Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: a multilevel meta-analysis

Jong, K. de; Conijn, J.M.; Gallagher, R.A.; Reshetnikova, A.S.; Heij, M.; Lutz, M.C.

Citation
Jong, K. de, Conijn, J. M., Gallagher, R. A., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: a multilevel meta-analysis. *Clinical Psychology Review*, *85*, 1-19. doi:10.1016/j.cpr.2021.102002

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/3250565

Note: To cite this publication please use the final published version (if applicable).

Clinical Psychology Review 85 (2021) 102002

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis

Kim de Jong [a,*], Judith M. Conijn [b], Roisin A.V. Gallagher [a], Alexandra S. Reshetnikova [a], Marya Heij [a], Miranda C. Lutz [a,c]

[a] Institute of Psychology, Leiden University, Leiden, the Netherlands
[b] Faculty of Behavioral and Movement Sciences, Vrije Universiteit, Amsterdam, the Netherlands
[c] Department of Psychology, Education & Child Studies, Erasmus University, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Progress feedback
Routine outcome monitoring
Psychotherapy
Outcome
Drop-out

ABSTRACT

Progress feedback is an intervention aimed at enhancing patient outcomes in routine clinical practice. This study reports a comprehensive multilevel meta-analysis on the effectiveness of progress feedback in psychological treatments in curative care. The short- and long-term effects of feedback on symptom reduction were investigated using 58 (randomized and non-randomized) studies, analyzing 110 effect sizes in a total of 21,699 patients. Effects of feedback on dropout rate, percentage of deteriorated cases, and treatment duration were also examined. Moderation analyses were conducted for study and feedback characteristics. A small significant effect of progress feedback on symptom reduction (d = 0.15, 95% CI: [0.10, 0.20]) was found, compared to control groups. This was also true for not-on-track cases (d = 0.17, 95% CI: [0.11, 0.22]). In addition, feedback had a small favorable effect on dropout rates (OR = 1.19, 95% CI: [1.03, 1.38]). The moderation analyses identified several potentially interesting variables for further research, including feedback instrument, outcome instrument, type of feedback, feedback frequency, treatment intensity, and country in which the study was conducted. Future studies should report on these variables more consistently so that we can obtain a better understanding of when and why feedback improves outcomes.

1. Introduction

Since its introduction in 2001 (Lambert et al., 2001), the use of routine outcome measurements as a method of providing feedback to clinicians about their patients' progress has increased substantially. Progress feedback typically uses standardized outcome instruments, which are administered regularly throughout therapy (Lambert, 2007). A wide variety of feedback systems is being used, although two systems have been studied more often than others, namely the Outcome Questionnaire System (OQ System; Lambert et al., 2004) and the Partners for Change Outcome Measurement System (PCOMS; Miller, Duncan, Sorrell, & Brown, 2005). By monitoring the patient's progress, the clinician can potentially adjust the treatment when a lack of progress occurs, so that poor outcomes might be prevented. Some studies only provide raw scores to clinicians (e.g., Puschner, Schöfer, Knaup, & Becker, 2009), whereas others benchmark the patients' scores against expected recovery trajectories or have added clinical support tools that assess the therapy processes after a patient is identified as not on track (e.g. Harmon, Hawkins, Lambert, Slade, & Whipple, 2005). Several prominent authors have stated that they consider progress monitoring to be one of the most promising interventions for improving outcomes in clinical practice (e.g., Bickman, 2008; Kazdin, 2008; Langkaas, Wampold, & Hoffart, 2018; Wampold, 2015). One of the reasons that progress feedback is important to use routinely is that without this information clinicians seem to be poor at identifying which patients are not progressing well (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010), and overestimate their own performance (Brosan, Reynolds, & Moore, 2008; Walfish et al., 2012).

2. Previous meta-analyses

Over the years, eight meta-analyses have been conducted on the effectiveness of providing clinicians (and clients) with progress feedback (Bergman et al., 2018; Kendrick et al., 2016; Knaup, Koesters, Schoefer, Becker, & Puschner, 2009; Lambert et al., 2003; Lambert, Whipple, & Kleinstäuber, 2018; Østergård, Randa, & Hougaard, 2018; Shimokawa,

* Corresponding author at: Leiden University, Institute of Psychology, Clinical Psychology Unit, PO Box 9555, 2300 RB Leiden, the Netherlands.
E-mail address: KJong@fsw.leidenuniv.nl (K. de Jong).

https://doi.org/10.1016/j.cpr.2021.102002
Received 20 December 2019; Received in revised form 4 February 2021; Accepted 23 February 2021
Available online 27 February 2021
0272-7358/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Lambert, & Smart, 2010; Tam & Ronan, 2017) and reported effects of feedback ranging between no effect and medium effects. Lambert et al. (2003), and Shimokawa et al. (2010) have found small to medium effects of feedback for cases at risk for negative outcomes – conceptualized as "not on track" (NOT) cases. Knaup et al. (2009) found a very small overall effect (Hedges' $g = 0.10$) of feedback in short-term therapies but not in long-term therapies. Kendrick et al. (2016) found no overall effect of feedback, but did report small effects in NOT cases (Hedges' $g = 0.22$). Kendrick et al. (2016) also reported an effect of feedback on treatment duration in on track (OT) cases. Lambert et al. (2018) reported very small overall effects of progress feedback on outcome ($d = 0.14$) for studies using the OQ System, and somewhat larger (but still small) effects for NOT cases ($d = 0.33$), especially when clinical support tools were used in addition to the feedback ($d = 0.49$). For studies using PCOMS only a small overall effect size was reported ($d = 0.40$). Østergård et al. (2018) analyzed only studies using PCOMS and found a small overall effect of feedback on outcome (Hedges'$g = 0.22$) and no effect of PCOMS in NOT cases. Bergman et al. (2018) and Tam and Ronan (2017) focused on youth mental health and had too small sample sizes to draw conclusions.

Variation in the study selection criteria may explain the different outcomes of the meta-analyses. Apart from Knaup et al. (2009) and the meta-analyses focusing on youth (Bergman et al., 2018; Tam & Ronan, 2017), the meta-analyses have exclusively focused on the OQ System and PCOMS. The Lambert et al. (2003) and Shimokawa et al. (2010) meta-analyses have included only studies conducted by their own research group, which limits its generalization to other contexts. Knaup et al. (2009) included several studies conducted with patients with severe mental illness, a group in which symptom change is unlikely due to the chronic character of their disorders. Furthermore, a majority of these studies only provided feedback once or twice, or only used an assessment to inform the diagnosis for treatment. Finally, Kendrick et al. (2016) focused on common mental disorders in primary and secondary care, and excluded a number of studies on patients with other disorders.

An additional limitation of the meta-analyses conducted to date is that most of them only included randomized controlled trials (RCTs) with a no feedback control group. Some studies with large effects of feedback have stopped making comparisons with no feedback control groups, but rather have compared different methods of feedback (e.g. Harmon et al., 2007; Slade, Lambert, Harmon, Smart, & Bailey, 2008). Although Lambert et al. (2018) take these two studies into account, they do not correct for dependence between the corresponding effect size estimates that is caused by both studies using the same archival control group. Moreover, studies that have used a cohort design have been excluded in previous meta-analyses. As a result, a substantial part of the available evidence has not been taken into account (Kazdin, 2008). Furthermore, the meta-analyses so far have only assessed a handful of potential moderators, despite the substantial differences between studies that could explain the differential effects of progress feedback. Finally, most feedback studies report on multiple outcome variables. Previous meta-analyses have only assessed the effect of feedback on one primary outcome variable.

3. Mechanisms and moderators

There is no consensus on which factors influence the effect of progress feedback on treatment outcomes, although several authors have offered potential explanations (e.g., Davidson, Perry, & Bell, 2014; Krägeloh, Czuba, Billington, Kersten, & Siegert, 2015; W. Lutz, De Jong, & Rubel, 2015). Feedback intervention theory assumes that when a clinician is provided with feedback about a patient' progress, a comparison is made between the feedback and a goal or standard (Kluger & DeNisi, 1996). When a discrepancy is noted between the goal and feedback, clinicians will be motivated to adapt the therapy (Riemer & Bickman, 2011). This explains why previous meta-analyses have found larger effects for NOT cases. After all, when the feedback message is that a patient is progressing well, feedback is not likely to improve outcomes.

In order for feedback to be effective, the discrepancy between the feedback and goal also needs to be observed by the clinician. As a result, it is likely that feedback systems that use explicit standards, such as expected recovery curves, are more effective than feedback systems that present raw scores. Feedback interventions in general also seem to be more effective when they are timely and provide more specific information (Kluger & DeNisi, 1996). Thus, immediate feedback should outperform (one week) delayed feedback, and systems using clinical support tools are expected to outperform systems that do not use these.

It is also plausible that differences in feedback effectiveness exist between countries, as culture, care systems, allocation of health care expenditure between the public and private sectors (Greene, 2004), as well as therapist training differ substantially across the world. In addition, it is possible that the effectiveness of feedback is moderated by treatment intensity. Davidson et al. (2014) conclude that feedback seems more effective in treatment settings with patients with a relatively mild severity. Similarly, Østergård et al. (2018) conclude in their meta-analysis that PCOMS is effective in counseling settings, but not in psychiatric settings.

An additional factor that is being discussed in the feedback literature is the degree of implementation. De Jong, van Sluis, Nugter, Heiser, and Spinhoven (2012) reported that only half of the therapists actively used the feedback reports in their clinical practice, and feedback was only effective for the group of active users. Bickman et al. (2016) conducted an RCT at two different sites, and implementation failed at one of the sites, which resulted in no effect of the feedback on that site. One important factor in implementation might be whether therapists enter into the study voluntarily or implementation is conducted department wide.

4. Study aim

In this article, we aim to report the most comprehensive meta-analysis on the effectiveness of progress feedback to date. The primary outcome variable is the effect of progress feedback on symptom reduction. Additionally, long-term effects of feedback on symptom reduction will be investigated, as this topic has been neglected so far in the literature (Davidson et al., 2014). In addition to its effect on symptom reduction, feedback has also been found to affect the percentage of deteriorated cases (e.g. Shimokawa et al., 2010) and treatment duration (e.g. Delgadillo et al., 2017; Kendrick et al., 2016) in some studies. Therefore, the secondary outcome variables are the effect of feedback on dropout rate, the percentage of deteriorated cases, and treatment duration. Moderation analyses will be conducted in order to understand which factors may explain why progress feedback is highly effective in some studies, but has negligible effects in others. In particular, this study aims to explore the moderating effects of study characteristics (e.g. treatment setting, study design, treatment intensity) and feedback characteristics (e.g. type and frequency of the feedback).

5. Method

5.1. Search strategy

A systematic literature search was undertaken on September, 30, 2020 in PsychINFO, PubMed, and Web of Science (see Online Supplement for search strings). Additional searches were performed in the Current Controlled Register Trials Register and Google Scholar in order to supplement studies that might have been missed. Furthermore, the eight meta-analyses referred to in the introduction and several review articles (Carlier et al., 2012; Davidson et al., 2014; Krägeloh et al., 2015) were screened for relevant publications. *E*-mails asking for unpublished studies were sent out to the listserv of the Society for Psychotherapy Research and to researchers known to publish in this area. The search was conducted by the first and last author. The first author ran the

search and screened the titles, and the first and last actors independently screened the abstracts and full texts and discussed inclusion.

5.2. Inclusion criteria

Progress feedback was defined as "providing information on treatment progress from standardized measures to a clinician and/or patient on a regular basis throughout the course of treatment". Inclusion criteria consisted of studies in the English, German, or Dutch language that met all of the following conditions: (a) the study examined effects of feedback interventions on outcome, comparing one or more feedback groups and a no feedback control group or cohort; (b) the study focuses on psychological interventions in a psychotherapy, psychiatry, or counseling setting; (c) patients with mental health or substance abuse problems were treated; and (d) outcomes were assessed and fed back on at least three moments during treatment.[1] Publications were excluded if the study population was limited to patients meeting criteria for severe mental illnesses,[2] if the publication did not report on outcomes (e.g., study protocols) or only reported outcomes on follow-up data, and if data for computing effect sizes was irretrievable after (failed or successful) author contact. Since progress feedback was introduced in 2001, the search was restricted to articles dating from that year or later.

5.3. Data extraction, coding, and supplementary data collection

Data extraction and coding of the study characteristics were conducted by three graduate level psychologists, under supervision of the first author. All data that was extracted, was checked by the first and last author, and disagreements were discussed in the larger research group. Data-extraction on unpublished results of studies in which the first author was involved were always conducted and checked by another author. Studies were coded for feedback characteristics (feedback instrument, type of feedback, timing, frequency, feedback recipient, training, check on feedback use), and study characteristics (treatment duration, treatment form, treatment setting, treatment intensity, age group, country, participation by therapists, independent outcome assessment). Authors were contacted if information was omitted, or if data was presented in a way that prohibited data extraction directly from the publication. In total, 40 authors were contacted. Five authors could not be reached; four authors responded that they were not able to provide us with additional information, and 31 provided additional information about their studies.

5.4. Outcome variables

5.4.1. Primary outcome variable

The primary outcome variable was the difference in post-therapy symptom reduction on a standardized outcome measure between patients who received treatment as usual (control group) and patients who received treatment that was supplemented with progress feedback (feedback group). Since a portion of the studies were cohort studies, and thus no randomization had taken place, equal pre-treatment scores across the groups could not be assumed. We therefore used the standardized mean difference in change score (d_{ppc}) as the effect size measure, as this measure takes the pre-treatment scores into account. The *pre-post-control* design effect size d_{ppc} was computed as the mean pre-post change in the feedback group minus the mean pre-post change in the control group, divided by the pooled pre-test standard deviation (Morris, 2008).[3] Computation of the corresponding sampling variance (Morris, 2008), requires an estimate of the correlation between the pre-score and post-score. Since these estimates were only known for the studies with available original data was available, the average of these values was used as an estimate of the pre-post correlation for each of the studies.[4]

For the five college counseling center studies by Lambert and colleagues (Harmon et al., 2007; Lambert, Whipple, et al., 2001; Lambert et al., 2002; Slade et al., 2008; Whipple et al., 2003), one effect size was computed over all the samples using the original data, since the studies used (parts of) the same control cohort as a comparison group. For four studies (Davidsen et al., 2017; Puschner et al., 2009; Schmidt et al., 2006; Trudeau, 2000) the available pre-score and post-score data (*M* and *SD*) were not based on the same sample, rendering them inappropriate as an effect size measure. However, since these studies were RCTs, we could use the standardized mean difference in post-treatment scores, the *independent groups* effect size (d_{IG}) as an alternative effect size measure (Morris & DeShon, 2002). If standardized mean differences in post-treatment scores are computed for studies that are RCTs, the two effect size measures, d_{ppc} and d_{IG}, can be aggregated in a single meta-analysis (Morris & DeShon, 2002).

Follow-up data on symptom reduction was analyzed if data was collected at least one month after the end of treatment or the end of the study. To quantify effect sizes at follow up, the standardized mean difference in change score (d_{ppc}) between pre-treatment and follow-up measurements was used.

5.4.2. Secondary outcome variables

Secondary outcomes included post-treatment differences between the control group and feedback group in the dropout rate, the rate of deteriorated cases, and treatment duration. Dropout rate was defined as the difference between cases that entered the study and analyzed cases, divided by the number of cases that entered the study. Given the naturalistic nature of the studies, most studies did not differentiate between different types of attrition. Our definition of drop-out therefore includes treatment dropout as well as study dropout, and cases that did not have three or more sessions of treatment. The rate of deteriorated cases was based on the Clinical Significance and Reliable Change criteria by Jacobson and Truax (1991), using a worsening in symptoms of the *reliable change index* or more to classify a patient as deteriorated (e.g. $\geq$ 14 point increase in score on the Outcome Questionnaire 45-item version [OQ-45], part of the OQ System, classifies a patient as deteriorated). Treatment duration was defined as the number of sessions. Inpatient and day patient samples were excluded for this analysis. To quantify effect sizes corresponding to differences in dropout rate and the rate of deteriorated cases, we used the log-odds ratio (log_{OR}), which were afterwards converted back or odds ratios (*OR*) to enhance interpretation. To quantify the effect sizes corresponding to treatment duration, we used the standardized mean difference in the number of sessions between the feedback and control condition (d_{IG}).

5.5. Study quality

Risk of bias in articles was rated based on guidelines provided by the Cochrane Handbook for Systematic Reviews of Interventions, version 6.1 (Higgins, Savović, Page, Elbers, & Sterne, 2020). RCTs were rated

[1] This criterion was chosen so that feedback on progress was provided at least once (start of treatment to feedback moment), and the effect of the feedback could be assessed (e.g., end of treatment).

[2] Given the negligible effects of psychological treatments in these patients, it seems unlikely to expect an effect of feedback. In addition, the majority of outcome monitoring in severe mental disorders is focused on measuring quality of life, rather than symptom reduction.

[3] In Bickman et al. (2016) the pre-treatment standard deviation of the feedback group was missing at baseline. Given that this was a randomized trial, we have assumed a similar standard deviation as in the control group.

[4] The estimated correlations for the control group were 0.70,0.76,0.70,0.71, 0.81, 0.58, and 0.68 ($M = 0.71$, $SD = 0.07$) for the control group and 0.72, 0.70,0.67,0.64,0.66,0.62,0.64,0.78,0.73,0.83,0.53,0.58, 0.68 ($M = 0.68$, $SD = 0.08$) for the feedback group. These were averaged into an overall estimate of 0.69

using the RoB 2 tool (Sterne et al., 2019) in combination with the RoB crib sheet. RoB 2 assesses five domains of potential bias: bias arising from the randomization process (D1), bias due to deviations from intended interventions (D2), bias due to missing outcome data (D3), bias in measurement of the outcome (D4), and bias in selection of the reported result (D5). Cohort studies were rated using the Risk Of Bias in Non-randomized Studies of Interventions (ROBINS-I) template (Sterne et al., 2016). ROBINS-I assesses seven domains of potential bias: bias due to confounding (D1), bias in selection of participants (D2), bias in classification of interventions (D3), bias due to deviations from intended interventions (D4), bias due to missing data (D5), bias in measurement of outcomes (D6), bias in selection of the reported result (D7). In addition to the domain scores, both systems also provide an overall bias rating per study. All bias criteria were rated on a three point rating scale: low, moderate, and high risk. Results were summarized using the robvis Shiny web app (McGuiness & Higgins, 2020).

Since the latest Cochrane tools are described in much detail and are highly structured, they allowed for bias ratings by one person per study. A subset of 8 studies (14%) was rated by two raters and differences were discussed until consensus was reached. Three post-graduate psychologists rated all studies (KDJ, ML, RG). The final ratings were assessed for potential systematic differences between raters by the first author and discussed with the other authors. In addition to the Cochrane criteria, potential for allegiance bias was assessed. This variable was coded 'yes' if the developer of the feedback system was involved as a co-author in the study.

5.6. Moderating variables

5.6.1. Feedback characteristics

Feedback type was coded as 'raw data', 'expected recovery trajectories (ERT)' when patients progress was checked against a benchmark, or 'clinical support tools' when a combination of ERT and additional information on the treatment process (e.g. motivation, therapeutic alliance) was used for NOT patients. Timing of feedback was coded as 'timely' when feedback was provided before or within the session, and as 'delayed' if the feedback took place after the session (often 1–7 days later). Frequency of feedback was coded as 'continuous' if feedback was given each session or weekly (or daily in inpatient settings) and as 'intermittent' if feedback was provided less frequent. Feedback to patients was coded 'yes' if patients either received feedback independent from the therapist, or if therapists were explicitly instructed to discuss the feedback with the patient at each feedback moment, and otherwise 'no'. Training was coded 'yes' if therapists were provided with training in the feedback tool.

5.6.2. Study characteristics

The country in which the study took place was coded as 'US' versus 'other countries' (Europe and Australia). Study year equaled the publication year. For studies that were current unpublished, 2020 was used as study year. Participation of therapists was rated as 'voluntary' if therapists could decide to participate in the study by themselves, and as 'mandatory' if the study was rolled out department wide. Treatment intensity was coded as 'mild', 'moderate', or 'severe', based on a combination of the treatment setting and the presented problems. Studies that were conducted in a college counseling center or in primary care were coded as mild, if the predominant problems treated were adjustment, depressive, and/or anxiety disorders. Studies in inpatient settings and emergency care were coded as severe. All other studies were coded as moderate. The setting of the study was coded as 'inpatient' or 'outpatient', the age group for treatment was rated as 'adults' or 'youth', and the treatment form was rated as 'individual/couple', 'group', or 'mixed' treatment of individual and group therapy. Treatment length was coded 'fixed' (i.e., preset) or 'flexible'. Independent outcome assessment was coded as 'yes' if an outcome instrument was used in addition to the feedback instrument, and 'no' if the outcome instrument was also the feedback instrument.

5.7. Meta-analytic approach

5.7.1. Primary outcome variables

Most of the included studies had multiple primary outcome variables, for example because multiple scales or multiple raters were used to assess symptom reduction between pre- and post-treatment. Multiple effect sizes were computed for these studies. To account for the dependency of effect sizes within studies we estimated three-level meta-analytic models (Cheung, 2014; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013), using the R package *metafor* (Viechtbauer, 2010). The three-level model takes three different levels of variability in effect sizes into account: the sampling variance of the individual extracted effect sizes (level 1), the variance between effect sizes extracted from the same study (level 2), and the variance between effect sizes extracted from different studies (level 3). Variability on levels 2 and 3 is estimated, while level-1 variability is assumed to be known and computed as the observed sampling variance of the extracted effect sizes.

We first estimated a random effects three-level meta-analytic model, to provide an estimate of the overall effect size and variance estimates. Likelihood ratio tests were used to assess whether the variance in effect size estimates between studies and within studies differed significantly from zero. Sensitivity analyses were conducted by excluding each of the effect sizes in turn (*leave-one-out* method) and re-estimating the overall effect size estimate and variance estimates. Effect sizes that were three standard deviations higher or lower than the mean ($z \leq -3$ or $z \geq 3.0$) were considered outliers. In a second step, we used a three-level mixed effects model to assess whether study-level moderator variables and bias criteria could explain variability in effect sizes between studies. Wald tests based on a t-distribution were used to test the fixed effects of continuous moderators variables. For categorical variables with more than two categories, an omnibus test based on the *F*-distribution was performed. For categorical moderators, the model was parameterized such that we could estimate a separate effect size for each category. We used an alpha level of 0.05. As moderator analyses in meta-analysis typically have low power (i.e., due to the limited number of studies included) we did not use a correction for testing multiple moderators (e.g., Bonferroni correction). The three-level random effects analyses were conducted twice; first for full samples (both NOT and OT cases) using data from all studies, and secondly for the NOT cases, using a subsample of studies providing data for NOT cases. The definition of NOT could vary over studies.

The I^2 statistic is used as a standardized measure of heterogeneity for the three-level meta-analysis model (Nakagawa and Santos, 2012). The I^2 values for the between-study variance and within-study variance can be interpreted as the percentage of the total variability in the outcome that is due to between- and within-study heterogeneity, respectively.

5.7.2. Secondary outcome variables

Since the secondary outcome variables (dropout, % of deteriorated cases, treatment duration) do not have nested data, standard random and mixed effects meta-analysis models were conducted. Moderation analyses were conducted using the same moderators used in the models for the primary outcome variables. Random effects were estimated with the DerSimonian and Laird method (DerSimonian & Laird, 1986).

5.7.3. Publication bias

Egger's regression test of funnel plot asymmetry (Egger, Smith, Schneider, & Minder, 1997; Sterne & Egger, 2001) was used to evaluate publication bias for the results of the standard random effect models. The regression test assesses whether there is a relationship between the effect sizes and a measure of study precision (in our case the sampling variance). For assessing publication bias in the results of the three-level random effect models, we used an adapted version of Egger's regression

test (Egger et al., 1997; Sterne & Egger, 2001; Viechtbauer, 2017) by re-estimating the three-level models and including the sampling variance as a moderator. Significant effects of the sampling variance are an indication of publication bias. Additionally, for the overall effect of feedback, the fail-safe N was estimated. The fail-safe N refers to the number of potentially missing studies with a z-value of 0 that should be added in order to make the overall effect size statistically insignificant.

6. Results

6.1. Sample

The search resulted in 1157 records. After removal of duplicates, 983 records were screened for eligibility, which resulted in excluding 863 records. After reading the full text of the remaining 120 articles, an additional 62 articles were excluded. The final sample consisted of 58 studies, which resulted in 110 effect sizes (see Fig. 1).

Studies were predominantly conducted in adults (91%), in outpatient psychotherapy or counseling settings (84%). Treatments were mostly individual or couples therapies (71%), compared to group therapy (10%), or a combination of group and individual therapies (19%). Studies took place in Europe (52%, led by the Netherlands [$n = 10$], Norway [$n = 5$], and Germany [$n = 4$]), the US (45%), and Australia (3%). The most frequently used feedback systems were PCOMS (36%), and the OQ System (38%); other feedback systems were only used in one or two studies each. The type of feedback provided were raw scores (38%), a comparison with expected recovery trajectories (ERT; 45%), or ETR supplemented with clinical support tools (CST; 17%). Over half of the studies provided immediate feedback (59%), rather than one week delayed (41%). The bulk of studies provided session by session feedback (88%), with only some studies using less frequent measurements (12%). The characteristics of the studies are presented in Table 1.

6.2. Study quality

Study quality assessments showed substantial sources of potential bias in the studies. Summaries of the bias criteria can be found in the Online Supplement (see Figs. I to IV). For the RCTs, the overall risk was high for 31% of the studies. With the exception of D4 (bias in measurement of the outcome), studies had a low risk on the individual domains and high risk scores were relatively rare. For the cohort studies, the overall risk was high for all studies. This was because all studies had a high risk on domain D1 (bias due to confounding). For most of the other domains the cohort studies had a low risk, with the exception of D5 (bias due to missing data) and D6 (Bias in measurement of outcomes). On D5 three out of nine studies had a moderate risk, whereas on D6 all nine studies had a moderate risk of bias.

Effect sizes for studies with a high and medium risk of bias were somewhat lower ($d = 0.15$, 95% CI [0.07, 0.23]) for both risk levels) than for studies with a low risk of bias ($d = 0.20$, 95% CI [0.01, 0.39]), although the difference was not significant, $F(1,106) = 0.15$, $p = 0.86$.

Studies in which a developer of a feedback system was involved (d = 0.21, 95% CI [0.12, 0.31]) had somewhat higher effect sizes than studies in which no feedback developers were in involved (d = 0.12, 95% CI [0.06, 0.19], although this potential allegiance effect was not significant, $F(2, 107) = 2.29$, $p = 0.13$.

6.3. Primary outcomes in full sample

6.3.1. Symptom reduction at post-treatment

The mean sample size across 58 studies was 321.4 ($SD = 433.6$; Mdn = 184),[5] and the total sample size across all studies was 21,699 cases. Between 1 and 9 outcome instruments were administered per study, with a mean of 2.04 ($SD = 1.73$). The three-level random effects model showed an overall effect size estimate that was significantly different from zero and equaled $d = 0.16$, 95% CI [0.11, 0.22], favoring the feedback condition. Patzig and Schiepek (2015) reported one extreme outlying effect size ($z = 7.47$). We decided to exclude this effect size from further analyses, which resulted in a similar overall effect size of $d = 0.15$, 95% CI [0.10, 0.20] (see Fig. 2). After excluding the Patzig and Schiepek effect size from the analysis (but keeping the other effect sizes reported in that study), the between-study variance I^2 equaled 72.63, meaning that approximately 72% of the total variability in outcome could be attributed to between-study heterogeneity. The within-study variance equaled zero. The remaining 28% of the total variance could therefore be attributed to sampling variance. Fig. 2 shows a forest plot including the average estimated effect size for each study.

Sensitivity analysis revealed that excluding single studies using leave-one-out did not result in significant changes in the effect size and variance. Analyses were also repeated excluding the eight studies that did not have a RCT design. The estimated overall effect size slightly increased to $d = 0.18$ (95% CI: [0.12, 0.24], $p < 0.001$). Excluding the five studies which had (low frequency) routine outcome measurement in the control group with information accessible to clinicians also resulted in a slightly increased effect size, $d = 0.17$ (95% CI: [0.11, 0.22], $p < 0.001$). The estimates of within-study variance and between-study variance in effect sizes were unaffected in both cases. Two studies had a much lower frequency of feedback than other studies (Lutz, Boehnke, Köck, & Bitterman, 2011; Schöttke, Unrath, & Uhlmann, 2019). However, excluding these had no effect on the effect size.

6.3.2. Moderators

We tested whether study and feedback characteristics moderated the effectiveness of progress feedback. For the categorical moderators, Table 2 shows a separate effect size for each level of the variable. Five moderator variables proved to significantly affect outcomes, namely the outcome and the feedback instrument(s) that were used in the study, the country in which the study had been conducted, and the year in which the study was published. Studies that used the Outcome Rating Scale (ORS; part of the PCOMS system; $d = 0.34$) had significantly larger overall effect sizes than studies using the OQ-45 ($d = 0.11$) or other outcome instruments ($d = 0.12$), $F(2, 106) = 8.63$, $p < 0.001$. Similarly, studies using the PCOMS feedback system ($d = 0.24$) also had larger effect sizes than studies using the OQ System ($d = 0.13$) or other feedback systems ($d = 0.07$), $F(1, 106) = 3.42$, $p = 0.04$. Studies conducted in the US ($d = 0.23$) resulted in higher effect sizes than studies conducted elsewhere ($d = 0.11$), $F(1, 107) = 4.57$, $p = 0.03$. Studies published in later years were found to report on average 0.02 lower effect sizes per year since the first study in 2001 ($\beta = 0.02$, $SE = 0.01$, $F(1, 107) = 6.70$, $p = 0.01$). Studies in which independent outcome instruments were used were more likely to report smaller effect sizes ($d = 0.08$), than studies in which the feedback instrument was the outcome instrument ($d = 0.19$), $F(1, 107) = 4.06$, $p = 0.047$.

6.3.3. Effects at follow-up

Four studies presented data on follow up measurements, resulting in ten effect sizes (see Fig. I, Online Supplement). The three-level random effects model provided an overall non-significant effect size estimate of 0.18 (95% CI [−0.03, 0.39]), favoring feedback. Sensitivity analyses showed that excluding the Anker, Duncan, and Sparks (2009) study substantially reduced the pooled effect size to 0.09, 95% CI [−0.13, 0.32].

6.4. Primary outcomes in Not on Track cases

6.4.1. Symptom reduction at post-treatment

For 27 studies, data for computing 43 effect sizes were available for the subgroup of NOT cases. Figure 3 shows the average estimated effect size for each study. The overall effect size derived from the three-level

[5] Excluding the combined Lambert CCC studies

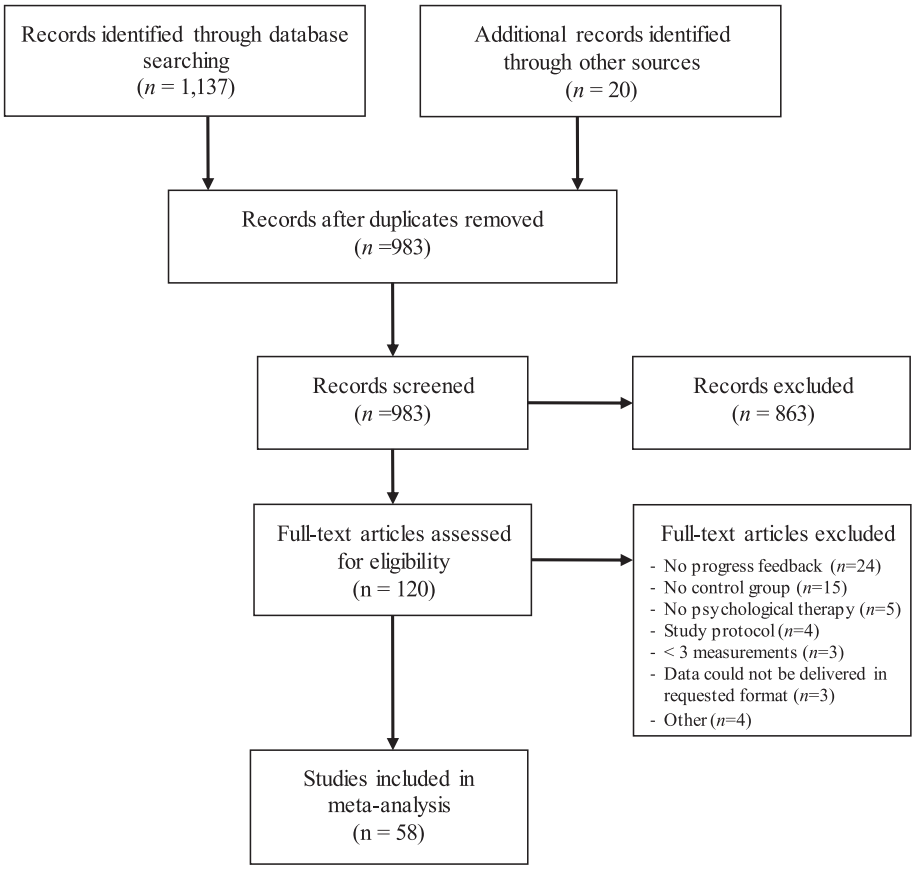

Fig. 1. Prisma Flow Diagram.

random effects model was significantly different from zero and equaled $d = 0.17$ (95% CI [0.09, 0.25]). Sensitivity analyses showed that removal of studies would not result in substantial variation in the effect size. Excluding studies that were not RCTs resulted in an overall effect size of 0.20, 95% CI (0.11, 0.28). Similarly, excluding studies that used a control group in which some form of outcome monitoring took place, also slightly increased the effect size to 0.20 (95% CI: [0.12, 0.27]). Statistic I^2 equaled 27.85 for the between-study variance and 36.01 for the within study variance, meaning that about 28% of the total variability in the outcome could be attributed to between-study heterogeneity and 36% could be attributed to the within study heterogeneity. The remaining 36% of the total variability could be attributed to sampling variance.

6.4.2. Moderators

Feedback type was found to be a significant moderator in the NOT sample. Studies using CSTs ($d = 0.36$) were more effective than feedback systems that presented ETRs ($d = 0.12$) or raw scores ($d = 0.04$) for this subgroup, $F(2, 40) = 5.08$, $p = 0.01$ (See Table 2).

6.5. Secondary outcomes

6.5.1. Dropout

For a subgroup of 39 studies information on dropout rate was available per condition. The mean overall dropout rate for control groups was 24.5%, whereas the mean overall dropout rate for feedback groups was 20.9%, which corresponds with a 20% increased chance of dropout in the control conditions compared to the feedback conditions, $OR = 1.19$, 95% CI [1.03, 1.38], $p < 0.01$, $k = 39$ (see Fig. 4), with moderate heterogeneity across studies ($I^2 = 47.75$). Sensitivity analyses showed that the largest drop in log odds ratio would be found by excluding Simon et al. (2013), which would reduce the odds ratio to 1.15, 95% CI [1.0, 1.32]. In RCTs the *OR* was 1.18, 95% CI [0.98, 1.39].

The effect of feedback on dropout was moderated by feedback instrument, $Q_M(2) = 16.33$, $p = 0.0003$. Studies using PCOMS ($OR = 1.48$) found significantly higher effects of feedback on reducing dropout than studies using the OQ System ($OR = 1.21$) and studies using other feedback systems ($OR = 1.08$). Additionally, in studies conducted outside the US ($OR = 1.07$), the effect of feedback on reducing dropout was significantly lower than in studies inducted in the US ($OR = 1.77$), $Q_M(1) = 8.99$, $p = 0.003$.

6.5.2. Percentage of deteriorated cases

For 26 studies data was available on the percentage of deteriorated cases per condition. On average 5.4% of patients deteriorated in the control conditions, whereas an average of 4.6% of patients deteriorated in the feedback conditions. There was no significant effect of feedback on the rate of deteriorated cases, $OR = 1.16$, 95% CI [0.99, 1.35], $p = 0.07$, $k = 26$ (see Fig. 5). Sensitivity analysis showed that excluding Delgadillo et al. (2018) would lower the effect size to 1.10, and excluding Lutz et al. (2011) would increase the effect size to 1.23. Since both studies were not considered outliers, they were kept in the sample. The heterogeneity I^2 across studies equaled 7.09, indicating low heterogeneity across studies.

Two significant moderators were found. Studies in which therapists had received training in the feedback system ($OR = 1.28$) had larger effect sizes than studies in which no training was provided ($OR = 0.81$), $Q_M = 5.45$, $p = 0.02$. In addition, studies that used feedback systems presenting ETRs ($OR = 1.36$) had larger effect sizes than studies using CSTs (OR = 1.29) or raw scores ($OR = 0.81$), $Q_M = 6.70$,$p = 0.04$.

6.5.3. Treatment duration

There was no significant overall effect of feedback on the number of sessions, $d = -0.04$, 95% CI [−0.06, 0.15], $p = 0.45$, $k = 26$ (see Fig. II,

Table 1
Characteristics of studies included in main analyses.

Study (first author, year(s))	Country	n	Setting	Presenting problems; Treatment intensity	Tx duration (M)	Design; N intervention /control groups; Training; Check feedback use	Feedback instrument(s); Outcome instrument(s)	Feedback type; timing; recipient	Frequency of feedback
Amble, Gude, Stubdal, Andersen, and Wampold (2015); Amble, Gude, Ulvenes, Stubdal, and Wampold (2016)	Norway	259	Inpatients and outpatients	Mood, anxiety, substance abuse disorders; severe	9.9 sessions	RCT;1/1;Y;N	OQ-45	ERT; timely; clin pt	Every session
Anker et al. (2009)	Norway	410	Outpatients: family counseling center	Relationship problems; mild	4.6 sessions	RCT;1/1;Y;N	PCOMS	ERT; timely; clin	Every session
Berking, Orth, and Lutz (2006)	Switzerland	118	Inpatients	Mood, anxiety, adjustment disorders; severe	39.0 days	RCT;1/1;N;Y	FEP	ERT; delayed; clin	Day 1, 3, weekly after that
Bickman, Douglas Kelley, Breda, Andrade, and Riemer (2011)	USA	340	Private behavioral health: home-based treatment; youth	Various (not specified); moderate	11.0 sessions	RCT;1/1;Y;Y	CFS (includes SFSS); SFSS	RS; timely; clin	Every session
Bickman et al. (2016)	USA	257	Outpatients; youth	Various (not specified); moderate	11.0 sessions	RCT;1/1;Y;Y	CFS (includes SFSS); SFSS,	RS; timely; clin	Every session
Bovendeerd et al., (2020)	Netherlands	1733	Outpatients	Mood, anxiety, psychoso-matic disorders; mild	188.3 days	RCT;1/1;Y;Y	PCOMS; OQ-45, MHC-SF	RS; timely; clin	Every session
Brattland et al. (2018)	Norway	113	Outpatients	Mood, anxiety, personality; moderate	12.9 sessions	RCT;1/1;Y;Y	PCOMS; BASIS-32	ERT; timely; clin	Every session
Connolly Gibbons et al. (2015)	USA	100	Outpatients: community mental health	Mood, post-traumatic stress disorders; moderate	6.0 sessions	RCT;1/1;N;Y	BASIS-24	ERT; delayed; clin	Every session
Crits-Christoph et al. (2012)	USA	304	Outpatients: community mental health service	Substance abuse disorders; moderate	7.8 sessions	Cohort;1/1;Y;N	OQ-45, ASC; + alcohol use, drug use	ERT; timely; clin	Every session
Davidsen et al. (2017)	Denmark	159	Outpatients; group	Eating disorders; moderate	12 sessions	RCT; 1/1;Y;N	PCOMS; + EDE, WHO-5, SCL-90	ERT; timely; clin pt	Every session
De Jong et al. (2012)	Netherlands	413	Outpatients	Mood, anxiety, adjustment disorders; moderate	10.2 sessions	RCT;1/1;Y;Y	OQ-45	RS; delayed; clin	Sessions 1–5, 10, 15, 20, 25, 30, 35
De Jong et al. (2014)	Netherlands	475	Outpatients	Personality, mood, adjustment disorders; moderate	32.3 sessions	RCT;1/1;N;N	OQ-45	RS; timely; clin pt	Every session
De Jong, Segaar, Ingenhoven, van Busschbach, and Timman (2017)	Netherlands	206	Inpatients	Personality disorders; severe	272.0 days	RCT;2/1;Y;N	OQ-45	RS; timely; clin pt	Weekly
De Jong et al., 2020	Netherlands	347	Outpatients	Mood, anxiety, somatoform; moderate	9.0 sessions	RCT;2/1;Y;Y	OQ-45, ASC	ERT; timely; clin	Every session up to session 15
Delgadillo et al. (2017)	UK	594	Outpatients: primary care	Mood, anxiety disorders; mild	8.7 sessions	Cohort;1/1;Y;N	PHQ-9, GAD-7	ERT; delayed; clin	Every session
Delgadillo et al. (2018)	UK	2233	Outpatients, primary care	Mood, anxiety disorders; mild	6.5 sessions	RCT, 1/1;Y;N	PHQ-9, GAD-7; + WSAS	ERT; RS; timely; clin	Every session
Errázuriz and Zilcha-Mano (2018)	Chile	547	Outpatient mental health center	Mood, adjustment disorder; moderate	7.8 sessions	RCT;4/1;N;Y	OQ-30, WAI, SCS; OQ-45	RS; timely; clin	Every session
Galvinhill (2001)[a]	USA	96	Outpatients: college counseling center	Adjustment, mood, anxiety disorders	–	RCT;1/1;Y;N	OQ-45	ERT; delayed; clin	Weekly
Grizzell et al. (2016)	USA	30	Outpatients; group	Vocational rehabilitation	5.6 sessions	RCT;1/1;Y;Y	OQ-45, ASC	ERT; delayed; clin pt	Weekly
Hansen et al. (2015)	Australia	73	Outpatients; youth	Mood, anxiety disorders	8.0 sessions	Cohort;1/1;Y;Y	PCOMS	ERT; timely; clin	Every session
	Sweden	262	Outpatients		15.8 sessions	RCT;1/1;Y;N	OQ-45	RS; timely; clin pt	Weekly

(*continued on next page*)

Table 1 (*continued*)

Study (first author, year(s))	Country	n	Setting	Presenting problems; Treatment intensity	Tx duration (M)	Design; N intervention /control groups; Training; Check feedback use	Feedback instrument(s); Outcome instrument(s)	Feedback type; timing; recipient	Frequency of feedback
Hansson, Rundberg, Österling, Öjehagen, and Berglund (2013)				Mood, anxiety, personality disorders					
Hawkins et al. (2004)	USA	201	Outpatients: hospital based clinic	Mood, anxiety disorders	8.2 sessions	RCT;2/1;N;N	OQ-45	ERT; delayed; clin pt	Every session
Janse et al. (2017)	Netherlands	1006	Outpatients	Mood, anxiety, somatoform	15.0 sessions	Cohort;1/1;Y;Y	PCOMS; + SCL-90	ERT; timely; clin	Every session
Janse et al. (2020)	Netherlands	368	Outpatients	Mood, anxiety, somatoform	15.3 sessions	RCT;1/1;Y;Y	PCOMS; + SCL-90	ERT; timely; clin	Every session
Kellybrew-Miller (2014)	USA	91	Outpatients; community mental health center	Mood, anxiety; moderate	2,2 sessions	RCT;1/1;Y;Y	PCOMS; SOS-10	RS; timely; clin	Every session
Koementas-de Vos et al. (2018)	Netherlands	259	Oupatient; group	Mood and anxiety disorders	10.8 sessions	Cohort;1/1;Y;Y	OQ-45, Group CST; OQ-45	RS, delayed, clin pt	Every session
Kremer (2018)	USA	108	Private practice; youth	Various (not specified), mild	7.5 sessions	RCT;1/1;Y;Y	PCOMS; PSC, PSC-Y	RS; timely; clin	Every session
Lambert CCC studies:	USA	5085	Outpatients: college counseling center	Mood, adjustment disorders	5.4 sessions	–	OQ-45	ERT; –; clin	Every session
- Lambert et al. (2001)		609			3.3 sessions	RCT;1/1; Y;Y		delayed	
- Lambert et al. (2002)		1020			5.3 sessions	Cohort;1/1;Y; N		delayed	
- Whipple et al. (2003)		981			6.2 sessions	RCT; 1/1; Y; N	+ ASC	delayed	
- Harmon et al. (2007)		1374			4.6 sessions	RCT; 2/0; N; N	+ ASC	delayed; +pt	
- Slade et al. (2008)		1101			5.8 sessions	RCT; 2/0; N; N	+ ASC	timely; +pt	
Lester (2012)	USA	118	Inpatients; acute psychiatric care; youth	Mood disorders	1.8 sessions	RCT;1/1;Y;Y	PCOMS; + Y-OQ	ERT; delayed; clin pt	Every individual session
Lutz et al. (2011)	Germany	349	Outpatients	Mood, anxiety, adjustment disorders	39.8 sessions	RCT;1/1;N;Y	BSI, IIP, BDI	RS; delayed; clin pt	Sessions 10, 20, 40/45, 55, 75, 95
McClintock, Perlman, McCarrick, Anderson, and Himawan (2017)	USA	56	Outpatient, counseling center	Depression; mild	4.13 sessions	RCT;1/1;N;Y	CFF; BDI-II, SOS	CST, delayed, clin	Every session
Murphy, Rashleigh, and Timulak (2012)	Ireland	110	Outpatients: college counseling center	Various (not specified)	3.7 sessions	RCT;1/1;Y;N	PCOMS	ERT; timely; clin pt	Every session
Newnham, Hooke, and Page (2010)	Australia	1308	Inpatients and day patients; group	Mood, anxiety disorders	10.0 days	Cohort;1/2;N;N	WHO-5; +DASS, SF-14, HoNOS	ERT; delayed; clin pt	Day 5, 10
Owen (2019)	USA	34	Outpatients, university clinic	Various (not specified), mild	17.6 sessions	RCT; 1/1;Y/N	PCOMS, PHQ-9, SOS-10	RS, timely, clin	Every session
Patzig and Schiepek (2015)	Austria	96	Inpatients	Chronic alcohol addiction	103.9 days	RCT;1/1;Y;Y	EKF, TBP—S; +SCL-90, BDI-II	ERT; delayed; clin	Biweekly
Probst et al. (2013); Probst, Lambert, Dahlbender, Loew, and Tritt (2014)	Germany	252	Inpatients; psychosomatic clinic	Mood, somatoform, anxiety disorders	33.6 days	RCT;1/1;N;N	OQ-45, ASC	ERT; delayed; clin	Weekly
Puschner et al. (2009)	Germany	294	Inpatients	Psychotic, mood disorders	59.9 days	RCT;1/1;N;Y	OQ-45	ERT; delayed; clin pt	Weekly
Reese et al. (2009; study 1)	USA	74	Outpatients: college counseling center	Various (not specified)	5.9 sessions	RCT;1/1;Y;N	PCOMS	ERT; timely; clin pt	Every session
Reese et al. (2009; study 2)	USA	74	Outpatients: training clinic	Various (not specified)	6.9 sessions	RCT;1/1;Y;N	PCOMS	ERT; timely; clin pt	Every session
Reese, Toland, Slone, and Norsworthy (2010)	USA	92	Outpatients; couples	Relationship problems	5.9 sessions	RCT;1/1;Y;N	PCOMS	ERT; timely; clin pt	Every session
Reeves (2010)	USA	220	Outpatients: college counseling center	Various (not specified)	7.8 sessions	Cohort;1/1;Y;N	OQ-45	ERT; timely; clin	Every session
	Norway	75				RCT;1/1;Y;Y			Every session

(*continued on next page*)

Table 1 (*continued*)

Study (first author, year(s))	Country	n	Setting	Presenting problems; Treatment intensity	Tx duration (M)	Design; N intervention /control groups; Training; Check feedback use	Feedback instrument(s); Outcome instrument(s)	Feedback type; timing; recipient	Frequency of feedback
Rise, Eriksen, Grimstad, and Steinsbekk (2012)			Outpatients; mental health hospital	Anxiety, mood disorders	3.8 sessions		PCOMS; + BASIS-32, SF-12 MHC	ERT; timely; clin pt	
Schmidt et al. (2006)	UK	33	Outpatients: eating disorders clinic	Eating disorders	10.0 sessions	RCT;1/1;Y;N	SEED; + TREAT-EAT, HADS	ERT; timely; pt	Sessions 1, 5, 10
Schottke et al. (2019)	Germany	230	Outpatients	Common mental disorders; moderate	30 sessions	RCT;1/2;Y;Y	FEP-2; OQ-30	RS; delayed; clin	Every three months
Schuman, Slone, Reese, and Duncan (2015)	USA	263	Outpatients: military; group	Substance use	5.0 sessions	RCT;1/1;N;N	PCOMS; ORS	ERT; timely; clin	Every session
She et al. (2018)	China	186	University counseling center	Various (e.g. interpersonal problems, mood)	5.1 sessions	RCT;1/1;Y;Y	PCOMS; ORS	ERT; timely; clin	Every session
Simon et al. (2012)	USA	370	Outpatient: hospital based clinic	Mood, anxiety disorders	6.6 sessions	RCT;1/1;N;N	OQ-45, ASC	ERT; delayed; clin pt	Every session
Simon et al. (2013)	USA	133	Inpatients	Eating disorders	12.6 sessions	RCT;1/1;N;N	OQ-45, ASC	ERT; timely; clin	Every session
Slone, Reese, Mathews-Duvall, and Kodet (2015)	USA	84	Outpatients: college counseling center; group	Anxiety, adjustment, mood disorders	7.3 sessions	RCT;1/1;Y;N	PCOMS; CCAPS-34 DI	ERT; timely; clin	Every session
Tilden et al. (2019)	Norway	328	Outpatients; couples and family therapy centers	Partner, family and individual problems; moderate	5 sessions	RCT;1/1; Y; N	STIC; RDAS, BDI-II, BAI, FAD, SDQ, SF36, OQ-45	RS; timely; clin	Every session
Trudeau (2000)[b]	USA	127	Outpatients: community mental health	Mood, anxiety, adjustment disorders	6.7 sessions	RCT;1/2;N;N	OQ-45; + RAND36	ERT; delayed; clin	Every session
Tzur Bitan et al. (2019)	Israel	123	Outpatient, group and individual	Various (no exclusion of diagnosis), severe	133.9 days	RCT;1/1;N;N	HSCL; SAI; OQ-45; PWB	RS; delayed; clin	Every session
Van Oenen et al. (2016)	Netherlands	370	Outpatients: psychiatric emergency center	Various (e.g. psychosis, depression)	9.3 sessions	RCT;1/1;Y;Y	PCOMS; + OQ-45, BSI	ERT; timely; clin pt	Every session
Winkelhorst, Hafkenscheid, and De Groot (2013)	Netherlands	69	Outpatients	Personality disorders; moderate	11.6 sessions	Cohort;1/1;Y;Y	PCOMS; OQ-45	RS; timely; clin	Every session

Note. RCT = randomized controlled trial; Y = yes; N = no; RS = raw score; ERT = expected recovery trajectories; CST = Clinical Support Tool, clin = clinician; pt. = patient; ASC = Assessment for Signal Clients; BAI = Beck Anxiety Inventory; BASIS = Behavior and Symptom Identification Scale; BDI = Beck Depression Inventory; BSI = Brief Symptom Inventory; CCAPS = Counseling Center Assessment of Psychological Symptoms; CFF = Common Factors Feedback system; CFS = Contextualized Feedback System; DASS = Depression Anxiety Stress Scale; EDE = Eating Disorder Examination; EKF = Emotionale Kompetenz Fragebögen [emotional competence questionnaire]; EMI = Emotion Inventory; FAD = Family Assessment Device; FEP = Fragenbogens zur Evaluation von Psychotherapieverläufen [psychotherapy evaluation questionnaire]; GAD-7 = Generalized Anxiety Disorder scale; HADS = Hospital Anxiety and Depression Scale; HoNOS = Health of National Outcome Scale; IIP = Inventory of Interpersonal Problems; MHC-SF = Mental Health Checklist Short Form; OQ-45 = Outcome Questionnaire; ORS = Outcome Rating Scale; PCOMS = Partners for Change Outcome Management System (consists of ORS and Session Rating Scale[SRS]); PHQ-9 = patient health questionnaire; PSC = Pediatric Symptom Checklist; RAND36 = RAND Health Survey; RDAS = Revised Dyadic Adjustment Scale; SCS = Self Concealment Scale; SCL-90 = Symptom Checklist; SDS = Sheehan Disability Scale; SDQ = Strengths and Difficulties Questionnaire; SEED = Short Evaluation of Eating Disorders; SF = Short Form health survey; SFSS = Symptoms and Functioning Severity Scale; SOS = Schwartz Outcome Scale; TBP-S = Therapie-Prozessbogen Sucht [therapy process questionnaire for addiction]; WAI = Working Alliance Inventory; WHO-5 = World Health Organization Wellbeing Index; WSAS = Wellbeing and Social Adjustment Scale; Y = Youth version;
Data was collected in individual therapy in adults, unless otherwise specified.

[a] Data from the oral feedback and written feedback groups was pooled.

[b] Data from the control group and no-feedback group was pooled.

Online Supplement), and considerable heterogeneity across studies was found, $I^2 = 86.26$. Contrary to what had been found by Kendrick et al. (2016) and Lambert et al. (2003), for the OT cases no significant effect of feedback on the number of sessions was found, $d = 0.02$, 95% CI [−0.12, 0.17], $p = 0.75$, $n = 8$. Additionally, a non-significant effect of feedback on treatment duration was found for the NOT cases, $d = 0.18$, 95% CI [−0.24, 0.60], $p = 0.85$, $n = 9$. This positive effect was mainly driven by one study (Koementas-de Vos, Nugter, Engelsbel, & De Jong, 2018), in which a very large effect was found on treatment duration in NOT cases ($d = 1.72$; 95% CI: [1.44, 2.01]). Although this study was not a significant outlier, removing the study resulted in an effect size of $d = -0.04$ (95% CI: [−0.09, 0.01]), which indicates that this effect is not stable. No significant moderators of the effect of feedback on treatment duration were found (see Table 3, number of sessions).

6.6. Publication bias

We did not find an indication of publication bias for any of the estimated meta-analytic models. In the three-level mixed effect models, there were no significant effects of sampling variance on the observed outcome (*p*-values equaled 0.55, 0.58, and 0.40 for the full sample, the NOT sample, and the follow-up data, respectively). Similarly, for none of the secondary outcomes measures Egger's regression test showed a significant effect of the sampling variance on observed outcome (*p*-

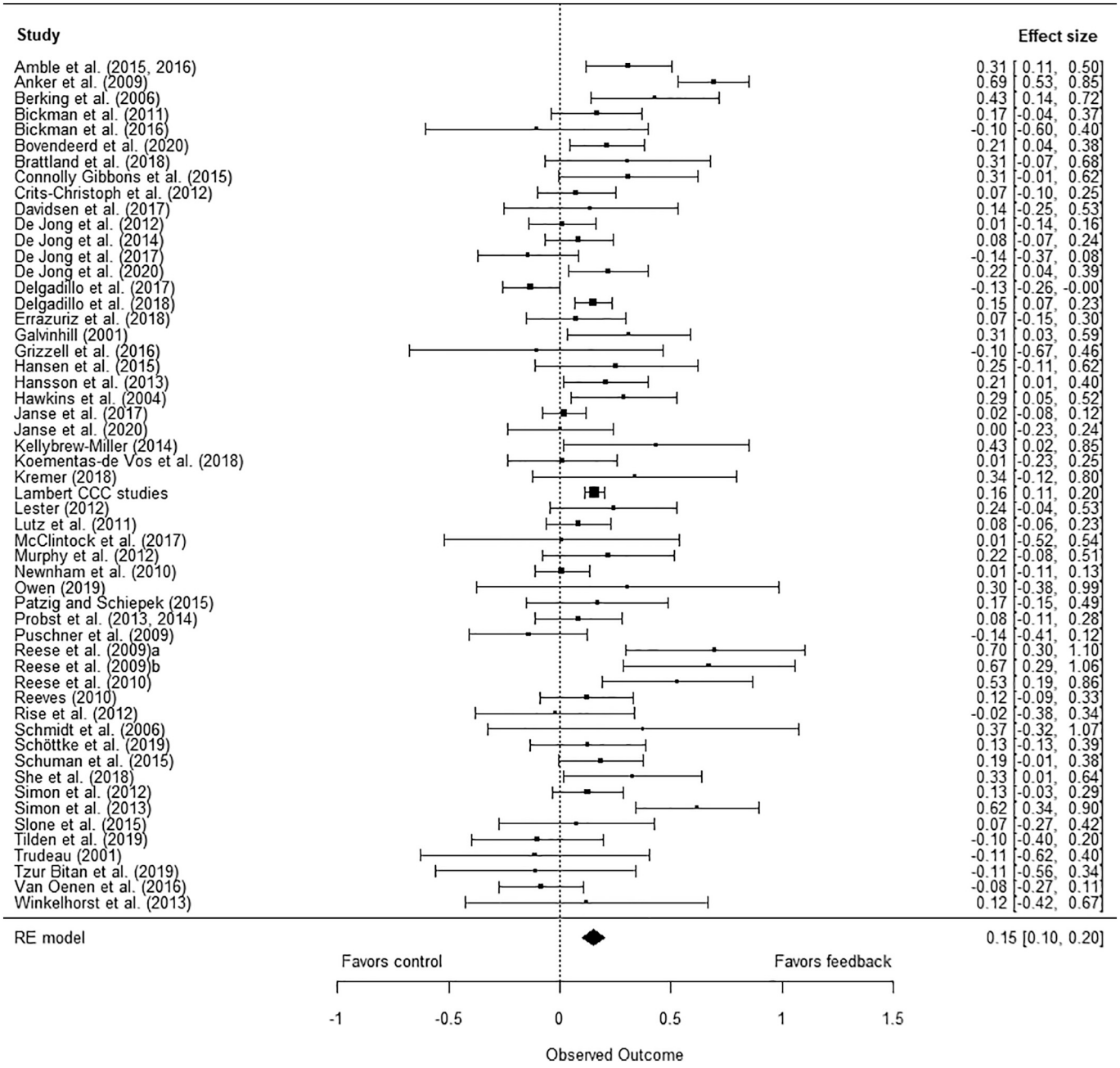

Fig 2. Effects of feedback in full group (OT and NOT combined)
Note. This forest plot is a graphical representation of the average effect sizes per study. Data was analyzed in a multilevel model, in which effect sizes per outcome measure were nested within studies. The RE model summary represents the outcome of this model.

values ranged from 0.28 to 0.66). The fail-safe N for the effect of feedback in the full sample was 3695, suggesting that there would need to be more than 3500 studies with no effect ($z = 0.00$) before the effect would become statistically non-significant.

6.7. Post-hoc analyses

Contrary to our expectations, we only found a slightly larger effect of feedback in the NOT cases ($d = 0.17$), compared to the full sample ($d = 0.15$). Previous meta-analyses by Lambert et al. (2003), Shimokawa et al. (2010), Kendrick et al. (2016), and Lambert et al. (2018) all reported larger effects of feedback in this subgroup. Østergård et al. (2018) analyzed studies using PCOMS feedback instrument and reported no effect of feedback in NOT cases. Post-hoc analyses based on our results and those of previous studies were conducted in order to enhance the interpretation of our results for this subgroup.

First, we looked at the differences between the subgroup of studies that reported on NOT cases and the subgroup that did not report on NOT cases. Studies that used the OQ system as feedback instrument were more likely to report NOT analyses than studies using other feedback systems: 78% of studies using the OQ System reported subgroup analyses on NOT cases versus 43% of studies using PCOMS, and 27% of studies using other feedback instruments, $\chi^2(2) = 9.25, p = 0.01$. Studies using the ORS (75%) or OQ-45 (65%) as outcome instrument were also more likely to report on NOT cases than studies using other outcome instruments, $\chi^2(2) = 7.65$, $p = 0.02$. A similar effect was found for studies using more advanced feedback systems: two third of studies using expected recovery curves and three quarter of studies using

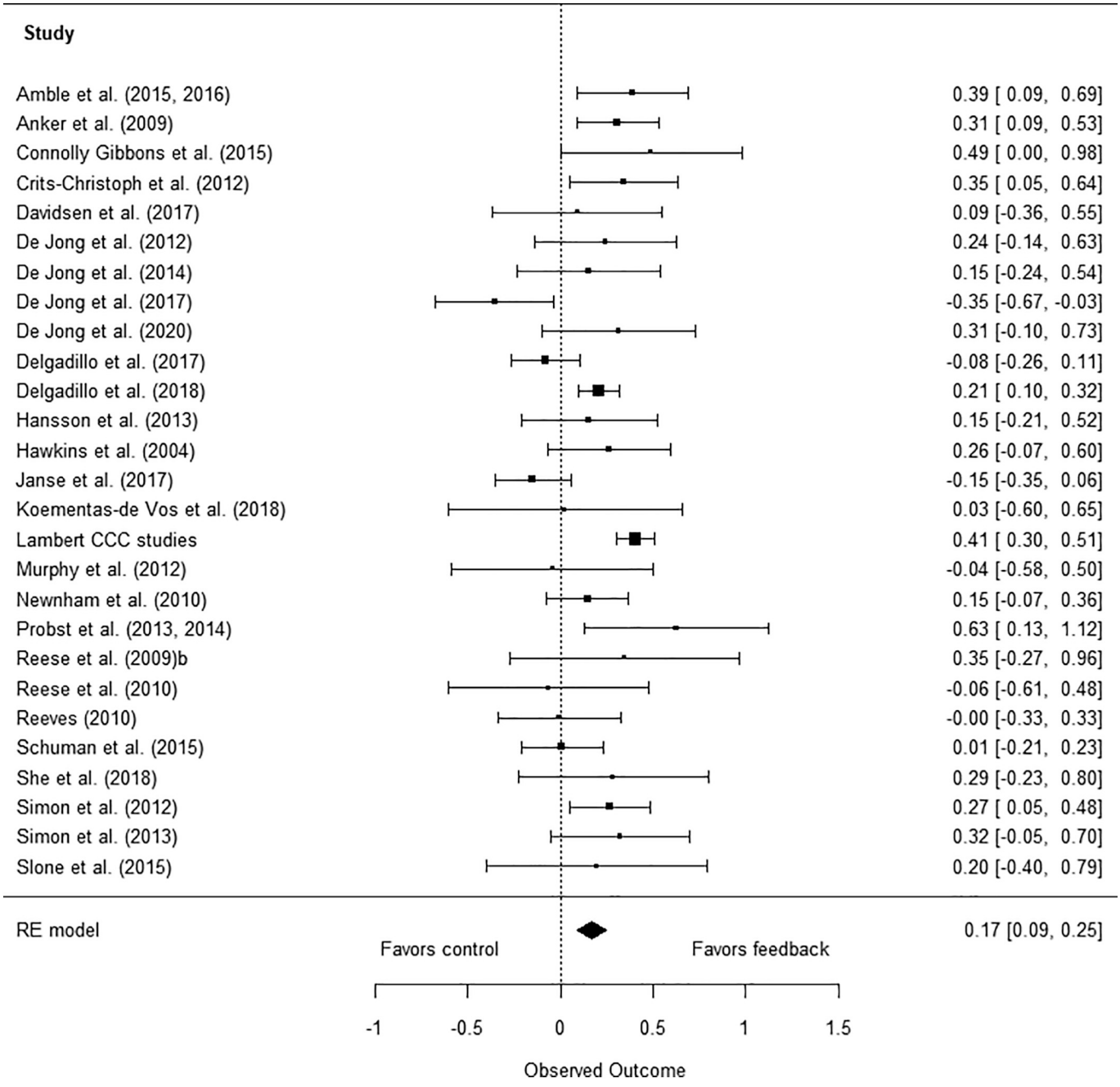

Fig. 3. Effect of feedback in NOT subgroup.
Note. This forest plot is a graphical representation of the average effect sizes per study. Data was analyzed in a multilevel model, in which effect sizes per outcome measure were nested within studies. The RE model summary represents the outcome of this model.

clinical support tools reported NOT subgroup analyses, compared to 23% of studies using raw scores, $\chi^2(2) = 11.21$, $p = 0.004$). No differences were found on other variables. Second, we checked whether the studies that reported NOT results had smaller effects in the full sample, and thus might be more likely to look for additional effects in subgroups. This was not the case ($F(1,107) = 0.75$, $p = 0.39$). Third, we checked whether our method of analysis might be of influence, as we are the only meta-analysis so far that uses three-level meta-analytic models to account for the dependency of multiple effect sizes from the same study. We replicated Lambert et al. (2018)[6] by selecting the studies they reported on and found an effect size of 0.36 (95% CI [0.18, 0.54]) in NOT cases for this subset. This is similar to the effect size reported by Lambert et al. ($d = 0.33$). A replication of Østergård et al. (2018), including a subset of studies using the PCOMS, resulted in an effect size of 0.03 (95% CI [−0.19, 0.25]), supporting their finding that no significant effect was found for NOT cases in PCOMS. We concluded that differences in results could not be ascribed to the use of different analytic models.

An additional post-hoc analysis was run in order to increase our

[6] The Lambert et al. (2018) meta-analyses reports separately on all of the college counseling center studies in their group, whereas we have pooled these studies.

Table 2
Moderation analyses in three-level random effects models.

	Full sample			Not on Track sample		
Moderator	*k*	*g*	*95% CI*	*k*	*g*	*95% CI*
Study characteristics						
Outcome instrument[1]						
- OQ-45	23	**0.11**	0.04, 0.18	14	0.22	0.09, 0.34
- ORS	13	**0.34**	0.24, 0.44	8	0.15	−0.05, 0.35
- Other	74	**0.12**	0.06, 0.18	23	0.14	0.02, 0.26
Treatment duration						
- Fixed (preset)	10	0.09	−0.03, 0.22	7	0.12	−0.03, 0.27
- Flexible	44	0.17	0.11, 0.22	20	0.19	0.09, 0.30
Country						
- USA	22	**0.23**	0.14, 0.31	11	*0.25*	0.13, 0.38
- Other	32	**0.11**	0.05, 0.17	16	*0.12*	0.03, 0.22
Treatment form						
- Individual/couple	39	*0.19*	0.13, 0.26	19	0.2	0.09, 0.30
- Group	6	*0.07*	−0.08, 0.23	5	0.1	−0.08, 0.28
- Mixed	9	*0.07*	−0.05, 0.18	3	0.16	−0.12, 0.45
Treatment setting						
- Outpatient	45	0.16	0.10, 0.21	23	0.18	0.09, 0.27
- Day or inpatient	9	0.13	−0.01, 0.26	4	0.13	−0.07, 0.34
Age group						
- Adults	49	0.15	−0.01, 0.34	27	-	
- Youth	5	0.17	0.09, 0.21	0	-	
Treatment intensity						
- Mild	20	*0.21*	0.12, 0.30	11	0.13	0.00, 0.26
- Moderate	24	*0.15*	0.07, 0.22	12	0.22	0.09, 0.35
- Severe	10	*0.06*	−0.06, 0.18	4	0.16	−0.04, 0.36
Outcome assessment						
- Independent measure	20	*0.08*	−0.00, 0.17	5	0.13	−0.03, 0.30
- Feedback instrument	34	*0.19*	0.13, 0.26	22	0.19	0.09, 0.29
Participation therapists						
- Mandatory	25	0.17	0.09, 0.24	10	0.17	0.02, 0.32
- Voluntary	29	0.14	0.07, 0.21	17	0.17	0.07, 0.27
Feedback characteristics						
Feedback instrument						
- OQ-45	18	**0.13**	0.04, 0.22	14	*0.25*	0.13, 0.36
- PCOMS	21	**0.24**	0.15, 0.32	9	*0.08*	−0.07, 0.23
- Other	15	**0.07**	−0.02, 0.17	4	*0.13*	−0.02, 0.27
Feedback type						
- Raw score	22	*0.10*	0.01, 0.18	5	**0.04**	−0.18, 0.25
- ERT	23	*0.20*	0.12, 0.28	16	**0.12**	0.04, 0.21
- CST	8	*0.16*	0.02, 0.30	6	**0.36**	0.22, 0.50
Timing						
- Delayed	21	0.10	0.02, 0.19	10	0.18	0.05, 0.31
- Immediate	33	0.19	0.12, 0.25	17	0.17	0.06, 0.27
Frequency						
- Intermittent	7	0.13	−0.01, 0.28	2	0.18	−0.07, 0.43
- Continuous	47	0.15	0.10, 0.21	25	0.17	0.08, 0.26
Feedback recipient[2]						
- Clinician only	32	0.15	0.09, 0.22	15	0.2	0.09, 0.31
- Clinician and/or patient	21	0.15	0.06, 0.24	12	0.13	0.00, 0.26
Training						
- Yes	39	0.16	0.10, 0.22	19	0.15	0.05, 0.25
- No	15	0.13	0.02, 0.23	8	0.24	0.07, 0.42

Note. Bold *d* values indicate significant differences between groups. Italic *d* values indicate an observed difference of more than 0.10 between categories. *k* = number of studies; ERC = Expected Recovery Curves; CST = Clinical Support Tool. 1. Studies may use multiple outcome instruments 2. Lambert CCC studies could not be coded because two studies provide feedback to patients and three do not.

understanding of why the ORS outcome instrument yielded larger effect sizes compared to other instruments (see Table 2). Østergård et al. (2018) found that the ORS results in larger effect sizes than other outcome instruments used in the same studies. To test whether this was the case in our sample as well, we ran an analysis on a subset of studies using PCOMS, comparing effect sizes on the ORS and other outcome instruments used within these same studies. Indeed, significantly larger effect sizes were found for ORS ($d = 0.35$) than for other outcome instruments ($d = 0.15$), $F(1, 36) = 9.29$, $p = 0.004$. This suggests that the ORS may either be more sensitive to change than other instruments, or that it overestimates effect sizes of feedback.

7. Discussion

This study provides the most comprehensive meta-analysis to date on the effect of progress feedback on psychotherapy outcomes, by including both randomized and non-randomized trials, looking at symptom reduction, as well as dropout and treatment duration, and by assessing moderating variables based on study and feedback characteristics. A small positive effect of progress feedback on symptom reduction was found both in the full sample ($d = 0.15$) and in the NOT subsample ($d = 0.17$). The size of the feedback effect on symptom reduction at follow up is yet unclear, as the number of studies that analyzed follow up data was very small, the effect found was not significant, and primarily driven by one study. As for the secondary outcomes measures, no effects of feedback were found on treatment duration and the percentage of deteriorated cases at the end of treatment. A small effect was found for the dropout rate.

A number of sources of potential bias were identified in different subsamples and sub analyses. For the full sample, studies in which the feedback instrument was also the outcome instrument were more likely to find large effect sizes than studies in which an independent outcome measure was used to assess the effectiveness of feedback. In the NOT subsample researchers' allegiance was found to bias results, as studies in which no developers of feedback systems were involved found smaller effects than studies in which developers were involved. In the subsamples for dropout and the percentage of deteriorated clients, it was found that studies that checked whether feedback was used as intended had smaller effect sizes than studies in which there was no check on feedback use by clinicians.

In addition, several moderators of feedback were identified. In the full sample, four significant moderators of feedback were found: outcome and feedback instruments, country in which the study had been conducted, and the year in which the study was published. Studies using the ORS as outcome instrument, the PCOMS feedback system, and were conducted in the US were found to have higher effect sizes. The effect of feedback also slightly reduced over time, with a 0.02 reduction of effect size per year. In the NOT subsample, feedback type was found to moderate feedback effects, with studies using CSTs being particularly effective. Effects on dropout were moderated by feedback instrument (studies using PCOMS reporting higher effect sizes) and country (studies conducted in the US reporting higher effect sizes). Additionally, in the deteriation subsample, training (studies providing training reporting higher effect sizes) and type of feedback (studies using ETRs reporting higher effect sizes) moderated feedback effects.

7.1. Comparison with previous meta-analysis

Regarding the estimated feedback effects, the results of this meta-analysis are only partially in line with our expectations and previous meta-analyses. The overall effect of feedback in the full sample is smaller than was reported by Lambert et al. (2003) and Østergård et al. (2018), but larger than the effect that was found by Knaup et al. (2009) and similar to the effect size reported by Lambert et al. (2018). The main differences between our study and previous meta-analyses are the broader inclusion criteria, having used unpublished data directly

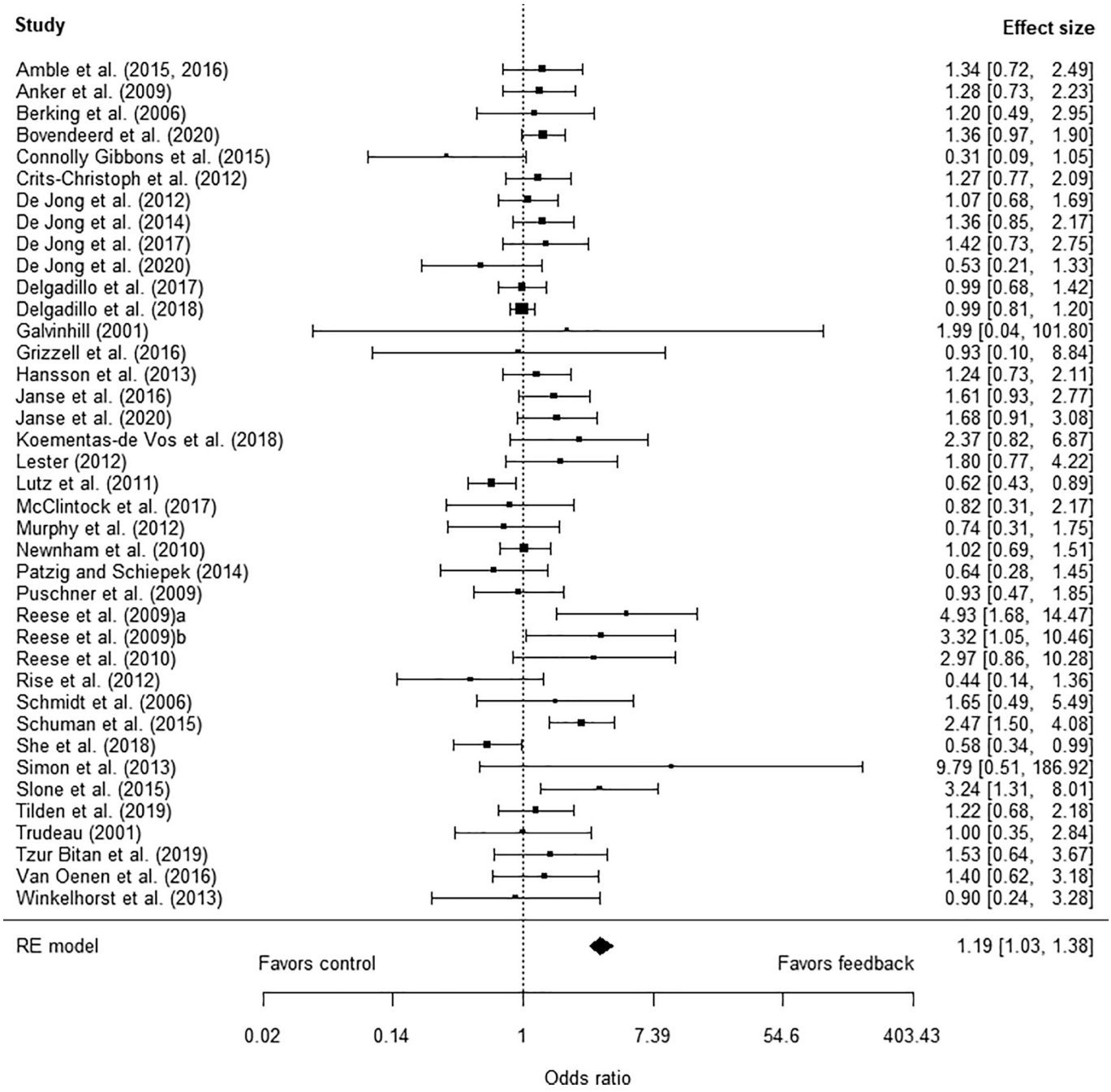

Fig. 4. Effect of feedback on dropout.
Note. This forest plot shows the back-transformed odds ratios. The analysis was conducted on the logOR.

obtained from researchers, and using a three-level multilevel approach to account for dependencies in the effect size data. Most previous meta-analyses only assessed studies that used PCOMS or the OQ System as a feedback tool, whereas we have also included studies using other feedback instruments. We have also included a few very small studies (e.g., Grizzell, Smart, Lambert, & Fargo, 2016; Hansen, Howe, Sutton, & Ronan, 2015; Lester, 2012) that were not always included in other meta-analyses, which may have introduced a larger variance between studies. With the exception of Østergård et al. (2018), previous meta-analyses have also not included non-randomized trials. Although our inclusion criteria may have resulted in a smaller overall effect of feedback than has been found by others, we do feel that it is important to include all the available evidence in a meta-analysis. At the same time, the increase of heterogeneity that results may make it more difficult to find significant moderators of the feedback effect on outcome.

Contrary to our expectations, we did not find a much stronger effect of feedback in the NOT cases than in the full sample, as has been reported by three other meta-analyses (Kendrick et al., 2016; Lambert et al., 2018; Shimokawa et al., 2010) and would be expected based on feedback theory. Only 27 out of 58 studies reported on separate outcomes for NOT cases. Not surprisingly, studies that reported on NOT cases were significantly more likely to use the OQ-45 or ORS as an outcome instrument, as well as providing clinical support tools and expected recovery curves as part of the feedback than studies not reporting on NOT cases. Given that the OQ-45 and ORS are the most commonly used feedback instruments, they likely have the largest databases on which ETRs can be calculated, and determining whether a case is NOT is often based on comparing ETRs to actual change.

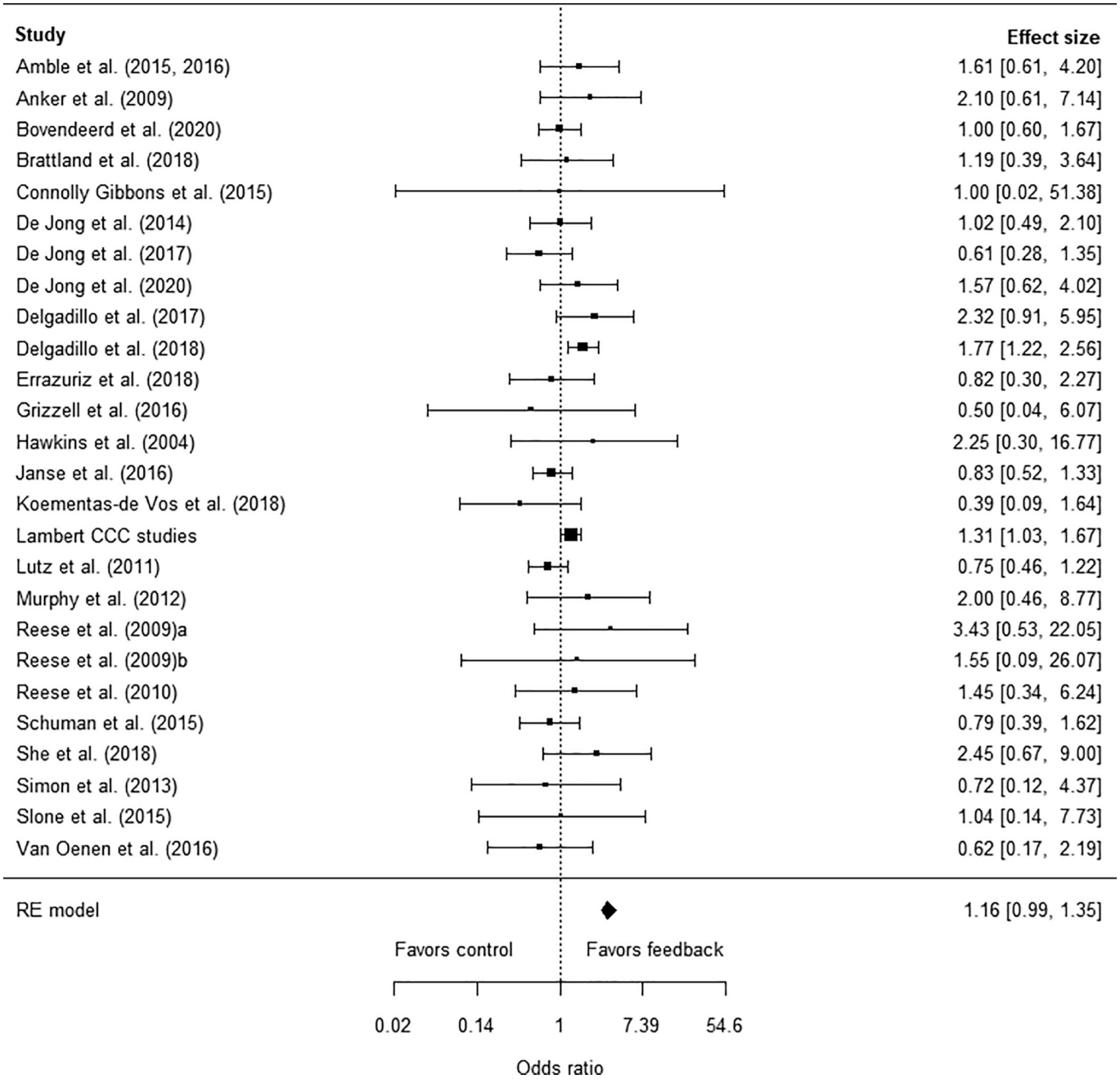

Fig. 5. The effect of feedback on the rate of deteriorated cases.

Replications of the analyses of two previous meta-analyses (Lambert et al., 2018; Østergård et al., 2018) further revealed that our analytic method could not explain the lack of a bigger difference in effect between the full sample and the NOT subgroup. It appears that not all feedback systems are equally effective in all patients. Studies using PCOMS have larger effect sizes in the full sample, but have negligible effect sizes in the NOT subgroup. For this meta-analysis, Duncan and Reese conducted new analyses on the NOT subgroup for six studies using PCOMS. Consequently, the number of studies with NOT cases using PCOMS is higher in our meta-analysis than in previous ones, which may dilute the effect in NOT cases. The OQ System seems more effective in NOT cases, especially when it is used in combination with CSTs, but seems to be doing less well in the full sample. Thus, PCOMS seems to be more effective in OT cases, whereas the OQ System works better in NOT cases. This is in line with how these feedback systems have been designed. The OQ System aims to give feedback signals for patients that did not progress well and strives to improve treatment outcomes for these patients (Lambert, 2007), whereas PCOMS has been constructed to be completed and discussed in session, thereby promoting better communication between patient and therapist.

The effect of feedback on dropout found in this meta-analysis is interesting, as the effect has not been reported on by most studies. Although the effect is small ($OR = 1.20$), it is similar in size as other known predictors of dropout, such as age, race, marital status and employment (Swift & Greenberg, 2012). What complicates the interpretation of the results to some degree is that some of the studies failed to differentiate between study dropout and treatment dropout, and reported on a combination of the two. Data on dropout was usually obtained from the authors, as the majority of the articles did not report on dropout per condition. If a combination of study dropout and treatment

Table 3
Moderation analyses for secondary outcomes.

	Dropout			Number of sessions			% Deteriorated		
Moderator	*k*	*OR*	*95% CI*	*k*	*d*	*95% CI*	*k*	*OR*	*95% CI*
Study characteristics									
Outcome instrument									
- OQ-45	14	1.23	0.94, 1.62	8	0.04	−0.16, 0.24	11	1.05	0.76, 1.44
- ORS	8	1.38	0.96, 1.97	4	−0.07	−0.39, 0.25	7	1.48	0.85, 2.58
- Other	17	1.11	0.89, 1.38	16	0.09	−0.07, 0.25	8	1.13	0.79, 1.61
Treatment duration									
- Fixed (preset)	9	1.35	0.99, 1.86	4	−0.14	−0.48, 0.21	5	0.86	0.51, 1.44
- Flexible	30	1.15	0.97, 1.36	25	0.06	−0.05, 0.18	21	1.20	0.97, 1.48
Country									
- USA	13	**1.77**	1.31, 2.41	15	−0.05	−0.22, 0.11	10	1.20	0.80, 1.78
- Other	26	**1.07**	0.92, 1.25	14	0.12	−0.02, 0.26	16	1.12	0.88, 1.42
Treatment form									
- Individual/couple	26	1.10	0.92, 1.32	23	0.02	−0.10, 0.15	20	1.22	0.98, 1.52
- Group	5	1.77	1.16, 2.69	4	0.14	−0.19, 0.48	4	0.69	0.34, 1.38
- Mixed	8	1.21	0.87, 1.68	2	0.08	−0.34, 0.51	2	0.92	0.45, 1.87
Treatment setting									
- Outpatient	31	1.22	1.02, 1.45	25	0.04	−0.08, 0.17	24	1.18	0.97, 1.44
- Inpatient	8	1.09	0.76, 1.57	4	0.05	−0.26, 0.36	2	0.63	0.27, 1.45
Age group									
- Adults	38	1.19	1.01, 1.38	24	0.03	−0.09, 0.15	26	1.14	0.93, 1.39
- Youth	1	1.80	0.63, 5.16	5	0.14	−0.15, 0.44	0	-	-
Treatment intensity									
- Mild	15	1.30	1.01, 1.68	13	0.07	−0.11, 0.26	13	1.35	1.08, 1.69
- Moderate	15	1.09	0.84, 1.43	13	0.01	−0.15, 0.18	10	0.89	0.64, 1.24
- Severe	9	1.19	0.85, 1.65	3	0.07	−0.28, 0.42	3	0.84	0.46, 1.55
Outcome assessment									
- Independent measure	16	1.27	1.00, 1.60	10	0.15	−0.07, 0.38	5	0.91	0.61, 1.36
- Feedback instrument	23	1.14	0.94, 1.39	19	0.01	−0.12, 0.14	21	1.22	0.98, 1.52
Participation therapists									
- Mandatory	18	1.27	1.00, 1.62	13	0.03	−0.15, 0.21	11	1.07	0.80, 1.43
- Voluntary	21	1.14	0.94, 1.38	16	0.05	−0.09, 0.20	15	1.21	0.90, 1.62
Feedback characteristics									
Feedback instrument									
- OQ-45	13	**1.21**	0.94, 1.55	8	0.04	−0.16, 0.24	9	1.09	0.77, 1.56
- PCOMS	15	**1.48**	1.17, 1.84	14	0.04	−0.14, 0.23	12	1.10	0.77, 1.55
- Other	11	**0.92**	0.74, 1.15	7	0.04	−0.18, 0.26	5	1.26	0.84, 1.90
Feedback type									
- Raw score	15	1.12	0.87, 1.43	11	0.07	−0.11, 0.26	6	0.81	0.58, 1.13
- ERC	19	1.30	1.04, 1.62	15	0.05	−0.11, 0.21	16	1.36	1.04, 1.78
- CST	5	0.98	0.57, 1.72	3	−0.08	−0.42, 0.26	4	1.29	0.88, 1.88
Timing									
- Delayed	15	1.16	0.90, 1.49	10	0.03	−0.14, 0.21	7	0.91	0.64, 1.28
- Immediate	24	1.22	1.00 1.48	19	0.05	−0.10, 0.20	19	1.29	1.00, 1.68
Frequency									
- Intermittent	6	0.9	0.65, 1.25	2	−0.09	−0.45, 0.28	1	0.75	0.40, 1.41
- Continuous	33	1.27	1.08, 1.48	27	0.06	−0.06, 0.18	25	1.19	0.98, 1.45
Feedback recipient									
- Clinician	21	1.07	0.90, 1.30	19	0.01	−0.12, 0.15	13	1.15	0.88, 1.51
- Clinician and/or Patient[a]	18	1.43	1.10, 1.84	9	0.17	−0.04, 0.38	12	1.02	0.67, 1.56
Training									
- Yes	28	1.21	1.01, 1.45	21	0.05	−0.09, 0.19	18	1.28	1.05, 1.57
- No	11	1.15	0.84, 1.55	8	0.03	−0.17, 0.23	8	0.81	0.57, 1.16

Note. Bold values indicate significant differences between groups. *k* = number of studies. *OR* = odds ratio. ERC = Expected Recovery Curves. CST = Clinical Support Tool.

[a] Lambert's CCC studies could not be coded because two studies provide feedback to patients and three do not.

dropout was reported, the effect of feedback on dropout might partially be explained by the fact that being in the active condition (feedback) is sometimes associated with lower dropout rates in RCTs than being in the control condition. Thus, results should be interpreted with care, but if they are confirmed by future studies, this could be a very important effect of feedback for psychotherapy.

Additionally, in the current meta-analysis a small effect of feedback was found (*OR* = 1.19) on the percentage of deteriorated cases, suggesting that feedback may reduce negative treatment outcomes at the end of treatment. The effect found was somewhat smaller than the effect reported by Shimokawa et al. (2010) and Lambert et al. (2018), but is in the same direction.

7.2. Moderator effects

In the full sample, the feedback and outcome instrument were significant moderators of the effect of feedback on symptom reduction. In the NOT sample, feedback type was found to be a significant moderator. Overall, using PCOMS and ORS results in larger effect sizes than other feedback systems and other outcome instruments in the full group. This seems to be partially caused by the ORS yielding larger effects than other outcome instruments, which is consistent with findings by Østergård et al., 2018. This could be caused by a higher sensitivity to change, but could also be the result of bias, for instance because the instrument is often completed by the patient in the presence of the therapist. However, one could also argue that scoring the questionnaire in the presence of the

therapist may make it more likely to be used directly in treatment, having positive effect on the implementation rate, which is known to influence the effectiveness of feedback (e.g. Bickman et al., 2016; De Jong et al., 2012). Studies using PCOMS as the feedback instrument were also found to have lower percentages of dropout, potentially because completing the feedback forms in the room may create more of a collaborative care approach.

The results also suggest that certain types of feedback systems may work better for specific groups of patients. Østergård et al. also reported that PCOMS was more effective in milder (counseling) settings, compared to more severe (psychiatric) settings. This is indirectly supported by the fact that PCOMS was the more effective feedback instrument in the full group, but that in the NOT subsample, which consists of more severe cases, the CST feedback seems to be doing better. It makes sense that simply providing a feedback signal is enough in milder cases, but that more information may be needed in more severe cases. The CST offers concrete suggestions for alternative strategies, based on the assessment of the treatment process, which may help the therapist adjust their treatment more specifically.

The negative temporal trend is a commonly found outcome in new interventions (e.g., Johnsen & Friborg, 2015), and it looks like feedback is no exception. Additionally, outcome measurement is often mandatory these days (Joint Commission, 2011), causing some newer studies to have control conditions in which a light version of outcome monitoring takes place, which may reduce the effect size of the progress feedback condition (Delgadillo et al., 2017; Delgadillo et al., 2018; Janse, De Jong, Van Dijk, Hutschemaekers, & Verbraak, 2017).

For some of the outcome variables, an effect of country was found, where studies conducted in the US found larger effects of feedback in symptom reduction and dropout. We hypothesize that this difference in effectiveness is largely due to differences in care systems. For instance, mental health care has a higher co-pay for clients in the US, than for those in the other countries in which feedback was studied (predominantly North and Western Europe and Australia), which may result in difference in access to mental health care. In addition, the training system for therapists is widely different across countries and may affect the results. Finally, culture and language may influence how questionnaires are being scored, and subsequently affect effect sizes.

7.3. Limitations and suggestions for further research

Although the main strength of this meta-analysis is that it is the most inclusive so far, as a result the sample is also more heterogeneous, making it potentially more difficult to find (moderator) effects. As such, our meta-analysis might be underpowered to find small differences between studies, especially in the moderation analyses. Because of a potential lack of power to find an effect of moderator variables, we did not correct for multiple testing. The downside of this is that the chance of false positive findings is substantial, given the large number of moderation analyses that were conducted. Another issue with the moderators is that study and feedback characteristics were often clustered within studies and difficult to disentangle (e.g., PCOMS is mainly studied in mild to moderate populations; only studies using the OQ System use clinical support tools). Furthermore, the moderation analysis was necessarily limited to moderators which had been investigated in a sufficient number of previous studies. Thus, our analysis did not include moderators with previously reported significant effects in only one or few studies, such as therapists' active use of feedback (De Jong et al., 2012), degree of implementation (Bickman et al., 2016), therapist expectations (De Jong et al., 2020), therapists' self-efficacy (De Jong et al., 2012), and regulatory focus (De Jong & De Goede, 2015). Future studies should report on a wider range of potential moderators, so that meta-analyses can analyze them more systematically in the future.

Additionally, we tested for many moderator effects, including explorative effects, without controlling for Type I error. Consistent with other meta-analyses, we did not control for Type I error because this would have further reduced the limited power that moderator tests in meta-analysis generally have. Particularly in our case, Type I error control would result in a too conservative testing because the different moderator effects we tested for were not dependent (e.g. the NOT data is a subset of the full data, and different moderators were related). We suggest that our results – especially the explorative moderators – are further explored in future studies.

An additional limitation of this meta-analysis is that by including cohort studies as well as RCTs, the methodological rigor of the included studies might be reduced, which may influence the results. However, the number of cohort studies was small, and excluding cohort studies had a negligible effect on the overall effect size. Therefore, adding cohort studies as additional sources of information seems justified when aiming to achieve a comprehensive review of the literature in this area.

Finally, we found that the quality of studies was sometimes found to be low and should be improved in future studies. Many studies do not use separate outcome instruments, which is problematic because patients and clinicians may discuss scores in session, which may result in bias on the outcome instrument. We also found that there are very few studies that check whether the feedback is actually used as intended, and several studies do not mention whether therapists were trained in using the feedback, both of which are especially important since previous studies have shown that feedback is not always properly used by clinicians (De Jong et al., 2012; Simon, Lambert, Harris, Busath, & Vazquez, 2012). Furthermore, potential allegiance issues were also found to be fairly common. All of these potential sources of bias moderated one or more outcome variables in our meta-analysis. Future studies should focus on reducing these potential sources of bias by taking them into account in their study design, for instance by including adversarial collaborators (e.g., from competing feedback systems) in the research team to avoid allegiance issues.

8. Conclusion

The results of the current meta-analysis imply there is a small effect of progress feedback on symptom reduction, both in the full sample and in the NOT subsample. The effect size varies substantially between studies, ranging from negligible to large effects. The fail-safe N suggests that the effect of feedback is relatively robust. Furthermore, feedback seems to reduce dropout rates. While these small effects may not seem meaningful, it should be considered that feedback is a relatively small and simple intervention within the full context of psychotherapy, and can be viewed as an add-on intervention for enhancing treatment outcomes in routine practice. Implementing feedback systems in routine care is relatively affordable, and has been found to be cost-effective (Delgadillo et al., 2017).

Based on significance or estimated effect sizes, the results of this meta-analysis suggest that the feedback instrument, type of feedback, feedback frequency, and treatment intensity are feedback characteristics that might be worth investigating further in future research. The outcome instrument used to evaluate feedback effectiveness, as well as the country in which the study is conducted might be context variables that could influence study results. Future studies should report on these and other relevant variables more consistently, in order to obtain a better understanding of when and why feedback improves outcomes. Finally, most studies so far have primarily focused on assessing the effectiveness of different feedback systems, and only a few studies have investigated the mechanisms of feedback in the context of psychotherapy. Studies on how progress feedback works are much needed in order to achieve a better understanding of how it can be used in the most optimal way to enhance outcomes for clients in routine practice.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors would like to thank Marc Molendijk for his feedback on the first draft of the manuscript. We would also like to thank Morten Anker, Leonard Bickman, Heidi Brattland, Annika Davidsen, Erik de Groot, Jaime Delgadillo, Barry Duncan, Paul Crits-Christoph, Mary Beth Gibbons, Saara Grizzell, Helena Hansson, Pauline Janse, Marjolein Koementas-de Vos, Michael Lambert, Wolfgang Lutz, Andrew McClintock, Kevin Murphy, Jesse Owen, Andrew Page, Judith Patzig, Thomas Probst, Jeff Reese, Julian Rubel, Henning Schöttke, Norah Slone, Terje Tilden, Ladislav Timulak, Linda Trudeau, Dana Tzur Bitan, Flip Jan van Oenen, and Sigal Zilcha Mano for providing us with additional data on their studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2021.102002.

References

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. https://doi.org/10.1080/10503307.2014.928756.

Amble, I., Gude, T., Ulvenes, P., Stubdal, S., & Wampold, B. E. (2016). How and when feedback works in psychotherapy: Is it the signal? *Psychotherapy Research, 26*(5), 545–555. https://doi.org/10.1080/10503307.2015.1053552.

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*(4), 693. https://doi.org/10.1037/a0016062.

Bergman, H., Kornør, H., Nikolakopoulou, A., Hanssen-Bauer, K., Soares-Weiser, K., Tollefsen, T. K., & Bjørndal, A. (2018). Client feedback in psychological therapy for children and adolescents with mental health problems. *Cochrane Database of Systematic Reviews, 8*. https://doi.org/10.1002/14651858.CD011729.pub2.

Berking, M., Orth, U., & Lutz, W. (2006). How effective is systematic feedback of treatment progress to the therapist? An empirical study in a cognitive-behavioural-oriented inpatient setting. *Zeitschrift fur Klinische Psychologie und Psychotherapie: Forschung und Praxis, 35*, 21–29. https://doi.org/10.1026/1616-3443.35.1.21.

Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(10), 1114–1119. https://doi.org/10.1097/CHI.0b013e3181825af8.

Bickman, L., Douglas Kelley, S., Breda, C., De Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on youth mental health outcomes: A randomized cluster design. *Psychiatric Services, 62*, 1423–1429. https://doi.org/10.1176/appi.ps.002052011.

Bickman, L., Douglas, S. R., De Andrade, A. R. V., Tomlinson, M., Gleacher, A., Olin, S., & Hoagwood, K. (2016). Implementing a measurement feedback system: A tale of two sites. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 410–425. https://doi.org/10.1007/s10488-015-0647-8.

Bovendeerd, B., De Jong, K., De Groot, E., Moerbeek, M., & De Keijser, J. (2020). *The effect of systematic client feedback: a cluster randomized trial in basic mental health care. (in preparation)*.

Brattland, H., Koksvik, J. M., Burkeland, O., Gråwe, R. W., Klöckner, C., Linaker, O. M., … Iversen, V. C. (2018). The effects of routine outcome monitoring (ROM) on therapy outcomes in the course of an implementation process: A randomized clinical trial. *Journal of Counseling Psychology, 65*(5), 641–652. https://doi.org/10.1037/cou0000286.

Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*(5), 581.

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J. A., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: Evidence and theory. *Journal of Evaluation in Clinical Practice, 18*, 104–110. https://doi.org/10.1111/j.1365-2753.2010.01543.x.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211. https://doi.org/10.1037/a0032968.

Connolly Gibbons, M. B., Kurtz, J. E., Thompson, D. L., Mack, R. A., Lee, J. K., Rothbard, A., … Crits-Christoph, P. (2015). The effectiveness of clinician feedback in the treatment of depression in the community mental health system. *Journal of Consulting and Clinical Psychology, 83*(4), 748–759. https://doi.org/10.1037/a0039302.

Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J. L., Lambert, M. J., Gallop, R., McClure, B., … Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of Substance Abuse Treatment, 42*(3), 301–309. https://doi.org/10.1016/j.jsat.2011.09.003.

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. https://doi.org/10.1037/ccp0000173.

Davidson, K., Perry, A., & Bell, L. (2014). 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: Theory, Research and Practice, 88*(1), 21–37. https://doi.org/10.1111/papt.12032.

De Jong, K., & De Goede, M. (2015). Why do some therapists not deal with outcome monitoring feedback? A feasibility study on the effect of regulatory focus and person–organization fit on attitude and outcome. *Psychotherapy Research, 25*(6), 661–668. https://doi.org/10.1080/10503307.2015.1076198.

De Jong, K., Nugter, M. A., La Croix, K., De Groot, E., Berns, P., & Spinhoven, P. (2020). *What works for whom in feedback? A randomized controlled trial comparing two types of feedback in relation to patient complexity. (in preparation)*.

De Jong, K., Segaar, J., Ingenhoven, T., van Busschbach, J., & Timman, R. (2017). Adverse effects of outcome monitoring feedback in patients with personality disorders: A randomized controlled trial in day treatment and inpatient settings. *Journal of Personality Disorders*, 1–21. https://doi.org/10.1521/pedi_2017_31_297.

De Jong, K., Timman, R., Hakkaart-Van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., & Busschbach, J. V. (2014). The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychotherapy Research, 24*(6), 629–639. https://doi.org/10.1080/10503307.2013.871079.

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. https://doi.org/10.1080/10503307.2012.673023.

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., … McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5*(7), 564–572. https://doi.org/10.1016/S2215-0366(18)30162-7.

Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., … de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy, 99*(Supplement C), 89–97. https://doi.org/10.1016/j.brat.2017.09.011.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. Retrieved from https://doi.org/10.1016/0197-2456(86)90046-2.

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. https://doi.org/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. https://doi.org/10.1037/ccp0000277.

Galvinhill, P. R. (2001). *The effects of providing feedback regarding client symptomatology on psychotherapy outcome*. University of Cincinnati. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=ucin992361796.

Greene, W. (2004). Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics, 13*(10), 959–980. https://doi.org/10.1002/hec.938.

Grizzell, S., Smart, J., Lambert, M. J., & Fargo, J. (2016). The use of feedback in group counseling in a state vocational rehabilitation setting. *Journal of Applied Rehabilitation Counseling, 47*(4), 10–19.

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.

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), 617–619. https://doi.org/10.1016/j.psychres.2015.05.007.

Hansson, H., Rundberg, J., Österling, A., Öjehagen, A., & Berglund, M. (2013). Intervention with feedback using outcome questionnaire 45 (OQ-45) in a Swedish psychiatric outpatient population. A randomized controlled trial. *Nordic Journal of Psychiatry, 67*(4), 274–281. https://doi.org/10.3109/08039488.2012.736534.

Harmon, C. S., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J. L. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *Journal of Clinical Psychology, 61*, 175–185. https://doi.org/10.1002/jclp.20109.

Harmon, C. S., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychotherapy Research, 17*(4), 379–392. https://doi.org/10.1080/10503300600702331.

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 17*(1), 25–32.

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 14*(3), 308–327. https://doi.org/10.1093/ptr/kph027.

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. C. (2020). Chapter 8: Assessing risk of bias in a randomized trial (editors). In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane*

Handbook for Systematic Reviews of Interventions, version 6.1 (updated September 2020). Cochrane. Available from www.training.cochrane.org/handbook.

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. https://doi.org/10.1037/0022-006x.59.1.12.

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. https://doi.org/10.1080/10503307.2016.1152408.

Janse, P. D., De Jong, K., Veerkamp, C., Van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2020). *The effect of feedback-informed cognitive behavioral therapies on treatment efficiency and dropout: A randomised controlled trial. (submitted)*.

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*(4), 747–768. https://doi.org/10.1037/bul0000015.

Joint Commission. (2011). *Comprehensive accreditation manual for behavioral health care*. Oakbrook Terrace, Illinois: Joint Commission Resources.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146–159. https://doi.org/10.1037/0003-066X.63.3.146.

Kellybrew-Miller, A. (2014). The impact of systematic client feedback on client outcomes in a community mental health center (Order No. 10010515). Available from ProQuest Dissertations & Theses Global. (1766579664). Retrieved from https://search.proquest.com/docview/1766579664?accountid=12045.

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*. https://doi.org/10.1002/14651858.CD011119.pub2. Art. No.: CD011119.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254.

Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *The British Journal of Psychiatry, 195*(1), 15–22. https://doi.org/10.1192/bjp.bp.108.053967.

Koementas-de Vos, M. M., Nugter, M. A., Engelsbel, F., & De Jong, K. (2018). Does progress feedback enhance the outcome of group psychotherapy? *Psychotherapy, 55*(2), 151.

Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in mental health services: A scoping study and typology. *Psychiatric Services, 66*(3), 224–241. https://doi.org/10.1176/appi.ps.201400141.

Kremer, C. M. (2018). Systematic evidence-based assessment and intervention of client feedback in youth mental health treatment in a private practice setting. Retrieved from https://scholarcommons.sc.edu/etd/4913.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17*(1), 1–14. https://doi.org/10.1080/10503300601032506.

Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., Shimokawa, K., … Reisinger, C. W. (2004). *Administration and scoring manual for the OQ-45.2 (outcome questionnaire)* (3 ed.). Wilmington, DE: American Professional Credentialing Services LLC.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288–301.

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520–537. https://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*(1), 49–68. https://doi.org/10.1080/713663852.

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. https://doi.org/10.1002/cpp.324.

Langkaas, T. F., Wampold, B. E., & Hoffart, A. (2018). Five types of clinical difference to monitor in practice. *Psychotherapy, 55*(3), 241–254. https://doi.org/10.1037/pst0000194.

Lester, M. C. (2012). *The effectiveness of client feedback measures with adolescents in an acute psychiatry inpatient setting*. PhD. Conway, AK: University of Central Arkansas (3522254).

Lutz, W., Boehnke, J. R., Köck, K., & Bitterman, A. (2011). Diagnostik und psychometrische Verlaufsrückmeldungen im Rahmen eines Modellprojektes zur Qualitätssicherung in der ambulanten Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie, 40*, 283–297. https://doi.org/10.1026/1616-3443/a000125.

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. https://doi.org/10.1080/10503307.2015.1079661.

McClintock, A. S., Perlman, M. R., McCarrick, S. M., Anderson, T., & Himawan, L. (2017). Enhancing psychotherapy process with common factors feedback: A randomized, clinical trial. *Journal of Counseling Psychology, 64*(3), 247.

McGuinness, L. A., & Higgins, J. P. T. (2020). Risk-of-bias VISualization (robvis): An R package and shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods.*. https://doi.org/10.1002/jrsm.1411.

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61*(2), 199–208. https://doi.org/10.1002/jclp.20111.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364–386. https://doi.org/10.1177/1094428106291059.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105–125. https://doi.org/10.1037/1082-989X.7.1.105.

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

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*(5), 1253–1274.

Newnham, E. A., Hooke, G. R., & Page, A. C. (2010). Progress monitoring and feedback in psychiatric care reduces depressive symptoms. *Journal of Affective Disorders, 127*(1–3), 139–146. https://doi.org/10.1016/j.jad.2010.05.003.

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*, 1–18. https://doi.org/10.1080/10503307.2018.1517949.

Owen, J. (2019). *Unpublished data. Received 13 October*.

Patzig, J., & Schiepek, G. (2015). Emotionsregulation und emotionsfokussiertes Prozessmonitoring in der Suchttherapie. In I. Sammet, G. Dammann, & G. Schiepek (Eds.), *Der Psycho-Therapeutische Prozess: Forschung für die Praxis* (pp. 125–142). Stuttgard: Kohlhammer.

Probst, T., Lambert, M. J., Dahlbender, R. W., Loew, T. H., & Tritt, K. (2014). Providing patient progress feedback and clinical support tools to therapists: Is the therapeutic process of patients on-track to recovery enhanced in psychosomatic in-patient therapy under the conditions of routine practice? *Journal of Psychosomatic Research, 76*(6), 477–484. https://doi.org/10.1016/j.jpsychores.2014.03.010.

Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75*(3), 255–261. https://doi.org/10.1016/j.jpsychores.2013.07.003.

Puschner, B., Schöfer, D., Knaup, C., & Becker, T. (2009). Outcome management in in-patient psychiatric care. *Acta Psychiatrica Scandinavica, 120*(4), 308–319. https://doi.org/10.1111/j.1600-0447.2009.01397.x.

Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training, 46*(4), 418–431. https://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: Theory, Research, Practice, Training, 47*(4), 616–630. https://doi.org/10.1037/a0021182.

Reeves, E. K. (2010). *The use of psychotherapy feedback on client progress: Implication for training and clinical practice*. PhD. New York, NY: New York University, 3390464.

Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation*. New York: Guilford Press.

Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2012). The short-term effect on alliance and satisfaction of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *BMC Health Services Research, 12*(1), 348. https://doi.org/10.1186/1472-6963-12-348.

Schmidt, U., Landau, S., Pombo-Carril, M. G., Bara-Carril, N., Reid, Y., Murray, K., … Katzman, M. (2006). Does personalized feedback improve the outcome of cognitive-behavioural guided self-care in bulimia nervosa? A preliminary randomized controlled trial. *British Journal of Clinical Psychology, 45*(1), 111–121. https://doi.org/10.1348/014466505X29143.

Schöttke, H., Unrath, M., & Uhlmann, C. (2019). The Effect of Patient Progress Feedback on Psychotherapy Outcome. *Verhaltenstherapie*, 1–11. https://doi.org/10.1159/000501176. Retrieved from https://www.karger.com.

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*(4), 396–407. https://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*(6), 727–737. https://doi.org/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*, 298–311. https://doi.org/10.1037/a0019247.

Simon, W., Lambert, M. J., Busath, G., Vazquez, A., Berkeljon, A., Hyer, K., … Berrett, M. (2013). Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomized controlled study. *Psychotherapy Research, 23*(3), 287–300. https://doi.org/10.1080/10503307.2013.787497.

Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on

patients at risk of treatment failure. *Psychotherapy Research, 22*(6), 638–647. https://doi.org/10.1080/10503307.2012.698918.
Slade, K., Lambert, M. J., Harmon, S., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology & Psychotherapy, 15*, 287–303. https://doi.org/10.1002/cpp.594.
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*(2), 122. https://doi.org/10.1037/gdn0000026.
Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*(10), 1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8.
Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., et al. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ, i4919*, 355. https://doi.org/10.1136/bmj.i4919.
Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., et al. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, l4898*, 366. https://doi.org/10.1136/bmj.l4898.
Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(4), 547.
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. https://doi.org/10.1016/j.cpr.2017.04.005.
Tilden, T., Wampold, B. E., Ulvenes, P., Zahl-Olsen, R., Hoffart, A., Barstad, B., … Håland, Å. T. (2019). Feedback in couple and family therapy: A randomized clinical trial. *Family Process*. https://doi.org/10.1111/famp.12485. ePub ahead of print.
Trudeau, L. S. (2000). *Effects of a clinical feedback system on client and therapist outcomes in a rural community mental health center*. PhD Dissertation. Iowa State University. Retrieved from https://lib.dr.iastate.edu/rtd/12367.
Tzur Bitan, D., Kivity, Y., Ganor, O., Biran, L., Grossman-Giron, A., & Bloch, Y. (2019). The effect of process and outcome feedback in highly distressed outpatients: A randomized controlled trial. *Psychotherapy Research*, 1–12. ePub ahead of print https://doi.org/10.1080/10503307.2019.1627014.
Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6.
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*(1), 110. https://doi.org/10.1186/s12888-016-0811-z.
Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.
Viechtbauer, W. (2017). Metafor package: bias and sensitivity diagnostics. Retrieved from https://stats.stackexchange.com/questions/155693/metafor-package-bias-and-sensitivity-diagnostics.
Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*(2), 639–644.
Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges. *Psychotherapy, 52*(4), 458–462. https://doi.org/10.1037/pst0000037.
Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59–68. https://doi.org/10.1037/0022-0167.50.1.59.
Winkelhorst, Y., Hafkenscheid, A., & De Groot, E. (2013). Verhoogt routine process monitoring (RPM) de effectiviteit van behandeling? *Een vergelijkende pilotstudy. Tijdschrift voor Psychotherapie, 39*(3), 146–156.

Kim de Jong is an Assistant Professor at the Institute of Psychology, Leiden University, the Netherlands. She is a leading expert in progress feedback.

Judith M. Conijn works as a post-doctoral research fellow at the Faculty of Behavioral and Movement Sciences at the Vrije Universiteit in Amsterdam. Before this position, she has worked at the University of Amsterdam and Leiden University. She has a PhD in Psychological Methods.

Roisin A.V. Gallagher and **Marya Heij** both have an MSc from Leiden University and are currently working as clinicians.

Alexandra S. Reshetnikova is a former bachelor student in psychology at Leiden University and is currently doing a master in international studies on media, power, and difference at Pompeu Fabra University in Barcelona.

Miranda C. Lutz is a former master student at Leiden University and is currently a PhD student at the Department of Psychology, Education & Child Studies at Erasmus University in Rotterdam.