

Journal of Consulting and Clinical Psychology

The Effect of Feedback-Informed Cognitive Behavioral Therapy on Treatment Outcome: A Randomized Controlled Trial

Pauline D. Janse, Kim de Jong, Carola Veerkamp, Maarten K. van Dijk, Giel J. M. Hutschemaekers, and Marc J. P. M. Verbraak

Online First Publication, July 13, 2020. <http://dx.doi.org/10.1037/ccp0000549>

CITATION

Janse, P. D., de Jong, K., Veerkamp, C., van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2020, July 13). The Effect of Feedback-Informed Cognitive Behavioral Therapy on Treatment Outcome: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*. Advance online publication. <http://dx.doi.org/10.1037/ccp0000549>

The Effect of Feedback-Informed Cognitive Behavioral Therapy on Treatment Outcome: A Randomized Controlled Trial

Pauline D. Janse

Pro Persona Research, Wolfheze, the Netherlands

Kim de Jong

Leiden University

Carola Veerkamp

HSK Groep, Arnhem, the Netherlands

Maarten K. van Dijk

Dimence, Deventer, the Netherlands

Giel J. M. Hutschemaekers and Marc J. P. M. Verbraak

Pro Persona Research, Wolfheze, the Netherlands, and Radboud University

Objective: Previous studies have shown that feedback-informed treatment can improve outcomes of psychological treatments. This randomized controlled effectiveness trial evaluated the effect of progress feedback on treatment duration, symptom reduction, and dropout in individual cognitive behavioral therapies (CBTs). A control condition where CBT was combined with low-intensive monitoring of progress was compared to an experimental condition where CBT was combined with a high-intensive form of feedback. **Method:** Data of 368 outpatients (57.9% female, mean age 41.4 years, $SD = 12.2$) in secondary care were analyzed using multilevel analyses. Treatment duration was assessed with the number of sessions clients received. Symptom reduction was measured with the Symptom Checklist Revised. Possible moderators of the effect of intensive progress feedback on outcome were explored. **Results:** Clients achieved the same amount of symptom reduction in significantly fewer sessions in the high-intensive feedback condition. Additionally, dropout was significantly lower in the high-intensive feedback condition. Post hoc analyses assessing clients' diagnoses as a possible moderator showed that clients with personality disorders (mainly Cluster C) achieved more symptom reduction in fewer sessions when high-intensity feedback was provided. Also, a high degree of implementation within the experimental condition was associated with fewer treatment sessions. **Conclusion:** In sum, the use of high-intensive client feedback reduced treatment duration and reduced dropout of CBT. Thus, feedback-informed CBTs seem to be a promising adaptation of conventional CBT.

What is the public health significance of this article?

This study suggests that frequently providing progress feedback can reduce treatment duration of individual cognitive behavioral therapies.

Keywords: feedback-informed treatment, outcome monitoring, cognitive behavioral therapies

Supplemental materials: <http://dx.doi.org/10.1037/ccp0000549.sup>

Cognitive behavioral therapies (CBTs) are effective for the treatment of various psychological disorders, particularly anxiety disorders (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012).

However, randomized controlled trials comparing CBT to treatment as usual or a placebo yielded only small to moderate effect sizes (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016).

✉ Pauline D. Janse, Pro Persona Research, Wolfheze, the Netherlands;
✉ Kim de Jong, Institute of Psychology, Leiden University; Carola Veerkamp, HSK Groep, Arnhem, the Netherlands; ✉ Maarten K. van Dijk, Dimence, Deventer, the Netherlands; ✉ Giel J. M. Hutschemaekers and ✉ Marc J. P. M. Verbraak, Pro Persona Research and Behavioural Science Institute, Radboud University.

The data reported in this article will be combined with data from a previous study to investigate the influence of therapist characteristics (including self-efficacy and personality traits of therapists) on the effect of

feedback-informed cognitive behavioral therapy on treatment outcome. This manuscript will be submitted soon. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare no conflict of interest. We thank the HSK Groep and the therapists and clients who participated in this study. We would also like to thank Jaime Delgadillo for his feedback.

Correspondence concerning this article should be addressed to Pauline D. Janse, Pro Persona Research, Postbus 27, 6870 AA, Wolfheze, the Netherlands. E-mail: p.janse@propersona.nl

Also, a meta-analysis shows that more than a quarter of clients dropped out of treatment (Fernandez, Salem, Swift, & Ramtahal, 2015) and thus did not receive the care they might need. Furthermore, the efficiency of CBT shown in randomized controlled trials was not always achieved in routine care, where CBT lasts significantly longer (Lutz, Schiefele, Wucherpfennig, Rubel, & Stulz, 2016). One possible explanation for the “inefficiency” of routine care treatments is that therapists may not be delivering CBT with fidelity to evidence-based treatment protocols, also known as “therapist drift” (Waller & Turner, 2016). At the same time, in many countries, health care systems have been under pressure to reduce costs (Stuckler, Reeves, Loopstra, Karanikolos, & McKee, 2017). Thus, there is pressure not only to provide effective treatments but also to do so efficiently in as little time as possible. Providing effective and efficient treatments is therefore important not only for clients and therapists but also on a larger scale for services and governments.

One possible way to improve the effectiveness and efficiency of CBT could be through the use of routine outcome monitoring and feedback on client progress (Waller & Turner, 2016). Feedback-informed treatment (FIT) is a method that can be utilized across a variety of treatment modalities to evaluate with the client whether therapy is helping them and, if not, to discuss what is necessary to make it work. FIT entails using standardized measures to monitor treatment outcomes such as symptom reduction or well-being, sometimes combined with monitoring the therapeutic relationship. Based on promising results of initial studies on the effect of feedback systems, the American Psychological Association recommends using feedback and monitoring treatment outcomes (Kazdin, 2008; Norcross & Wampold, 2011). The two most commonly used feedback systems, the OQ Analyst (Lambert et al., 2013) and the Partners for Change Outcome Management System (PCOMS; Miller & Duncan, 2004), have even been granted the status of evidence-based interventions by the Substance Abuse and Mental Health Services Administration in the United States. Feedback systems have been implemented within mental health organizations worldwide, especially in countries in which routine outcome monitoring has become mandatory for service providers (e.g., the Netherlands, the United Kingdom, Australia).

Meta-analyses suggest that FIT can improve symptom reduction or well-being (Lambert, Whipple, & Kleinstäuber, 2018; Østergård, Randa, & Hougaard, 2018; Shimokawa, Lambert, & Smart, 2010), especially for clients who are not progressing well—so-called “not on track” (NOT) clients (Carlier et al., 2012; Delgadillo et al., 2018; Gondek, Edbrooke-Childs, Fink, Deighton, & Wolpert, 2016; Kendrick et al., 2016). Moreover, FIT has been found to improve treatment efficiency by attaining similar levels of symptom reduction as for clients in a control group, but in fewer sessions (e.g., Delgadillo et al., 2017; Janse, De Jong, Van Dijk, Hutschemaekers, & Verbraak, 2017; Koementas-de Vos, Nugter, Engelsbel, & De Jong, 2018; Reese, Norsworthy, & Rowlands, 2009). In addition, a recent meta-analysis demonstrated that FIT also significantly reduced treatment dropout (De Jong, Conijn, Lutz, Gallagher, & Heij, 2020).

Despite these encouraging results, some studies show modest effects, with FIT only having a small effect on symptom reduction for NOT cases (Kendrick et al., 2016), as well as inconsistent findings in regard to treatment efficiency (e.g., Gondek et al., 2016) and dropout (e.g., Delgadillo et al., 2018). Various feedback

studies have methodological limitations (Gondek et al., 2016; Kendrick et al., 2016), such as a lack of independent outcome measures, and it stands to reason that variability in studies in aspects such as actual treatment methods used and type and intensity of feedback makes it difficult to form general conclusions about the effects of FIT.

Many studies have investigated feedback applied in the context of “eclectic” and “integrative” therapy approaches (e.g., De Jong et al., 2014), some of which may be more or less effective, making it difficult to assess the added effectiveness of FIT. Combining feedback with a homogeneous treatment, such as a manualized CBT, might prove to be a solution. In CBT, the self-monitoring of symptoms, behaviors, or thoughts is standard practice (see, e.g., Kazantzis, Dattilio, Cummins, & Clayton, 2013; King & Boswell, 2019), which in theory makes it very suitable to combine with a feedback system. Although the number of studies specifically focusing on the additional value of feedback for CBT is still limited, studies on feedback-informed cognitive behavioral therapies (FI-CBT) have shown promising results. For instance, a randomized controlled trial conducted in a setting that provided CBT by Delgadillo et al. (2018), in which 2,233 clients participated, showed that NOT clients had less severe symptoms of depression or anxiety after treatment if they were allocated to the FIT group. Also, two quasi-experimental studies showed that the addition of FIT to CBT improved treatment efficiency; in the study by Janse et al. (2017) in which 1,006 clients participated, the same treatment outcome was achieved in fewer sessions (mean difference was two sessions), and these results were replicated in a study by Delgadillo et al. (2017) in which 594 clients participated and providing FIT resulted in not only fewer sessions (almost four sessions) but also a significant reduction in treatment costs.

Few moderators of feedback effects have been found so far (De Jong et al., 2020). One potential moderator is diagnosis. For instance, Janse et al. (2017) found that feedback had an effect on depressed clients but not on clients with other diagnoses. Delgadillo et al. (2018) found that feedback did have an effect on clients with depression and anxiety, but only for those that were classed as NOT. Mixed results have been found for clients with eating disorders, showing both positive and null effects of feedback (Davidson et al., 2017; Simon et al., 2013), and inpatients and day patients with Cluster B personality disorders even experienced an initial negative effect of feedback (de Jong, Segaaar, Ingenhoven, van Busschbach, & Timman, 2018). Another factor that possibly influences the effect of FIT is the intensity of feedback. Studies have different designs with respect to feedback intensity: some provide feedback only once, others at three points in time during treatment, and others at every session. Additionally, several studies have struggled with implementation issues, with patients not consistently completing the measures, as well as therapists not utilizing available FIT systems in their practice (e.g., de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; Simon, Lambert, Harris, Busath, & Vazquez, 2012). Indeed, Bickman and colleagues (2016) have found that the extent of implementation significantly moderated the effectiveness of FIT.

In summary, because there is still room for improvement in the effectiveness of CBT, an intensive feedback system might be a useful addition to routine practice. However, more research is needed to assess the value of FI-CBT. The present study compared the results of CBT combined with low-intensity outcome monitor-

ing versus CBT combined with a high-intensity feedback system. This study is a follow-up to a previous quasi-experimental study carried out by the same research group, in which intensive feedback had a positive effect on treatment duration and efficiency (Janse et al., 2017). The design of this previous study limited the generalizability of the results. The added value of the present study is that it is a randomized controlled trial, with an independent measure of treatment outcome and in which therapists provided manualized CBT. Also, the high-intensity feedback system not only provided feedback on the client's well-being but also provided feedback on the quality of the therapeutic relationship. The main research question of the present study was this: Does an intensive form of client feedback improve treatment duration and/or effectiveness of standardized CBT in comparison to CBT with a less intensive form of feedback? We hypothesized that FI-CBT will reduce treatment duration compared to the CBT control group with low-frequency feedback. Additionally, we expected a lower dropout rate in the intensive feedback condition compared to the control condition. Furthermore, post hoc analyses were performed to explore whether the clients' diagnoses moderated the effect of FIT on outcomes and whether the degree of implementation of feedback influenced treatment duration and effectiveness.

Method

Design

This effectiveness study was a parallel-group, two-arm, superiority trial (trial registration doi.org/10.1186/ISRCTN14466125). The primary outcome measures were the number of sessions clients received and the reduction of symptoms as measured with the Symptom Checklist Revised (SCL-90-R; Derogatis, 1994); the secondary measure was the percentage of dropout from treatment. The ethical review board of Radboud University approved this study (ECSW2017-1303-49).

Participants

Recruitment and inclusion of clients took place between September 2013 and April 2015. The trial ended in February 2017.

Clients were treated at one of the six participating treatment locations of an outpatient mental health care institution. Clients who were acutely suicidal or psychotic or had severe substance abuse were excluded from treatment and referred elsewhere. A total of 395 clients were approached to participate. Figure 1 shows a CONSORT diagram describing the flow of trial participants. Clients who required specialized secondary care were included; those who did not want to participate or only needed general primary care were excluded from the study ($N = 27$). Data from 368 clients were used for the intention to treat (ITT) analyses. The mean age of participants was 41.4 ($SD = 12.2$), ranging between the age of 18 and 71. Overall, more women participated in the study than men (57.9% vs. 42.1%). In the subsequent per protocol (PP) analyses (clients receiving the intended treatment in both conditions), PP was operationalized as clients having completed a minimum of three measurements on the Outcome Rating Scale (ORS) and Sessions Rating Scale (SRS; the measures used in the high-intensive feedback condition) between the start and completion of the treatment (i.e., dropout from treatment and insufficient use of feedback was excluded). The mean age of clients included in the PP analyses was 42.9 years ($SD = 11.9$), and 56.8% of these clients were female.

At intake, all participants were diagnosed by therapists using the semistructured clinical interview for *DSM-IV* disorders, the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). Also, clients were asked to complete the Assessment for *DSM-IV* Personality Disorders (ADP-IV; Schotte & de Doncker, 1994). The most prevalent diagnoses were somatoform disorder ($N = 148$, 40.2%), followed by mood disorders ($N = 111$, 30.2%) and anxiety disorders ($N = 81$, 22.0%). Based on the outcome on the ADP-IV in combination with the clinical interview at intake, a little over one third of the clients were diagnosed with comorbid personality disorders (34.5%, $N = 127$). Of the clients with personality disorders, Cluster C traits (avoidant, dependent, obsessive-compulsive traits) were the most prevalent ($N = 93$, 73.2%), followed by a mix of traits (not otherwise specified, $N = 20$, 15.8%) and Cluster B traits (more specifically, borderline traits, $N = 15$, 11.8%).

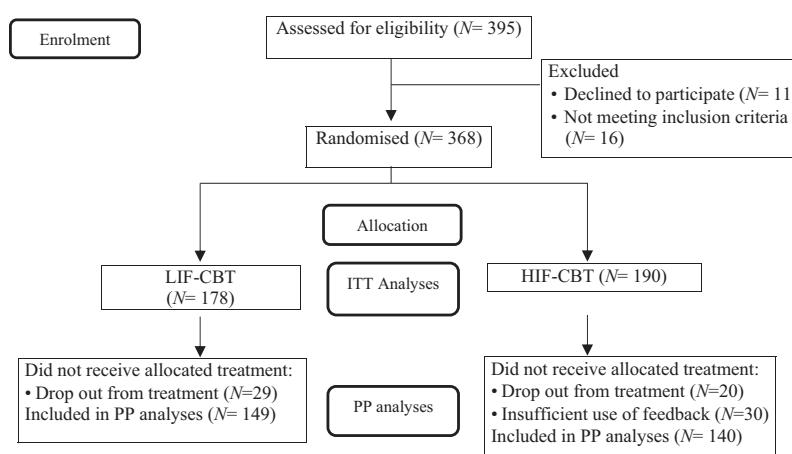


Figure 1. Flow chart of patient inclusion in the study. LIF-CBT = low intensity feedback cognitive behavioral therapies; HIF-CBT = high intensity feedback cognitive behavioral therapies.

On average, clients had a high severity level of symptoms at intake compared to norms of the normal Dutch population and an average level of symptoms compared to an outpatient psychiatric norm group (Arrindell & Ettema, 2003), with a mean Global Severity Index (GSI) of 1.01 ($SD = .49$) on the SCL-90. The average number of sessions they received was 15 sessions ($SD = 9.2$, ranging between 3 and 49 sessions), and the average length of treatment was 33.3 weeks ($SD = 21.5$). Table 1 provides further information on characteristics of clients in both conditions in the total sample.

Therapists

A total of 66 therapists participated in this study. The mean number of clients per therapist in this study was 5.6 ($SD = 4.3$, median = 5). Most therapists were female ($N = 45$, 83.3%). The mean age of therapists was 32.6 ($SD = 6.8$), ranging from 24 to 58 years. They had between 1 and 13.5 years of experience as a therapist, with a mean of 5.5 years ($SD = 2.6$). All therapists had a master's degree in psychology and had received basic training in CBT and the use of specific treatment protocols (Keijsers, van Minnen, Verbraak, Hoogduin, & Emmelkamp, 2017) and training in the use of the MINI for making diagnoses when they started to work at this organization. They received supervision twice a week during the first 9 months of their appointment. Thereafter, therapists received advanced training and supervision (a total of 50 hr of training and 50 hr of supervision) in CBT and became registered CBT therapists. Of the total number of therapists, 26 therapists (39.4%) were junior psychologists who had received basic training in CBT, 18 therapists (27.3%) had received advanced training in CBT, and 30 therapists (55.6%) were considered senior psychologists.

Intervention

Low-intensity feedback CBT (LIF-CBT). The control condition consisted of manualized CBT combined with a low-

frequency monitoring of symptoms. Therapists asked clients to fill in the SCL-90 at intake, every fifth session, and at the end of treatment. This is part of treatment as usual within this mental health care institution. Therapists had access to a brief report of the scores and the interpretation of the scores on the SCL-90 in the electronic patient file and were instructed to discuss the results with the client. Specific treatment protocols, based on CBT techniques, were used by therapists during treatment. A previous study showed that with regard to the provision of CBT, treatment integrity within the same organization in which the present study was performed was good, with 80% of the interventions from the treatment manuals being applied on average (Dijkgraaf-Hartland, Brals, Verbraak, & Van Dijk, 2013). Also, as a quality control system, therapists were required to discuss their caseload, specifically clients who had 10 or more sessions, every 6–8 weeks with two colleagues to check whether CBT was provided as intended, if the treatment was effective, and, if not, what changes needed to be made to the treatment plan. Although treatments were not restricted to a maximum number of sessions restriction to the maximum number of sessions, therapists were encouraged by management, in supervision, and caseload discussions to provide not only effective treatments but also efficient treatments. Performance indicators such as average effect sizes of individual therapists (based on the SCL-90) as well as average treatment duration were accessible to everyone and discussed in the yearly performance interviews. In the event of a poor performance, extra supervision on delivering CBT would be provided. This quality control system was also used in the high-intensity feedback CBT.

High-intensity feedback CBT (HIF-CBT). As an add-on to LIF-CBT, the PCOMS (Duncan, 2012) was used to generate feedback in the experimental condition. It consists of two measures: the ORS and SRS (Miller & Duncan, 2004). All therapists who participated in this study were trained in the use of PCOMS and the web-based program FIT-Outcomes (2013). Most therapists (62.1%) already had experience using PCOMS because they had participated in the previous study from our research group (Janse

Table 1
Clients' Characteristics in the LIF-CBT and HIF-CBT Conditions

Sample characteristic	Condition			
	LIF-CBT		HIF-CBT	
	N	%	N	%
Sex				
Male	81	45.5	74	38.9
Female	97	54.5	116	61.1
Age	$M = 42.6$, $SD = 12.3$		$M = 40.4$, $SD = 11.9$	
Diagnoses Axis I				
Somatoform disorders	73	41.0	75	39.5
Mood disorders	55	30.9	56	29.5
Anxiety disorders	37	20.8	44	23.2
Other	13	7.3	14	7.4
Comorbid diagnoses Axis I	63	35.4	64	33.7
Diagnoses Axis II				
Cluster C traits/disorder	53	28.8	40	21.1
Cluster B traits/disorder	9	5.1	6	3.1
Not otherwise specified	7	3.9	13	6.8
GSI at intake	$M = 1.05$, $SD = 0.48$		$M = 0.99$, $SD = 0.49$	

Note. LIF-CBT = low-intensity feedback cognitive behavioral therapies; HIF-CBT = high-intensity feedback cognitive behavioral therapies; GSI = Global Severity Index.

et al., 2017). Therapists received several follow-up group supervisions on the use of PCOMS. Therapists were instructed to use FIT-Outcomes on a session-by-session basis with every new client in the HIF-CBT condition and to evaluate with their clients whether enough progress was made during treatment. A client was considered NOT if they did not have an improvement of 5 points on the ORS within the first five sessions or did not meet the expected recovery curve—that is, the scores on the ORS were below the 50th percentile compared to the scores of other clients in the FIT-Outcomes database. Therapists were instructed to discuss with their client why they were NOT and what needed to change. The cutoff score on the SRS that was used was 34 points, based on the study of Janse, Boezen-Hilberdink, Van Dijk, Verbraak, and Hutschemaekers (2014). Therapists were instructed to discuss with their clients how to improve the therapeutic alliance if the score dropped below the cutoff point.

Measurements

Mini International Neuropsychiatric Interview (MINI). The MINI-Plus 5.0.0.-R (Sheehan et al., 1998) was used for classifying Axis-I disorders according to the *DSM-IV*. The reliability and validity have been reported to be sufficient (Lecrubier et al., 1997). The Dutch translation of the MINI-Plus 5.0.0-R was used in an electronic version (Van Vliet, Leroy, & Van Megen, 2000). The MINI is part of the standard intake procedure on all six research locations. The MINI has good psychometric properties, with test-retest reliability ranging from 0.76–0.93, interrater reliability ranging from $k = .88$ –1.00, and sufficient validity (Lecrubier et al., 1997).

Assessment for DSM-IV Personality Disorders (ADP-IV). The ADP-IV (Schotte & de Doncker, 1994) is a self-report questionnaire and is used for the screening for diagnoses of *DSM-IV* Axis-II disorders. It consists of 94 items and measures personality traits and the degree of distress clients experience from this trait. Clients give a score on a scale from 1 (*totally disagree*) to 7 (*totally agree*) on the trait items and then rate the amount of distress it causes on a scale from 1 (*totally not*) to 3 (*most certainly*). The inclusion of amount of distress mitigates the risk of overdiagnosing Axis-II disorders based upon trait items alone; the ADP-IV leads to the same number of diagnoses as when using the Structured Clinical Interview for *DSM-IV* Axis-II Disorders (Schotte, de Doncker, & Courjaret, 2007). The ADP-IV has proven to be a valid and reliable measure for screening for personality disorders: the internal validity of the categorical scales ranges from $\alpha = .64$ to .88 (Schotte & de Doncker, 2000).

Symptom Checklist Revised (SCL-90-R). The SCL-90-R (Derogatis, 1994) measures various symptoms of psychopathology and is a widely used questionnaire on symptom severity and treatment outcome. It consists of 90 items that are scored on a 5-point Likert scale ranging from 1 to 5. The Dutch SCL-90-R has eight subscales: anxiety, agoraphobia, depression, somatization, insufficiency, hostility, paranoia and insomnia, and a residual category of miscellaneous items. The GSI, which reflects general psychological and psychosomatic well-being, is calculated by the sum of the scores of all the subscales divided by the number of items. The psychometric properties of the Dutch SCL-90-R are adequate (Arrindell & Ettema, 2003), with good internal consistency of the subscales ($\alpha = .97$) and test-retest reliability ($r = .82$).

In the present study, the internal consistency (α) of subscales on the SCL-90 ranged from .62 to .90. The internal consistency of the total scale was $\alpha = .96$.

Outcome Rating Scale and Session Rating Scale (ORS and SRS). The ORS and SRS (Miller & Duncan, 2004) are visual analogue scales and both consist of four items each. The scales of each item are 10 cm long. On each item, the client makes a mark on the scale, and the score on the item is derived by measuring to the nearest millimeter. The total scores are calculated by summing the scores on the four items. The ORS measures functioning and well-being individually (personal well-being), interpersonally (family, close relationships), and socially (work, school, friendships) and is an overall assessment of daily functioning. The SRS measures the quality of the therapeutic alliance: Clients rate the relationship between them and their therapist, consensus on goals/topics, and approach or method and give a global assessment of the treatment session. The psychometric properties of the English versions are adequate (Duncan et al., 2003; Miller, Duncan, Brown, Sparks, & Claud, 2003), as are those of the Dutch versions (Janse et al., 2014). The internal consistency of the ORS and SRS at the first session were respectively $\alpha = .85$ and $\alpha = .84$ in the present study.

FIT-Outcomes. FIT-Outcomes (2013) is a web-based outcome management system for measuring the ORS and SRS. The application produces graphs with expected treatment response (ETR)—that is, trajectories of change, which serve as a reference for the therapist to compare the clients' progress. The trajectories of the ORS are based on a sample of 427,744 administrations of the ORS, provided by 2,354 different clinicians. The algorithm uses initial scores of the client on the ORS to generate the ETR. The cutoff score of the SRS is 36 in the English version (Miller & Duncan, 2004); we used the cutoff of 34 for the Dutch version (Janse et al., 2014), meaning that scores below this cutoff point are considered inadequate and should be discussed with the client.

The number of sessions clients received during treatment was determined by including face-to-face sessions only (e-mail and telephone contact excluded). Dropout was defined by clients indicating they did not want to continue treatment, even though according to the therapist the treatment had not been completed, and clients not showing up for treatment sessions anymore.

Procedure

Clients were referred for treatment by their general physician or company physician. At intake, the clients' psychological histories were taken, the MINI (Sheehan et al., 1998) was administered, and clients completed the ADP-IV (Schotte & de Doncker, 1994) and SCL-90 (Derogatis, 1994). The diagnosis and treatment plan were then discussed with the client. After intake, clients who thereafter were to receive treatment were asked to participate in the study. They received written information on the study and were asked to give consent for the use of their data for research purposes. After giving consent, clients were randomly assigned to either the LIF-CBT or the HIF-CBT treatment condition using Excel 2010 (Microsoft, Redmond, WA). The method of randomization was a simple randomization procedure. A research assistant assisted with the allocation and informing therapists. Therapists were then informed about which treatment condition their client was assigned to. Clients were asked to complete the SCL-90 at intake, every

fifth session, and at the end of treatment in both treatment conditions, and therapists were provided the results. Both conditions completed the SCL-90 on the same schedule. The ORS and SRS were used to monitor the progress on a session-to-session basis in the HIF-CBT treatment condition only.

Statistical Analyses

Pretreatment differences between the conditions were examined with *t* tests and chi-square tests. To test our hypotheses, two main analyses were performed: ITT analyses, in which all allocated clients after randomization were included, and PP analyses. Post hoc analyses investigated if there were differences in the effect of feedback between diagnostic categories (with diagnosis as a moderator) and whether a high or low degree of implementation (based on the implementation index) within the HIF-CBT condition was related to the effect feedback might have.

The main analyses were performed using multilevel modeling, which takes into account that measurements are nested within clients and clients are nested within therapists (see, e.g., Hox, 2010). Concerning the hypothesis about treatment duration, two-level models were used (the number of sessions on the level of the client, clients nested within therapists). The hypothesis of the effect of high-intensive feedback on symptom reduction over time was tested with repeated measures on the GSI on the first level, clients on the second level, and therapists on the third level. The full maximum likelihood method was used with an unstructured variance-covariance matrix to estimate random effects. The following steps were taken within the multilevel analyses: First, an unconditional model was tested to investigate the amount of variance in outcome (sessions or symptom reduction) explained by either clients or therapists by calculating the initial interclass correlation. Second, it was tested whether random effects on both the client and therapist level were a better fit than only including random effects on either the client or therapist level by looking at the Akaike information criterion (AIC; with a smaller number indicating a better model). Third, the final model was determined to test whether feedback had an effect on outcome.

The secondary research questions concerning diagnoses as a moderator were tested by creating dummy variables for the most prevalent diagnoses (depression, anxiety, somatoform disorders, other disorders, and comorbid personality disorders), which were then incorporated as moderators in the multilevel analyses as described above. The exploration of the effect of the degree of implementation was investigated by comparing two extremes in implementation index (low or high implementation, based on the quartiles of the distribution of frequencies of the implementation index) in treatment outcome within the HIF-CBT condition, using multilevel analyses. The implementation index Bickman et al. (2016) devised consisted of the rate of completion of the questionnaire by respondents and the clinicians' feedback viewing rate. Because the ORS and SRS were filled in during the session with the therapist present, only the number of completed ORS and SRS scores divided by the number of total sessions the client received were used to calculate the implementation index.

To determine the effect size of treatment outcome, Cohen's *d* (Cohen, 1988) was calculated. The pooled standard deviation was adjusted with weights for the sample sizes because the groups had different sample sizes. To calculate the within-subjects effect size,

Morris and DeShon's (2002) equation was used, which takes the correlation between the pre- and posttest into account.

The reliable change index (RCI) was calculated to determine the amount of change a client must show on the GSI between measurement occasions for that change to be reliable—that is, larger than that reasonably expected due to measurement error alone. The RCI of the GSI was calculated based on the formula proposed by Christensen and Mendoza (1986), multiplying *s_{diff}* (standard error of the *t*₁–*t*₂ difference) by the *z* value of the requisite significance level ($1.96, p < .05$). This resulted in an RCI of .28. The cutoff score of the GSI was calculated based on Criterion C from Jacobson and Truax (1991), using the means and standard deviations of clients in the present study at intake and the means and standard deviations of a nonclinical population (Arrindell & Ettema, 2003). This resulted in a cutoff score of .65. Based on the RCI and the cutoff score for the GSI, four outcome classifications were distinguished: "clinical significant change" (a reliable improvement and a GSI score at the end of treatment below the cutoff score), "reliable change" (a reliable improvement but not a GSI score at the end of treatment below the cutoff score), "no change" (no reliable change), or "deterioration" (a reliable increase in symptoms at the end of treatment).

Results

Intention to Treat Analyses

Preliminary analyses. Pretreatment differences between clients in both conditions were examined. The clients in the HIF-CBT and LIF-CBT conditions did not differ significantly in age, $t(366) = 1.77, p = .078, 95\% \text{ CI} [-.25, 4.70]$. Although there were relatively more women in the HIF-CBT group, the difference was not significant, $\chi^2(1, 368) = 1.62, p = .203$. Also, the treatment conditions did not significantly differ in symptoms at intake as measured by the GSI, $t(361) = 1.14, p = .256$.

Number of sessions. In the LIF-CBT condition, clients received an average of 16.3 sessions ($SD = 10.16$), and in the HIF-CBT condition, 14.4 sessions ($SD = 8.0$). In the two-level unconditional model, between-groups differences were statistically significant ($b = 7.30, z = 2.05, p = .040, 95\% \text{ CI} [0.32, 14.27]$), with 8.77% of the variance in the number of sessions due to therapist effects (i.e., attributable to differences between therapists, irrespective of the random allocation). High-intensity feedback had a significant effect on the number of sessions in the two-level analysis (see Table 2), with an effect size of $d = .22$. This was also the case when clients who had dropped out of therapy were excluded ($b = -2.68, t = -2.78, p = .006, \text{CI} [-4.58, -.79], d = .28$).

Symptom reduction. The mean number of improvements on the GSI in the LIF-CBT condition was .61 points ($SD = .49$), and in the HIF-CBT condition, .62 points ($SD = .47$). The three-level unconditional model showed that 0.03% of symptom reduction over time was due to therapist effects, and 24.99% of the variance was due to differences between clients (e.g., case-mix features). There were no significant differences between therapists in the intercept (symptom severity at intake; $b = .00, z = -.02, p = .986, 95\% \text{ CI} [-.01, .01]$). Clients did differ in symptom severity at intake ($b = .07, z = 5.14, p < .01, \text{CI} [.04, .10]$). The slope variance between clients (differences in rate of change) was also

Table 2
Multilevel Model of the Effect of High-Intensity Feedback on the Number of Sessions

Variable	b	SE	95% CI
Fixed effects			
Intercept	16.35**	0.77	[14.82, 17.87]
HIF-CBT	-2.15*	0.94	[-4.00, -0.30]
Random effects			
Intercept	7.68*	3.62	[0.59, 14.77]
Residual	74.47**	5.89	[64.15, 87.52]

Note. HIF-CBT = high-intensity feedback cognitive behavioral therapies; SE = standard error.

* $p < .05$. ** $p < .01$.

significant ($b = -.09$, $z = -2.31$, $p = .021$, CI $[-.17, -.01]$), whereas variance between therapists in rate of change was not ($b = -.02$, $z = -1.21$, $p = .228$, CI $[-.04, .01]$). The model with random intercepts and slopes on the client level only had the best goodness of fit ($AIC = 955.5$); therefore, in the subsequent analysis of the effect of feedback, random intercepts and slopes on the therapist level were not included. The model of the effect of high-intensity feedback on symptom reduction over time shows that there was no significant difference compared to the control group (see Table 3). The between-groups effect size was $d = .01$.

Outcome classifications. Based on the RCI and the cut-off score for the GSI, clients were grouped into outcome classifications. There were no significant between-groups differences in outcome in terms of percentages of clinically significant or reliable change and no change or deterioration, $\chi^2(3, N = 274) = .97$, $p = .807$, $V = .06$.

Dropout. In the HIF-CBT group, there was a significantly lower dropout rate (10.6%) compared to the LIF-CBT group (16.6%), $\chi^2(2, 364) = 6.90$, $p = .032$, $V = .14$. On average, dropout occurred after 8.6 sessions ($SD = 5.5$) in the LIF-CBT condition and after 8.9 sessions ($SD = 7.2$) in the HIF-CBT condition. This difference was not significant, $t(47) = -.15$, $p = .879$, 95% CI $[-3.94, 3.38]$.

Per Protocol Analyses

The same pattern of results and effect sizes were found in the PP analyses as in the ITT analyses. For further details of the results of the PP analyses, we refer the reader to the online supplemental materials.

Exploratory Post Hoc Analyses

Diagnosis as moderator. We explored whether diagnostic category as a moderator would influence the effect that intensive feedback might have on treatment outcomes. Table 4 provides information on the number of sessions and symptom reduction on the GSI per diagnostic category and treatment condition. Clients with comorbid personality disorders received significantly fewer sessions when high-intensity feedback was used, $b = 4.75$, $t(359) = -2.47$, $p = .014$, 95% CI $[-8.52, -.97]$, $d = .49$. No other significant differences were found in the effect of high-intensity feedback between diagnostic categories concerning the number of sessions.

Furthermore, high-intensity feedback had a positive effect on symptom reduction with clients with comorbid personality disorders compared to other disorders, $b = 0.55$, $t(347) = -2.34$, $p = .020$, 95% CI $[-1.00, -.09]$. The decrease in symptom reduction with clients with comorbid personality disorders occurred sooner in the high-intensity feedback condition, $b = -0.09$, $z = -2.36$, $p = .018$, CI $[-.17, -.02]$. The between-groups effect size of symptom reduction (i.e., GSI improvement) was $d = .30$. There were no other significant differences between diagnostic categories in terms of symptom reduction.

Implementation within the HIF-CBT condition. The mean implementation index (the number of completed ORS and SRS scores divided by the number of total sessions the client received) was 68.86% ($SD = 29.41$). There were no significant differences between the low and high implementation categories in symptom severity at intake, $t(68) = -.40$, $p = .692$, 95% CI $[-.30, .20]$. The unconditional two-level model showed that there were significant differences between therapists in the degree of implementation, $b = 218.92$, $z = 2.16$, $p = .031$, CI $[20.02, 417.18]$. Of the variance in the implementation index, 25.43% was explained by therapist effects. Clients with a low implementation index (the lowest 25%) had an average of 16.78 sessions ($SD = 8.99$). Those with an average implementation index of 15.67 ($SD = 7.7$) and those with a high implementation index (the highest 25%) had an average of 12.82 ($SD = 6.2$). In a two-level analysis with implementation category as a predictor, with the low implementation category as reference group, the difference in sessions between high and low implementation was significant, $b = -4.39$, $t(73) = -2.40$, $p = .019$, CI $[-8.03, -.74]$, $d = .51$. Although more symptom reduction was achieved in the high implementation group ($M = .72$, $SD = .55$) compared to low implementation ($M = .62$, $SD = .45$), these differences were not statistically significant, $b = -.23$, $t(68) = -.90$, $p = .371$, CI $[-.74, .28]$.

Discussion

The aim of this randomized controlled trial was to investigate whether using an intensive form of feedback using the PCOMS system improves the treatment duration and/or effectiveness of structured CBT. This effectiveness study compared two forms of feedback: basic feedback, which only monitors symptom reduction

Table 3
Multilevel Model of the Effect of High-Intensity Feedback on Symptom Reduction

Variable	b	SE	95% CI
Fixed effects			
Intercept	1.08**	.04	[1.01, 1.15]
Time	-1.23**	.08	[-1.39, -1.07]
HIF-CBT	-0.08	.05	[-0.18, 0.03]
HIF-CBT \times Time	0.02	.11	[-0.20, 0.24]
Random effects			
Intercept	0.14**	.02	[0.10, 0.18]
Slope	-0.09*	.04	[-0.16, -0.01]
Covariance	0.07	.12	[-0.16, 0.31]
Residual	0.10**	.01	[0.09, 0.13]

Note. HIF-CBT = high-intensity feedback cognitive behavioral therapies; SE = standard error.

* $p < .05$. ** $p < .01$.

Table 4
Amount of Sessions, GSI Scores, and Within-Subject Effect Sizes Per Diagnosis

Diagnosis	Sessions <i>M (SD)</i>	GSI intake <i>M (SD)</i>	GSI fifth <i>M (SD)</i>	GSI end <i>M (SD)</i>	Change <i>M (SD)</i>	ES
Mood (<i>N</i> = 114)						
LIF-CBT	17.7 (11.7)	1.25 (.46)	1.05 (.47)	.55 (.56)	.68 (.66)	1.26
HIF-CBT	13.8 (7.3)	1.04 (.51)	0.81 (.53)	.50 (.50)	.53 (.48)	1.02
Anxiety (<i>N</i> = 81)						
LIF-CBT	16.4 (11.6)	0.97 (.49)	0.81 (.42)	.45 (.39)	.49 (.39)	1.09
HIF-CBT	16.2 (9.7)	1.01 (.51)	0.85 (.50)	.36 (.28)	.69 (.43)	1.57
Somatoform (<i>N</i> = 148)						
LIF-CBT	15.9 (7.7)	0.99 (.45)	0.76 (.47)	.31 (.23)	.65 (.40)	1.84
HIF-CBT	14.0 (7.1)	0.97 (.49)	0.65 (.41)	.35 (.34)	.67 (.49)	1.16
Other (<i>N</i> = 27)						
LIF-CBT	13.0 (11.3)	0.70 (.43)	0.39 (.27)	.25 (.27)	.40 (.28)	0.99
HIF-CBT	13.4 (9.6)	0.87 (.42)	0.80 (.27)	.58 (.51)	.36 (.36)	1.74
Comorbid personality (<i>N</i> = 127)						
LIF-CBT	20.2 (11.7)	1.07 (.48)	0.92 (.50)	.54 (.51)	.49 (.56)	0.95
HIF-CBT	15.0 (9.5)	1.14 (.49)	0.81 (.42)	.53 (.52)	.66 (.56)	1.15

Note. GSI = Global Severity Index; ES = within-subject effect size pre-post GSI, corrected for dependence between means of GSI intake and GSI end; LIF-CBT = low-intensity feedback cognitive behavioral therapies; HIF-CBT = high-intensity feedback cognitive behavioral therapies.

every five sessions, and an intensive feedback system, which provided feedback not only on a session-by-session basis on the clients' well-being but also on the quality of the therapeutic relationship. Our first hypothesis was that treatment duration would be reduced by using intensive feedback. The results showed that, overall, clients received fewer sessions when high-intensive feedback was provided, with an average of 14 sessions in the HIF-CBT condition versus 16 sessions in the LIF-CBT condition. The difference was significant in both the ITT and PP analyses; however, the effect sizes were small ($\sim d = .22$ to $.29$). Furthermore, no significant differences were found between the HIF-CBT and LIF-CBT groups in symptom reduction, as measured on the GSI of the SCL-90. This is in line with a meta-analysis on the effect of PCOMS, which showed no effect on the clients' symptoms in psychiatric settings (Østergård et al., 2018), although a recent study did show a significant beneficial effect of PCOMS on treatment outcome of brief therapy (Bovendeerd, 2018). In the present study, high-intensity feedback possibly did not have an effect on symptom reduction because the within-group effect sizes were already large for the control condition (ranging from $d = 0.95$ to 1.84 for the different diagnoses), and as a basic form of feedback was provided in both conditions, the additional effect of the intensive feedback system is harder to measure. Nonetheless, the results suggest that, overall, the same amount of symptom reduction could be achieved with fewer sessions and therefore support the hypothesis that treatment duration can be reduced and that efficiency can be improved when using client feedback continuously. Furthermore, our second hypothesis that dropout rates would be reduced by using intensive feedback was also supported.

Post hoc analyses were performed to explore whether certain diagnoses moderated the effect high-intensive feedback might have on treatment outcome. Contrary to the findings of the quasi-experimental study on PCOMS (Janse et al., 2017), clients with depression did not benefit significantly more from intensive feedback than those with other diagnoses in terms of number of sessions or symptom reduction. However, clients with personality disorders achieved more symptom reduction in fewer sessions

when high-intensity feedback was used. It is important to mention that in the present study, personality disorders consisted mainly of Cluster C personality disorders, such as avoidant, dependent, and obsessive-compulsive personality traits. Previous research has shown that patients with Cluster C personality disorders can benefit from feedback, while those with Cluster B disorders and those with personality disorders not otherwise specified experienced adverse effects when feedback was provided (de Jong et al., 2018). This is possibly due to patients with Cluster C traits being relatively more stable than those with Cluster B traits and experiencing fewer strong negative emotions when feedback indicates that they are not showing the expected positive change (de Jong et al., 2018). Also, for clients with Cluster C personality traits, who are often fearful of rejection or expressing themselves, using a feedback system during therapy might be a way to practice being assertive in a safe environment.

Furthermore, we explored whether the degree of implementation of feedback had an effect on treatment outcomes. The results indicate that treatments in which the implementation index was high were significantly shorter (fewer sessions) than those with a low implementation index, while achieving a similar amount of symptom reduction. It should be stressed that this was an exploratory research question in order to generate hypotheses for further research. As such, no definitive conclusions can be made concerning the required degree of implementation of feedback. What the results do show, however, is that it is important to take implementation into account when researching the effect of feedback. As Bickman et al. (2016) also stated, at the very least, future studies should report an implementation index (or something similar) providing insights into how often the measures were administered or how they might be viewed by the therapists (and/or clients). The next step could then be to investigate if there is a minimum degree of implementation for a positive effect of a feedback system, which not only could be helpful in designing future studies but also might sway clinicians to be mindful of how often measuring feedback is required. The results in the current study showed that there were significant differences between therapists in how inten-

sive feedback is used. Although our impressions of why these differences exist are only anecdotal, some therapists have remarked during the study that although they find the measures useful, it costs too much session time to use every session, or they think it's unnecessary to use them so frequently. Therefore, if future studies confirm the hypothesis that feedback is more effective when it is provided almost every session, it might help motivate clinicians to use these feedback systems as intended. However, more research is needed to pinpoint the tipping point when feedback is effective (which might differ per context). In the Bickman study, an effect was found at 34%, whereas in the current study, an effect on the number of sessions was found at 94% implementation, which is a much higher degree of implementation of feedback. Again, this difference might be due to a low frequent form of feedback already being provided in the control condition.

The analyses in the present study were performed using multi-level modeling, which can take into account possible differences between therapists. The amount of variance explained by therapists in number of sessions found in this study (8.77% and 5.85%) is in line with 5–8% of the variation in outcome due to therapist factors found by Baldwin and Imel (2013). It is, however, noteworthy that the variance explained by therapists in symptom reduction was very small in the present study. This could be due to therapists' adherence to specific treatment protocols based on CBT techniques. The adherence in this study has not been examined, but as stated in the "Method" section, a previous study in the same institution indicated a high level of adherence (Dijkgraaf-Hartland et al., 2013), which might result in fewer differences between therapists in how they provide treatments and, subsequently, their treatment outcomes.

One of the strengths of this study is that even though it is an effectiveness study—that is, in a "real-world" setting—the setting was a relatively controlled treatment context due to the extensive training of therapists in CBT and the use of treatment manuals and attention to adherence in providing CBT. Another strength of this study is the "culture of feedback" within the organization: Therapists were used to evaluating the results of treatments, there was transparency concerning therapists' effectiveness, and the majority of therapists who participated in this study had previously worked with PCOMS. Although an implementation index of 68.86% for a randomized controlled trial might seem modest, the implementation of progress feedback in routine clinical practice is much lower; a study by Jensen-Doss and colleagues (2018) showed that only 13.9% of clinicians use progress measures monthly.

A limitation of this study is a relatively small number of participants. Even though our main hypothesis of the effect of feedback was confirmed, when investigating diagnoses as moderators, it is possible that lack of power might make it less likely to find significant results. Therefore, conclusions of the moderation analyses should be made with some caution. However, the main results of this study are meaningful since they are in accord with earlier studies (Delgadillo et al., 2017; Janse et al., 2017).

As this was an effectiveness study, another limitation of this study is that even though the diagnoses were made based on a structured interview and questionnaires, the reliability of the diagnoses was not tested in this study. A final limitation of this study is that the treatment integrity has not been studied in depth. Therapists used PCOMS in the HIF-CBT condition; however, it was not systematically recorded how they used the feedback, what

they discussed with clients when using PCOMS, and how this influenced the use of CBT treatment protocols. An interesting follow-up question is thus whether therapists adhere better to or deviate from the treatment protocols based on the feedback from clients when insufficient progress is made, as well as what they do differently.

Thus, future research should consist of a larger sample size when investigating possible moderators such as diagnoses, should control for treatment integrity, and should investigate therapist factors when using high-intensive feedback. Much remains unknown on how feedback works and what therapists do with the feedback they receive. Qualitative studies thus far showed that feedback technology can enhance collaboration with the clients (Delgadillo et al., 2017) and that therapists react and learn from feedback in different ways (Brattland et al., 2018). To fully understand how feedback works requires more mixed-method studies—that is, qualitative combined with quantitative research.

In conclusion, this study showed that the same amount of symptom reduction was achieved in significantly fewer sessions and that dropout was reduced when high-intensive progress feedback was provided, thus reducing treatment duration and improving efficiency of CBT. Also, the level of implementation of feedback is a factor that can influence whether feedback has an effect or not. As such, the results of this study provide another step in understanding when and for whom feedback works. High-intensive feedback seems to be a promising way to improve the results of CBT.

References

- Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90, Handleiding bij een multidimensionale psychopathologie indicator* [SCL-90, Manual for a multidimensional indicator of psychopathology]. Lisse, the Netherlands: Swets & Zeitlinger.
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). Hoboken, NJ: Wiley.
- 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, 410–425. <http://dx.doi.org/10.1007/s10488-015-0647-8>
- Bovendeerd, B. (2018, June). *When does PCOMS work? Preliminary results of a PCOMS study in basic mental health care*. Poster session presented at the International Federation of Psychotherapie (IFP) Congres, Amsterdam.
- Brattland, H., Høiseth, J. R., Burkeland, O., Inderhaug, T. S., Binder, P. E., & Iversen, V. C. (2018). Learning from clients: A qualitative investigation of psychotherapists' reactions to negative verbal feedback. *Psychotherapy Research*, 28, 545–559. <http://dx.doi.org/10.1080/10503307.2016.1246768>
- 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. <http://dx.doi.org/10.1111/j.1365-2753.2010.01543.x>
- Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, 17, 305–308. [http://dx.doi.org/10.1016/S0005-7894\(86\)80060-0](http://dx.doi.org/10.1016/S0005-7894(86)80060-0)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, 15, 245–258. <http://dx.doi.org/10.1002/wps.20346>
- Davidsen, A. H., Poulsen, S., Lindschou, J., Winkel, P., Tróndarson, M. F., Waadgaard, M., & Lau, M. (2017). Feedback in group psychotherapy for eating disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 85, 484–494. <http://dx.doi.org/10.1037/ccp0000173>
- De Jong, K., Conijn, J. M., Lutz, M. C., Gallagher, R. A. V., & Heij, M. (2020). *Using progress feedback to improve outcomes and reduce drop-out: A multilevel meta-analysis*. Manuscript submitted for publication.
- de Jong, K., Seghaar, J., Ingenhoven, T., van Busschbach, J., & Timman, R. (2018). 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*, 32, 393–413. http://dx.doi.org/10.1521/pedi_2017_31_297
- De Jong, K., Timman, R., Hakkaart-Van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., & Van Busschbach, J. (2014). The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychotherapy Research*, 24, 629–639. <http://dx.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, 464–474. <http://dx.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, 564–572. [http://dx.doi.org/10.1016/S2215-0366\(18\)30162-7](http://dx.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 (Suppl. C), 89–97. <http://dx.doi.org/10.1016/j.brat.2017.09.011>
- Derogatis, L. R. (1994). *Symptom Checklist 90-R: Administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.
- Dijkgraaf-Hartland, W., Brals, C., Verbraak, M., & Van Dijk, M. (2013). Comorbide persoonlijkheidsproblematiek en de effectiviteit van protocollaire behandeling. [Co-morbid personality problems and the effectiveness of manualized treatment]. *Directieve Therapie*, 33, 14–30. Retrieved from <https://www.directievetherapie.nl/artikelen/jaargang33/comorbide-persoonlijkheidsproblematiek-en-de-effectiviteit-van-protocollaire-behandeling-33-1-4/>
- Duncan, B. (2012). The partners for change outcome management system (PCOMS): The heart and soul of change project. *Canadian Psychology/Psychologie canadienne*, 53, 93–104. <http://dx.doi.org/10.1037/a0027762>
- Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The session rating scale: Preliminary psychometric properties of a “working” alliance measure. *Journal of Brief Therapy*, 3, 3–12. Retrieved from <https://scottdmiller.com/wp-content/uploads/documents/SessionRatingScale-JBTv3n1.pdf>
- Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *Journal of Consulting and Clinical Psychology*, 83, 1108–1122. <http://dx.doi.org/10.1037/ccp0000044>
- Fit-Outcomes [Computer software]. (2013). Retrieved from <https://www.fit-outcomes.com/>
- Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43, 325–343. <http://dx.doi.org/10.1007/s10488-015-0710-5>
- Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 36, 427–440. <http://dx.doi.org/10.1007/s10608-012-9476-1>
- Hox, J. (2010). *Multilevel analysis* (2nd ed.). New York, NY: Routledge. <http://dx.doi.org/10.4324/9780203852279>
- 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, 12–19. <http://dx.doi.org/10.1037/0022-006X.59.1.12>
- Janse, P. D., Boezen-Hilberdink, L., Van Dijk, M. K., Verbraak, M. J. P. M., & Hutschemaekers, G. J. M. (2014). Measuring feedback from clients: The psychometric properties of the Dutch Outcome Rating Scale and Session Rating Scale. *European Journal of Psychological Assessment*, 30, 86–92. <http://dx.doi.org/10.1027/1015-5759/a000172>
- Janse, P. D., De Jong, K., Van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2017). Improving the efficiency of cognitive-behavioural therapy by using formal client feedback. *Psychotherapy Research*, 27, 525–538. <http://dx.doi.org/10.1080/10503307.2016.1152408>
- Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R., Lewis, C. C., Stanick, C. F., & Hawley, K. M. (2018). Monitoring Treatment Progress and Providing Feedback is Viewed Favorably but Rarely Used in Practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 45, 48–61. <http://dx.doi.org/10.1007/s10488-016-0763-0>
- Kazantzis, N., Dattilio, F. M., Cummins, A., & Clayton, X. (2013). Homework assignments and self-monitoring. In D. Dozois (Ed.), *The Wiley handbook of cognitive behavioral therapy* (Vol. 1, pp. 311–330). New York, NY: Wiley-Blackwell. <http://dx.doi.org/10.1002/9781118528563.wbcbt14>
- 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, 146–159. <http://dx.doi.org/10.1037/0003-066X.63.3.146>
- Keijzers, G., van Minnen, A., Verbraak, M., Hoogduin, K., & Emmelkamp, P. (2017). *Protocollaire behandelingen voor volwassenen met psychische klachten, deel 1–3* [Protocolized treatments for adults with psychological problems, part 1–3]. Amsterdam, the Netherlands: Boom.
- 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*, 2016, CD011119. <http://dx.doi.org/10.1002/14651858.CD011119.pub2>
- King, B. R., & Boswell, J. F. (2019). Therapeutic strategies and techniques in early cognitive-behavioral therapy. *Psychotherapy*, 56, 35–40. <http://dx.doi.org/10.1037/pst0000202>
- Koementas-de Vos, M. M. W., Nugter, M. A., Engelsbel, F., & De Jong, K. (2018). Does progress feedback enhance the outcome of group psychotherapy? *Psychotherapy*, 55, 151–163. <http://dx.doi.org/10.1037/pst0000164>
- Lambert, M. J., Kahler, M., Harmon, C., Burlingame, G. M., Shimokawa, K., & White, M. M. (2013). *Administration and scoring manual: Outcome Questionnaire OQ-45.2*. Salt Lake City, UT: OQMeasures.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55, 520–537. <http://dx.doi.org/10.1037/pst0000167>

- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., . . . Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI): A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12, 224–231. [http://dx.doi.org/10.1016/S0924-9338\(97\)83296-8](http://dx.doi.org/10.1016/S0924-9338(97)83296-8)
- Lutz, W., Schiefele, A. K., Wucherpfennig, F., Rubel, J., & Stulz, N. (2016). Clinical effectiveness of cognitive behavioral therapy for depression in routine care: A propensity score based comparison between randomized controlled trials and clinical practice. *Journal of Affective Disorders*, 189, 150–158. <http://dx.doi.org/10.1016/j.jad.2015.08.072>
- Miller, S. D., & Duncan, B. L. (2004). *The Outcome and Session Rating Scale: Administration and scoring manual*. Chicago, IL: Institute for the Study of Therapeutic Change.
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J., & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analogue measure. *Journal of Brief Therapy*, 2, 91–100. Retrieved from <https://scottdmiller.com/wp-content/uploads/documents/OutcomeRatingScale-JBTv2n2.pdf>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7, 105–125. <http://dx.doi.org/10.1037/1082-989X.7.1.105>
- Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy*, 48, 98–102. <http://dx.doi.org/10.1037/a0022161>
- 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*, 30, 195–212. <http://dx.doi.org/10.1080/10503307.2018.1517949>
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy*, 46, 418–431. <http://dx.doi.org/10.1037/a0017901>
- Schotte, C., & de Doncker, D. (1994). *ADP-IV Questionnaire*. Antwerp, Belgium: Antwerp University Hospital.
- Schotte, C., & de Doncker, D. (2000). De ADP-IV: Een vragenlijst voor therapeutisch georiënteerde diagnostiek van persoonlijkheidsstoornissen. [The ADP-IV: A questionnaire for therapy orientated diagnostics of personality disorders]. *Psychopraxis*, 2, 124–130. <http://dx.doi.org/10.1007/BF03071865>
- Schotte, C. K. W., de Doncker, D., & Courjaret, J. (2007). De ADP-IV: Vragenlijst voor DSM-IV As II én voor gedragstherapeutische diagnos- tiek [The ADP-IV: Questionnaire for DSM-IV and behavioral therapeutic diagnostics]. *Gedragstherapie*, 40, 111–123.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M. I. N. I.): The development and validation of a structured diagnostic psychiatric interview for *DSM-IV* and ICD-10. *The Journal of Clinical Psychiatry*, 59 (Suppl. 20), 22–33. Retrieved from <https://www.psychiatrist.com/JCP/article/Pages/1998/v59s20/v59s2005.aspx?sclick=1>
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78, 298–311. <http://dx.doi.org/10.1037/a0019247>
- 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, 287–300. <http://dx.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, 638–647. <http://dx.doi.org/10.1080/10503307.2012.698918>
- Stuckler, D., Reeves, A., Loopstra, R., Karanikolos, M., & McKee, M. (2017). Austerity and health: The impact in the U. K. and Europe. *European Journal of Public Health*, 27(Suppl. 4), 18–21. <http://dx.doi.org/10.1093/eurpub/ckx167>
- Van Vliet, I. M., Leroy, H., & Van Megen, H. J. G. M. (2000). *M. I. N. I. International Neuropsychiatric Interview: Dutch Version 5.0.0-R*. Utrecht, the Netherlands: Department of Psychiatry, University Medical Center Utrecht.
- Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, 77, 129–137. <http://dx.doi.org/10.1016/j.brat.2015.12.005>

Received December 18, 2019

Revision received March 26, 2020

Accepted April 30, 2020 ■