TOP scores was extremely important to her and provided comfort. She was informed that she might experience emotional dysregulation; this was consistent with her experience, and she watched the numbers validate this over time.

It is difficult to draw a firm conclusion based on a single case anecdote, but the use of the TOP in this instance positively impacted the individual patient and trainee.

Use of the TOP at the Penn State University Clinic not only helps students to acquire and develop their clinical skills, but can also allow them to conduct clinically relevant studies (including Masters' and doctoral theses). Having been transformed into a PRN infrastructure, this clinic has been the source of more than 20 studies and more than 30 publications and presentations (Castonguay, Pincus, & McAleavey, 2015). A number of these studies have included the TOP, which again was completed as part of clients' clinical routine and as an intrinsic component of students' training—thereby seamlessly confounding clinical, training, and research tasks (Castonguay, 2011). As one study example, Nordberg et al. (2014) found that clients who began treatment with a high level of severity on the TOP Depression scale split into two groups over the course of treatment—those who remained high in symptom severity and those who demonstrated rapid change. The use of other TOP scores (e.g., Social Conflict, Hostility, Sexual Functioning, Suicidality) also enabled the identification of empirically based predictors of who was likely to change and who was not likely to change while being treated at the clinic. The potential training and health care management (e.g., triage, referrals, provision of more intensive or longer treatment) implications of these findings are numerous. For example, given that individuals with elevated scores on both Depression and Social Conflict were more likely to be in the nonresponder group, such clients may require more intensive or adjunctive treatment. In addition, practicum supervisors may prefer to assign less complicated cases to new trainees. Because these implications were derived from data collected at the clinic, there are no concerns about whether or not they will generalize to clients seen and therapists trained in this setting (or psychotherapy that is delivered in similar CMHC settings).

In addition, standardized feedback has the potential to help novice clinicians move beyond the typical "imposter syndrome." TOP reports provide profiles for the types of patients with whom a clinician (or system of care) works best, as well as those patients with whom they need to improve their clinical skills. Angtuaco, Castonguay, and Kraus (2005) compared TOP scores from PPA-PRN clinicians with trainee-therapists at the Penn State Psychological Clinic. Results showed that graduate trainee-therapists were particularly effective at improving patients' functioning (e.g., Sexual Functioning), as well as decreasing Suicidality and Violence. Trainees were pleased to learn that full-time clinicians could stand to learn something from them.

Another example of the TOP's ability to facilitate science–practice integration in a nonhierarchical collaboration among stakeholders can be found at Sundown Ranch, Inc., in Canton, Texas. Sundown Ranch is a residential adolescent substance abuse care facility that has used the TOP for over a decade. Benchmarked outcome reports documented that Sundown was achieving less than desired outcomes in the area of violence and anger. Consequently, the center sought additional training in rational–emotive behavior therapy (REBT) strategies for addressing anger and brought these back to the treatment center. For agencies in similar situations, researchers at Penn State have built a library of evidence-based treatments and principles building on the work of Castonguay and Oltmanns (2013) and Castonguay and Beutler (2006). For Sundown Ranch, over time, violence outcomes were observed to improve, lending some support to the REBT training intervention (Adelman, McGee, Power, & Hanson, 2005). Encouraged, the lead Sundown researcher/practitioner became interested in what might have facilitated these improvements. This led to a collaboration between the TOP developer, Sundown Ranch, and researchers (a graduate student and a faculty member at Penn State University; see Adelman et al., 2015) on a study that examined patient attachment and working alliance, with the TOP functioning as the primary outcome assessment. Zack et al. (2015) showed that a patient's attachment to the primary caregiver moderated the impact of the working alliance on treatment outcome, such that for youth with the poorest attachment history, working alliance had a stronger relationship with outcome.

Benchmarking and Outcome Variability

Potentially the most innovative features of the TOP include its application of risk adjustment and benchmarking capability. Variables that are beyond the control of the therapeutic process but nonetheless influence outcome are defined as case mix variables (Goldfield, 1999). Naturalistic outcome research typically lacks the controls used in efficacy research, and without measuring or controlling for case mix (e.g., medical morbidity, life stress, initial severity), comparisons can be misleading (Hsu, 1989). Case-mix-based risk adjustment greatly enhances the scientific and administrative value of naturalistic outcome data. It can facilitate fairer comparisons (see Saxon & Barkham, 2012) and benchmarking.

Numerous studies have uncovered the clinical reality of outcome variability among systems of care and individual clinicians (Baldwin & Imel, 2013; Wampold & Imel, 2015). In one example, D. R. Kraus et al. (2011) investigated the naturalistic treatment outcomes over multiple TOP domains (e.g., Depression, Substance Abuse, Sleep, Quality of Life) of 6,960 patients being treated by nearly 700 clinicians. The average initial z score for the sample was 1.36 standard deviations above the normative population (range $= -0.02$ for Mania to 2.26 for Depression). All patients completed the TOP on multiple occasions. Results demonstrated a differential pattern of provider performance depending on the problem domain; some providers demonstrated substantial effectiveness in Depression reduction, while others evidenced particular effectiveness in the Substance Abuse domain. Many providers demonstrated effectiveness over multiple problem domains, yet no providers demonstrated reliable effectiveness across all domains. A small, but notable 4% of the provider sample failed to demonstrate positive outcomes on any domain. These findings have been replicated in another large provider and patient sample (D. R. Kraus et al., 2015). Even with risk adjustment, substantial outcome variability was observed among therapists. Furthermore, hierarchical linear modeling–based correlations demonstrated stability in therapist performance across subsequent patients. With the exception of Mania, all domains documented significantly positive correlations between past and future performance. Those therapists whose outcomes were significantly above their peer-group average (labeled as *exceptional*) reliably remained above the 50th percentile with their next 30 patients. Most clinicians (88%) had areas in which they were exceptional. We believe that clinician burnout and patient outcomes can be improved by matching patient needs with clinicians who have corresponding strengths. Similarly, we believe that identified areas of relative poor performance highlight potentially important continuing education experiences. Finally, as previously argued by others (e.g., Baldwin & Imel, 2013; Castonguay, Barkham, Lutz, & McAleavey, 2013), areas of relative strong performance highlight the opportunity to learn more about what consistently high-performing clinicians are doing to achieve superior outcomes through sorely needed process research.

Child Welfare

The Annie E. Casey Foundation and the Duke Endowment have funded the development of a nonprofit charity (Kids Insight) that uses the TOP in child welfare to measure well-being, track outcomes, and make referrals for abused and neglected children. In a statewide child welfare system, the TOP was used to assess the effectiveness of residential facilities by measuring positive and negative outcomes over time on specific domains (D. R. Kraus, 2015). Sites in Ohio, North Carolina, Delaware, and Colorado have collected data to document that the vast majority of children in child welfare are improving while in care (e.g., 82% on Depression), and TOP profiles are significantly different for children, depending on the type of suspected abuse (Alexander, Baxter, Bentley, & Kraus, 2015). Of particular benefit to child welfare agencies, TOP scores were able to predict which children and adolescents were more likely to experience frequent moves in the following year (Alexander et al., 2015). For example, children with elevated Conduct, Violence, and Suicide scores were more likely to have multiple moves within the next year.

Implications and Future Directions

Many view routine outcome monitoring primarily as a transtheoretical tool for enhancing the benefit of psychological interventions for the individual patient. Although far from mutually exclusive motivations, other stakeholder groups (e.g., third-party payers, policymakers) view outcome/performance assessment as a means of ensuring accountability and "quality of care." The various stakeholder priorities seem to highlight a larger issue. Even if unfounded, there may be an implicit tension between enhancing benefit for the individual patient and experiencing pressure to demonstrate a certain magnitude of change (typically, in a relatively short period of time) on an identified outcome marker. Regardless, health care reform appears to have triggered considerably more rhetoric about accountability and cost containment (Boswell, Constantino, Kraus, Bugatti, & Oswald, 2015) that places routine outcome and performance assessment at the center. In considering the role of progress monitoring and feedback in the next decade, we focus on two areas: implementation and decision making.

Implementation

Similar to any new health care service practice, routine outcome monitoring and feedback involves a complex implementation process (Boswell et al., 2015). Individual clinicians or systems of care must first choose to adopt this practice. Once adopted, successful implementation is typically marked, at a minimum, by a demonstration of adequate fidelity and sustainability (Proctor et al., 2011). Over time, research has demonstrated increasingly positive clinician attitudes toward routine outcome monitoring and feedback (Bickman et al., 2000; Boswell, Constantino, & Kraus, 2014; Hatfield & Ogles, 2004). For example, in a large practitioner survey, Hatfield and Ogles (2004) found that 37.1% of respondents reported using some form of outcome assessment. More recently, Boswell et al. (2014) found that, in a sample of 22 CMHC clinicians, 19 expressed "strong agreement" with the statement that outcome monitoring tools provide rich clinical information. However, endorsed values and beliefs regarding routine outcome monitoring and feedback may not be consistent with its actual utilization (Hatfield & Ogles, 2007; Riemer, Rosof-Williams, & Bickman, 2005). In the work of Hatfield and Ogles (2004), among the reasons cited for not using outcome measures, practitioners most strongly endorsed "Adds too much paperwork," "Takes too much time," "Extra burden on clients," and "Feel it is not helpful." In a subsequent study, Hatfield and Ogles (2007) once again found that practical barriers (time, paperwork, burden) were more strongly endorsed by nonusers than was the utility of outcome measures and knowledge regarding their use. In other words, it appears as though practitioners value outcome monitoring in theory, yet many continue to struggle with perceived implementation barriers.

Nevertheless, when the pathway toward adoption is paved by internal (e.g., hospital administration) or external (e.g., insurance panel) mandates, clinician-level outcomes such as satisfaction, adherence, and competence cannot be taken for granted. Lack of attention to these implementation outcomes (facilitators and barriers) may ultimately undermine the impact of this intervention on individual patient outcomes. Consequently, we predict (and hope) that more research will be conducted in the area of routine outcome monitoring and feedback implementation at both the individual provider and systems level. Certainly more research is needed on stakeholder and system-level facilitators and barriers to progress monitoring feedback adoption, fidelity, and sustainability. It would be informative to investigate the impact of including frontline clinicians in discussions and planning regarding adoption and implementation of routine progress and outcome monitoring in specific service delivery settings. An innovative approach in health care practice has been the use of "communities of practice" (collaborative working and learning groups) that emphasize postimplementation sustainability in addition to initial adoption and adherence (Li et al., 2009; Ranmuthugala et al., 2011). Sustainable new practices are adapted and improved by clinicians in collaboration with knowledgeable experts, based on the needs of their population and organizational features (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005).

We hypothesize that clinician participation in such a collaborative working group would enhance organizational responsiveness to identified implementation barriers and promote increased fidelity and sustainability. For example, De Jong, van Sluis, Nugter, Heiser, & Spinhoven (2012) found that individual clinician factors—such as one's natural tendency to seek feedback, self-efficacy regarding the use of feedback, and expressed commitment to using feedback—moderated the impact of routine outcome-monitoring feedback on patient outcomes. At the practitioner level, we believe that "corrective experiences" observed and shared among clinicians in a "community of practice" would likely increase the use and value of systematic and regular empirical assessment of change. A corrective experience might involve a clinician who was previously apprehensive about outcome monitoring learning about a major clinical problem on the feedback report (e.g., sexual dysfunction) that her patient was uncomfortable disclosing face-to-face in session, and then sharing this experience with a colleague. It will be crucial to identify and help facilitate these transformational experiences.

At the systems level, we predict that more attention will be given to developing infrastructures that support the integration of routine outcome assessment with other health service process and outcome indicators (e.g., other electronic medical records). We also predict that infrastructures that foster collaboration and promote shared goals among different stakeholders (including clinicians) in determining how and for what purposes data should be collected (e.g., determining level of care, individual patient progress monitoring, program evaluation, training resources) are likely to foster the implementation and use of outcome monitoring (Castonguay, Youn, Xiao, Muran, & Barber, 2015; McAleavey, Lockard, Castonguay, Hayes, & Locke, 2015).

Treatment Decision Making

As argued elsewhere (Boswell, Constantino, et al., 2015), we believe that the implications of routine outcome monitoring and feedback can and should be expanded to enhance treatment responsiveness beyond the level of the individual patient and clinician to the level of the health care setting and broader system of care (Constantino, Boswell, Bernecker, & Castonguay, 2013). Health care systems are highly motivated to integrate service assessment protocols that can inform data-driven decision making (Berwick et al., 2008). When collected over many patients, being treated by many clinicians in diverse treatment settings, routine progress and outcome monitoring become psychotherapy's version of "big data" that can (and should) supplant financial claims data as a proxy for "outcome" data. We believe that behavioral health care systems and payers will increasingly rely on routinely collected outcome indicators to make key decisions regarding level of care, whom is selected to provide care, and duration of care. A pitfall to this approach is the potential overreliance on a single indicator (e.g., severity) to determine key health care decisions. In addition, the utility of any measurement-based care initiative will need to be carefully evaluated.

Sound decision making will be based, at a minimum, on the use of reliable and valid standardized measures and the application of risk-adjustment models. The critical significance of these minimum criteria is highlighted by increasing attention to pay-for-performance (Rosenthal, Frank, Li, & Epstein, 2005). Scanlon, Lindrooth, and Christianson (2008) offered a systems-level example; they studied the health care–seeking behavior of employees at a large manufacturing company. This company offered an incentive (no coinsurance payment) to employees who sought care at a preferred hospital (tiered hospital network). A preferred hospital was determined on the basis of several hospital-level safety and efficiency indicators established by the Leapfrog Group (http://www.leapfroggroup.org/). Results indicated that employees who were offered the incentive did indeed seek care at preferred hospitals at a significantly higher rate than they did at nonpreferred hospitals. Similarly, Chernew, Gowrisankaran, and Scanlon (2008) found that consumers who were given health plan report cards were willing to pay more (although only slightly) to avoid health plans with poor ratings. Importantly, these studies did not involve direct dissemination of provider performance information to either employers or consumers. Survey studies suggest that mental health consumers are interested in accessing such information (Boswell et al., 2014; Hibbard & Jewett, 1996). However, two Cochrane Reviews (Henderson & Henderson, 2010; Ketelaar et al., 2011) found mixed support for direct-to-consumer dissemination of provider performance information, and none of the studies included or cited in these reviews involved mental health care.

Henderson and Henderson (2010) conducted a systematic review of studies involving the provision of surgeon performance data to people considering elective surgery. Although a number of studies were available for initial review, no studies were deemed to be of sufficient quality to meet inclusion criteria (i.e., randomized controlled trial [RCT], quasi-randomized controlled trial, or controlled prepost design). In a subsequent review of studies involving the public release of medical provider performance data (Ketelaar et al., 2011), only four studies were deemed methodologically suitable (i.e., RCT, quasi-RCT, interrupted time series, or controlled prepost design). The release of provider performance data was linked to small improvements in acute myocardial infarction mortality rates, and a small increase in quality improvement activity within care organizations.

Despite robust evidence to support the integration of patient-focused routine outcome monitoring and feedback in psychotherapy, very few studies have used rigorous designs to test the impact of routine outcome data-driven systems-level interventions (including pay-for-performance). Randomized controlled trials are certainly needed, but other sources of evidence make an extremely convincing case. For example, research on therapist effects has consistently demonstrated variability in outcomes among clinicians (Baldwin & Imel, 2013), and the outcomes of higher- and lower-performing clinicians tend to be stable across patients (Wampold & Brown, 2005). Consequently, researchers can relatively reliably predict the likelihood that a given patient will experience a treatment benefit depending on the clinician to whom he or she is assigned or referred (D. R. Kraus et al., 2015; Wampold & Imel, 2015). DeRubeis et al. (2014) demonstrated the utility of predictive models (Personalized Advantage Index) derived from archival trial data for aiding patient-centered treatment selection. The same logic that is applied to treatment selection can be applied to provider selection. If differences among clinicians generally account for more outcome variance than do the interventions being delivered (M. S. Krause, Lutz, & Saunders, 2007), then patient-centered provider selection may ultimately yield even more (or at least incremental) benefit.

These are extremely complicated issues that should not be addressed with haste. For example, although considerable attention has been paid to consumer choice in recent years, it does not necessarily follow that performance "report cards" should be disseminated directly to consumers in a vacuum. If some clinicians struggle to use routine outcome feedback effectively (De Jong et al., 2012), then it is reasonable to expect that patients may struggle with this information as well (Hibbard, Peters, Slovic, Finucane, & Tusler, 2001). For example, direct-to-consumer dissemination of provider report cards may simply result in making the already stressful process of finding a clinician more confusing ("What do these numbers mean?"). On the other hand, patients may feel disconnected from important health care decisions if such information is being used by their insurance carrier, yet is being withheld from them. The role of the patient in outcome data informed shared decision making underscores the importance of future work in the area of implementation.

The Big Vision

We envision a near-future, integrated outcome management service for behavioral health providers, consumers, and referring professionals. A primary care physician (PCP) would routinely screen all of their patients (at least annually) for behavioral health concerns with data flowing into dashboards and reports that shape the annual physical conversation regarding quality of life, sleep, stress, coping, and if present, substance abuse, depression and other psychological problems. With outcome benchmarks, the PCP would know whether he or she has a stable track record of helping each patient him- or herself. If a referral were indicated, the outcome report would provide options of at least three area clinicians who have documented superior outcomes in treating clients with similar issues. Once referred, the new clinician would continue monitoring outcomes with the same system, and if elected by the patient, the outcome reports would flow to both PCP and clinician to help coordinate care and encourage collaboration.

By simply using such a system, PCPs and behavioral health professionals would naturally focus their practice in the areas where they achieve superior outcomes, thus enhancing population-level outcomes for the community. Aggregate data would help therapists and clinics identify the areas where they have the greatest opportunity to improve outcomes through continuing education, supervision, and online trainings. By monitoring what educational opportunities therapists use, the system would quickly document the relative outcomes of various educational opportunities with thousands of real-world, rigorous, controlled A-B additive designs. Feeding this information back to the network would create a multitiered learning system that would help each patient, therapist, clinic, and network naturally and seamlessly improve the quality of care.

Finally, providers who participate in such a system should be paid at higher rates because they are delivering more value to the self-insured employers and federal government that pay for this care. The value delivered through greater work productivity and lower medical care costs would more than offset the higher reimbursement rates to clinicians. The outcome-monitoring system data linked to claims data would be used to demonstrate and advocate for just this.

References

Adelman, R. W., Castonguay, L. G., Kraus, D. R., & Zack, S. E. (2015). Conducting research and collaborating with researchers: The experience of clinicians in a residential treatment center. *Psychotherapy Research, 25,* 108–120. http://dx.doi.org/10.1080/10503307.2014.935520

Adelman, R., McGee, P., Power, R., & Hanson, C. (2005). Reducing adolescent clients' anger in a residential substance abuse treatment facility. *Joint Commission Journal on Quality and Patient Safety, 31,* 325–327.

Alexander, P. C., Baxter, L., Bentley, J., & Kraus, D. R. (2015). *The use of the TOP in child welfare*. Manuscript in preparation.

Angtuaco, L. A., Castonguay, L. G., & Kraus, D. R. (2005, June). *The TOP: A core battery for the assessment of psychotherapy outcome*. Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Montreal, Quebec, Canada.

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* (pp. 258–297). New York: Wiley.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation.

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs, 27,* 759–769. http://dx.doi.org/10.1377/hlthaff.27.3.759

Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., & Karver, M. S. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice, 31,* 70–74. http://dx.doi.org/10.1037/0735-7028.31.1.70

Boswell, J. F., Constantino, M. J., & Kraus, D. R. (2014, April). *Promoting context-responsiveness with outcomes monitoring*. Paper presented at the Society for the Exploration of Psychotherapy Integration Conference, Montreal, Quebec, Canada.

Boswell, J. F., Constantino, M. J., Kraus, D. R., Bugatti, M., & Oswald, J. M. (2015). The expanding relevance of routinely collected outcome data for mental health care decision making. *Administration and Policy in Mental Health and Mental Health Services Research*. Advance online publication. http://dx.doi.org/10.1007/s10488-015-0649-6

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

Boswell, J. F., Kraus, D. R., Nordberg, S., & Castonguay, L. G. (2009, June). *The Treatment Outcome Package (TOP): An investigation of its validity*. Poster presented at the Society for Psychotherapy Research Conference, Santiago, Chile.

Boswell, J. F., & McHugh, R. K. (2015). Learning from research. In J. C. Norcross, G. R. VandenBos, & D. K. Freedheim (Eds.), *APA handbook of clinical psychology: Volume V. Education and profession*. Washington, DC: American Psychological Association Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2001). *Minnesota Multiphasic Personality Inventory—2 (MMPI–2): Manual for administration and scoring* (rev. ed.). Minneapolis, MN: University of Minnesota Press. http://dx.doi.org/10.1016/B0-08-043076-7/01294-8

Castonguay, L. G. (2011). Psychotherapy, psychopathology, research and practice: Pathways of connections and integration. *Psychotherapy Research, 21,* 125–140. http://dx.doi.org/10.1080/10503307.2011.563250

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: Approaches and applications. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of therapeutic change* (6th ed.). New York: Wiley.

Castonguay, L. G., & Beutler, L. E. (2006). *Principles of therapeutic change that work*. New York, NY: Oxford University Press.

Castonguay, L. G., Boswell, J. F., Zack, S. E., Baker, S., Boutselis, M. A., Chiswick, N. R., . . . Holtforth, M. G. (2010). Helpful and hindering events in psychotherapy: A practice research network study. *Psychotherapy: Theory, Research, Practice, Training, 47,* 327–344. http://dx.doi.org/10.1037/a0021164

Castonguay, L. G., & Oltmanns, T. F. (2013). *Psychopathology: From science to clinical practice*. New York, NY: Guilford Publications.

Castonguay, L. G., Pincus, A. L., & McAleavey, A. A. (2015). Practice research network in a psychology training clinic: Building an infrastructure to foster early attachment to the scientific-practitioner model. *Psychotherapy Research, 25,* 52–66. http://dx.doi.org/10.1080/10503307.2013.856045

Castonguay, L. G., Youn, S. J., McAleavey, A. A., Boswell, J. F., Boutselis, M., Braver, M., . . ., & O'Leary Wiley, M. (2014, June). *The use of empirically based techniques in day-to-day practice: Does it matter,* *and if so, for whom?* Paper presented at the Society for Psychotherapy Research Conference, Copenhagen, Denmark.

Castonguay, L. G., Youn, S. J., Xiao, H., Muran, J. C., & Barber, J. P. (2015). Building clinicians–researchers partnerships: Lessons from diverse natural settings and practice-oriented initiatives. *Psychotherapy Research, 25,* 166–184. http://dx.doi.org/10.1080/10503307.2014.973923

Chernew, M., Gowrisankaran, G., & Scanlon, D. P. (2008). Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics, 144,* 156–174. http://dx.doi.org/10.1016/j.jeconom.2008.01.001

Constantino, M. J., Boswell, J. F., Bernecker, S. L., & Castonguay, L. G. (2013). Context-responsive integration as a framework for unified psychotherapy and clinical science: Conceptual and empirical considerations. *Journal of Unified Psychotherapy and Clinical Science, 2,* 1–20.

De Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22,* 464–474. http://dx.doi.org/10.1080/10503307.2012.673023

Derogatis, L. R. (1975). *Brief Symptom Inventory*. Baltimore: Clinical Psychometrics Research.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations. A demonstration [Advance online publication]. *PLoS ONE, 9,* e83875. http://dx.doi.org/10.1371/journal.pone.0083875

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publ. no. 231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.

Foa, E. B., & Rothbaum, B. A. (1998). *Treating the trauma of rape: Cognitive behavioral therapy for PTSD*. New York: Guilford Press.

Goldfield, N. (Ed.). (1999). *Physician profiling and risk adjustment*. Frederick, MD: Aspen.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35,* 485–491. http://dx.doi.org/10.1037/0735-7028.35.5.485

Hatfield, D. R., & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health and Mental Health Services Research, 34,* 283–291. http://dx.doi.org/10.1007/s10488-006-0110-y

Henderson, A., & Henderson, S. (2010). Provision of a surgeon's performance data for people considering elective surgery. *Cochrane Database of Systematic Reviews, 11,* CD006327. http://dx.doi.org/10.1002/14651858.CD006327.pub2

Hibbard, J. H., & Jewett, J. J. (1996). What type of quality information do consumers want in a health care report card? *Medical Care Research and Review, 53,* 28–47. http://dx.doi.org/10.1177/107755879605300102

Hibbard, J. H., Peters, E., Slovic, P., Finucane, M. L., & Tusler, M. (2001). Making health care quality reports easier to use. *Joint Commission Journal on Quality Improvement, 27,* 591–604.

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61,* 678–685. http://dx.doi.org/10.1037/0022-006X.61.4.678

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

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57,* 131–137. http://dx.doi.org/10.1037/0022-006X.57.1.131

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67,* 300–307. http://dx.doi.org/10.1037/0022-006X.67.3.300

Ketelaar, N. A. B. M., Faber, M. J., Flottorp, S., Rygh, L. H., Deane, K. H. O., & Eccles, M. P. (2011). Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database of Systematic Reviews, 11,* CD004538.

Koerner, K., & Castonguay, L. G. (2015). Practice-oriented research: What it takes to do collaborative research in private practice. *Psychotherapy Research, 25,* 67–83. http://dx.doi.org/10.1080/10503307.2014.939119

Kraus, D. R. (2015, March). *Measuring and improving child well-being*. Invited address at the Annie E. Casey Foundation Conference, New Orleans, LA.

Kraus, D. R., Bentley, J., Alexander, P., Boswell, J. F., Constantino, M. J., Castonguay, L. G., & Baxter, L. (2015). *Predicting therapist effectiveness*. Manuscript in preparation.

Kraus, D., & Castonguay, L. G. (2010). Treatment Outcome Package (TOP): Development and use in naturalistic settings. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 155–174). New York: Wiley. http://dx.doi.org/10.1002/9780470687994.ch7

Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research, 21,* 267–276. http://dx.doi.org/10.1080/10503307.2011.563249

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61,* 285–314. http://dx.doi.org/10.1002/jclp.20084

Krause, M. S., Lutz, W., & Saunders, S. M. (2007). Empirically certified treatments or therapists: The issue of separability. *Psychotherapy: Theory, Research, Practice, Training, 44,* 347–353. http://dx.doi.org/10.1037/0033-3204.44.3.347

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48,* 72–79. http://dx.doi.org/10.1037/a0022238

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Use of communities of practice in business and health care sectors: A systematic review [Advance online publication]. *Implementation Science, 4,* 27. http://dx.doi.org/10.1186/1748-5908-4-27

McAleavey, A. A., Lockard, A. J., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2015). Building a practice research network: Obstacles faced and lessons learned at the Center for Collegiate Mental Health. *Psychotherapy Research, 25,* 134–151. http://dx.doi.org/10.1080/10503307.2014.883652

McHorney, C. A., Ware, J. E., Jr., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care, 31,* 247–263. http://dx.doi.org/10.1097/00005650-199303000-00006

Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology, 70,* 886–903. http://dx.doi.org/10.1002/jclp.22077

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38,* 65–76. http://dx.doi.org/10.1007/s10488-010-0319-7

Ranmuthugala, G., Plumb, J. J., Cunningham, F. C., Georgiou, A., Westbrook, J. I., & Braithwaite, J. (2011). How and why are communities of practice established in the healthcare sector? A systematic review of the literature. *BMC Health Services Research, 11,* 273. http://dx.doi.org/10.1186/1472-6963-11-273

Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child and Adolescent Psychiatric Clinics of North America, 14,* 241–254.

Rosenthal, M. B., Frank, R. G., Li, Z., & Epstein, A. M. (2005). Early experience with pay-for-performance: From concept to practice. *Journal of the American Medical Association, 294,* 1788–1793. http://dx.doi.org/10.1001/jama.294.14.1788

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80,* 535–546. http://dx.doi.org/10.1037/a0028898

Scanlon, D. P., Lindrooth, R. C., & Christianson, J. B. (2008). Steering patients to safer hospitals? The effect of a tiered hospital network on hospital admissions. *Health Services Research, 43,* 1849–1868. http://dx.doi.org/10.1111/j.1475-6773.2008.00889.x

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (Eds.). (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery*. Washington, DC: American Psychological Association Press. http://dx.doi.org/10.1037/10232-000

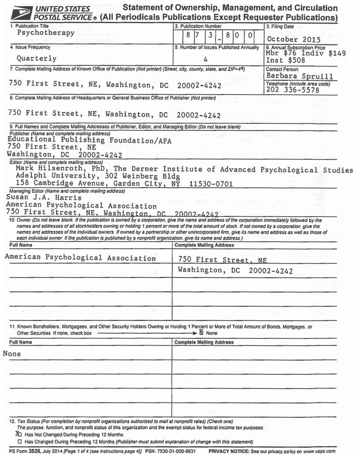
Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923. http://dx.doi.org/10.1037/0022-006X.73.5.914

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

Youn, S. J., Kraus, D. R., & Castonguay, L. G. (2012). The treatment outcome package: Facilitating practice and clinically relevant research. *Psychotherapy, 49,* 115–122. http://dx.doi.org/10.1037/a0027932

Zack, S. E., Castonguay, L. G., Boswell, J. F., McAleavey, A. A., Adelman, R., Kraus, D. R., & Pate, G. A. (2015). Attachment history as a moderator of the alliance outcome relationship in adolescents. *Psychotherapy, 52,* 258–267. http://dx.doi.org/10.1037/a0037727

Received June 10, 2015
Accepted June 12, 2015 ■



UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation (All Periodicals Publications Except Requester Publications)

1. Publication Title Psychotherapy	2. Publication Number 8 7 3 - 8 0 0	3. Filing Date October 2015
4. Issue Frequency Quarterly	5. Number of Issues Published Annually 4	6. Annual Subscription Price Mbr $76 Indiv $149 Inst $508
7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4®)* 750 First Street, NE, Washington, DC 20002-4242		Contact Person Barbara Spruill Telephone *(Include area code)* 202 336-5578

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
750 First Street, NE, Washington, DC 20002-4242

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
Educational Publishing Foundation/APA
750 First Street, NE
Washington, DC 20002-4242

Editor *(Name and complete mailing address)*
Mark Hilsenroth, PhD, The Derner Institute of Advanced Psychological Studies
Adelphi University, 302 Weinberg Bldg
158 Cambridge Avenue, Garden City, NY 11530-0701

Managing Editor *(Name and complete mailing address)*
Susan J.A. Harris
American Psychological Association
750 First Street, NE, Washington, DC 20002-4242

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

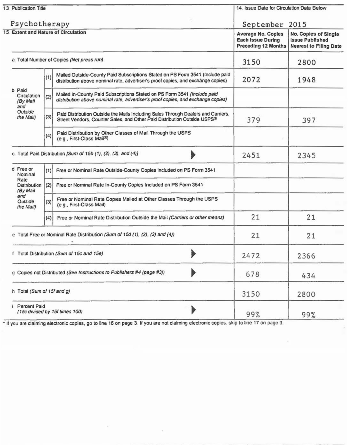
Full Name	Complete Mailing Address
American Psychological Association	750 First Street, NE Washington, DC 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address
None	

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, July 2014 *[Page 1 of 4 (see instructions page 4)]* PSN: 7530-01-000-9931 **PRIVACY NOTICE:** See our privacy policy on *www.usps.com*

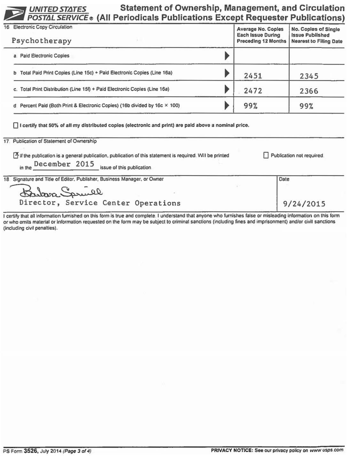


13. Publication Title Psychotherapy		14. Issue Date for Circulation Data Below September 2015

15. Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*			3150	2800
b. Paid Circulation *(By Mail and Outside the Mail)*	(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	2072	1948
	(2)	Mailed In-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*		
	(3)	Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	379	397
	(4)	Paid Distribution by Other Classes of Mail Through the USPS (e.g., First-Class Mail®)		
c. Total Paid Distribution *[Sum of 15b (1), (2), (3), and (4)]* ▶			2451	2345
d. Free or Nominal Rate Distribution *(By Mail and Outside the Mail)*	(1)	Free or Nominal Rate Outside-County Copies included on PS Form 3541		
	(2)	Free or Nominal Rate In-County Copies Included on PS Form 3541		
	(3)	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g., First-Class Mail)		
	(4)	Free or Nominal Rate Distribution Outside the Mail *(Carriers or other means)*	21	21
e. Total Free or Nominal Rate Distribution *(Sum of 15d (1), (2), (3) and (4))*			21	21
f. Total Distribution *(Sum of 15c and 15e)* ▶			2472	2366
g. Copies not Distributed *(See Instructions to Publishers #4 (page #3))* ▶			678	434
h. Total *(Sum of 15f and g)*			3150	2800
i. Percent Paid *(15c divided by 15f times 100)* ▶			99%	99%

* If you are claiming electronic copies, go to line 16 on page 3. If you are not claiming electronic copies, skip to line 17 on page 3.

PS Form **3526**, July 2014 *(Page 2 of 4)*



UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation (All Periodicals Publications Except Requester Publications)

16. Electronic Copy Circulation Psychotherapy	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Paid Electronic Copies ▶		
b. Total Paid Print Copies (Line 15c) + Paid Electronic Copies (Line 16a) ▶	2451	2345
c. Total Print Distribution (Line 15f) + Paid Electronic Copies (Line 16a) ▶	2472	2366
d. Percent Paid (Both Print & Electronic Copies) (16b divided by 16c × 100) ▶	99%	99%

☐ **I certify that 50% of all my distributed copies (electronic and print) are paid above a nominal price.**

17. Publication of Statement of Ownership
☒ If the publication is a general publication, publication of this statement is required. Will be printed in the December 2015 issue of this publication ☐ Publication not required.

18. Signature and Title of Editor, Publisher, Business Manager, or Owner
Barbara Spruill
Director, Service Center Operations
Date
9/24/2015

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form **3526**, July 2014 *(Page 3 of 4)* **PRIVACY NOTICE:** See our privacy policy on *www.usps.com*