

# When Meta-Analysis Misleads: A Critical Case Study of a Meta-Analysis of Client Feedback

Barry L. Duncan

Better Outcomes Now, Jensen Beach, Florida

Jacqueline A. Sparks

University of Rhode Island

Consumers of psychotherapy outcome literature consider meta-analysis the gold standard for assessing the efficacy of interventions across disparate studies. Many assume that findings are valid, especially when published in journals with research credentials. Uncritical acceptance, however, can result in real-world consequences, including whether interventions attain evidence-based status or become marginalized or are considered for implementation in public service arenas. This article examines one meta-analysis, “The Effect of Using the Partners for Change Outcome Management System as Feedback Tool in Psychotherapy—A Systematic Review and Meta-Analysis” (Østergård, Randa, & Hougaard, 2018). The findings are at odds with both the empirical record of routine outcome management as well as professional taskforce recommendations and thus provide an ideal exemplar of the risks of uncritically accepting the conclusions of a meta-analysis. Using guidelines from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011) and a qualitative case study methodology, this article examines Østergård et al.’s (2018) study selection, quality of evidence, and appropriateness of interpretation, emphasizing the link between flawed method and the ultimate validity of its conclusions. The method illustrated in this case study can be used to assess the legitimacy of meta-analytic findings to inform practice, funding, and policy decisions as well as how rhetoric minimizes flaws and bolsters believability. Our analysis revealed that half of the selected studies of the meta-analysis contained significant limitations, including inadequate dose of treatment and/or adherence problems, thereby calling into question its conclusions.

**Keywords:** meta-analysis, systematic client feedback, routine outcome monitoring (ROM), measurement-based care, Partners for Change Outcome Management System


Systematic reviews select, synthesize, and provide an integrated evaluation of relevant scientific studies to address specific research questions related to an intervention (Gertler & Cameron, 2018; Higgins & Green, 2011). The heart of any systematic review is the selection of methodologically rigorous studies that can provide a scientifically sound basis for the review’s conclusions (Higgins & Green, 2011). Thus, a key strategy is the development of preset inclusion criteria to screen out poor-quality studies that would jeopardize validity.

Meta-analysis is one statistical data-analysis technique used to estimate the effects of multiple studies in a systematic review (Higgins & Green, 2011). Meta-analytic strategies are meant to provide a statistically accurate assessment of an intervention’s efficacy not possible with a single study. As a type of systematic review, meta-analysis aims to synthesize disparate data to facilitate informed, practical decisions across the spectrum of health-care service delivery and use.

From seminal meta-analytic studies (e.g., Smith & Glass, 1977) to more recent analyses (e.g., Flückiger, Del Re, Wampold, & Horvath, 2018), meta-analysis has played a pivotal role in defining key points of debate regarding treatment outcome and what accounts for change in psychotherapy. The primary strategy is the conversion of outcomes from distinct studies into a common effect size that allows critical comparisons of interventions across studies. This statistical collation provides consumers a more accurate understanding of the true efficacy of the intervention being studied (Comer & Kendall, 2013). Because of its sophistication, meta-analysis has become the preferred mechanism for translating findings from multiple sources into treatment guidelines and evidence-based lists that drive clinical choice and third-party funding.

Although meta-analysis aims to simplify research for consumers, the esoteric nature of its method can be a barrier for those wishing to evaluate the soundness of a given meta-analytic study. On its face, meta-analysis can be intimidating, and acceptance in a

---

 Barry L. Duncan, Better Outcomes Now, Jensen Beach, Florida; Jacqueline A. Sparks, Department of Human Development and Family Studies, College of Health Sciences, University of Rhode Island.

Barry L. Duncan is a co-holder of the copyright of the Partners for Change Outcome Management System (PCOMS) instruments. The measures are free for individual use, but Duncan receives royalties from licenses issued to organizations. In addition, the web application of PCOMS, <http://betteroutcomesnow.com>, is a commercial product, and he receives profits based on sales.

The authors are indebted to Arthur C. Bohart, who independently evaluated our manuscript for how we managed our stated bias for PCOMS.

Correspondence concerning this article should be addressed to Barry L. Duncan, Better Outcomes Now, P.O. Box 6157, Jensen Beach, FL 34957. E-mail: [barryduncan@betteroutcomesnow.com](mailto:barryduncan@betteroutcomesnow.com)

peer-reviewed journal offers a ready-made rationale for assuming its validity. Unfortunately, such uncritical acceptance has a cost. The conclusions of a meta-analysis compromised by undetected faulty procedures can become accepted dogma, driving far-reaching decisions, such as what approaches are allowed or reimbursed in given practice settings and which become marginalized.

Since its inception in the early 1990s, the Cochrane Collaboration (<https://www.cochrane.org/>), recognized as the premier reference for conducting systematic reviews, has developed guidelines for establishing the parameters for study quality and inclusion. Despite these efforts, flawed meta-analyses continue to be published, with the flaws being not so much in the method itself but in how the method is carried out (Wampold et al., 2017). It is argued here that it is crucial for consumers of psychotherapy research to know how to critically approach meta-analyses and determine credibility relative to researchers' claims.

This article aims to extend efforts to make meta-analyses more transparent and more easily evaluated by practitioners, funders, and policymakers. To do so, we critically examine a recent publication in *Psychotherapy Research*, "The Effect of Using the Partners for Change Outcome Management System as Feedback Tool in Psychotherapy—A Systematic Review and Meta-Analysis" (Østergård, Randa, & Hougaard, 2018). The purpose of the study was to examine the effects of the Partners for Change Outcome Management System (PCOMS), a systematic client-feedback protocol (also called routine outcome monitoring [ROM] or measurement-based care) and explore the possible roles of moderators in that effect. It reported a small effect in counseling settings and no effect in psychiatric settings. Østergård et al. concluded that studies finding effects were likely affected by researcher allegiance and the use of only one outcome measure, the Outcome Rating Scale (ORS; Miller, Duncan, Brown, Sparks, & Claud, 2003), which was likely influenced by social desirability.

We selected this study for several reasons. First, we have a vested interest in any study addressing PCOMS, given that we are its developers and researchers. We acknowledge our allegiance to the system and our commitment to educating practice settings regarding its value in improving outcomes and promoting client voice and choice. Second, we are intimately familiar with the research included in the meta-analysis and aware of the methodology of each study, giving us a broader perspective on both the research itself and conclusions that can be reached. Finally, Østergård et al.'s (2018) conclusions are at odds with both the established record of ROM as well as professional taskforce recommendations (cf., Lambert, 2017) and thus provide an ideal exemplar of the risks of uncritically accepting the conclusions of a meta-analysis.

This case study offers a counterstory to Østergård et al.'s (2018) findings for those implementing or considering implementing PCOMS or similar systems in routine settings. In a broader sense, we offer an example of the use of Cochrane guidelines to assess meta-analyses across psychotherapy interventions, making this form of research accessible to everyday clinicians.

## Method

The method employed here is a single instrumental case study. This strategy entails an in-depth exploration of one bounded entity to explicate and address an area of concern (Creswell, Hanson,

Clark Plano, & Morales, 2007; Stake, 1995). It also is critical in that it seeks to challenge accepted modes of research through an analysis of rhetorical claims, distortions, or omissions that mislead, whether intentional or not.

The structure of our investigation uses three parameters derived from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011). These are (a) study selection, (b) quality of the body of evidence, and (c) appropriateness of interpretation. These are defined in the following discussion. We examine the study of Østergård et al. (2018) along each of these dimensions to assess its overall validity. We conclude with implications for practice and policy and future recommendations for consumers of meta-analyses.

Consistent with qualitative methods of transparency (e.g., Hiles, 2008), we further state our position related to the purpose of this article. As founders and researchers of PCOMS, we have a vested interest in ensuring that research of the system is sound. Recognizing this, we have employed two means to minimize bias. First, we chose to structure our critique using guidelines from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011), a widely respected source for conducting valid meta-analyses. We have endeavored to be as precise as possible in weighing concerns regarding the validity of Østergård et al.'s (2018) conclusions based on this framework. Second, we invited an outside reader versed in ROM research and meta-analysis yet not a member of any research team involved in ROM or, more importantly, PCOMS trials or publications to serve as a knowledgeable outsider in evaluating the article. The intent of utilizing an outside reader is to identify areas of bias and bolster the credibility of our analysis. Investigator triangulation (Denzin, 1978) is a widely used strategy in qualitative research that increases the credibility and trustworthiness of a study. Based on this reader's feedback, we were able to reexamine our conclusions.

## Analysis

### Study Selection

Study selection determines the validity of any meta-analysis. Study selection hinges on preset eligibility criteria. According to the *Cochrane Handbook* (Higgins & Green, 2011), these criteria include the extent to which selected studies address the question posed by the meta-analysis and include appropriate participants and procedures to do so. In the instance of the Østergård et al. (2018) study, this involves factors such as randomization and proper implementation of the experimental intervention as designed. Although the *Cochrane Handbook* provides guidance on locating unpublished studies, the criteria for selection emphasize that studies be of sufficient quality, whether published or unpublished. The studies selected by Østergård et al. are examined herein based on their overall quality, particularly whether they accurately represent the actual implementation of the experimental condition.

**Randomized or nonrandomized.** The *Cochrane Handbook* (Higgins & Green, 2011) states: "a Cochrane review would typically seek all rigorous studies (e.g., randomized trials) of a particular comparison of interventions in a particular population of participants, irrespective of the outcomes measured or reported" (para. 5.1.2). The predefined inclusion criteria of Østergård et al. (2018) specified that both randomized clinical trials (RCTs) and

nonrandomized trials (N-RCTs) would be eligible for inclusion. Østergård et al. included 18 heterogeneous studies, a mixture of investigations of individual, couple, and group intervention as well as both inpatient and outpatient settings and published and unpublished studies. Of the 18 studies, 14 were RCTs, and 4 were N-RCT (group comparison without randomization) studies.

**Published or unpublished.** Unpublished studies were included in the search, resulting in 3 unpublished studies in the final selection of 18 studies. One of the unpublished studies, that by Chow and Huixian (2015), was a poster session and two, those by Kellybrew-Miller (2015) and Lester (2012), were dissertations. Omitted was a published large-group comparison study ( $N = 288$ ) of children, caretakers, and teachers (Cooper, Stewart, Sparks, & Bunting, 2013).

Quality studies that fail to find an effect may not be pursued or accepted for publication. It is appropriate for meta-analytic researchers to consider all worthy studies with positive and negative results, published and unpublished, to gain a true measure of the available evidence for a given intervention and to avoid the “file drawer problem” (Rosenthal, 1979, p. 638). On the other hand, studies may not be published due to significant weaknesses in design. Unpublished and published studies are analyzed herein for their overall methodological rigor.

**Defining types of interventions.** According to the *Cochrane Handbook* (Higgins & Green, 2011), “it is useful to consider exactly what is delivered, at what intensity, how often it is delivered, who delivers it, and whether people involved in delivery of the intervention need to be trained” (para. 5.2). Moreover, the guidelines state that “review authors should also consider whether variation in the intervention (i.e. based on dosage/intensity, mode of delivery, frequency, duration etc.) is so great that it would have substantially different effects on the participants and outcomes of interest, and hence may be important to restrict” (Higgins & Green, 2011, para. 5.2). Østergård et al. (2018) state that “the experimental condition had to include PCOMS as an ‘add on intervention’ to an intervention without PCOMS in the control condition” (p. 4). As described later in the article, there was considerable variability in the quality of the studies included related to whether the experimental condition (PCOMS) was delivered as intended and outlined in published protocols (e.g., Duncan & Sparks, 2002, 2010, 2018).

## Quality of the Body of Evidence

The quality of the body of evidence in the study by Østergård et al. (2018), specifically as it relates to the implementation of the experimental intervention, is examined here along three critical dimensions: number of sessions, adherence, and completeness of data. Table 1 summarizes these dimensions across studies that did not find a PCOMS advantage on the primary measure.

**Number of sessions (dose of treatment).** Four studies of the six RCTs not finding a feedback effect in the meta-analysis averaged less than four sessions: Lester (2012; 1.7 sessions); Kellybrew-Miller (2015; 2.2 sessions); Murphy, Rashleigh, and Timulak (2012; 3.7 sessions); and Rise, Eriksen, Grimstad, and Steinsbekk (2016; 3.8). All RCTs finding a feedback effect (eight RCTs) averaged above four sessions in the experimental condition.

**Adherence to experimental condition.** According to the three editions of the manual for PCOMS implementation (Duncan

Table 1  
Studies Without Significant Findings on the Primary Measure

Study	Dose (No. of sessions)	Study type	Adherence to PCOMS/Missing data	Significant findings
Murphy, Rashleigh, and Timulak (2012) <sup>a</sup>	3.7	RCT	Omitted SRS	YOQ-P, <sup>b</sup> alliance
Lester (2012)	1.7	RCT	Therapists found PCOMS not useful and unnecessary	ORS; % CSS <sup>b</sup>
Kellybrew-Miller (2015)	2.2	RCT	66% use; missing data	PAM
Rise, Eriksen, Grimstad, and Steinsbekk (2016)	3.8	RCT	92% use	
Davidson et al. (2017)	12.4	RCT	Therapists did not use PCOMS as intended; flexibility discouraged	
van Oenen et al. (2016)	9.3	RCT	67% use with 70% of clients	
Janse, De Jong, Van Dijk, Hutschemaekers, and Verbraak (2017)	15.0	N-RCT	76.8% use (unspecified amount); 23.2% no evidence of use; 11% of charts missing	ORS; RC in fewer sessions
Hansen, Howe, Sutton, and Ronan (2015)	8.0	N-RCT	Used 50% of time; undetermined intervals of use; 70% missing data on a primary measure	HoNOSCA <sup>b</sup>
Winkelhorst, Hafkenscheid, and de Groot (2013)	15.0	N-RCT	“Partially integrated”; not used at every encounter	

Note. PCOMS = Partners for Change Outcome Management System; RCT = randomized clinical trial; N-RCT = not RCT; SRS = Session Rating Scale; YOQ-P = Youth Outcome Questionnaire-Parent; ORS = Outcome Rating Scale; % CSS = percentage of clinically significant change; PAM = Patient Activation Measure; RC = reliable change; HoNOSCA = Health of the Nation Outcome Rating Scale for Children and Adolescents.

<sup>a</sup> ORS was the primary measure. <sup>b</sup> Finding was on a primary outcome measure.

& Sparks, 2002, 2010, 2018) as well as other published accounts (Duncan, 2014; Duncan & Reese, 2015), measures are administered at every session: outcome instruments at the start of the meeting and alliance instruments at the end. The therapist and client collaboratively discuss scores at each session. Client progress or lack thereof directs what happens in the session thereafter. The studies included by Østergård et al. (2018) that did not find a feedback effect report incorrect and noteworthy lapses in the implementation of PCOMS. For example, Murphy et al. (2012) omitted the alliance component of PCOMS and exerted no control over the therapist's ability to see the results of both the feedback and no-feedback conditions. Moreover, four of the six RCTs not finding a feedback effect (Davidsen et al., 2017; Kellybrew-Miller, 2015; Lester, 2012; van Oenen et al., 2016) contained significant issues with adherence and/or therapist perceptions of usefulness. Davidsen et al. (2017) reported that most therapists did not use PCOMS to individualize treatment or alter treatment in any way based on the data and that organizational factors also limited therapist flexibility. Kellybrew-Miller (2015) found that therapists did not use the measures one third of the time, and one therapist accounted for over half of the clients in the study. Lester (2012) stated that therapists found the measures more disturbing and unnecessary than helpful. van Oenen et al. (2016) noted that only 67% of therapists said they used PCOMS with 70% of their clients.

Regarding N-RCTs, three of four (Hansen, Howe, Sutton, & Ronan, 2015; Janse, De Jong, Van Dijk, Hutschemaekers, & Verbraak, 2017; Winkelhorst, Hafkenscheid, & de Groot, 2013) had issues with adherence and/or therapist perceptions of usefulness; the fourth, a poster presentation by Chow and Huixian (2015), was unavailable for scrutiny. Janse et al. (2017) reported that 76.8% of charts had evidence (unspecified how much) of PCOMS use. Hansen et al. (2015) noted that PCOMS measures were used only half the time and at undetermined intervals. Winkelhorst et al. (2013) stated that the PCOMS intervention was only partially integrated into treatment and was not conducted at every session. It is important to note that adherence to the PCOMS protocol goes beyond the percentage of sessions that include the measures; fidelity includes using the data to identify clients who are not benefiting and collaboratively altering treatment.

**Missing data.** The study by Kellybrew-Miller (2015), an RCT finding no effect on the primary measure, reported significant missing data and a 3-month hiatus of any data collection. In the Janse et al. (2017) study, 11% of the charts were missing post-treatment, and of the remaining charts, 23.2% had no evidence of PCOMS whatsoever. Hansen et al. (2015) reported that only 30% of the data on a primary outcome measure, the Strength and Difficulties Questionnaire (SDQ; Goodman, 2001), was collected.

Despite the noted problems, seven of the nine total studies not finding a feedback effect on a primary measure reported mixed or trending positive results: one found a trend toward significance on the Outcome Questionnaire 45.2 (OQ; Winkelhorst et al., 2013), two found significant effects on the ORS (Janse et al., 2017; Kellybrew-Miller, 2015), two trended toward significance on the ORS (Murphy et al., 2012; Rise et al., 2016), and four found significant results on other measures (Hansen et al., 2015; Janse et al., 2017; Lester, 2012; Rise et al., 2016). Lester (2012) did not find an effect on the adolescent-completed Youth Outcome Questionnaire (YOQ; Wells, Burlingame, & Lambert, 1996), but there

was an advantage for the PCOMS condition on the parent-completed YOQ.

### Appropriateness of Interpretation

The *Cochrane Handbook* (Higgins & Green, 2011) cautions that “when there is inconclusive evidence, it is wrong to claim that it shows that an intervention has ‘no effect’ or is ‘no different’ from the control intervention” (para. 12.7.4); further, a common mistake made by authors of meta-analyses is “to reach conclusions that go beyond the evidence” (para. 12.7.4). The following interpretations, or conclusions, are examined considering these cautions.

**No evidence of effects in psychiatric settings.** Østergård et al. (2018) conclude that they found no evidence to support a feedback effect for the PCOMS in psychiatric settings. Of the six RCTs classified as conducted in psychiatric settings, one found a feedback effect on two measures (Brattland et al., 2018), and five did not. Of the five studies not finding a feedback effect, three were less than four sessions in duration (Kellybrew-Miller, 2015; Lester, 2012; Rise et al., 2016), and four of the five (Davidsen et al., 2017; Kellybrew-Miller, 2015; Lester, 2012; van Oenen et al., 2016) reported significant adherence problems and/or poor therapist perceptions of usefulness.

Of the N-RCTs in psychiatric settings, three of the four investigations reported significant adherence problems: Hansen et al. (2015), 50% adherence; Janse et al. (2017), 23.2% of the charts showed no evidence of PCOMS; Winkelhorst et al. (2013), therapists used the PCOMS partially, and attrition was substantial, starting with 35 in the feedback group and ending with 19 for the final comparison. Information on Chow and Huixian (2015) was not available. Of note, the study conducted in a substance abuse treatment center with mandated individuals diagnosed with substance abuse disorders (Schuman, Slone, Reese, & Duncan, 2015) was classified as a “counseling” setting as opposed to outpatient psychiatric treatment.

**Sole reliance on the ORS diminishes the validity of findings of effects.** Østergård et al. (2018) concluded that the lack of use of outcome measures besides the ORS compromised the validity of studies finding an effect for PCOMS. Although raising the important issue of the sole use of the ORS as a measure of outcome and the need for comparison with more symptom-based instruments, the authors fail to note the studies with significant findings on the ORS that have used other indices of outcomes. Table 2 summarizes these studies. For example, Brattland et al. (2018) found positive results on both the Behavior and Symptom Identification Scale 32 (BASIS-32; Eisen, Wilcox, Leff, Schaefer, & Culhane, 1999) and the ORS. Additionally, five other RCTs used different measures of outcomes and found an advantage for the PCOMS, thereby corroborating the reported effects on the ORS. First, Anker, Duncan, and Sparks (2009) used separation/divorce rates and the Locke–Wallace (Locke & Wallace, 1959) in addition to the ORS at 6-month follow-up. They found significantly lower separation/divorce rates for the PCOMS condition and trending to better outcomes on the Locke–Wallace (the power was insufficient given the lower number of intact couples at follow-up). Second, the group investigation of mandated soldiers with substance abuse disorders (Schuman et al., 2015) found a significant advantage for PCOMS on therapist ratings of outcome, blinded commander ratings of positive reintegration, the number of sessions attended,



Table 2  
*RCTs With Significant Findings on the Outcome Rating Scale as Primary Measure*

Study	Dose (No. of sessions)	Other significant findings	Allegiance/Mitigation strategies
Reese, Norsworthy, and Rowlands (2009, Study 1)	6.3	RC in fewer sessions	None
Reese, Norsworthy, and Rowlands (2009, Study 2)	8.0	RC in fewer sessions	None
Anker, Duncan, and Sparks (2009)	4.7	Separation/divorce rate	Yes/assessment and independent analysis
Reese, Toland, Slone, and Norsworthy (2010)	5.9	Improvement more rapidly	None
Schuman, Slone, Reese, and Duncan (2015)	4.2	Blinded commander ratings; therapist ratings; attendance; dropout	Yes
Slone, Reese, Mathews-Duvall, and Kodet (2015)	8.0	Attendance	Yes
She et al. (2018)	4.8	CSC in fewer sessions; dropout; alliance posttreatment	Yes
Brattland et al. (2018)	12.0	BASIS 32 <sup>a</sup> ( $d = .42$ )	None

*Note.* RCT = randomized clinical trial; RC = reliable change; CSC = clinically significant change; BASIS 32 = Behavioral and Symptom Identification Scale 32.

<sup>a</sup> Primary outcome measure.

and premature termination. Third, in addition to findings on the ORS, Slone, Reese, Mathews-Duvall, and Kodet (2015) reported that clients in the PCOMS condition attended significantly more sessions. Fourth, She et al. (2018) found a significant advantage for the PCOMS condition in premature termination and posttreatment alliance scores. Although not standardized outcome measures, so-called “real-world” measures (divorce rates, blinded commander ratings of successful reintegration, dropout, and attendance) speak to both the internal and external validity of the findings on the ORS. In addition, although based on the ORS, several studies reported reliable or clinically significant change in significantly fewer sessions in the PCOMS condition (Reese, Norsworthy, & Rowlands, 2009, Studies 1 and 2; Reese, Toland, Slone, & Norsworthy, 2010; She et al., 2018). Finally, an inspection of Table 1 reveals that Lester (2012) found an advantage on the parent-completed YOQ (Wells et al., 1996) and Rise et al. (2016) reported an advantage on the Patient Activation Measure (Hibbard, Stockard, Mahoney, & Tusler, 2004).

Regarding N-RCTs, despite the flaws noted previously, all three studies available for scrutiny either realized an effect or trended toward an effect on a measure other than the ORS. Hansen et al. (2015) found significant change on the Health of the Nation Outcome Scale for Children and Adolescents (HoNOSCA; Gowers et al., 1999). Winkelhorst et al. (2013) found an 11.3-point difference on the OQ at Session 15, although only 19 clients remained, restricting the possibility of significance. Janse et al. (2017) found a feedback effect on the ORS and the number of sessions to achieve change. Finally, the omitted N-RCT, that by Cooper et al. (2013), found a feedback effect on its primary measure, the Strength and Difficulties Questionnaire, with both parents and teachers of 7- to 11-year-old children.

**An allegiance effect/bias diminishes the validity of findings of an effect.** Eight of the 14 RCTs found a significant feedback effect (see Table 2). Seven of these eight were conducted by those currently affiliated with Better Outcomes Now (formally the Heart and Soul of Change Project). The remaining study (Brattland et al., 2018) was conducted independently and found a feedback effect on the primary measure (BASIS-32) as well as the ORS. Included in the seven conducted by those currently affiliated with Better Outcomes Now were three early studies by R. Jeffrey Reese

(Reese et al., 2009, Studies 1 and 2; Reese et al., 2010). These studies were independently conducted before Reese was affiliated with the Heart and Soul of Change Project, diminishing Østergård et al.’s (2018) conclusion that these studies were most likely biased. Second, although researcher bias or allegiance is difficult to quantify and control, Anker et al. (2009) examined this factor by measuring therapist allegiance to feedback. They found that over the course of the study, the therapists, who were naïve to feedback at the start of the study, did not increase their belief that feedback would improve their outcomes. Although not conclusive, this provided some indication that researcher bias was minimized in that therapists were not positively influenced. Also of note, the analyses of Anker et al. were conducted independently by researchers without allegiance to PCOMS, a fact located in the study’s acknowledgments section. Consequently, five of the eight studies finding a feedback effect either did not have researcher bias or mitigated its effects.

**Social desirability in the use of the ORS skews results.** Østergård et al. (2018) suggest that social desirability is inherent in the use of the ORS and leads to unreliable and inflated results. This is not supported by the evidence of other measures finding an effect for the PCOMS condition (noted previously) in the examined studies. Further evidence that social desirability may not be a significant factor in studies using the ORS is provided in Anker et al. (2009). Six months after termination, clients in this study were mailed follow-up questionnaires, including the ORS; clients completed the measures at home without contact with, or the potential influence of, their therapists. The fact that the feedback effect was maintained on the ORS and corroborated on other measures at follow-up both supports the effects of feedback and casts doubt on the likelihood that social desirability played a role in the original findings. Of note also is that the PCOMS measures in the She et al. (2018) trial were discussed per protocol but not administered by therapists in session, and a feedback effect still occurred. Finally, social desirability issues are typically irrelevant to measures of global distress. Although clients may hide things from their therapists, they are more likely to withhold an immediate negative reaction to the therapist or session than to hide or misrepresent their level of distress (Farber, 2003; Lambert, 2017).

## Conclusions

Our analysis points to significant problems in the target study by Østergård et al. (2018). It poses three questions: (a) Are the search criteria flawed? (b) What is the overall quality of the evidence? (c) What is the overall validity of the conclusions? The answers to these questions raise concerns about the dissemination of flawed information into practice and policymaking settings.

Ideally, meta-analyses include RCTs, as recommended by the *Cochrane Handbook* (Higgins & Green, 2011). Østergård et al. (2018) included both RCTs and N-RCTs as well as published and nonpublished studies, including unpublished dissertations. It is understandable that researchers want sufficient numbers to strengthen their statistical analysis, and we recognize that there are trade-offs in accomplishing this goal. Given this, it is unclear why the study by Cooper et al., 2013, a well-designed, published N-RCT using a different outcome measure, was missed in the selection process. Regardless, of the 18 studies selected, 6 were either N-RCTs, unpublished, or both.

More important, however, is the quality of the studies that were eventually selected. Although there is no such thing as a perfect study, some studies included by Østergård et al. (2018) contained methodological issues so egregious that they ought to have been excluded. Specifically, the inclusion criteria did not specify the number of sessions and did not account for fidelity/adherence or therapist perceptions of usefulness. Because eligibility criteria allowed methodologically unsound studies to be included, we assert that it is likely that PCOMS, in fact, was not implemented in many instances, not only for occurrences of nonadherence but also when clinicians who used the system omitted key definitional components or held negative views about its worth.

Four of the six RCTs that found no or mixed results did not meet a minimal threshold for adequate treatment; two, that by Kellybrew-Miller (2015) and Lester (2012), not published, averaged but 2.2 and 1.7 sessions, respectively. In contrast, all RCTs in the experimental condition finding a feedback effect (eight) averaged more than four sessions. Similarly, four of the six RCTs not finding an effect contained significant adherence problems, ranging from the results not being discussed with clients or used to alter treatment (Davidsen et al., 2017) to PCOMS being used only about two thirds of the time (Kellybrew-Miller, 2015; van Oenen et al., 2016) and/or substantial negative therapist perceptions of PCOMS usefulness (Davidsen et al., 2017; Lester, 2012). Likewise, all three N-RCTs (no information available for Chow & Huixian, 2015) reported adherence issues, ranging from “partially integrated” (Winkelhorst et al., 2013) to 50% use (Hansen et al., 2015) to 76.7% of charts showing some (not specified) use of PCOMS (Janse et al., 2017).

In total, nine of the selected studies (no information available for Chow & Huixian, 2015) demonstrated an inadequate dose of the intervention and/or significant adherence problems. A meta-analysis holds up only to the extent that the studies analyzed are methodologically sound and, thus, support the validity of conclusions. The selection of studies by Østergård et al. (2018) does not meet criteria sufficient to qualify for this standard of “quality of evidence.” This initial misstep generates a chain of necessarily flawed interpretations that overreach the data upon which they are based.

Østergård et al.’s (2018) study also contains dubious conclusions derived from those studies we would classify as legitimate for inclusion, specifically related to the use of only one measure, bias/allegiance, and social desirability. Nine studies included in the analysis demonstrated a feedback effect on measures other than the ORS (Anker et al., 2009; Brattland et al., 2018; Hansen et al., 2015; Janse et al., 2017; Lester, 2012; Rise et al., 2016; Schuman et al., 2015; She et al., 2018; Slone et al., 2015). Of studies finding an effect on the ORS, Østergård et al. do not mention additional measurements of outcome that support positive findings of PCOMS. Some of these are valid outcome instruments, whereas others take the form of real-world outcomes. A comprehensive analysis would, at a minimum, mention real-world outcomes as supportive of the study’s overall findings. The authors also fail to report on efforts of the researchers of one major positive study (Anker et al., 2009) to minimize bias and failed to adequately determine several other instances of the independence of study investigators (Brattland et al., 2018; Reese et al., 2009, Studies 1 and 2; Reese et al., 2010). Finally, claiming that social desirability plays a significant role in diminishing findings of effect in positive studies is overemphasized. As noted, there are indications on other measures in the studies themselves that suggest this was not the case. Additionally, evidence regarding this factor in the empirical literature downplays its interference with accurate self-report assessment (Farber, 2003; Lambert, 2017).

Rhetorically, emphasis on the use of only one measure, bias/allegiance, and social desirability for studies finding an effect for PCOMS tends to lessen, if not erase, the view that the experimental intervention is effective. For example, the authors suggest that allegiance effects offer one explanation of the different findings of PCOMS studies. Given that allegiance effects were monitored or mitigated in five of the eight studies finding a feedback effect, the noted methodological flaws seem to provide a better explanation for the nonfindings in other studies. These same flaws (especially number of sessions and adherence issues) draw into question Østergård et al.’s (2018) conclusions that PCOMS is not likely to be effective in psychiatric settings. Moreover, the distinction between “counseling” and “psychiatric” settings seems arbitrary. Except for the unusually low intake scores reported by van Oenen et al. (2016), all the treatment settings included in the analysis reported similar levels of distress in participating clients. Further muddying the waters regarding “psychiatric” versus “counseling” settings, the relationship between diagnoses and outcomes has not been well established (Reese, Duncan, Bohanske, Owen, & Minami, 2014). Finally, evidence from benchmarking studies demonstrating that PCOMS does result in change comparable to RCTs in “psychiatric” settings in both inpatient (Reese et al., 2017) and outpatient clinics (Reese et al., 2014) contradicts Østergård et al.’s findings.

## Discussion

Østergård et al. (2018) conclude:

Based on this meta-analysis . . . the overall effect of using PCOMS is small . . . the positive effect in counseling settings may be biased due to positive researcher allegiance and use of the ORS as the only outcome measure. The ORS score is likely to be influenced by social desirability when completed in therapy. (p. 13)

Supporting these conclusions was a recent meta-analysis of 17 ROM investigations (Kendrick et al., 2016) that included three PCOMS studies (Murphy et al., 2012; Reese et al., 2009, Studies 1 and 2). This review considered studies of client-reported outcome measures for improving treatment and suggested that the evidence base for these methods was weak.

Østergård et al.'s (2018) conclusions, however, contradict other reviews and meta-analyses of PCOMS (Fortney et al., 2017; Lambert & Shimokawa, 2011; Lambert, Whipple, & Kleinstäuber, 2018). A review of 51 studies, including 5 PCOMS investigations (Anker et al., 2009; Murphy et al., 2012; Reese et al., 2009, Studies 1 and 2; Reese et al., 2010), by Fortney et al. (2017) concluded that virtually all randomized clinical trials with frequent and timely feedback of client-reported symptoms to the practitioner significantly improved treatment outcomes. This conclusion was echoed in the current analysis of studies included by Østergård et al. Whereas the Lambert and Shimokawa (2011) meta-analysis that reported a strong effect ( $r = .23$ ) was based on only three PCOMS trials (Anker et al., 2009; Reese et al., 2009, Studies 1 and 2), Lambert et al. (2018) aggregated nine PCOMS studies. Unlike Østergård et al., Lambert et al. included only RCTs and omitted the two unpublished dissertations (Kellybrew-Miller, 2015; Lester, 2012), a study primarily focused on the effects of the PCOMS on the alliance (Rise et al., 2016), van Oenen et al. (2016), and the recently published works by She et al. (2018) and Brattland et al. (2018). Lambert included studies criticized here for inadequate dose of treatment (Murphy et al., 2012) and significant adherence issues (Davidsen et al., 2017; Janse et al., 2017). Nevertheless, Lambert et al. (2018) concluded:

In sum, aggregated findings from the nine studies indicate that PCOMS rests on a growing empirical base that boosts confidence in its use as an ROM system. Practitioners can expect that PCOMS feedback will enhance client outcomes with an average effect size of .40, and at 95% CI, it will be between .29 and .51 . . . it is recommended that psychotherapists use either the OQ-System or PCOMS with adults across treatment modalities (e.g., individual, couple, and group) and clinical settings. (pp. 532–534)

It is unclear why Østergård et al. (2018) did not discuss these important reviews and meta-analyses and their divergence from their own findings.

Responding to evidence from reviews and meta-analyses, professional bodies highly recommend systematic client feedback (ROM or measurement-based care). The American Psychological Association (American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006) has recommended that ROM be a part of effective psychological services. Concurring, the American Association for Marriage and Family Therapy Task Force on Core Competencies advises therapists to solicit and use client feedback throughout the therapeutic process (Nelson et al., 2007). In addition, the Association of State and Provincial Psychology Boards (2015) has recommended that client feedback be a part of competency-based supervision. Two systems, PCOMS (Duncan, 2014) and the Outcome Questionnaire System (Lambert, 2015), were listed in the Substance Abuse and Mental Health Services Administration's National Registry of Evidence-Based Programs and Practices (<https://betteroutcomesnow.com/wp-content/uploads/2017/12/BON-PCOMS-SAMHSA-designated-evidence-based-practice.pdf>).

Østergård et al. (2018) do not discuss the discrepancy between their findings and the evidence of previous reviews leading to recommendations from professional bodies. Nor do they mention the problems identified in this article or the limitations mentioned in the respective articles themselves. These omissions give greater emphasis to the take-home conclusions of minimal efficacy for the experimental condition. The net effect is that readers likely will be ill-served in making decisions about whether to instigate, or continue, PCOMS or a similar ROM system in their practice settings. As demonstrated by Brattland et al. (2018), the implementation of systematic client feedback in routine care requires resources, largely related to training at the managerial, supervisory, and frontline-practitioner levels. Buy-in is critical, and ongoing support and monitoring of efforts are required for these systems to be properly incorporated into everyday practices. If the effort and cost are perceived as not worth the benefit, those in positions to make decisions may opt out.

With the exception of Schuman et al. (2015), a replication of earlier Lambert "red dot" investigations that neither trained nor followed up with therapists and gave only simple color-coded feedback to clinicians (e.g., Lambert et al., 2001), PCOMS trials that included adherence checks and reinforcement of PCOMS use via supervision, graph-checking, and data review have found a significant feedback effect. For example, given the noted negative impact of adherence and therapist perceptions of usefulness, a recent study, that by She et al. (2018; included in the Østergård et al., 2018, meta-analysis) attempted to mitigate the effects of both nonadherence and negative therapist perceptions of usefulness by incorporating ongoing graph review and identification of not-on-track clients. At the end of the study, 100% of the clients had fully completed graphs of ORS scores. A poststudy survey revealed that 83.3% of the therapists perceived the ORS/Session Rating Scale (SRS) information as useful, in sharp contrast to the studies reviewed previously, which reported that therapists either didn't use the PCOMS consistently and/or rated the usefulness of the measures relatively low.

The studies included in the meta-analysis critiqued here for adherence and fidelity issues perhaps point to a problem of both meta-analysis in general and the PCOMS intervention in particular. Meta-analysis is not intended to assess implementation viability or effectiveness in diverse usual-care settings. It is possible that although PCOMS has been demonstrated to be effective in settings with high adherence, the understanding that it requires that high level of fidelity may mean that some, if not many, settings cannot implement it effectively.

Adherence may be particularly important to the PCOMS feedback effect. PCOMS is intended to be used to discuss outcomes and alliance with clients in every session. It is therefore not only a monitoring system to inform the therapist but also requires discussion and collaboration with clients (Sparks & Duncan, 2018). Such a process creates a higher demand on the therapist to incorporate the feedback. Fidelity and therapist perceptions of usefulness of feedback speak to the importance of sustainability in real-world clinical settings. Initial training combined with a lack of organizational commitment, as demonstrated by Davidsen et al. (2017), will not sustain implementation or result in therapist perceptions of usefulness. Success requires an organizational commitment to data collection, timely identification of not-on-track clients, and dissemination of the data to clinicians and supervisors,



as well ongoing attention to adherence and data integrity (Duncan, 2014; Duncan & Reese, 2015; She et al., 2018).

We believe that the evidence to date across well-designed RCTs and meta-analyses makes a strong case that the resources required would, in fact, be worth it—a conclusion supported by other reviews and analyses (Fortney et al., 2017; Lambert et al., 2018), professional bodies, and real-world applications. Agencies that have successfully implemented PCOMS have realized benefits to both outcomes and efficiency. Three benchmarking studies conducted in a large public behavioral health setting that had implemented PCOMS as a quality improvement strategy found outcomes comparable with RCTs with adults and children (Kodet, Reese, Duncan, & Bohanske, 2019; Reese et al., 2014), as well as with patients in an acute psychiatric inpatient unit (Reese et al., 2017). Regarding efficiency, Bohanske and Franczak (2010) surveyed PCOMS results at several public agencies and reported substantial improvements in client retention, therapist productivity, and length of stay.

In the interest of transparency, we again acknowledge that we approach this critique with an inevitable bias. Our roles in developing, articulating, and researching PCOMS necessarily entail a commitment to ensuring accuracy in how it is represented in the media, educational settings, scholarly papers, and research. Following standard qualitative methods, we invited an outside reader to evaluate how we managed this bias. The reader, a meta-analytic researcher, has worked with us on prior projects but has not been involved in PCOMS-related research or articles. After reading both Østergård et al.'s (2018) article and our article, he commented that we had followed best practice in qualitative research by openly discussing our bias in the article. He also noted that our arguments were grounded in data, allowing readers to come to their own conclusions. He stated his belief that many of the concerns we raised, in particular the number of sessions in many of the included trials, were valid. However, he expressed discomfort with the implication that the meta-analysis was uniquely flawed. In his opinion, the study was sound in its analysis of the data. However, one of its primary shortcomings, in his view, was its failure to discuss salient limitations and how these contextualize conclusions. Finally, the reader suggested that well-reasoned critiques are “part of the back and forth that make scientific investigation work” (A. C. Bohart, personal communication, June 26, 2019). Ultimately, this dialogical process serves to inform consumers regarding how best to utilize findings.

We recommend that meta-analytic researchers pay close attention to the quality of the studies they choose to include in their analyses, that they go beyond the simple aggregation of data and calculation of effect sizes. The *Cochrane Handbook* (Higgins & Green, 2011) is an important resource for ensuring that included studies are of sufficient quality to offer valid findings regarding the intervention being examined. Often, studies are not published for a good reason. We also advise that researchers not overreach their data by taking care to make interpretations that appropriately reflect the limitations of the studies included. At the very least, the most common limitations of the included studies should be aggregated and discussed so that the reader can draw his or her own conclusions from the reported data. Finally, we hope that consumers of research will be aided by this analysis to enhance their own understanding of the meta-analytic literature as they seek to improve psychotherapy services and outcomes.

In closing, any system that is dedicated to bringing clients' views to the forefront has the potential to transform traditional psychotherapy practices in ways that empower consumers to determine the trajectories of their own change (Duncan & Reese, 2015; Sparks & Duncan, 2018). We hope that our critical analysis will be of service to those making decisions about the practical and ideological value of such systems.

## References

- American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271–285. <http://dx.doi.org/10.1037/0003-066X.61.4.271>
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, *77*, 693–704. <http://dx.doi.org/10.1037/a0016062>
- Association of State and Provincial Psychology Boards. (2015). *Supervision guidelines for education and training leading to licensure as a health service provider*. Retrieved from [https://cdn.ymaws.com/www.asppb.net/resource/resmgr/guidelines/supervision\\_guidelines\\_for\\_g.pdf](https://cdn.ymaws.com/www.asppb.net/resource/resmgr/guidelines/supervision_guidelines_for_g.pdf)
- Bohanske, R., & Franczak, M. (2010). Transforming public behavioral health care: A case example of consumer directed services, recovery, and the common factors. In B. Duncan, S. Miller, B. Wampold, & M. Hubble (Eds.), *The heart and soul of change: Delivering what works* (2nd ed., pp. 299–322). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/12075-010>
- Brattland, H., Koksvis, J. M., Burkeland, O., Gråve, R. W., Klöckner, C., Linaker, O. M., . . . Iversen, V. C. (2018). The effects of routine outcome monitoring (ROM) on therapy outcome in the course of an implementation process: A randomized clinical trial. *Journal of Counseling Psychology*, *65*, 641–652.
- Chow, D., & Huixian, S. (2015). *The use of routine outcome monitoring in an Asian outpatient psychiatric setting*. Poster session presented at the World Federal of Mental Health Conference, Singapore.
- Comer, J. S., & Kendall, P. C. (2013). Methodology, design, and evaluation in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy change* (6th ed., pp. 21–48). New York, NY: Wiley.
- Cooper, M., Stewart, D., Sparks, J., & Bunting, L. (2013). School-based counseling using systematic feedback: A cohort study evaluating outcomes and predictors of change. *Psychotherapy Research*, *23*, 474–488.
- Creswell, J. W., Hanson, W. E., Clark Plano, V. L., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist*, *35*, 236–264. <http://dx.doi.org/10.1177/0011000006287390>
- Davidson, A. H., Poulsen, S., Lindschou, J., Winkel, P., Tróndarson, M. F., Waadegaard, M., & Lau, M. (2017). Feedback in group psychotherapy for eating disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *85*, 484–494. <http://dx.doi.org/10.1037/ccp0000173>
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York, NY: Praeger.
- Duncan, B., & Sparks, J. (2010). *Heroic clients, heroic agencies: Partners for change* (2nd ed.). Jensen Beach, FL: Author.
- Duncan, B., & Sparks, J. (2018). *The Partners for Change Outcome Management System: An integrated elearning manual for everything PCOMS*. Jensen Beach, FL: Author.
- Duncan, B. L. (2014). *On becoming a better therapist: Evidence based practice one client at a time* (2nd ed.). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/14392-000>
- Duncan, B. L., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client's frame of refer-



- ence. *Psychotherapy*, 52, 391–401. <http://dx.doi.org/10.1037/pst0000026>
- Duncan, B. L., & Sparks, J. A. (2002). *Heroic clients, heroic agencies: Partners for change*. Jensen Beach, FL: Author.
- Eisen, S. V., Wilcox, M., Leff, H. S., Schaefer, E., & Culhane, M. A. (1999). Assessing behavioral health outcomes in outpatient programs: Reliability and validity of the BASIS-32. *The Journal of Behavioral Health Services & Research*, 26, 5–17. <http://dx.doi.org/10.1007/BF02287790>
- Farber, B. A. (2003). Patient self-disclosure: A review of the research. *Journal of Clinical Psychology*, 59, 589–600. <http://dx.doi.org/10.1002/jclp.10161>
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55, 316–340. <http://dx.doi.org/10.1037/pst0000172>
- Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, 68, 179–188. <http://dx.doi.org/10.1176/appi.ps.201500439>
- Gertler, P., & Cameron, I. D. (2018). Making sense of data analytic techniques used in a Cochrane systematic review. *Brain Impairment*, 19, 81–90. <http://dx.doi.org/10.1017/BrImp.2017.27>
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337–1345. <http://dx.doi.org/10.1097/00004583-200111000-00015>
- Gowers, S. G., Harrington, R. C., Whitton, A., Beevor, A. S., Lelliott, P., Wing, J. K., & Jezzard, R. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *British Journal of Psychiatry*, 174, 413–416.
- Hansen, B., Howe, A., Sutton, P., & Ronan, K. (2015). Impact of client feedback on clinical outcomes for young people using public mental health services: A pilot study. *Psychiatry Research*, 229, 617–619. <http://dx.doi.org/10.1016/j.psychres.2015.05.007>
- Hibbard, J. H., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the Patient Activation Measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Services Research*, 39, 1005–1026. <http://dx.doi.org/10.1111/j.1475-6773.2004.00269.x>
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). Retrieved from <http://www.handbook.cochrane.org>
- Hiles, D. (2008). Transparency. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative research methods* (pp. 891–893). Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412963909.n467>
- Janse, P. D., De Jong, K., Van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2017). Improving the efficiency of cognitive-behavioural therapy by using formal client feedback. *Psychotherapy Research*, 27, 525–538. <http://dx.doi.org/10.1080/10503307.2016.1152408>
- Kellybrew-Miller, A. (2015). *The impact of systematic client feedback on client outcomes in a community mental health center* (Unpublished doctoral dissertation). University of Central Arkansas, Conway, AR.
- Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., . . . Moore, M. (2016). Routine use of patient reported measures (PROMs) for improving treatment of common mental disorders in adults. *Cochrane Database of Systematic Reviews*, 2016(7), CD011119.
- Kodet, J., Reese, R., Duncan, B., & Bohanske, R. (2019). Psychotherapy for depressed youth in poverty: Benchmarking outcomes in a public behavioral health settings. *Psychotherapy*, 56, 254–259.
- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy*, 52, 381–390. <http://dx.doi.org/10.1037/pst0000027>
- Lambert, M. J. (2017). Implementing routine outcome monitoring (ROM) in clinical practice. *Journal of Health Service Psychology*, 43, 55–59.
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48, 72–79. <http://dx.doi.org/10.1037/a0022238>
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55, 520–537. <http://dx.doi.org/10.1037/pst0000167>
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11, 49–68. <http://dx.doi.org/10.1080/173663852>
- Lester, M. C. (2012). *The effectiveness of client feedback measures with adolescents in an acute psychiatric inpatient setting* (Unpublished doctoral dissertation). University of Central Arkansas, Conway, AR.
- Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage & Family Living*, 21, 251–255. <http://dx.doi.org/10.2307/348022>
- Miller, S., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of reliability, validity, and feasibility of a brief visual analogue measure. *Journal of Brief Therapy*, 2, 91–100.
- Murphy, K. P., Rashleigh, C. M., & Timulak, L. (2012). The relationship between progress feedback and therapeutic outcome in student counseling: A randomised control trial. *Counselling Psychology Quarterly*, 25, 1–18. <http://dx.doi.org/10.1080/09515070.2012.662349>
- Nelson, T. S., Chenail, R. J., Alexander, J. F., Crane, D. R., Johnson, S. M., & Schwallie, L. (2007). The development of core competencies for the practice of marriage and family therapy. *Journal of Marital and Family Therapy*, 33, 417–438. <http://dx.doi.org/10.1111/j.1752-0606.2007.00042.x>
- Østergård, O. K., Randa, H., & Hougaard, E. (2018). The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy: A systematic review and meta-analysis. *Psychotherapy Research*. Advance online publication. <http://dx.doi.org/10.1080/10503307.2018.1517949>
- Reese, R. J., Duncan, B. L., Bohanske, R. T., Owen, J. J., & Minami, T. (2014). Benchmarking outcomes in a public behavioral health setting: Feedback as a quality improvement strategy. *Journal of Consulting and Clinical Psychology*, 82, 731–742. <http://dx.doi.org/10.1037/a0036915>
- Reese, R. J., Duncan, B. L., Kodet, J., Brown, H. M., Meiller, C., Farook, M. W., . . . Bohanske, R. T. (2017). Patient feedback as a quality improvement strategy in an acute care, inpatient unit: An investigation of outcome and readmission rates. *Psychological Services*, 15, 470–476.
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy*, 46, 418–431. <http://dx.doi.org/10.1037/a0017901>
- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy*, 47, 616–630. <http://dx.doi.org/10.1037/a0021182>
- Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2016). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *Patient Education and Counseling*, 99, 164–168. <http://dx.doi.org/10.1016/j.pec.2015.07.016>
- Rosenthal, R. (1979). File drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641. <http://dx.doi.org/10.1037/0033-2909.86.3.638>
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research*, 25, 396–407. <http://dx.doi.org/10.1080/10503307.2014.900875>

- She, Z., Duncan, B. L., Reese, R. J., Sun, Q., Shi, Y., Jiang, G., . . . Clements, A. L. (2018). Client feedback in China: A randomized clinical trial in a college counseling center. *Journal of Counseling Psychology, 65*, 727–737.
- Slone, N. C., Reese, R. J., Mathews-Duvall, S., & Kodet, J. (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice, 19*, 122–136. <http://dx.doi.org/10.1037/gdn0000026>
- 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>
- Sparks, J. A., & Duncan, B. L. (2018). The Partners for Change Outcome Management System: A both/and system for collaborative practice. *Family Process, 57*, 800–816. <http://dx.doi.org/10.1111/famp.12345>
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- van Oenen, F. J., Schipper, S., Van, R., Schoevers, R., Visch, I., Peen, J., & Dekker, J. (2016). Feedback-informed treatment in emergency psychiatry; a randomised controlled trial. *BMC Psychiatry, 16*, 110. <http://dx.doi.org/10.1186/s12888-016-0811-z>
- Wampold, B. E., Flückiger, C., Del Re, A. C., Yulish, N. E., Frost, N. D., Pace, B. T., . . . Hilsenroth, M. J. (2017). In pursuit of truth: A critical examination of meta-analyses of cognitive behavior therapy. *Psychotherapy Research, 27*, 14–32. <http://dx.doi.org/10.1080/10503307.2016.1249433>
- Wells, M. G., Burlingame, G. M., & Lambert, M. J. (1996). The Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training, 33*, 275–283.
- Winkelhorst, Y., Hafkenscheid, A., & de Groot, E. (2013). Verhoogt Routine Process Monitoring (RPM) de effectiviteit van behandeling [Increasing the effectiveness of treatment with Routine Process Monitoring (RPM)]. *Tijdschrift voor Psychotherapie, 39*, 146–156. <http://dx.doi.org/10.1007/s12485-013-0028-2>

Received January 2, 2019

Revision received September 3, 2019

Accepted September 20, 2019 ■