Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: http://www.tandfonline.com/loi/tpsr20

The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis

Ole Karkov Østergård, Hilde Randa & Esben Hougaard

To cite this article: Ole Karkov Østergård, Hilde Randa & Esben Hougaard (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, DOI: 10.1080/10503307.2018.1517949

To link to this article: https://doi.org/10.1080/10503307.2018.1517949

Published online: 13 Sep 2018.

Psychotherapy Research, 2018
https://doi.org/10.1080/10503307.2018.1517949

EMPIRICAL PAPER

The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis

OLE KARKOV ØSTERGÅRD, HILDE RANDA, & ESBEN HOUGAARD

Department of Psychology and Behavioural Sciences, Aarhus University, Aarhus C, Denmark

(Received 15 May 2018; revised 16 August 2018; accepted 16 August 2018)

Abstract
Objective: The aims of the study were to evaluate the effects of using the Partners for Change Outcome Management System (PCOMS) in psychotherapy and to explore potential moderators of the effect. **Method**: A comprehensive literature search including grey literature was conducted to identify controlled outcome studies on the PCOMS, randomized (RCTs), or non-randomized trials (N-RCT). **Results**: The literature search identified 18 studies, 14 RCTs, and four N-RCTs, including altogether 2910 participants. The meta-analysis of all studies found a small overall effect of using the PCOMS on general symptoms (g = 0.27, p = .001). The heterogeneity of the results was substantial. Moderation analyses revealed no effect of the PCOMS in psychiatric settings (g = 0.10, p = .144), whereas a positive effect was found in counseling settings (g = 0.45, p < .001), although almost all of these studies were characterized by a positive researcher allegiance and using the PCOMS Outcome Rating Scale (ORS) as the only outcome measure. **Conclusion**: The meta-analysis revealed a small overall effect of using the PCOMS, but no effect in psychiatric settings. The positive results in counseling settings might be biased due to researcher allegiance and use of the ORS as the only outcome measure.

Keywords: client feedback; Partners for Change Outcome Management System (PCOMS); Routine Outcome Monitoring (ROM); psychotherapy outcome; meta-analysis

Clinical or methodological significance of this article: This first meta-analysis solely on the Partners for Change Outcome Management System (PCOMS) questions the evidence base of one of the most used Routine Outcome Management systems. We found no indication of using the PCOMS in psychiatric settings and the moderate effect in counseling settings might be biased due to positive researcher allegiance and outcome measure. There is a need for more studies, especially studies in counseling settings, using other outcome measures than the PCOMS Outcome Rating Scale.

Psychotherapy is generally found to be helpful for clients; however, many do not benefit, and dropouts are frequent. Although about 70% of all psychotherapy clients achieve reliable change, less than 50% can be considered "recovered" after treatment in the sense that they are now more similar to a normal than to a clinical population (Lambert, 2013). In a meta-analysis, Swift and Greenberg (2012) showed that, on average, 19.7% of all clients dropped out of psychotherapy. Previous research has indicated that a lack of early progress in therapy is a risk factor for not improving or dropping out of therapy (Lambert, 2013). Therapists are generally poor at predicting negative outcomes of psychotherapy. For example, in a study by Hannan et al. (2005), 48 therapists only predicted one of 40 student clients who were deteriorated at the end of therapy, whereas an algorithm based on questionnaire feedback was able to predict 77% of all cases correctly. Among clinicians and researchers, such findings have spurred an interest in the collection of systematic client feedback in psychotherapy.

Routine Outcome Monitoring (ROM) is a general term covering a range of different client feedback

Correspondence concerning this article should be addressed to Ole Karkov Østergård Department of Psychology and Behavioural Sciences, Aarhus University, Bartholins Allé 9, 8000 Aarhus C, Denmark. E-mail: karkov@psy.au.dk

systems established to monitor the client's progress during psychotherapy through repeated measurements (Castonguay, Barkham, Lutz, & McAleavy, 2013; Lutz, de Jong, & Rubel, 2015). Treatment progress can thus be compared with a rationally or empirically derived expected treatment response (ETR), which makes it possible to identify clients who are not improving as expected and hence at risk of a poor treatment outcome (Castonguay et al., 2013; Lutz et al., 2015). This feedback offers clinicians, and sometimes also clients, an opportunity to reflect on the course of therapy with the possibilities of making changes; for instance, trying to strengthen the therapeutic alliance, shift focus, revisit goals, or alter interventions to prevent client non-response, deterioration or dropout. ROM is rooted in a broader movement within patient-focused research (Howard, Moras, Brill, Martinovich, & Lutz, 1996) and practice-based evidence (Barkham & Margison, 2007) sharing the crucial goal of making psychotherapy in clinical practice a research-supported intervention (Lutz et al., 2015).

During the last 15–20 years, a variety of feedback systems have been developed and implemented in different places in the world; 10 of the most widely used systems are described in Drapeau (2012). In England, the Clinical Outcomes in Routine Evaluation (CORE) is widely used (Barkham et al., 2001) and in 2002 the "Child Outcomes Research Consortium" (2018) was founded with the goal of collecting and using evidence to improve children and young people's mental health. Although not launched as a ROM system, therapists in the program "Improving Access to Psychological Therapies" (IAPT) are supposed to receive weekly outcome-informed supervision based on session-wise measurements (Clark, 2018). In Germany, at least two systems have been developed (Kordy, Hannöver, & Richard, 2001; Lutz, Böhnke, & Köck, 2011), and in Australia, a nationwide program of ROM was implemented in 2000 (Burgess et al., 2009). In the Netherlands, ROM is also widely used, for example, within the fields of anxiety and affective disorders (de Beurs et al., 2011) and psychosis (Tasma et al., 2016). The Outcome Questionnaire 45 (OQ-45; Lambert et al., 2004) has been widely used and researched in the USA (Lambert et al., 2002), in the Netherlands (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012), as well as in Norway (Amble, Gude, Stubdal, Andersen, & Wampold, 2015). The Partners for Change Outcome Management System (PCOMS; Miller & Duncan, 2004) is also widely researched and used, both in the USA (e.g., Reese, Norsworthy, & Rowlands, 2009a, 2009b), Norway (e.g., Anker, Duncan, & Sparks, 2009), the Netherlands (e.g., Janse, de Jong, van Dijk, Hutschemaekers, & Verbraak, 2017), Ireland (Murphy, Rashleigh, & Timulak, 2012), Australia (Hansen, Howe, Sutton, & Ronan, 2015), China (She et al., 2018), and Denmark (Davidsen et al., 2017). In February 2018, a total of 92 partners or certificated PCOMS trainers from 14 different countries were registered at the homepage of the S. D. Miller initiated project the "International Center for Clinical Excellence" (2018) or at the homepage of the B. L. Duncan launched "Heart and Soul of Change Project" (2018). B. L. Duncan (personal communication, February 7, 2018) stated that their system had about 30000 individual licenses, more than 1000 group licenses, and over a million administrations in databases across 20 countries. S. D. Miller (personal communication, February 7, 2018) informed that about 45000 therapists were registered users of their system, excluding group licenses.

The PCOMS makes use of two four-item feedback measures, the Outcome Rating Scale (ORS), focusing on the client's level of general well-being or distress; and the Session Rating Scale (SRS), focusing on the therapeutic alliance. The ORS measures well-being on four dimensions (individual, interpersonal, social, and overall), and the SRS measures the alliance on four dimensions (relationship, goal, method, and overall). Both measures consist of four visual analog scales with 10 cm long lines; the client places a mark on a paper-and-pencil or a tablet (electronic) version of the scales. The scales are summed up to a total ORS or SRS score ranging from 0 to 40. The ORS is administered at the beginning of each session by asking the client to cover the last week (first session) or period since the last session. The client completes the SRS at the end of each session. It takes 1–3 minutes to administer each scale.

Both PCOMS's scales have acceptable psychometric properties. For the ORS, the internal consistency has been reported in at least 14 studies with Cronbach's alphas ranging from .81 (Seidel, Andrews, Owen, Miller, & Buccino, 2017) to .93 (Miller, Duncan, Brown, Sparks, & Claud, 2003). The convergent validity between the ORS total score and the OQ-45 total score has been estimated in at least four studies with correlations between .57 (Bringhurst, Watson, Miller, & Duncan, 2006) and .76 (Campbell & Hemsley, 2009). The test-retest reliability has been reported in at least four non-clinical samples as correlations ranging from .66 (Miller et al., 2003) to .80 (Bringhurst et al., 2006). The SRS has psychometric properties comparable to that of the ORS (See Duncan & Reese, 2015). The PCOMS manual (Miller & Duncan, 2004) established a clinical cutoff score for adults at 25 points for the ORS total scale and a reliable change index (RCI) of 5 points according to the criteria of Jacobson

and Truax (1991). This makes it possible to use the following classification of clients: clinically significant change, reliable change, no change, or deterioration. The clients were defined as not-on-track (NOT) of a good outcome early in psychotherapy if they had not achieved at least a 5-point increase on the ORS total scale in session three. In addition, the ETR was empirically derived from the client's initial ORS score by comparing the score to the (average) treatment response of previously treated clients (as a function of their initial ORS scores). The main purpose of these calculations is to serve as a benchmark and to inform the therapist–client discussion of treatment progress. Different web-based systems have been developed in order to make the ETR immediately available for therapist and client. The apparent simplicity and ease of administration of the PCOMS makes it an attractive choice as a systematic client feedback tool in psychotherapy.

Most often ROM has been evaluated in its generic form without distinguishing between the different systems. Collecting client feedback in psychotherapy was recommended in the American Psychological Association's (2006) manifesto on evidence-based practice and considered "demonstrably effective" by the APA interdivisional task force on evidence-based therapy relationships (Norcross, 2011; Norcross & Wampold, 2011). However, the ROM systems differ in many ways. The PCOMS measures outcome as well-being on visual analog scales and not as symptoms on Likert scales. The PCOMS's scales are completed in the sessions and feedback is given immediately both to the therapist and client. The PCOMS measures the therapeutic alliance in each session, whereas Lambert's (2015) OQ-system only measures the alliance (plus motivation, social support, and stressful life events) when the client is NOT. Finally, the PCOMS does not offer specific clinical guidelines for NOT clients, as the Clinical Support Tools in the OQ-system. These distinctive characteristics of the PCOMS and its simplicity and widespread use are arguments for performing a meta-analysis specifically on the PCOMS.

The PCOMS has been included in three meta-analyses on feedback systems in psychotherapy. The meta-analysis by Lambert and Shimokawa (2011a, 2011b) was conducted as part of the interdivisional task force (Norcross, 2011) and included three studies on the PCOMS in addition to the six studies on the OQ-45 (previously analyzed in Shimokawa, Lambert, and Smart (2010)). The PCOMS enhanced post-treatment outcome with an incremental effect size (ES) of $g = 0.48$ for all treated clients, whereas the OQ-45 achieved an ES of 0.53 when only NOT clients were included. Both systems halved the number of clients who deteriorated. In their Cochrane Report on ROM in individual psychotherapy for common mental health disorders, Kendrick et al. (2016) included 12 RCTs, nine with the OQ-45 and three with the PCOMS (of which two were also in Lambert & Shimokawa, 2011a, 2011b). They found no significant differences in outcome between the feedback and the control group ($g = 0.09$, $p = .10$). No separate analysis was performed on the PCOMS. In their meta-analysis, Tam and Ronan (2017) included only feedback studies from youth samples (10–19 years) and found a small positive effect of feedback with an ES of $g = 0.20$ for the four RCTs (including two studies on the PCOMS). The PCOMS has been included in the National Registry of Evidence-based Programs and Practices of the Substance Abuse and Mental Health Administration (SAMHSA, 2017). The previous meta-analyses dealt with ROM in general and included only a few studies on the PCOMS.

Several moderators of the results of using client feedback, including the PCOMS, have been suggested. Thus, some authors have argued that it could be challenging to use the PCOMS in the treatment of more severely disordered clients, since such clients might prefer expert advice and find it difficult to reflect on and discuss the feedback (e.g., van Oenen et al., 2016). Other researchers have questioned the validity of the ORS as an outcome measure and found that the ORS yields larger ESs than other outcome measures (Janse, Boezen-Hilberdink, van Dijk, Verbraak, & Hutschemaekers, 2014; Rise, Eriksen, Grimstad, & Steinsbekk, 2016; Seidel et al., 2017). The ORS is usually completed in the presence of the therapist, which might lead to social desirability effects due to the clients' wish to please the therapist and portray themselves or the therapist in a more positive light. Therapist fidelity (adherence and competence) and amount of prior training or supervision during therapy have also been claimed to be crucial for the effectiveness of ROM (de Jong, 2016; Duncan & Reese, 2015; She et al., 2018). Finally, meta-analyses have consistently found a moderate association between researcher allegiance and outcome in psychotherapy research (e.g., Munder, Brütsch, Leonhart, Gerger, & Barth, 2013; $r = 0.26$).

The aims of the present meta-analysis are to evaluate the effects of using the PCOMS as a feedback tool in psychotherapy, overall and specifically for NOT clients, and to explore potential moderators of the overall effect. The following pre-specified moderators were planned to be investigated (Østergård, Randa, & Hougaard, 2017): (1) Study design, RCT vs. N-RCT, (2) use of the ORS as outcome measure, (3) treatment setting (psychiatric vs. counseling), (4) treatment format (individual vs. couple vs. group),

(5) therapists' theoretical approach, (5) researcher allegiance, (6) therapist training and supervision in the PCOMS, (7) client age, and (8) treatment duration.

Method

The study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations (Moher, Liberati, Tetzlaff, & Altman, 2009). A prestudy protocol was published (Østergård et al., 2017). All statistical analyses were performed with Comprehensive Meta-Analysis, version 3 (2014).

Inclusion Criteria

To be included in the meta-analysis, studies had to deal with the PCOMS and be either a randomized controlled trial (RCT) or a group comparison study without randomization (N-RCT). The experimental condition had to include the PCOMS as an "add-on intervention" to an intervention without the PCOMS in the control condition. The intervention could be of any theoretical orientation (including eclectic or not specified), modality (individual, group, family, internet/phone), and treatment setting (primary care, student counseling center, psychiatric ward). Participants should be clients seeking help for their problems with no restrictions as to age, gender, ethnicity, or diagnoses. Contrary to the initially published protocol, we decided to include two studies using the ORS without the SRS in the intervention arm and to run a sensitivity analysis to test whether full implementation of PCOMS (i.e., both the ORS and the SRS) would make a difference with regard to treatment effect.

Search Strategy

The electronic databases of PsycINFO, PubMed, SCOPUS, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) were searched using the following keywords: "PCOMS" OR "Partners for change" OR "Outcome Rating Scale" OR "client feedback" OR "feedback informed" OR "outcome informed" OR "routine outcome monitoring" OR "continuous outcome monitoring" OR "continuous outcome assessment". The search was restricted to studies published in 2000 or later as the PCOMS did not appear until 2000 (Duncan, 2012). The search was not restricted by language.

Grey literature, including dissertations, was searched through the ProQuest Dissertation & Theses database, the OpenSIGLE (former SIGLE), the database of the Healthcare Management Information Consortium (HMIC), Google Scholar (top 100 hits) and Google.com by use of the keywords: "PCOMS" OR "partners for change." Clinical trials were searched using the World Health Organization's trials portal (ICTRP), and ClinicalTrials.gov. Unpublished studies were requested from known researchers within the field: B. L. Duncan and the 26 leaders and trainers from the "Heart and Soul of Change Project" (2017), S. D. Miller and all subscribers to the discussion forum of the "International Center for Clinical Excellence" (2017), and the 30 founding members of the "International Network Supporting Psychotherapy Innovation and Research into Effectiveness" (INSPIRE, personal communication, August 22, 2017).

For the database search, the first and the second author independently conducted the literature search and study selection using the web-based systematic review software Covidence (2017). Disagreements ($n = 3$) were resolved through discussion. In addition, a backward search was conducted using reference lists of identified articles and systematic reviews together with a forward search using citation tracking until no additional relevant articles were identified (including unpublished and in-press citations). The first author performed this citation tracking and the search for grey literature.

The searches were conducted on June 15 and rerun on December 6, 2017, just before the final analyses. The first author e-mailed the corresponding authors to all the primary studies asking for relevant information not reported in the publications. The search strategies are available at Østergård et al. (2017).

Primary and Secondary Outcomes

Due to comparability, we preferred outcome measures in the form of general symptoms or distress scales. The primary outcome measures were prioritized as follows: (1) A generic outcome measure of general symptoms or distress, such as the Global Severity Index (GSI) on versions of the Symptom Checklist (SCL; Derogatis, 1992), or the mean score on OQ-45 (Lambert et al., 2004). (2) If the study did not report a generic outcome, a standardized mean of all symptomatic outcome measures was calculated. (3) If the study reported neither 1 nor 2, the ORS was used as the primary outcome.

Secondary outcomes were planned to include: (1) Number of dropouts, (2) number of deteriorated clients, (3) mean number of psychotherapy sessions, (4) social functioning, self-report, (5) social

functioning, objective measured, for example, completion or dropout of higher education, (loss of) employment, and divorce rate, and, finally, (6) costs, defined as direct costs of implementing and using the PCOMS.

Assessment of Risk of Bias

The first and third author independently assessed the risk of bias for the studies using the criteria from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins, Altman, & Sterne, 2011). The Cochrane risk of bias assessment tool has seven potential sources of bias: random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. We judged the risk of bias for each of these dimensions as high, low or unclear and justified our judgment in a risk of bias table. Disagreements between the raters ($n = 18$ (14.29%)) were discussed until a negotiated conclusion was reached.

Analytical Strategy

Effect Sizes. The ESs for continuous outcomes were calculated as the standardized difference in means according to the formula $d = (M_{pre1} - M_{post1}) - (M_{pre2} - M_{post2})/SD_{ChangePooled}$ where $(M_{pre1} - M_{post1})$ is the mean difference in the control group and $(M_{pre2} - M_{post2})$ is the mean difference in the PCOMS group, and $SD_{ChangePooled} = \sqrt{(((n_1 - 1) * SD^2_{Change1} + (n_2 - 1) * SD^2_{Change2})/(n_1 + n_2 - 2))}$. Since the calculation of SD_{change} requires knowledge of the pre–post correlations, which were not reported, we imputed a conservative estimate of $r = 0.5$. This calculation, based on differences between pre- and post-scores, is less sensitive to pre-treatment differences than ESs based on post-scores (Borenstein, Hedges, Higgins, & Rothstein, 2009). Hedges' g correction was used to correct for possible bias due to small sample size. A positive ES indicates that the PCOMS group did better than the control group. ES estimates were aggregated across studies, adopting a random effects model, which allows for a distribution of true effects across studies (Borenstein, Hedges, Higgins, & Rothstein, 2010). ES parameters for individual studies were treated as if they were a random sample from a larger population, thus allowing for generalizations beyond the observed studies (Hedges & Vevea, 1998). In the random effects model, the weight assigned to each study is based on the inverse variance including both the within-study and the between-study variance, thus assigning relatively more weight to small studies in the estimation of mean ES. For dichotomous outcomes (i.e., dropout, deterioration) ESs were calculated as odds ratio (OR). Completer data were used in the meta-analyses because an intention to treat analysis (ITT) was only reported in three primary studies (Davidsen et al., 2017; Rise et al., 2016; van Oenen et al., 2016).

Heterogeneity Analyses. Heterogeneity was examined using Q, I^2 and T^2 statistics and the prediction interval (PI). Q was calculated as the distance of each study from the mean effect and is used to test for significant heterogeneity. I^2 is an estimate of the proportion of true variance between studies compared to variation due to sampling error. Thus, I^2 tells us what proportion of variance would remain if we could remove sampling error (Higgins, Thompson, Deeks, & Altman, 2003). I^2 values of 0%, 25%, 50%, and 75% are usually considered negligible, low, moderate, and high, respectively (Higgins & Thompson, 2002). However, this interpretation has been questioned, also by Higgins in Borenstein, Higgins, Hedges, and Rothstein (2017). T^2 is an estimate of the variance of true effects. Hence, T is the standard deviation of true effects. T can be used to calculate the PI, which estimates how widely the true effect varies in different populations around the mean (Borenstein et al., 2017). Borenstein et al. (2017) argued that the PI should supplement, if not replace, I^2 as an estimation of heterogeneity. To avoid confusion, the PI is not the same as the confidence interval (CI). The CI tells us how precisely the mean ES has been estimated. It is a property of the sample and depends on the number of studies in the analyses (with more studies we can estimate the mean more precisely). By contrast, the PI is an index of dispersion that estimates how widely the effect varies across populations. The PI is not related to the number of studies. The true effect size falls within the range of the PI for 95% of all populations (randomly sampled from the same universe of populations as those included in the meta-analysis). Thus, in this meta-analysis, the PI will be used to evaluate whether the mean effect size can be generalized to all populations, or if a wide dispersion in effect across populations should be expected when using the PCOMS.

Moderation Analyses. Explorative moderation analyses were conducted with subgroup analyses for the categorical variables and with meta-regression for the continuous variables. The moderation analyses were restricted to the primary outcome. Subgroups were combined using a random effects model allowing for some true variation in effects

within subgroups. The within-group estimation of T^2 was pooled because we assumed that the true study-to-study dispersion was the same within all subgroups and because the number of studies within subgroups was too small to yield an acceptably accurate estimate of T^2 without pooling (Borenstein et al., 2009). The meta-regression analyses were based on a random effects model and estimated with the method of moments yielding a Q statistics comparable to that of the chosen method of subgroup analyses. For both categorical and continuous variables, the R^2 analog, defined as the total between-study variance explained by the moderator, was calculated based on the meta-regression matrix (Borenstein et al., 2009).

The analyses were performed only if at least three studies could be included in the regression analyses, and in each subgroup in the subgroup comparisons. Moderators were investigated for overlap (multi-collinearity) before deciding whether or not to conduct moderation analysis and multivariate regression analysis.

Researcher allegiance was defined as positive if at least one of the study authors was registered as partner or certificated trainer at the "International Center for Clinical Excellence" (2017) or the "Heart and Soul of Change Project" (2017), and otherwise as neutral. Treatment dropout was defined as unplanned endings (client canceled or failed to attend).

Sensitivity Analyses. The goal of the sensitivity analyses was to investigate how the result might have changed if different inclusion criteria had been used. Significant outcome and moderation analyses were performed with RCTs only. The primary outcome analysis was re-run with and without studies using the ORS only (and not the SRS), with and without unpublished studies, and, finally, with and without studies with Z-values ≥ 2 below or above the mean to test for the impact of outlier studies.

Type of Outcome Measure. *Post-hoc*, it was decided to explore a potential difference in ES depending on the type of outcome measure by comparing the ES calculated from the ORS and a generic symptom measures for all studies which reported both types of outcome.

Not-on-track Clients. The primary outcome was analyzed separately for NOT clients. The definition of NOT followed the slightly different definitions in the primary studies; two studies defined NOT as clients who at session three had not made a specified increase at the ORS, at least six points in She et al. (2018) and five points in Reese et al. (2009b) as suggested in the PCOMS manual (Miller & Duncan, 2004). Two other studies used the same five-point criteria, but Davidsen et al. (2017) also included clients who at some point in treatment had deteriorated by at least five ORS points, and Janse et al. (2017) calculated NOT in session four or five. One study defined NOT as an ORS score below percentile 50 in the ETR trajectory at session three (Murphy et al., 2012).

Publication Bias. Publication bias was visually inspected using funnel plots and statistically tested with Egger's test (Egger, Smith, Schneider, & Minder, 1997) and the Trim and Fill method by Duval and Tweedie (2000). Egger's test is a regression test of the funnel plot asymmetry testing the association between the estimated intervention effect and study size (measured as the within-study standard error of the effect). The Trim and Fill method was used to plot the number of missing studies needed to make the funnel plot symmetrical and, if necessary, to adjust the overall ES accordingly. A random effect model was used to look for missing studies to the left of the mean. Cumulative analysis was used to assess the potential impact of bias and to evaluate the robustness of the effect size (Borenstein et al., 2009). Cumulative meta-analysis was performed by cumulatively adding each study after sorting the studies by standard errors from the most precise to the least precise. If the effect size point estimate did shift when the least precise (typically smaller) studies were added, this was interpreted as an indication of bias.

Results

In total, 18 studies were included in the analyses. The study identification and selection process is illustrated in Figure 1.

Characteristics of Studies

Characteristics of studies are seen in Table 1. Fourteen studies were RCTs, and four studies were N-RCTs. In 14 of the studies, the same therapists treated patients in both the PCOMS and the control condition, whereas four studies (Reese et al., 2009b; Reese, Toland, Slone, & Norsworthy, 2010; Rise et al., 2016; Winkelhorst, Hafkenscheid, & de Groot, 2013) used different therapists in the two conditions. The total number of participants included in the meta-analysis was 2910. Study sample size ranged from 39 to 741 with a mean of 161.7. Ten studies were from psychiatric settings

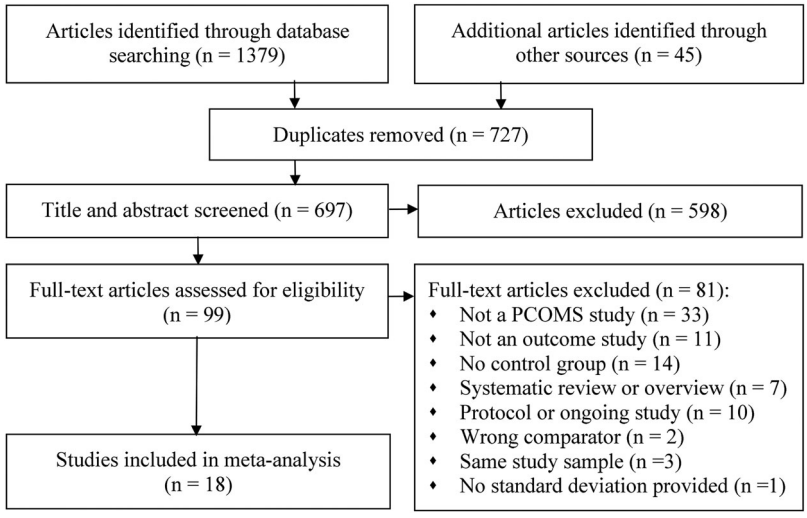

Figure 1. Flowchart of study identification and selection. PCOMS = Partners for Change Outcome Management System.

and included eight psychiatric outpatient services for adults and two for youths. Eight studies were from counseling settings and included four student counseling services, two university graduate training clinics, one family counseling clinic, and one military substance abuse counseling center. The participants had a variety of diagnoses and presenting problems with depression and anxiety being most frequent. The treatment modality was individual in 13 studies, group in three studies, and couple therapy in two studies. The treatment approach was described as eclectic or varying in 12 of the studies, as primarily systemic in two studies, and as cognitive–behavioral and interpersonal in one study each. Two studies did not report the treatment approach. The mean number of attended sessions ranged from 1.8 to 15.0 with a total mean of 7.8.

The ORS was used as the only outcome measure in nine studies, seven studies used both a generic symptom measure and the ORS, while two studies used a generic symptom measure without the ORS. Two studies used the ORS without the SRS in the PCOMS condition (Murphy et al., 2012; Schuman, Slone, Reese, & Duncan, 2015). Only five studies (Davidsen et al., 2017; Janse et al., 2017; Murphy et al., 2012; Reese et al., 2009b; She et al., 2018) reported the number of NOT clients with a study mean of 44.5% (range 25.7–81.3%) for the PCOMS condition and of 52.4% (range 37.9–78.5%) for the control condition. None of the studies measured observer-rated compliance with the PCOMS protocol, but four studies used a self-report treatment fidelity checklist about how the PCOMS was used (Davidsen et al., 2017; Kelly-brew-Miller, 2015; Lester, 2012; van Oenen et al., 2016). The amount of training before the intervention was reported in 14 studies and ranged from one to 12 hours and supervision during the intervention was reported in nine studies and varied from weekly to monthly (except Schuman et al., 2015 who reported zero hours of training and supervision). Nine studies were judged to have a positive researcher allegiance.

Assessment of Bias

Assessment of bias according to the Cochrane criteria is described in the Appendix. In summary, there was a high risk of bias in most studies, primarily due to lack of blinding of outcome assessment, incomplete outcome data, and other bias.

Primary Outcome

Overall Effect Size. The overall combined effect across the 18 studies comparing the PCOMS and the control condition was small ($g = 0.27$, CI [0.14, 0.41], $p < .001$; see Figure 2). A positive effect of the PCOMS was indicated for all the analyses by Hedges $g > 0$, or OR <1.

Follow-up analyses were not performed because only one study reported outcome at follow-up (Anker et al., 2009).

Heterogeneity. The heterogeneity among the 18 studies was substantial and significant ($Q = 48.13$, df = 17, $p < .001$, $I^2 = 64.68$, $T^2 = 0.05$). The PI ranged from −0.22 to 0.76.

Sensitivity Analyses. No significant difference in overall effect was found between the 14 RCTs

Table 1. Characteristics of studies.

Study	Setting and participants (% woman/M age), country	Treatment approach (allegiance)	Groups: n randomized/ analyzed	Duration in weeks (M sessions)	Outcome measures	Training (supervision) in hours*
Randomized controlled trials						
Anker et al., 2009	Couples in community family counseling (50.0/37.8), Norway	Eclectic couple therapy (positive)	Fb: 446/206 NFb: 460/ 204	NR (4.6)	**ORS**, deteriorated, sessions attended	11 (9)
Brattland et al., 2018	Psychiatric outpatients, primarily with mood and anxiety disorders, ADHD (63.3/34.1), Norway	Varied (neutral)	Fb: 85/55 NFb: 85/58	NR (12.0)	**BASIS-32**, ORS, deteriorated, sessions attended	6 (monthly)
Davidsen et al., 2017	Psychiatric outpatients with eating disorders (98.1/26.9), Denmark	Systemic group therapy (neutral)	Fb: 80/46 NFb: 79/51	20-25 (12.4)	**GSI**, ORS, EDE, SDS, WHO-5, SHI, sessions attended, *NOT*	6 (monthly)
Kellybrew-Miller, 2015	Psychiatric outpatients, primarily with mood and anxiety disorders (62.6/36.4), USA	Varied (neutral)	Fb: NR/43 NFb: NR/ 48	NR (2.2[a])	**SOS-10**[b], ORS, sessions attended	3 (NR)
Lester, 2012	Psychiatric youth inpatients, primarily with mood disorders (51.7/14.8), USA	Varied crisis interventions (neutral)	Fb: 81/58 NFb: 83/60	1-2 (1.8)	**Y-OQ-SR**, ORS, sessions attended	NR (monthly)
Murphy et al., 2012	University counseling clients (58.2/23.8), Ireland	Varied (neutral)	Fb[c]: NR/59 NFb: NR/ 51	NR (3.7)	**ORS**, deteriorated, NOT	NR (NR)
Reese et al., 2009a	University counseling clients (54.6/20.2), USA	Varied (positive)	Fb: 60/50 NFb: 53/24	NR (6.5)	**ORS**, deteriorated, sessions attended	1 (NR)
Reese et al., 2009b	Clients at a graduate training clinic (70.8/ 33.0), USA	Varied (positive)	Fb: 52/45 NFb: 44/29	NR (7.1)	**ORS**, deteriorated, sessions attended, NOT	1 (NR)
Reese et al., 2010	Couples at a graduate training clinic (50/ 30.2), USA	Systemic, eclectic couple therapy (positive)	Fb: 60/54 NFb: 50/38	2-17 (5.9)	**ORS**, deteriorated, *sessions attended*	1 (weekly)
Rise et al., 2016	Psychiatric outpatients, primarily with mood and anxiety disorders (62.7/29.9), Norway	Varied (neutral)	Fb: 37/22 NFb: 38/28	Max 48 (NR)	**BASIS-32**[d], ORS, PAM, TAS, CSQ, SF-12, PM, PP	12 (NR)
Schuman et al., 2015	Soldiers in counseling with substance abuse (12.0/27.1), USA	Eclectic process group therapy (positive)	Fb[c]: NR/137 NFb: NR/ 126	5 (3.9)	**ORS**, PPR, deteriorated, sessions attended	0 (0)
She et al., 2018	University counseling clients (78.5/21.4), China	Varied (positive)	Fb: 169/101 NFb: 163/ 85	NR (4.5)	**ORS**, deteriorated, sessions attended, NOT	1 (NR)
Slone et al., 2015[e]	University counseling clients (64.3/21.5), USA	Interpersonal process group therapy (positive)	Fb: NR/43 NFb: NR/ 41	10 (7.3)	**ORS**[f], deteriorated, sessions attended	1 (weekly)
van Oenen et al., 2016	Psychiatric outpatients, primarily with psychosis, mood disorder, PD (53.0/ 38.1), Netherlands	Varied (positive)	Fb: 149/72 NFb: 138/ 57	Max 24 (9.3)	**GSI (BSI)**[g], ORS, OQ-45, deteriorated, *sessions attended, dropout*	NR (regularly)

(*Continued*)

Table 1. Continued.

Study	Setting and participants (% woman/M age), country	Treatment approach (allegiance)	Groups: n randomized/ analyzed	Duration in weeks (M sessions)	Outcome measures	Training (supervision) in hours*
Non-randomized controlled trials						
Chow & Huixian, 2015[h]	Psychiatric outpatients, primarily with mood and anxiety disorders (51.4/35.3), Singapore	Eclectic (positive)	Fb: NR/79 NFb: NR/ 55	NR (4.5)	***ORS,*** *deteriorated, sessions attended*	*6* (NR)
Hansen et al., 2015	Psychiatric child & youth outpatients, primarily with mood and anxiety disorders (45.2/9-17), Australia	NR (neutral)	Fb: NR/38 NFb: NR/ 35	Max 12 (8)	**SDQY**[i], SDQP HoNOSCA, CGAS	NR (NR)
Janse et al., 2017	Psychiatric outpatients, primarily with somatoform, adjustment, mood and anxiety disorders (53.4/42.6), Netherlands	CBT (neutral)	Fb: 2009/346[j] NFb: 1960/ 395	NR (15.0)	**GSI**, ORS, deteriorated, sessions attended, NOT	4 (NR)
Winkelhorst et al., 2013	Psychiatric outpatients, primarily with personality disorders (71.0/33.3), Netherlands	NR (neutral)	Fb: 35/19 NFb: 32/20	NR (15.0)	**OQ-45**[k], *deteriorated,* dropout	6 (2)

Notes. The outcome measures in bold is used as the primary outcome in the meta-analyses. Outcome measures in italic originate from e-mail requests to researchers. Fb = Partners for Change Outcome Management System (PCOMS), NFb = control group, NR = not reported, ORS = Outcome Rating Scale, BASIS-32 = Behavior and Symptom Identification Scale 32, GSI = Global Severity Index of the Symptom Checklist (SCL-90-R), EDE = Eating Disorder Examination interview, SDS = Sheehan Disability Scale, WHO-5 = WHO-Five Well-being Index, SHI = Self-harm Inventory, NOT = not-on-track, SOS-10 = Schwartz Outcome Scale-10, Y-OQ-SR = Youth Outcome Questionnaire Self-Report, PAM = Patient Activation Measure, TAS = Treatment Alliance Scale, CSQ = Client Satisfaction Questionnaire-8, SF-12 = Short Form-12, PM = Patient Motivation, PP = Patient Participation, PPR = Patient Progress Report, rated by therapist and commander, PD = personality disorder, BSI = Brief Symptom Inventory, OQ-45 = Outcome Questionnaire 45, SDQY = Strengths and Difficulties Questionnaire Youth, SDQP = Strengths and Difficulties Questionnaire Parents, HoNOSCA = Health of the Nation Outcome Scale for Children and Adolescents, CGAS = Children's Global Assessment Scale, CBT = Cognitive Behavioral Therapy.

[a]The 10 (6.2%) clients reported as having 6+ sessions were included in the calculation with 6 sessions.

[b]One therapist collected 49.5% of the outcome data.

[c]Only the ORS (and not the SRS) was used in the intervention group.

[d]Outcome data collected 6 and 12 months after start of treatment, and not at the last session. 12 months used as outcome in the meta-analysis.

[e]Groups-as-a-whole (and not individuals) were randomized.

[f]Second outcome measure used in the primary study but not reported.

[g]Outcome data collected 6 and 12 weeks after treatment started (24 weeks for a subsample). 12 weeks used as outcome in the meta-analysis.

[h]Poster. Calculations based on raw data kindly provided by D. Chow.

[i]Outcome measures collected at discharge or after 3 months, even if the treatment had not ended.

[j]Number of analyzed (completers) from P. Janse (personal communication, August 13, 2018).

[k]Outcome data collected after eight and 15 sessions, and not at the last session. Outcome after 15 sessions used in the meta-analysis.

*One day calculated as 6 hours. Training defined as therapist training in the PCOMS before the study period and supervision as PCOMS supervision during the study period.

and the four N-RCTs ($Q = 2.32$, df = 1, $p = .128$, $R^2 = 0.33$), although the RCTs had a numerically higher ES ($g = 0.32$, CI [0.18, 0.46], $p < .001$) than the N-RCTs ($g = 0.10$, CI [−0.15, 0.35], $p = .446$).

Excluding the two studies using the ORS without the SRS in the intervention arm did not change the effect of the PCOMS ($g = 0.28$, CI [0.13, 0.44], $p < .001$), neither did exclusion of the three unpublished studies by Kellybrew-Miller (2015), Lester (2012), and Chow and Huixian (2015) ($g = 0.30$; CI [0.14, 0.46], $p < .001$). When excluding the six studies (Anker et al., 2009; Brattland et al., 2018; Reese et al., 2009a, 2009b; Reese et al., 2010; She et al., 2018) with Z-values ≥ 2 above the mean, the PCOMS effect was no longer significant ($g = 0.08$; CI [−0.01, 0.17], $p = .063$).

Type of Outcome Measure. *Post-hoc* analysis found that the seven studies, all in psychiatric settings (Brattland et al., 2018; Davidsen et al., 2017; Janse

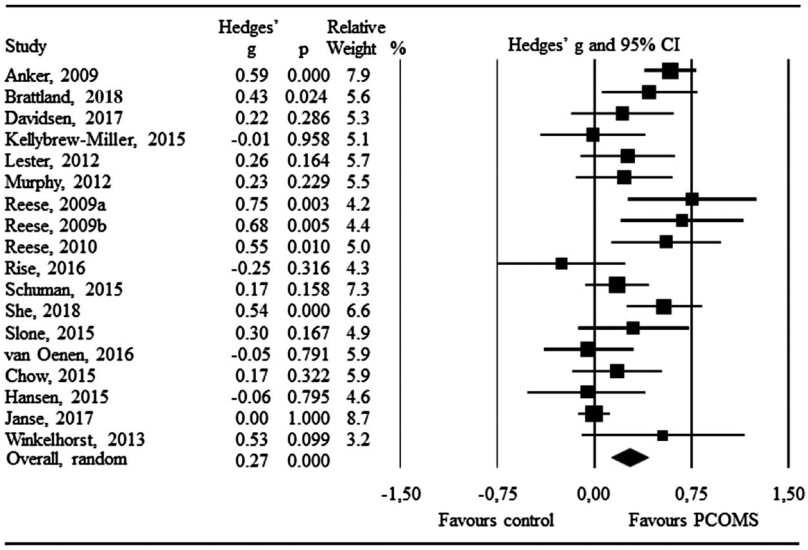

Figure 2. Post-treatment overall effect of the Partners for Change Outcome Management System (PCOMS) compared to control.

et al., 2017; Kellybrew-Miller, 2015; Lester, 2012; Rise et al., 2016; van Oenen et al., 2016) reporting outcome both on the ORS ($g = 0.19$, CI [0.02, 0.37], $p = .030$) and general symptoms ($g = 0.08$, CI [−0.07, 0.22], $p = .284$) had a difference in ES of 0.11. In five of these seven studies, the ORS was completed outside the therapy session in both the PCOMS and the control condition, whereas the ORS was completed in therapy with the presence of the therapist in six of the eight studies from counseling settings.

Not-on-track Clients. Based on data from five studies (Davidsen et al., 2017; Janse et al., 2017; Murphy et al., 2012; Reese et al., 2009b; She et al., 2018) the PCOMS condition was not significantly better than the control condition to help NOT clients ($g = 0.21$, CI [−0.11, 0.53], $p = .193$).

Publication Bias. A visual inspection of the funnel plot indicated no substantial publication bias (see Appendix, Figure A1). This was confirmed by an insignificant Egger's test ($t = 0.94$, df = 16, $p =$ 181 (1-tailed)), and the Trim and Fill method indicating that there was no need for imputing missing studies to make the funnel plot symmetrical. In the cumulative meta-analysis, adding studies from the most to the least precise studies, the ES only changed marginally; for example, when including the nine most precise studies, the ES was 0.26 and, when including the 14 most precise studies, the ES was 0.24 (compared to 0.27 in the total sample).

Secondary Outcomes

No effect of the PCOMS was found on the number of deteriorated clients (OR = 0.91, CI [0.67, 1.23], $p = .537$, $k = 13$); or number of psychotherapy sessions attended ($g = 0.06$, CI [−0.09, 0.21], $p = .418$, $k =$ 14).

The protocol pre-specified outcome analyses for treatment dropout, social functioning, and direct costs of using the PCOMS were not performed. The reason was that only two studies (van Oenen et al., 2016; Winkelhorst et al., 2013) reported on treatment dropout; and none of the studies reported on social functioning or costs of treatment.

Moderation Analyses

Treatment Setting. The difference between psychiatric and counseling settings was significant and reduced the between-study variance with 76% ($Q = 13.08$, df = 1, $p < .001$; $R^2 = 0.76$). Studies from the psychiatric settings did not reveal a significant effect of the PCOMS ($g = 0.10$, CI [−0.03, 0.23], $p = .144$), whereas studies from the counseling settings had a significant effect ($g = 0.45$, CI [0.31, 0.59], $p < .001$). The heterogeneity was substantially reduced, both within the psychiatric settings ($Q = 10.94$, df = 9, $p = .280$, $I^2 = 17.73$, $T^2 = 0.01$, PI [−0.14, 0.34]) and the counseling settings ($Q = 11.63$, df = 7, $p = .114$, $I^2 = 39.79$, $T^2 = 0.02$, PI [0.08, 0.83]). Sensitivity analysis found that setting was still a significant moderator when only the 14 RCTs were included ($Q = 7.35$, df = 1, $p = .007$), with no effect of the PCOMS in the psychiatric settings ($g = 0.12$, CI [−0.07, 0.31],

$p < .224$) and a significant effect in the counseling settings ($g = 0.46$, CI [0.30, 0.61], $p < .001$).

Other Moderators. No difference in the PCOMS effect was found between the 13 studies on individual therapy and the three studies on group therapy ($Q = 0.00$, df = 1, $p = .985$, $R^2 = 0.00$). Meta-regressions found neither a PCOMS effect of age ($Q = 1.24$, df = 1, $p = .266$, $R^2 = 0.06$, $k = 17$), nor of the amount (in hours) of therapist PCOMS training ($Q = 0.58$, df = 1, $p = .447$, $R^2 = 0.00$, $k = 14$).

The pre-specified moderation analyses on researcher allegiance and type of outcome measure (the ORS alone vs. other outcome measures) were not performed because of overlap between setting, allegiance, and outcome measure. Thus, nine out of 10 studies from psychiatric settings reported outcome on a generic symptom measure and eight had a neutral researcher allegiance, whereas all eight studies from a counseling setting used the ORS as the only outcome measure and seven had a positive researcher allegiance.

The protocol pre-specified moderation analyses on treatment approach, supervision in the PCOMS, couple therapy, and treatment duration were not performed. The reason for this was insufficient variation in treatment approaches, inadequate reporting on supervision in the PCOMS, and that only two studies dealt with couple therapy or reported mean treatment duration. Finally, meta-regression with more than one covariate was not performed because of the low number of studies and study overlap.

Discussion

The overall ES of the 18 included studies comparing the PCOMS to control conditions was significant and small ($g = 0.27$, CI [0.14, 0.41], $p < .001$). The PCOMS was not specifically better at helping NOT clients or preventing client deterioration. No substantial publication bias was found. The risk of bias was high in most studies, mainly due to lack of blinding of outcome assessment, incomplete outcome data and the use of the ORS as both intervention tool and outcome measure. In the overall analysis, the heterogeneity was large ($Q = 48.13$, df = 17, $p < .001$, $I^2 = 64.68$, $T^2 = 0.05$). The PI was in the range from −0.22 to 0.76. Thus, even though the mean ES of the 18 sampled studies was significantly above zero, the PI included zero, and its range indicated that a wide dispersion in effect across populations should be expected when using the PCOMS. This may reflect the fact that the participants from the 18 studies received very different kinds of treatment of varying duration, from 1 to 2 and up to 48 weeks, and were sampled from divergent populations.

The 10 studies from psychiatric settings did not reveal a significant effect of the PCOMS ($g = 0.10$, CI [−0.03, 0.23], $p = .144$), whereas the eight studies from counseling settings had a significant effect ($g = 0.45$, CI [0.31, 0.59], $p < .001$). The difference was significant and large ($Q = 13.08$, df = 1, $p < .001$; $R^2 = 0.76$). The heterogeneity was substantially reduced within both settings, which indicates that setting contributed to heterogeneity. A sensitivity analysis showed that these results were not affected by including RCTs only. Although methodological quality varied among studies from psychiatric settings (see appendix), nine of the 10 studies used a general outcome measure, and not only the ORS, and eight of the studies were conducted by researchers without positive allegiance. In their meta-analysis of ROM, Kendrick et al. (2016) did not find that setting moderated the effect of feedback. They included seven studies using the OQ-45 in "multidisciplinary mental health care" (resembling our psychiatric setting) and five studies, including three PCOMS and two OQ-45 studies, from "psychological therapy settings" (resembling our counseling setting). All three PCOMS studies from "psychological therapy settings" (Murphy et al., 2012; Reese et al., 2009a, 2009b) were also included in the present meta-analysis, which included five additional studies from counseling settings (and 10 other studies from psychiatry settings). The differences in included studies may explain the different results.

Factors related to patient and treatment characteristics may explain the lack of PCOMS effects in psychiatric settings. Psychiatric patients are likely to be more severely disordered than counseling clients, and severely disordered patients may prefer direct advice and guidance as they may have little motivation or lack capacity to reflect on feedback. In psychiatric settings, also, more patients with therapy-resistant symptoms and lack of early progress can be expected. Thus, in Davidsen et al.'s (2017) eating disorder sample, 79.8% were classified as NOT at least once during treatment. In ROM it may be demoralizing for psychiatric patients as well as for therapists to be confronted with the patient's low level of functioning and lack of progress. Research on another feedback system, the OQ-45, points in the same direction. In Chile, Errázuriz and Zilcha-Mano (2018) included 547 patients and found that feedback concerning negative progress (compared to positive progress) had a negative impact on outcome, specifically for "severe patients" who had high baseline symptomatology and previous psychiatric hospitalizations. In a systematic review with 11 studies (nine on the OQ-

45 and two on the PCOMS) a diminishing effect of feedback with more patient severity was found (Davidson, Perry, & Bell, 2015).

Organizational factors such as lack of time and limited treatment flexibility in psychiatric settings have also been proposed as potential barriers for feedback to have an effect (Davidsen et al., 2017; de Jong, 2016). Davidsen et al. (2017) suggested that their standardized treatment program prevented the therapist from using the feedback to alter the treatment. Moreover, in psychiatric settings, patients often receive supplementary treatment by caregivers not using the PCOMS, which can make it more difficult to find an incremental effect of the PCOMS.

The PCOMS was found to have a moderate effect in counseling settings with less severely disordered clients, such as student and family counseling centers. It seems plausible that the PCOMS would be more suitable for such clients who can be assumed to fit better into a "fast change model" with discussions of therapeutic alliance and progress. However, the results of studies on the PCOMS in counseling settings might be influenced by bias due to researcher allegiance and use of the ORS as the only outcome measure. All eight studies from counseling settings used the ORS as the only outcome measure, and seven of the studies were performed in cooperation with the "Heart and Soul of Change Project" (2017) with B. L. Duncan and/or R. J. Reese as coauthors. It was not possible to test the influence of these potential bias due to an overlap of setting, use of the ORS, and researcher allegiance.

No feedback effect was found for NOT clients. However, only five studies reported analyses based on NOT clients. The PCOMS assumes feedback to be helpful for all clients, and the PCOMS has not developed specific Clinical Support Tools for NOT clients, as the OQ-Analyst in Lambert's (2015) OQ-system. Shimokawa et al.'s (2010) meta-analysis found that the feedback effect was concentrated on NOT clients and that this effect was larger when the Clinical Support Tools were used. Although these results have to be replicated and tested in RCTs directly comparing the different feedback systems, they might indicate that the PCOMS does not offer enough help specifically to NOT clients. Moreover, in the PCOMS, between 25.7% and 81.3% of the clients were identified as NOT compared to about 20–40% on the OQ-45 (Lambert, 2015). Thus, the PCOMS might create too many "signal cases," which might blur the therapist's attention, or as discussed above, give rise to demoralization. Seidel et al. (2017) re-calculated the RCI taking the test–retest reliability into account and found an ORS raw score RCI of 10.7, compared to an RCI of 5 (Miller & Duncan, 2004) used in most of the included studies.

As mentioned in the introduction, several researchers have questioned the validity of the ORS as an outcome measure and found that the ORS yields larger ESs than other outcome measures. Seidel et al. (2017) found in a small clinical sample of 73 clients that the ES from session one to three was 0.83 with the ORS and 0.44 with the OQ-45. Moreover, a distressed subsample of 28 clients from a community sample achieved an ES of 0.59 on the ORS at time 3 without receiving any treatment. In the present meta-analysis, a direct *post-hoc* comparison between the seven studies that reported outcome on both the ORS and a generic symptom measure showed a difference in ES of 0.11. Moreover, in five of these seven studies (all from psychiatric settings), the ORS was completed outside therapy sessions, whereas the ORS was completed in therapy with the presence of the therapist in six of the eight studies from counseling settings (all used the ORS as the only outcome measure). Thus, in the majority of studies from counseling settings, the ES might be elevated because of social desirability effects. Only one study has investigated the effect of social desirability in the PCOMS. In a study with 102 participants, Reese et al. (2013) found that the client's rating on the SRS was not elevated by the presence of the therapist (compared to a condition where the therapist did not have access to the rating).

Therapist fidelity has been claimed to play a major role in the outcome of using the PCOMS (see the introduction). However, none of the studies in this meta-analysis had a direct assessment of therapist adherence or competency. The amount of therapist training was not a significant moderator of the PCOMS effect. The exact frequency of using the PCOMS in the sessions was only reported in four studies, seven studies only reported range or number of clients not completing the ORS/SRS in all sessions, while seven studies did not report on the in-session use of the PCOMS. The amount of supervision was also infrequently reported ($k = 9$), and, when reported, inaccurately as weekly, monthly, or regularly. Two studies investigated the effect of the PCOMS as a function of time of its implementation. Davidsen et al. (2017) found no difference in the effect of the PCOMS between a first and a second implementation phase, whereas Brattland et al. (2018) found that the superiority of the PCOMS over control increased significantly with time. Brattland et al. (2018) argued that one explanation could be that the PCOMS might have been used increasingly more effectively during the implementation, but this was not investigated.

Limitations

A major limitation of the meta-analysis is the small number of studies included, which is a special problem for its moderator and NOT analyses. To increase the number of included studies, we also included N-RCTs. Sensitivity analyses showed that this did not significantly change the major results. Several of the planned moderation analyses in the protocol could not be performed due to too few studies and/or study overlap. All analyses were based on completer data.

Conclusion

Based on this meta-analysis of 18 studies including 2910 participants, the overall effect of using the PCOMS is small ($g = 0.27$, CI = [0.14, 0.41]). The heterogeneity of studies was substantial. Even though the effect was significant, the prediction interval was in the range from −0.22 to 0.76 indicating that a wide dispersion in effect across populations should be expected when using the PCOMS. The 10 studies from psychiatric settings did not reveal an effect of the PCOMS, whereas the eight studies from counseling settings had a moderate effect. However, 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. There is a need for more studies, especially studies in counseling settings, using other outcome measures than the ORS. Future studies should also measure therapist adherence and report information about how the PCOMS is used to understand better the conditions under which the PCOMS might work and for whom. Further studies might also investigate the potential mechanisms of change in the PCOMS. It may be essential to adapt feedback systems to the requirements of clinical settings and specific client characteristics and needs.

Acknowledgements

The authors would like to express their thanks to the researchers who provided valuable additional data on their studies. We also wish to thank Annie Dolmer Kristensen who assisted in the proof-reading of the manuscript.

ORCID

Ole Karkov Østergård http://orcid.org/0000-0001-8909-1241
Hilde Randa http://orcid.org/0000-0003-1119-1095
Esben Hougaard http://orcid.org/0000-0002-1254-7674

References

In the review, the specifically asked for having the papers included in the meta-analysis marked by *.

Amble, I., Gude, T., Stubdal, S., Andersen, B. J., & Wampold, B. E. (2015). The effect of implementing the outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. *Psychotherapy Research*, *25*(6), 669–677. doi:10.1080/10503307.2014.928756

American Psychological Association. (2006). Evidence-based practice in psychology: APA presidential task force on evidence-based practice. *The American Psychologist*, *61*(4), 271–285. doi:10.1037/0003-066X.61.4.271

*Anker, M., Duncan, B. L., & Sparks, J. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, *77*(4), 693–704. doi:10.1037/a0016062

Barkham, M., & Margison, F. (2007). Practice-based evidence as a complement to evidence-based practice: From dichotomy to chiasmus. In C. Freeman & M. Power (Ed.), *Handbook of Evidence-Based Psychotherapies: A Guide for Research and Practice* (pp. 443–476). New York: John Wiley. doi:10.1002/9780470713242.ch25

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., … McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*(2), 184–196. doi:10.1037/0022-006X.69.2.184

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York: John Wiley. doi:10.1002/9780470743386

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. doi:10.1002/jrsm.12

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. doi:10.1002/jrsm.1230

*Brattland, H., Koksvik, J. M., Burkeland, O., Gråve, R. W., Klöckner, C., Linaker, O. M. … Iversen, V. C. (2018). The effects of Routine Outcomes Monitoring (ROM) on therapy outcome in the course of an implementation process. A randomized clinical trial. *Journal of Counseling Psychology*. Advance online publication. doi:10.1037/cou0000286.

Bringhurst, D. L., Watson, C. W., Miller, S. D., & Duncan, B. L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, *5*, 23–30. Retrieved from http://journalbrieftherapy.com/

Burgess, P. M., Pirkis, J. E., Slade, T. N., Johnston, A. K., Meadows, G. N., & Gunn, J. M. (2009). Service use for mental health problems: Findings from the 2007 National Survey of Mental Health and Wellbeing. *The Australian and New Zealand Journal of Psychiatry*, *43*, 615–623. doi:10.1080/00048670902970858

Campbell, A., & Hemsley, S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist*, *13*(1), 1–9. doi:10.1080/13284200802676391

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavy, A. (2013). Practice-oriented research: Approaches and applications. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 85–133). New York: John Wiley.

Child Outcomes Research Consortium. (2018, February 7). Retrieved from http://www.corc.uk.net/

*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..

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, *14*, 159–183. http://doi.org/10.1146/annurev-clinpsy-050817-084833

Comprehensive Meta-Analysis (Version 3) [Computer software]. (2014). Englewood, CO: Biostat. Retrieved from http://www.comprehensive.com

Covidence Systematic Review Software. (2017). Veritas Health Innovation. Retrieved from www.covidence.org.

*Davidsen, A. H., Poulsen, S., Lindschou, J., Winkel, P., Tróndarson, M. F., Waaddegaard, M., & Lau, M. (2017). Feedback in group psychotherapy for eating disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *85*(5), 484–494. doi:10.1037/ccp0000173

Davidson, K., Perry, A., & Bell, L. (2015). Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychology and Psychotherapy*, *88*(1), 21–37. http://doi.org/10.1111/papt.12032

de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S. … Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology & Psychotherapy*, *18*(1), 1–12. doi:10.1002/cpp.696

de Jong, K. (2016). Deriving implementation strategies for outcome monitoring feedback from theory, research and practice. *Administration and Policy in Mental Health*, *43*(3), 292. doi:10.1007/s10488-014-0589-6

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*(4), 464–474. doi:10.1080/10503307.2012.673023

Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring & procedures manual – II for the r(evised) version and other instruments of the psychopathology rating scale series* (2nd ed.). Towson, MD: Clinical Psychometric Research.

Drapeau, M. (2012). The value of progress tracking in psychotherapy. *Integrating Science and Practice*, *2*(2). Retrieved from file:///U:/dokumenter%2015_11_2017/Artikler/Artikel%20metaanalyse/Drapeau%2010%20outcome%20monotoring%20tools_2012.pdf

Duncan, B. L. (2012). The Partners for Change Outcome Management System (PCOMS): The heart and soul of change project. *Canadian Psychology/Psychologie Canadienne*, *53*(2), 93–104. doi:10.1037/a0027762

Duncan, B. L., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS): Revisiting the client's frame of reference. *Psychotherapy*, *52*(4), 391–401. doi:10.1037/pst0000026

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Errázuriz, P., & Zilcha-Mano, S. (2018). In psychotherapy with severe patients discouraging news may be worse than no news: The impact of providing feedback to therapists on psychotherapy outcome, session attendance, and the alliance. *Journal of Consulting and Clinical Psychology*, *86*(2), 125–139. doi:10.1037/ccp0000277

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, *61*(2), 155–163. doi:10.1002/jclp.20108

*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*(1–2), 617–619. doi:10.1016/j.psychres.2015.05.007

Heart and Soul of Change Project. (2017, August 22). Retrieved from https://heartandsoulofchange.com/content/community

Heart and Soul of Change Project. (2018, February 7). Retrieved from https://heartandsoulofchange.com/content/community

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. doi:10.1037/1082-989X.3.4.486

Higgins, J. P. T., Altman, D. G., & Sterne, J. A. C. (2011). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Ed.), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011, pp. 187–241). The Cochrane Collaboration, 2011. Retrieved from www.handbook.cochrane.org

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. doi:10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. doi:10.1136/bmj.327.7414.557

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, *51*(10), 1059–1064. doi:10.1037/0003-066X.51.10.1059

International Center for Clinical Excellence. (2017, August 22). Retrieved from https://www.centerforclinicalexcellence.com

International Center for Clinical Excellence. (2018, February 7). Retrieved from https://www.centerforclinicalexcellence.com

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. doi:10.1037/0022-006X.59.1.12

Janse, P. D., Boezen-Hilberdink, L., van Dijk, M. K., Verbraak, M. J. P. M., & Hutschemaekers, G. J. M. (2014). Measuring feedback from clients: The psychometric properties of the Dutch Outcome Rating Scale and Session Rating Scale. *European Journal of Psychological Assessment*, *30*(2), 86–92. doi:10.1027/1015-5759/a000172

*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*(5), 525–538. doi:10.1080/10503307.2016.1152408

*Kellybrew-Miller, A. (2015). Retrieved from https://search.proquest.com/docview/1766579664?accountid=14468

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., … Moore, M. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*, *7*, CD011119. doi:10.1002/14651858.CD011119.pub2

Kordy, H., Hannöver, W., & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology*, *69*(2), 173–183. doi:10.1037/0022-006X.69.2.173

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's*

handbook of psychotherapy and behavior change (6th ed., pp. 169–218). New York: John Wiley.

Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy*, *52*(4), 381–390. doi:10.1037/pst0000027

Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., Reid, R. C., … Burlingame, G. M. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)*. Salt Lake City, UT: OQ Measures, LLC.

Lambert, M. J., & Shimokawa, K. (2011a). Collecting client feedback. *Psychotherapy*, *48*(1), 72–79. doi:10.1037/a0022238

Lambert, M. J., & Shimokawa, K. (2011b). Collecting client feedback. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed. pp. 203–223). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199737208.003.0010

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology & Psychotherapy*, *9*(2), 91–103. doi:10.1002/cpp.324

*Lester, M. C. (2012). Retrieved from https://search.proquest.com/docview/1039541820?accountid=14468

Lutz, W., Böhnke, J. R., & Köck, K. (2011). Lending an ear to feedback systems: Evaluation of recovery and non-response in psychotherapy in a German outpatient setting. *Community Mental Health Journal*, *47*(3), 311–317. doi:10.1007/s10597-010-9307-3

Lutz, W., de Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research*, *25*(6), 625–632. doi:10.1080/10503307.2015.1079661

Miller, S. D., & Duncan, B. L. (2004). *The outcome and session rating scales: Administration and scoring manual*. Chicago, IL: Institute for the Study of Therapeutic Change.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, *2*, 91–100. Retrieved from http://journalbriefftherapy.com/

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. doi:10.1016/j.jclinepi.2009.06.005

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review*, *33*(4), 501–511. doi:10.1016/j.cpr.2013.02.002

*Murphy, K. P., Rashleigh, C. M., & Timulak, L. (2012). The relationship between progress feedback and therapeutic outcome in student counselling: A randomised control trial. *Counselling Psychology Quarterly*, *25*(1), 1–18. doi:10.1080/09515070.2012.662349

Norcross, J. C. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed.). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199737208.001.0001

Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy*, *48*(1), 98–102. doi:10.1037/a0022161

Østergård, O. K., Randa, H., & Hougaard, E. (2017). The effect of using the Partner for Change Outcome Management System (PCOMS) as a feedback tool in psychotherapy: protocol for a systematic review and meta-analysis. Retrieved from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017069867

Reese, R. J., Gillaspy, J. A., Owen, J. J., Flora, K. L., Cunningham, L. C., Archie, D., & Marsden, T. (2013). The influence of demand characteristics and social desirability on clients' ratings of the therapeutic alliance. *Journal of Clinical Psychology*, *69*(7), 696–709. doi:10.1002/jclp.21946

*Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009a). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*, *46*(4), 418–431. doi:10.1037/a0017901

*Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009b). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*, *46*(4), 418–431. doi: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*(4), 616–630. doi: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*(1), 164–168. doi:10.1016/j.pec.2015.07.016

SAMSHA. (2017, August 22). Substance Abuse and Mental Health Administration's (SAMHSA) National Registry of Evidence-based Programs and Practices. Retrieved from https://www.samhsa.gov/nrepp

*Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. L. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research*, *25*(4), 396–407. doi:10.1080/10503307.2014.900875

Seidel, J. A., Andrews, W. P., Owen, J., Miller, S. D., & Buccino, D. L. (2017). Preliminary validation of the Rating of Outcome Scale and equivalence of ultra-brief measures of well-being. *Psychological Assessment*, *29*(1), 65–75. doi:10.1037/pas0000311

*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*. Advance online publication. doi:10.1037/cou0000300

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, *78*(3), 298–311. doi:10.1037/a0019247

*Slone, N. C., Reese, R. J., Mathews-Duvall, S., & Kodet, J. (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics*, *19*(2), 122–136. doi:10.1037/gdn0000026

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(4), 547–559. doi:10.1037/a0028226

Tam, H. E., & Ronan, K. (2017). The application of a feedback-informed approach in psychological service with youth: Systematic review and meta-analysis. *Clinical Psychology Review*, *55*, 41–55. doi:10.1016/j.cpr.2017.04.005

Tasma, M., Swart, M., Wolters, G., Liemburg, E., Bruggeman, R., Knegtering, H., & Castelein, S. (2016). Do routine outcome monitoring results translate to clinical practice? A cross-sectional study in patients with a psychotic disorder. *BMC Psychiatry*, *16*, 107. doi:10.1186/s12888-016-0817-6

*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. doi:10.1186/s12888-016-0811-z

*Winkelhorst, Y., Hafkenscheid, A., & de Groot, E. (2013). Verhoogt Routine Process Monitoring (RPM) de effectiviteit van behandeling. *Tijdschrift voor Psychotherapie*, *39*(3), 146–156. https://doi.org/10.1007/s12485-013-0028-2

Appendix

Risk of Bias. Assessment of bias according to the Cochrane criteria is summarized in Table AI. In general, the RCTs had an adequate or unclear sequence generation and allocation concealment. None of the studies was able to blind the participants as to treatment since this is not possible with psychological interventions. As a substitute for client blindness of allocation in psychotherapy research, comparability of treatment credibility is sometimes checked, but this was not performed in any of the studies. Davidsen et al. (2017) used blinded assessment of their primary outcome (eating disorder symptoms), but this measure was not used in the present meta-analysis. Outcomes used in the analyses were all client-reported. The risk of incomplete outcome data (attrition bias) was high in six studies and unclear in seven studies, mainly due to a different number of missing data in the PCOMS and the control group and/or due to lack of information about how many participants were randomized to the PCOMS and the control group, respectively. In general, the risk of selective reporting was judged to be low because most studies reported on all outcome measures used.[1] The risk of other bias was rated as high in eight studies using the ORS as the only outcome measure and in three other studies. In summary, there was a high risk of bias in most studies, primarily due to lack of blinding of outcome assessment, incomplete outcome data, and other bias.

Note

[1] In contrast to the meta-analysis by Kendrick et al. (2016) we did not rate the risk of reporting bias as high if the SRS outcome was not reported because we consider the SRS to be a process, rather than an outcome measure.

Table AI. Risk of bias summary.

	Anker et al., 2009	Brattland, et al., 2018	Davidsen et al., 2017	Kellybrew-Miller, 2015	Lester, 2012	Murphy et al., 2012	Reese et al., 2009a	Reese et al., 2009b	Reese et al., 2010	Rise et al., 2016	Schuman et al., 2015	She et al., 2018	Slone et al., 2015	van Oenen et al., 2016	Chow & Huixian, 2015	Hansen et al., 2015	Janse et al., 2017	Winkelhorst et al., 2013
Random sequence generation	+	+	+	?	?	+	+	+	?	+	?	?	?	+	–	–	–	–
Allocation concealment	+	+	+	?	?	?	+	?	?	+	?	?	–	+	–	–	–	–
Blinding of participants	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Blinding of outcome assessment	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Incomplete outcome data	+	+	+	–	?	?	–	–	?	?	?	+	+	?	–	–	?	–
Selective reporting	?	+	+	+	+	+	+	+	+	+	+	+	–	+	+	?	+	+
Other bias	–	+	+	–	?	–	–	–	–	+	?	–	–	+	–	–	+	–

Note. +, Risk of bias low; ?, Risk of bias uncertain; –, Risk of bias high.

Funnel Plot

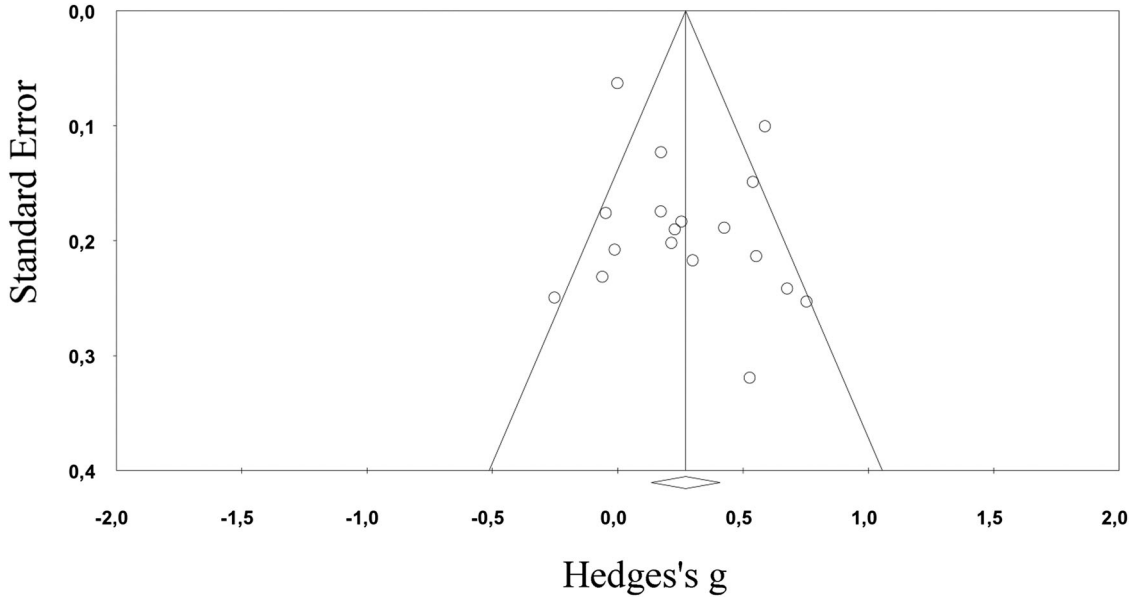

Figure A1. Funnel plot of Standard Error by Hedge's *g*.