

EMPIRICAL PAPER

The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial

KIM DE JONG¹, REINIER TIMMAN¹, LEONA HAKKAART-VAN ROIJEN¹,
PASCAL VERMEULEN², KEES KOOIMAN¹, JAN PASSCHIER³, &
JAN VAN BUSSCHBACH¹

¹Medical Psychology and Psychotherapy, Erasmus University Medical Center, Rotterdam, The Netherlands; ²Institute for Directive Interventions, Amsterdam, The Netherlands & ³Faculty of Psychology and Education, Vrije Universiteit, Amsterdam, The Netherlands

(Received 19 February 2013; revised 30 September 2013; accepted 25 November 2013)

Abstract

Objective. Outcome monitoring feedback has become popular, but its effect on treatment outcome has been mixed. Feedback seems most effective for patients who are not progressing well (“not on track” (NOT) cases). There are some indications that patient feedback has an additional effect and that feedback effects differentiate between short- and long-term therapy. This study aimed to evaluate the effect of outcome monitoring feedback to therapists and patients on outcome in short- and long-term psychotherapy. **Methods.** Patients ($n = 475$) were randomly assigned to three conditions: Feedback to therapist (FbT), feedback to therapist and patient (FbTP), and no feedback (NFb). Feedback consisted of progress charts based on the Outcome Questionnaire and a feedback message. **Results.** In short-term therapies (<35 weeks) FbT and FbTP was preventive of negative change for NOT cases. In long-term therapy only FbTP had a small positive effect on the rate of change. Feedback did not result in better outcomes at treatment ending, although there was a trend for FbTP to have fewer deteriorated cases. **Conclusions.** Benefits of feedback were strongest for cases that were not progressing well in short-term therapies when both the patient and therapist received feedback on the patients’ progress. Contrary to previous findings, we also found a small effect of feedback to therapists and patients in long-term therapies. Feedback to both patients and therapists may be more effective than feedback to therapists alone due to implementation issues or empowerment of the patient.

Keywords: psychotherapy research; feedback; outcome monitoring; outpatients

Introduction

Although the vast majority of research has demonstrated the general effectiveness of psychotherapy, it is not equally effective for all patients (e.g., Lambert & Ogles, 2004). According to Bickman (2008), providing outcome monitoring feedback to clinicians and patients is a good method of improving outcomes for individual patients. Using this method, the health status of the patient (e.g., symptom distress) is assessed frequently during therapy and therapists are provided with feedback on the patient’s progress. By assessing their patients’ progress, therapists can adapt their treatment approach if the patient is not

progressing well. Studies have demonstrated that based on clinical judgement alone, therapists are not very effective in detecting negative change (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010) and may need this type of feedback to signal patients being “not on track” to achieve positive outcomes. Similarly, patients may respond to a lack of progress by increasing their efforts in therapy or by discussing a treatment adaptation with their therapist.

Outcome monitoring feedback has become increasingly popular and has been adopted by many mental health care providers all over the world (e.g., Evans et al., 2002; Howard, Moras, Brill,

Correspondence concerning this article should be addressed to Kim de Jong, Erasmus University Medical Center, Section Medical Psychology and Psychotherapy, Rotterdam, The Netherlands. Email: KJong@fsw.leidenuniv.nl

Martinovich, & Lutz, 1996; Kraus, Seligman, & Jordan, 2005; Miller, Duncan, Sorrell, & Brown, 2005; Trauer, 2010). Research has shown that measuring outcomes and providing feedback as part of routine practice appears to have a positive impact on the accuracy of diagnosis (Carlier et al., 2012; Marshall, Haywood, & Fitzpatrick, 2006) and on communication between patient and clinician (Carlier et al., 2012) in both somatic and mental health care problems, but the impact on patient outcome is less consistent. Most meta-analyses and reviews from social psychology, educational science, and medicine demonstrate a small positive effect for feedback interventions compared to the control group (e.g., Hattie & Timperley, 2007; Jamtvedt, Young, Kristoffersen, O'Brien, & Oxman, 2006; Kluger & DeNisi, 1996; Veloski, Boex, Grasberger, Evans, & Wolfson, 2006). Sapyta (2004, in Sapyta, Riemer, & Bickman, 2005) found an average effect size of .21 (Cohen's *d*) for health status feedback compared to no feedback control. More specifically tailored to mental health care, Knaup, Koesters, Schoefer, Becker, and Puschner (2009) concluded that health status feedback compared to no feedback had a small positive effect on outcome in short-term treatments ($d = .10$), but not in longer term treatments ($d = -.06$). Lambert et al. (2003) and Shimokawa, Lambert, and Smart (2010) found much larger effects of feedback on outcome compared to no feedback, ranging between .28 and .70, but their meta- and mega-analyses included only studies from their own research group and were mainly conducted in the university counselling centre. More recent trials, both inside and outside Lambert's group, have shown more moderate but consistent effects of feedback (Bickman, Douglas Kelley, Breda, De Andrade, & Riemer, 2011; Byrne, Hooke, Newnham, & Page, 2012; Newnham, Hooke, & Page, 2010; Reese, Norsworthy, & Rowlands, 2009; Simon et al., 2013; Simon, Lambert, Harris, Busath, & Vazquez, 2012).

Feedback seems mainly effective for patients who are not doing well in therapy, the so called "not on track" (NOT) cases (Carlier et al., 2012; Lambert et al., 2003), although some studies have found feedback to be equally effective for both "on track" (OT) and NOT cases (Bickman et al., 2011; Simon et al., 2013). NOT cases are typically identified as being those individuals who fall out of range of a positive expected treatment response. There are also some indications that feedback is more effective in achieving positive outcomes when both the therapist and the patient receive feedback (Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004), but in other studies there was no significant additional effect of providing feedback to patients as well as therapists

(Harmon, Hawkins, Lambert, Slade, & Whipple, 2005; Slade, Lambert, Harmon, Smart, & Bailey, 2008). The meta-analysis by Knaup et al. (2009) showed differential effects for short-term and long-term therapy. However, the limitation of their analysis is that the studies in these two groups differed substantially in patient population and frequency of the feedback. The long-term therapy group consisted mainly of studies conducted in severe mental disorders that received case management rather than psychotherapy and in which infrequent feedback (once or twice) was used, whereas the short-term group consisted of studies in mood and anxiety disorders and personal concerns, and most studies used weekly feedback. A comparison of short- and long-term treatment with a more comparable population is necessary.

The current study investigates the effect of feedback in a sample of outpatients treated in mental health care institutions or private practices. Patients completed session-by-session questionnaires in a web-based application. The main research question was whether feedback improves outcomes and whether feedback to patients and therapists would be more effective than feedback to therapists alone. The secondary research question was whether feedback effects differed between short- and long-term therapies. There were three conditions: Feedback to therapists, feedback to patients and therapists, and a no-feedback control group. The feedback was expected to be mainly effective for NOT cases. Short-term and long-term therapies were defined post hoc by splitting on the median of treatment duration. We expected feedback to patients and therapists to be more effective than feedback to therapists alone. In addition, based on the literature, feedback was expected to be more effective in short-term therapies, although it should be noted that no prior research has been performed in long-term outpatient psychotherapies.

Method

Subjects

Patients. Data were collected in a web-based monitoring application in the period 1 July 2006 to 31 June 2011. The study design was approved by the Medical Ethical Committee of the Erasmus University Medical Center Rotterdam, as well as by the cooperating institutes. Participants were recruited in private psychotherapy practices and outpatient mental health institutes. Inclusion criteria were an age of 17 years or older and sufficient understanding of the Dutch language to complete questionnaires without assistance. A total of 790 patients were assessed for

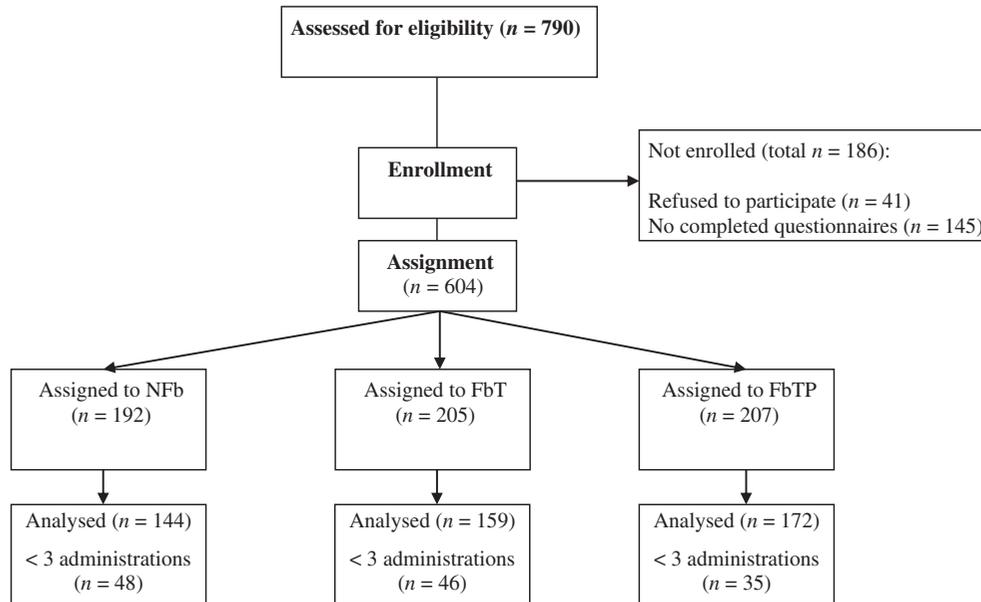


Figure 1. Flowchart of participants in the study.

eligibility, of whom 41 refused to participate and 145 patients never completed a single questionnaire. After agreeing to participate, 604 subjects were randomly assigned to one of three conditions: Feedback to therapists (FbT), feedback to both therapists and patients (FbTP), or a control group without feedback (NFb). Participants with fewer than three Outcome Questionnaire (OQ-45; Lambert et al., 2004) administrations were excluded from analyses, because two administrations are the absolute minimum to present feedback with a gain or decrease that can have an effect on treatment outcome at session three or later. As a result, 129 patients were excluded from analysis (see Figure 1).

Therapists. A total of 110 therapists participated in the study. In the analysed sample, therapists had between 1 and 34 patients participating in the study, with a mean of 4.3 patients per therapist ($SD = 6.4$). Approximately half of the therapists worked in private practice (49%) and most therapists were trained as psychologists (76%) or psychiatrists (15%). Years of experience after training varied from 0 to 36 years, with a mean of 16.9 ($SD = 9.5$) years. Therapists of all major therapy orientations participated in the study, although cognitive behavioural therapy (27%), client-centred therapy (24%) and psychodynamic therapy (14%) were most frequent.

Instruments

Outcome Questionnaire – 45 item version (OQ-45). The Dutch version of the Outcome Questionnaire – 45 item version (OQ-45) was used

to measure patient progress during treatment. The OQ-45 (Lambert et al., 2004) is a self-report instrument and has 45 items, nine of which are reversed, asking how the respondent has felt over the last week on a 5-point rating scale, ranging from 0 (never) to 4 (almost always). Higher scores reflect a higher level of dysfunctioning. The OQ-45 consists of three subscales that are aimed at assessing different domains of client functioning: Symptom Distress, Interpersonal Relations, and Social Role. The internal consistency (Cronbach's alpha) for the Total score of the Dutch OQ-45 ranges between .92 and .96 in university, community, patients, and community and patients combined samples. For the subscales the internal consistency is .90–.95 for the Symptom Distress (SD) scale, .74–.84 for the Interpersonal Relations (IR) subscale and .53–.72 for the Social Role (SR) subscale (De Jong, Nugter, Lambert, & Burlingame, 2009). In the current study, alpha was .93 for the Total score and .92 for the SD, .76 for the IR and .74 for the SR subscales.

Patient characteristics. Patients completed a basic background questionnaire after entering the study. The questionnaire consisted of six items on gender, age and email address of the patient, the name of the therapist, and the frequency of visits to the therapist.

Clinical diagnosis. A clinician-rated psychiatric classification according to the Diagnostic and Statistical Manual of Mental Disorders IV on all five axes was provided by the therapist in the online system.

Procedure

Therapists were instructed to inform patients who matched inclusion criteria about the study and ask them to participate. After being entered into the project by the researchers, the online feedback system allocated the patient to one of the three conditions, following a block randomization design in which patients were randomized within therapists.

In all conditions, patients filled out the OQ-45 online before each therapy session, though not more than once a week. At the first administration, the background questionnaire was also administered. Patients were provided with an individual login and password and were able to log in from any location, although most completed their questionnaires in the therapist's waiting room on a laptop that was provided to the therapist for the purpose of the study. In the Nfb condition neither the therapist nor the patient would receive feedback about the patient's progress. In the FbT and FbTP conditions, feedback was generated immediately for use in the therapy session. Therapists and patients could access the feedback either through email or by logging into the therapist/patient portal of the online feedback system. In both feedback conditions, therapists were given full autonomy on discussing the feedback messages with the patient. Feedback consisted of a progress graph and a message tailored to the status of patients. The graph represented the total OQ-45 score and the subscale scores at the various therapy sessions. A horizontal red line indicated the cut-off score (i.e., 58) between the normal and clinical population. In each feedback message a comparison was made between the current OQ-45 Total score, the baseline score, and the cut-off score for normal functioning. An example of a feedback message is: "Your patient shows a high level of complaints, but feels better than at the start of treatment. Your patient has a good chance of benefitting from further treatment." In the patient feedback, patients received the same feedback as the therapist, except the feedback messages used language that was directed towards the patient.

Statistical analysis

NOT cases were defined post hoc as having had a deterioration of at least the reliable change index (14 points; de Jong et al., 2009) compared to baseline. In order to prevent precipitate categorization as a NOT case by one negative outlier measure, this should have occurred on minimally two occasions during the course of therapy. Therapies were divided into short-term and long-term therapies post hoc using the median of the treatment duration (35 weeks), thus creating two groups of similar size.

Data were analysed with multilevel modelling, using the MIXED procedure in SAS (SAS Institute Inc., SAS 9.2. Cary, NC, 2008). Initially, three levels were postulated: Therapists as upper level, patients as second level, and time-points as lowest level. Bias caused by very long therapies was avoided by deletion of data after 2 years of therapy (104 weeks). The deviance statistic was used for testing the need for a three-level model over a two-level model. Saturated models were formulated with the natural logarithm of time, dummies for FbT, FbTP, and NOT, second-order interactions between feedback and NOT, and third-order interactions with time. Both intercept and slope were random. Non-significant predictors (p -value > .05) were removed until a parsimonious model was reached that did not significantly differ from the saturated model. Effect sizes were computed using Equation 1, in which the difference between the estimate at time point t and the baseline OQ-45 score was divided by the baseline OQ-45 standard deviation.

$$d = \frac{\text{estimate}_t - \text{estimate}_{\text{baseline}}}{sd_{\text{baseline}}} \quad (1)$$

Baseline differences were analysed using a one-way ANOVA with post-hoc Bonferroni correction.

Reliable and clinically significant changes (Jacobson & Truax, 1991) were computed using the cut-off score for normal functioning of the Dutch OQ-45 that was available at the start of this study (de Beurs, den Hollander-Gijsman, Buwalda, Trijsburg, & Zitman, 2005). The current cut-off score for the Dutch OQ-45 is 55 (De Jong et al., 2009) but, since feedback was provided based on the cut-off score of 58, in the calculations for clinically significant change 58 rather than 55 was used. This resulted in 13 patients (2.7%) being classified as recovered who would have been classified as improved if a cut-off of 55 had been used. End status functioning of patients was determined by the last available OQ. Last observation carried forward was used if the OQ-45 from the session immediately preceding treatment termination was not available. Differences in reliable and clinically significant change between conditions were tested using a chi square test, and standardized residuals evaluated as z-scores for the interpretation of individual cells.

Results

Patients

A total of 475 outpatients were analysed. Demographic characteristics and diagnoses for each condition are presented in Table I. Significantly higher OQ-45 baseline scores were found for FbTP

Table I. Patient characteristics per condition

	NFb		FbT		FbTP		Total	
	<i>n</i>	% or mean (<i>SD</i>)	<i>n</i>	% or mean (<i>SD</i>)	<i>n</i>	% or mean (<i>SD</i>)	<i>n</i>	% or mean (<i>SD</i>)
Female	144	65%	159	64%	172	74%	475	68%
Age	144	38.6 (11.8)	158	38.1 (11.6)	170	37.9 (12.5)	472	38.2 (12.0)
> High school	140	69%	157	71%	166	71%	463	72%
Diagnoses	127		135		154		417	
Mood disorder		26%		22%		31%		27%
Adjustment disorder		17%		21%		16%		18%
Anxiety disorder		15%		8%		8%		10%
Relational problems (V-codes)		15%		12%		15%		14%
Other ¹		18%		21%		17%		18%
Personality disorder		43%		40%		34%		39%
Comorbidity within axis 1		45%		42%		50%		46%
Comorbidity axis 1 and 2		42%		38%		33%		37%
Baseline OQ-45 score ²								
Included ≥ 3 administrations	144	65.1 (22.4)	159	69.3 (22.5)	172	72.4 (21.9)	475	69.2 (22.4)
Excluded < 3 administrations	48	68.4 (27.4)	46	59.8 (24.2)	35	59.2 (20.7)	129	62.9 (24.8)
Number of sessions	126	33.5 (40.5)	140	36.0 (56.7)	144	27.5 (17.2)	410	32.3 (41.4)
Number of OQ-45 administrations	144	15.7 (16.6)	159	15.8 (15.2)	172	17.4 (18.0)	475	16.4 (16.7)
Percent completeness per patient	126	55 (28)	140	54 (26)	144	57 (27)	410	57 (27)
Not on track	144	15%	159	21%	172	19%	475	18%

Note. NFb = no feedback; FbT = feedback to therapist; FbTP = feedback to therapist and patient.

¹ Other disorders include: Disorders usually first diagnosed in infancy, childhood or adolescence, impulse control disorders, eating disorders, dissociative disorders, sexual disorders, substance-related disorders and psychotic disorders (in order of frequency).

² Significant differences were found between conditions. Within included: NFb vs. FbTP, $p = .01$; included vs. excluded: FbTP, $p = .002$.

compared to NFb, $F(2, 472) = 4.41$, $p = .013$. This seemed to be caused by selective drop-out before session 3 in the FbT and FbTP groups, in which the OQ-45 baseline scores of included patients were somewhat higher than those of the excluded patients ($t(203) = -2.48$; $p = .014$ and $t(205) = -3.27$; $p = .001$ respectively).

The median therapy length was 35 weeks ($M = 43.0$, $SD = 31.1$), and was used to distinguish between short- ($n = 231$) and long-term ($n = 243$) therapies. Since the distinction between short- and long-term therapy was made post hoc, differences in demographic and therapy variables were assessed (see Table II). Results indicated that the groups were comparable on most variables, but significant differences were found on four variables: Age, percentage of completed OQ-45 administrations, the frequency of sessions (in number of days between sessions), and patient-initiated termination of therapy. Patients in the short-term therapy group were on average somewhat younger ($t(471) = -2.62$, $p = .009$), had less time between sessions ($t(343) = -3.08$, $p = .002$), a higher completion rate ($t(408) = -9.12$, $p < .001$) and more frequently initiated treatment termination ($\chi^2(6) = 31.06$, $p < .001$) than patients in the long-term therapy group. In addition, long-term therapies included more NOT patients ($\chi^2(1) = 13.52$, $p < .001$).

Rate of change

The effect of feedback on outcome was examined in two ways: Rate of change (speed of progress) and end-state functioning (final outcome). The rate of change refers to the steepness of the slope in the change model and indicates how much faster or slower patients change over time due to the factors investigated. Initially, we tested whether all three levels were required in the multilevel model. Analysis of the unconditional model showed that only 2% of the total variance was situated at the therapist level. In addition, the slope for the therapist level was not significant ($\chi^2(2) = 1.47$; $p = .48$). Therefore, the therapist level was dropped from subsequent analyses. The two-level model for all therapy lengths showed an overall significant small positive effect (Cohen, 1992; interpreted according to Thalheimer & Cook, 2002) of feedback to therapists and patients over time ($d = .16$ after 35 weeks and $d = .20$ after 78 weeks), but contrary to expectations no significant effect of feedback to therapists alone was found. Also, no significant interaction was found between feedback (either FbT or FbTP) and the patient being not on track (see Table III).

Next, short- and long-term therapies were analysed separately. In short-term therapies there was a significant three-way interaction between time, the status of the patient being not on track, and type of feedback. Figure 2a shows this effect, with the effect

Table II. Patient characteristics for short-term and long-term therapies

	<i>n</i>	Short-term % or mean (<i>SD</i>)	<i>n</i>	Long-term % or mean (<i>SD</i>)
Female	232	66%	243	70%
Age**	229	36.7 (11.7)	243	39.6 (12.1)
> High school	224	69%	239	75%
Diagnoses	208		209	
Mood disorder		26%		27%
Adjustment disorder		15%		20%
Anxiety disorder		10%		11%
Relational problems (V-codes)		14%		14%
Other ¹		19%		18%
Personality disorder		36%		42%
Comorbidity within axis 1		45%		47%
Comorbidity axis 1 and 2		34%		41%
Baseline OQ–45 score	232	69.3 (23.2)	243	68.9 (21.9)
Treatment modality	192		197	
Integrative		39%		40%
Cognitive behavioural		32%		30%
Psychoanalytic/psychodynamic		18%		19%
Supportive		6%		4%
System-oriented		3%		6%
Other		2%		1%
Number of OQ–45 administrations***	232	7.2 (4.0)	243	25.1 (19.3)
Percent completeness per patient***	200	44 (25)	210	66 (23)
Days between sessions**	232	11.0 (6.6)	243	14.2 (14.5)
Patient initiated termination***	231	17%	243	7%
Not on track***	232	12%	243	25%

¹ Other disorders include: Disorders usually first diagnosed in infancy, childhood or adolescence, impulse control disorders, eating disorders, dissociative disorders, sexual disorders, substance-related disorders and psychotic disorders.

** $p < .01$; *** $p < .001$.

size on the y -axis. Overall, NOT cases have a negative effect size over time. In the FbT and FbTP conditions receiving feedback respectively had a large ($d = .91$ at 35 weeks) and a very large effect ($d = 1.28$ at 35 weeks) for the NOT cases and was preventive of negative outcomes. The negative effect of being NOT was compensated by receiving feedback, but did not result in positive change. There was no effect of feedback in the on-track cases.

In the long-term therapy group there was a significant difference in OQ–45 scores at baseline for both feedback conditions compared to the NFb control group. Therefore, the baseline OQ–45 scores for the FbT and FbTP groups were included in the model as intercept predictors. The FbTP condition had a favourable small effect on the rate of change ($d = .24$ at 35 weeks and $d = .29$ at 78 weeks), equally for OT and NOT cases (see Figure 2b).

End-state functioning

Table IV shows the functioning of patients at the end of treatment based on reliable and clinically significant change. Although there were no overall significant differences between conditions ($\chi^2(6) = 8.01$, $p = .24$), there was a trend for the FbTP condition to have the best results: the lowest rate of deteriorated patients ($Z = -1.3$, $p = .097$) was in this condition.

Subgroup analysis of short- and long-term therapies showed similar results per subgroup, although recovery rates were somewhat better in the long-term therapy group than the short-term therapy group.

Discussion

In this study we aimed to demonstrate the effect of feedback on patient progress to therapists and patients. Feedback to both therapists and patients was most effective and was a significant predictor of the rate of change in both short- and long-term therapies. The benefits of feedback were strongest for cases that were not progressing well in short-term therapies. Feedback provided to the therapist alone was effective for NOT patients in short-term therapies, but not in long-term therapies. Feedback influenced the rate of change, but did not significantly improve end-state functioning, although there was a trend for the feedback to therapists and patients group to have the least deteriorated patients.

The effects in the short-term therapy group resemble results found by Lambert's group. His group was among the first to study the effect of feedback on patient outcomes and has performed the largest number of studies on the effect of feedback. Their studies typically demonstrate that feedback is most effective for NOT cases (Lambert et al., 2003;

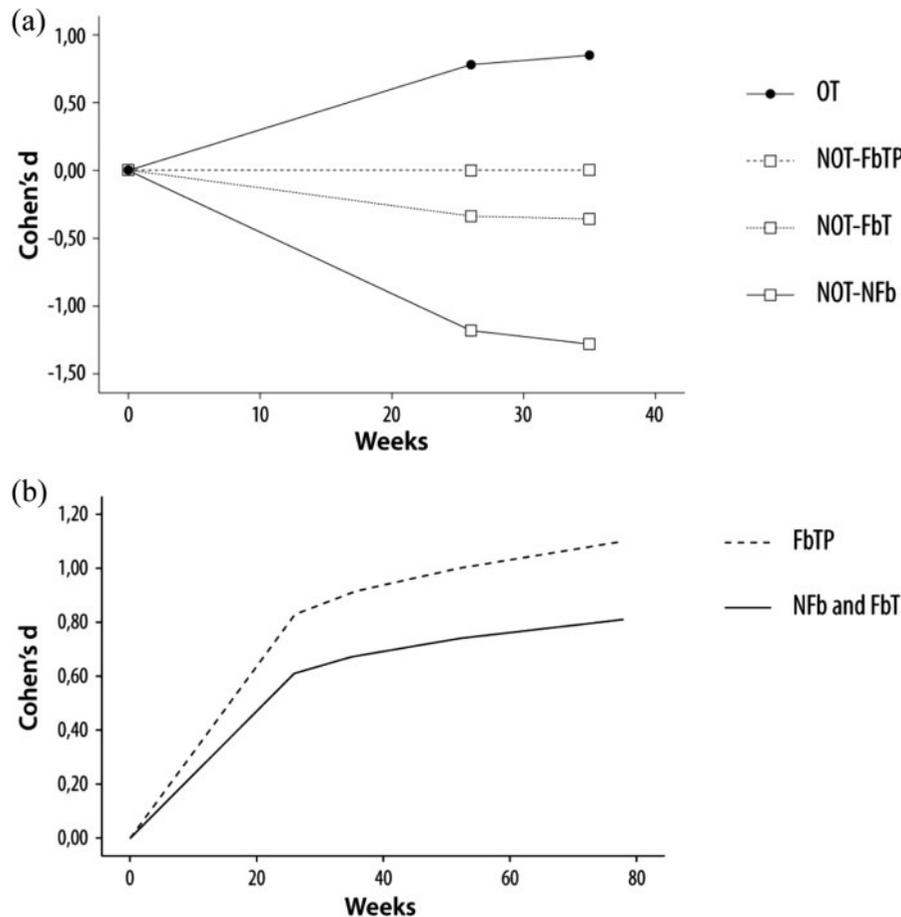


Figure 2. (a) Effect sizes per group for short-term therapies. (b) Effect sizes per group for long-term therapies. In (b) there was no difference between OT and NOT cases, nor was there a difference between the NFb and FbT conditions, therefore the effect size lines overlap.

Note. OT = on track (all conditions have equal patterns); NOT = not on track; NFb = no feedback; FbT = feedback to therapists; FbTP = feedback to therapists and patients.

Shimokawa et al., 2010). For feedback in long-term therapies, Knaup et al. (2009) found no significant effect in their meta-analysis. In contrast, we did find a small but significant effect ($d = .24$ after 35 weeks and $d = .29$ after 78 weeks). Our sample of long-term therapies differed in several ways from the long-term therapies they included in their analysis. They defined long-term effects of feedback as measured between 3 and 12 months after initial assessment, whereas in our long-term group treatment duration is much longer. In addition, the majority of the studies they included focused on a more chronic population that included patients with schizophrenia and chronic (bipolar) depression. Moreover, in three of the five studies the feedback was provided only once or twice. So their long-term group possibly did not include the most effective types of feedback and it may have included a subgroup of patients amongst whom not much progress might be expected.

These findings raise questions about the apparent differential effects of feedback interventions in short-

versus longer-term therapy. Differences in patient groups may be one explanation for these differential effects. Although patients did not differ significantly on diagnosis or treatment modality—variables on which one might expect differences between short- and long-term therapies—patients did differ on a few relevant variables. One important difference between the groups is that in the short-term therapy group more patients initiated treatment termination, indicating that most of the treatment drop-out took place before 35 weeks of therapy. Furthermore, patients in the long-term therapy group were somewhat older than patients in the short-term therapy group, hinting at the possibility of more chronic complaints in that group, although prior episodes or duration of complaints were not measured in this study. An alternative explanation could be that in longer therapy, with patients who have remained in therapy for at least 35 weeks, therapists have more opportunities to signal negative change themselves, even without feedback. It may also be that after a longer period of

Table III. Fixed and random effects for change trajectories

	All therapy lengths		Short-term therapies (<35 weeks)		Long-term therapies (≥35 weeks)	
	Estimate	SE	Estimate	SE	Estimate	SE
Short-term therapies						
Fixed effects						
Intercept	71.79****	1.09	70.89****	1.57	66.96****	2.55
Time	-4.10****	.31	-5.52****	.42	-4.05****	.40
FbT					7.37*	3.37
FbTP					9.21**	3.45
Time * FbTP	-1.03*	.51			-1.46*	.73
Time * NOT			13.81****	2.04		
Time * NOT * FbT			-5.93*	2.35		
Time * NOT * FbTP			-8.31**	3.09		
Random effects						
Intercept	500.97****	36.38	500.3****	53.44	472.0****	47.29
Slope	24.07****	2.23	17.0****	3.40	22.5****	2.50
Covariance	-43.02****	7.10	-29.1***	10.55	-46.3****	8.61
Residual	113.79****	2.02	113.0****	4.54	113.9****	2.26

Note. NFb = no feedback; FbT = feedback to therapist; FbTP = feedback to therapist and patient; NOT = not on track; time is the natural log of weeks + 1.

* $p < .05$; ** $p < .01$; *** $p < .001$.

therapy, therapists have better rapport with their patients, which may make feedback redundant.

In the current study, the strongest effect of feedback was found when both therapist and patient received feedback. Our findings may shed light on possible reasons why previous studies on patient feedback have shown mixed results. In the Hawkins et al. (2004) study feedback to patients and therapists outperformed feedback to therapists alone, but the studies by Slade et al. (2008) and Harmon et al. (2005) did not show significant effects of patient and therapist feedback over therapist feedback alone. The overall effect of these three studies resulted in no significant effect for patient feedback (Shimokawa et al., 2010). One of the explanations for the differential effects might be found in different populations. The study by Hawkins et al. took place in an outpatient centre, whereas the studies by both Harmon et al. and Slade et al. were done in a university counselling centre that provided therapy to students with personal concerns. The outpatient group had more severe patients as well as a more mature group (Shimokawa et al., 2010) and a less

controlled treatment situation and resembles our group more than the counselling centre sample does. It may be that feedback to patients is more effective in these circumstances.

One could wonder why feedback to therapist and patient shows a more pronounced effect than feedback to therapists alone. A couple of explanations are viable, for instance, it could be a matter of implementation. The therapist knows that the patient has direct access to the feedback, and this might encourage the therapist to look at the feedback as well. Research shows that taking time to look at the feedback can be a barrier for therapists to use feedback (Boswell, Kraus, Miller, & Lambert, 2013; De Jong, 2012). Alternatively, therapists may experience resistance to being evaluated (Riemer & Bickman, 2011) and might avoid looking at the feedback as a result. If the feedback is not seen by the patient, the therapist might assign looking at the feedback a lower priority than other tasks. Some of the therapists in our study indicated that patients may be more empowered when they receive feedback about their own progress in therapy. By receiving the

Table IV. Reliable and clinically significant change per condition

	All therapy lengths ($n = 475$)			Short-term therapies ($n = 232$)			Long-term therapies ($n = 243$)		
	NFb	FbT	FbTP	NFb	FbT	FbTP	NFb	FbT	FbTP
Recovered	37%	38%	43%	32%	30%	35%	41%	48%	50%
Improved	10%	8%	13%	11%	8%	12%	8%	8%	15%
No change	46%	42%	38%	47%	49%	47%	45%	35%	32%
Deteriorated	8%	11%	5%	10%	13%	7%	6%	9%	4%

Note. NFb = no feedback; FbT = feedback to therapist; FbTP = feedback to therapist and patient.

feedback, patients are in a position to actively discuss their (lack of) progress with their therapist and this may promote communication between patient and therapist. An alternative explanation is that if patients can track their own progress, they can also manipulate the results. It is impossible to filter out such an effect, and in our experience some patients will use the feedback to communicate with their therapist through the questionnaires, but the effect of this usually disappears after a few weeks.

The current study has some limitations that might influence study results. One of the problems we encountered was that baseline scores were higher for the feedback to therapist and patients group than for the no-feedback group. This difference was caused by excluding patients with fewer than three administrations of the OQ-45 and was most pronounced in the long-term therapies. Possibly feedback causes patients with higher complaint levels to stay in the study. We tried to compensate for this problem by adding the baseline scores of the OQ-45 to the multilevel model as a covariate, but there may still be selectiveness in our patient sample as a result.

A factor that might complicate the generalization of our results is that it is unclear to what extent our sample is selective. Contrary to our instructions, therapists may have made self-selections of patients they approached to participate in the study and we had no way of checking this. For patients who did participate in the study, but did not complete an OQ-45 at every session, systematic missing data may be an issue. For instance, some patients may have skipped completing the questionnaire when they did not feel well. However, missing data are not necessarily systematically missing. Some patients had more than one session per week, whereas administration of the OQ was restricted to once a week. Other non-systematic reasons for missing an administration are things like being late for therapy because of traffic or computer malfunctions.

Another issue that needs discussion is our definition of NOT cases. We decided to use a definition in which a patient needed to have a deteriorated score at least two times. This resulted in relatively low percentages of NOT cases (14–20%), whereas other studies resulted in NOT cases in 20–30% of the cases (Slade et al., 2008), and sometimes even up to 50% of the cases (Hawkins et al., 2004). We chose to have two deteriorations rather than one in order to rule out accidental high scores on the OQ-45 and to ensure that a patient was actually on a negative track.

Finally, our definition of short- and long-term therapies has some drawbacks. We divided therapies

into two groups post hoc, which resulted in equal group sizes and thus optimal power to detect an effect for both groups, but may cause problems when drawing inferences. For instance, it is possible that receiving the feedback had its influence on treatment duration, although we did not find significant differences in treatment duration between the conditions. In addition, it may be that patients who were not progressing well are over-represented in the long-term therapy group. We did indeed find that there were more NOT in the long-term therapy group, but this could also be due to higher chances of being NOT by having more sessions. Also, since we used treatment duration in weeks, this division does not tell us much about dosage. Ideally, patients would have been allocated to short-term and long-term therapies randomly or at least an a priori estimation of treatment duration could have been made, something we did attempt to obtain from the therapists. Unfortunately, it turned out to be unfeasible to collect this information. New studies, using a clear definition of short-term and long-term therapy, should provide more information on the effects of feedback in long-term psychotherapy.

The current study shows that feedback can be effective in improving the rate of change in outpatient mental health care. Although outcomes were not necessarily better when feedback was provided, progress was achieved faster, which may result in more cost-effective interventions and earlier diminution of suffering. Feedback effects were small in long-term therapy and OT cases. Consistent with previous studies (Lambert et al., 2003), the strongest effects of feedback in our study were found in NOT cases in short-term therapies, so providing feedback is mainly recommended in those cases.

Although more studies are emerging on the topic of outcome monitoring feedback, there is still much we do not know about the subject. There is still little known on how feedback works in clinical practice and why it improves outcomes in some situations, but does not in others. In addition, most feedback studies have been performed with outpatient adults, and we do not know what the results are in other treatment settings and other groups. Newnham et al. (2010) showed for instance that in an acute clinic, feedback was only effective in reducing depressive symptoms, not anxious or stress symptoms. A recent study by Bickman et al. (2011) in youth mental health care demonstrated differential effects for outcomes measured by clinicians, parents or caregivers and the youth themselves, with the clinicians being most optimistic about these effects and the youth the least. Feedback theory (Riemer & Bickman, 2011) might be able to provide us with a better framework to understand how feedback works. More research is

also needed on how therapists and patients use the feedback in therapy. Overall, this study provides us with more knowledge on the effectiveness of feedback to therapists and patients, for short- and long-term therapies.

Acknowledgment

The authors would like to thank Andrew Page for his comments on the draft of this manuscript.

Funding

This study was supported by a grant from The Netherlands Organization for Health Research and Development (ZonMW) [grant number 94506414].

References

- Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of American Academy of Child and Adolescent Psychiatry*, 47, 1114–1119. doi: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. doi:10.1176/appi.ps.002052011
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2013). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, Epub ahead of print.
- Byrne, S. L., Hooke, G. R., Newnham, E. A., & Page, A. C. (2012). The effects of progress monitoring on subsequent readmission to psychiatric care: A six-month follow-up. *Journal of Affective Disorders*, 137, 113–116. doi:10.1016/j.jad.2011.12.005
- 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*, 104–110. doi:10.1111/j.1365-2753.2010.01543.x
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. doi:10.1037/0033-2909.112.1.155
- de Beurs, E., den Hollander-Gijsman, M., Buwalda, V., Trijsburg, W., & Zitman, F. (2005). De Outcome Questionnaire (OQ-45): een meetinstrument voor meer dan alleen psychische klachten [The Outcome Questionnaire (OQ-45): An instrument that measures more than psychological distress alone]. *De Psycholoog*, 40, 53–63.
- De Jong, K. (2012). De rol van de behandelaar: de 'vergeten' factor in ROM [The role of the therapist: The 'forgotten' factor in routine outcome monitoring]. *Tijdschrift voor Psychiatrie*, 54, 197–201.
- De Jong, K., Nugter, M. A., Lambert, M. J., & Burlingame, G. M. (2009). *Handleiding voor afname en scoring van de Outcome Questionnaire (OQ-45)*. Salt Lake City, UT: OQ Measures.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., & Mellor-Clark, J. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry*, 180, 51–60. doi:10.1192/bjp.180.1.51
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., & Shimokawa, K. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61, 155–163. doi:10.1002/jclp.20108
- Harmon, C., 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. doi:10.1002/jclp.20109
- 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*, 17, 25–32. doi:10.1002/cpp.656
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. doi:10.3102/003465430298487
- 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, 308–327. doi:10.1093/ptr/kph027
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51, 1059–1064. doi:10.1037/0003-066X.51.10.1059
- 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. doi:10.1037/0022-006X.59.1.12
- Jamtvedt, G., Young, J. M., Kristoffersen, D. T., O'Brien, M. A., & Oxman, A. D. (2006). Audit and feedback: Effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews*, 2, CD000259.
- 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, 254–284. doi:10.1037/0033-2909.119.2.254
- Knaup, C., Koesters, M., Schofer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *The British Journal of Psychiatry*, 195, 15–22. doi:10.1192/bjp.bp.108.053967
- Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology*, 61, 285–314. doi:10.1002/jclp.20084
- Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., & Shimokawa, K. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)* (3rd ed.). Wilmington, DE: American Professional Credentialing Services.
- Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York: John Wiley & Sons.
- 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 & Practice*, 10, 288–301. doi:10.1093/clipsy.bpg025
- Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient-reported outcome measures on routine practice: a structured review. *Journal of Evaluation in Clinical Practice*, 12, 559–568. doi:10.1111/j.1365-2753.2006.00650.x
- Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology*, 61, 199–208. doi:10.1002/jclp.20111
- 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, 139–146. doi:10.1016/j.jad.2010.05.003
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*, 46, 418–431. doi:10.1037/a0017901
- 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.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, Research and practice. *Journal of Clinical Psychology: In Session*, 61, 145–153. doi:10.1002/jclp.20107
- 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. doi:10.1037/a0019247
- Simon, W., Lambert, M. J., Busath, G., Vazquez, A., Berkeljon, A., & Hyer, K. (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. doi: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. doi: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. doi:10.1002/cpp.594
- Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology Retrieved February 2006, from http://work-learning.com/effect_sizes.htm.
- Trauer, T. (Ed.). (2010). *Outcome measurement in mental health. Theory and practice*. Cambridge: Cambridge University Press.
- Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A., & Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher*, 28, 117–128. doi:10.1080/01421590600622665