



Universiteit  
Leiden  
The Netherlands

## **A chance for change : building an outcome monitoring feedback system for outpatient mental health care**

Jong, K. de

### **Citation**

Jong, K. de. (2012, April 17). *A chance for change : building an outcome monitoring feedback system for outpatient mental health care*. Retrieved from <https://hdl.handle.net/1887/18691>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/18691>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/18691> holds various files of this Leiden University dissertation.

**Author:** Jong, Kim de

**Title:** A chance for change : building an outcome monitoring feedback system for outpatient mental health care

**Date:** 2012-04-17

The effect of outcome monitoring  
feedback to clinicians and patients  
in outpatient mental health:  
randomized controlled trial

# Chapter 9

De Jong, K., Timman, R.,  
Hakkaart-van Roijen, L., Vermeulen, P.,  
Kooiman, K., Passchier, J., van Busschbach, J.

*Manuscript submitted for publication.*

**Background.** Outcome monitoring has become popular, but meta-analyses show mixed results. Feedback to so called 'not on track' (NOT) cases and to both patient and therapist seems most effective.

**Aims.** This study aimed to evaluate the effect of outcome monitoring feedback to therapists and patients on outcome.

**Method.** Patients ( $n = 474$ ) were randomly assigned to three conditions: feedback to therapist (FbT), feedback to therapist and patient (FbTP) and no feedback (NFb).

**Results.** In the full sample, no significant effect of the FbT condition was found. FbT did result in less negative change pattern for NOT cases in short-term therapies (<35 weeks). FbTP was preventive of negative change for NOT cases in short-term therapies ( $d = 1.28$  after 35 weeks) and had a small positive effect on the rate of change in long-term therapies ( $\geq 35$  weeks).

**Conclusions.** FbTP results in faster progress, especially for NOT cases. FbT was only effective in short-term therapies.

## Introduction

Providing outcome monitoring feedback to clinicians and patients has become an increasingly popular method to improve outcomes and has been adopted by many mental health care providers all over the world (e.g. Evans et al., 2002; Howard, Moras, Brill, Martinovich, & Lutz, 1996; Kraus, Seligman, & Jordan, 2005; Miller, Duncan, Sorrell, & Brown, 2005). 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., 2010; Marshall, Haywood, & Fitzpatrick, 2006) and on communication between patient and clinician (Carlier et al., 2010), but the impact on patient outcome and treatment duration is less consistent. Meta-analyses on the effect of feedback on outcome show mixed effects.

A meta-analysis by Knaup et al. (Knaup, Koesters, Schoefer, Becker, & Puschner, 2009) concluded that health status feedback has a small positive effect on outcome in short-term treatments ( $d = 0.10$ ), but not in longer term treatments ( $d = -0.06$ ). A problem in the meta-analysis by Knaup et al. is that compared short-term and long-term effects of feedback, but 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 and infrequent feedback (once or twice), whereas the short-term group consisted of studies in mood and anxiety disorders and personal concerns, and most studies used weekly feedback. Lambert et al. (Lambert, Whipple, Hawkins, Vermeersch, Nielsen & Smart, 2003) and Shimokawa et al. and (Shimokawa, Lambert, & Smart, 2010) found much larger effects of feedback on outcome, ranging between 0.28 and 0.70, but their meta- and mega-analysis included only studies from their own research group and were mainly conducted in the university counselling center. Feedback seems mainly effective for patients who are not doing well in therapy, the so called 'not on track' (NOT) cases (Carlier et al., 2010; Lambert et al., 2003). Not on track cases are typically identified as being those individuals who fall below a cut-off indicating an expected treatment response. There are also some indications that feedback is more effective 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 (Harmon, Hawkins, Lambert, Slade, & Whipple, 2005; Slade, Lambert, Harmon, Smart, & Bailey, 2008).

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. 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. We were interested in the effects of feedback in both short-term and long-term therapies. Short-term and long-term therapies were defined post-hoc by splitting on the median of treatment duration (35 weeks).

## Method

### Subjects

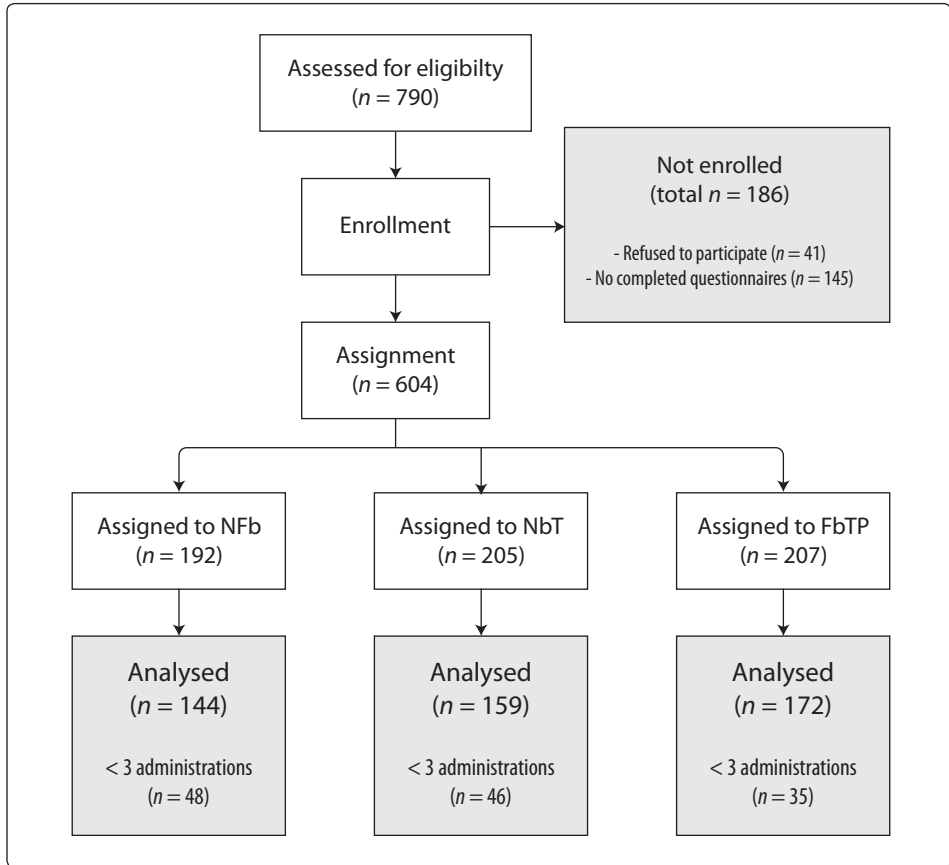
#### Patients

Data were collected in a web-based monitoring application in the period of July 1, 2006 to June 31, 2011. 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 answer questionnaires without assistance. Therapists asked their patients to participate in the study at intake. After agreeing to participate, 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). A randomized block design, with blocks within therapists, was applied. The study was approved of by the Medical Ethical Committee of the Erasmus University Medical Center Rotterdam, as well as by the cooperating institutes. All participants signed an informed consent form.

Participants with less than three OQ-45 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. Figure 1 shows the flow of participants through each stage of the trial.

#### Therapists

A total number of 110 therapists participated in the study on a voluntary basis. In the analysed sample, therapist 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 and most therapists were originally trained as psychologists or psychiatrists. 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, client-centred therapy and psychodynamic therapy were most frequent. Table 1 shows the characteristics of the participating therapists.

**Figure 1** Flowchart of participants

## 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, 9 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 for the Total score of the Dutch OQ-45 ranges between 0.92 and 0.96 in university, community, patients and community and patients combined samples. For the subscales the consistency is 0.90-0.95 for the Symptom Distress scale, 0.74-0.84 for the Interpersonal Relations subscale and 0.53-0.72 for the Social Role subscale (De Jong, Nugter, Lambert, & Burlingame, 2009).

**Table 1** Therapists characteristics

	<i>n</i>	
Female	110	64%
Age	98	$M = 47.8, SD = 10.9$
Member of therapists association(s)	101	30 %
Private practice	108	49 %
Number of therapy hours/week	98	$M = 26.6, SD = 11.1$
Years of practice	87	$M = 16.9, SD = 9.5$
Education	103	
Psychology		63 %
Medicine		14 %
Educational sciences		8 %
Other university degree		16 %
Other education		15 %
Most important professional orientation	98	
Cognitive behavioural		27%
Client centered		24%
Psychodynamic		14%
Integrative		8 %
Eclectic		7 %
Systemic		7 %
Other		9 %

### Patient characteristics

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

### Clinical diagnosis

A psychiatric classification according to the Diagnostic and Statistic Manual of Mental Disorders IV on all five axes was provided by the therapist in the online system.

### Procedure

The background questionnaire was administered prior to the first therapy session. Before each therapy session, though not more than once a week, the patient filled out the OQ-45 online, through a secure internet connection. 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 waiting room of the therapist. Feedback was generated immediately for use in the therapy session. Therapists could access the feedback either through e-mail or by logging into the therapist portal of the online feedback system. Therapists and patients were free to discuss the feedback messages or not. 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. Messages to therapists included suggestions on the level of complaints and continuation of the therapy, for instance “Your patient shows a high level of complaints, but feels better than at the start of treatment. Your patient has a good chance to benefit 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 On Track cases were defined by a deterioration of at least the reliable change index (14 points) compared to baseline at least twice in the course of the therapy, to ensure that a patient was really not on track and not just had one negative outlier once. 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 analyzed with multilevel modelling, using the MIXED procedure in SAS (SAS Institute Inc, SAS 9.2. Cary, NC, USA. 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 ( $p\text{-out} > 0.05$ ) 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. Baseline differences were analyzed using a one-way ANOVA with post-hoc Bonferroni correction.

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

Clinically significant and reliable change (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 based on preliminary analyses (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., 2007), 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. End status functioning of patients was determined by the last available OQ-45. Last observation carried forward was used if the OQ-45 from the session immediately preceding treatment termination was not available. Differences in clinical significant and reliable change between conditions were tested using a Chi Square test.

## Results

### Patients

A number of 475 outpatients met the requirements for inclusion in the study and had at least three administrations. Demographic characteristics and diagnoses in each condition are presented in Table 2. There were no significant baseline differences between conditions on the OQ-45 for all participants. However, after excluding patients with less than three OQ-45 administrations, small differences occurred: within the FbT and FbTP groups, the OQ-45 baseline scores of included patients were somewhat higher than those of the excluded patients ( $t(203) = -2.48$ ;  $p = 0.014$  and  $t(205) = -3.27$ ;  $p = 0.001$  respectively). This resulted in higher baseline scores for FbTP than the control group in the final sample ( $F(2, 472) = 4.41$ ,  $p = 0.013$ , especially for long-term therapies.

The median therapy length was 35 weeks, and was used to distinguish between short ( $n = 231$ ) and long-term ( $n = 243$ ) therapies. Long-term therapies included more NOT patients ( $\chi^2(1) = 13.52$ ,  $p < 0.001$ ). Baseline differences for the short and long-term therapy group were not significant, except for age. Patients in the long term therapy group were somewhat older ( $M = 43.1$ ,  $SD = 12.2$ ) than patients in the short term therapy group ( $M = 40.1$ ,  $SD = 11.7$ ),  $t(472) = 2.71$ ,  $p = 0.007$ .

### 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. Participants did not complete the OQ-45 on every therapy session, but compliance was reasonably good, given that on average more than half of the administered questionnaires were

completed by the patients (see Table 1).

The analyses began by testing if all three levels were required in the multilevel model. The intraclass correlation for the therapist level was computed on an empty model and had a value of 0.02, meaning that only 2% of the total variance in the data was situated at the therapist level. In the three-level model the slope for therapists level was not significant ( $\chi^2(2) = 1.47, p = 0.48$ ). Therefore, the therapist level was dropped from subsequent analyses. Table 3 shows the results of the multilevel models on the effect of feedback on symptom reduction on the OQ-45. Table 4 and Figure 2a, 2b and 2c shows the effect sizes of the feedback in the different models after 26, 35, 52 and 78 weeks of treatment.

Then a model for all therapy lengths was analysed. There was an overall significant small positive effect (according to Cohen's criteria (Cohen, 2002)) of feedback to therapists and patients over time, but contrary to expectations no significant effect of feedback to therapists alone was found. Also, no significant interaction was found

**Table 2** Patient characteristics

	NFb		FbT		FbTP		Total	
	<i>n</i>	% or Mean (SD)	<i>n</i>	% or Mean (SD)	<i>n</i>	% or Mean (SD)	<i>n</i>	% or Mean (SD)
Female	144	65 %	159	64 %	172	74 %	475	68 %
Age	144	42.3 (11.9)	158	41.6 (11.7)	170	41.2 (12.4)	475	41.7 (12.0)
> High school	140	69 %	157	71 %	166	71 %	463	72 %
Diagnoses	121		128		151		400	
Mood disorder		26 %		21 %		31 %		26 %
Adjustment disorder		17 %		20 %		16 %		18 %
Anxiety disorder		15 %		8 %		9 %		10 %
Relational problems (V-codes)		15 %		11 %		15 %		14 %
Other <sup>1</sup>		26 %		40 %		29 %		32 %
Personality disorder		41 %		38 %		34 %		37 %
Co morbidity within axis 1		46 %		44 %		51 %		47 %
Co morbidity axis 1 and 2		43 %		35 %		32 %		36 %
Baseline OQ-45 score								
Included $\geq 3$ administrations <sup>2</sup>	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)
% 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 %
Short (< 35 weeks) term	71	7 %	84	17 %	77	10 %	232	12 %
Long (>35 weeks) term	73	22 %	75	25 %	95	26 %	243	25 %

Note. NFb = No Feedback; FbT = Feedback to Therapist; FbTP = Feedback to Therapist and Patient

<sup>1</sup> 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).

**Table 3** Fixed and random effects for change trajectories

	All therapy lengths		Short term therapies (<35 weeks)		Long term therapies (≥35 weeks)	
	<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>
Short term therapies						
Fixed effects						
Intercept	71.79 <sup>****</sup>	1.09	70.89 <sup>****</sup>	1.57	66.96 <sup>****</sup>	2.55
Time	-4.10 <sup>****</sup>	0.31	-5.52 <sup>****</sup>	0.42	-4.05 <sup>****</sup>	0.40
FbT					7.37 <sup>*</sup>	3.37
FbTP					9.21 <sup>**</sup>	3.45
Time * FbTP	-1.03 <sup>*</sup>	0.51			-1.46 <sup>*</sup>	0.73
Time * NOT			13.81 <sup>****</sup>	2.04		
Time * NOT * FbT			-5.93 <sup>*</sup>	2.35		
Time * NOT * FbTP			-8.31 <sup>**</sup>	3.09		
Random effects						
Intercept	500.97 <sup>****</sup>	36.38	500.3 <sup>****</sup>	53.44	472.0 <sup>****</sup>	47.29
Slope	24.07 <sup>****</sup>	2.23	17.0 <sup>****</sup>	3.40	22.5 <sup>****</sup>	2.50
Covariance	-43.02 <sup>****</sup>	7.10	-29.1 <sup>***</sup>	10.55	-46.3 <sup>****</sup>	8.61
Residual	113.79 <sup>****</sup>	2.02	113.0 <sup>****</sup>	4.54	113.9 <sup>****</sup>	2.26

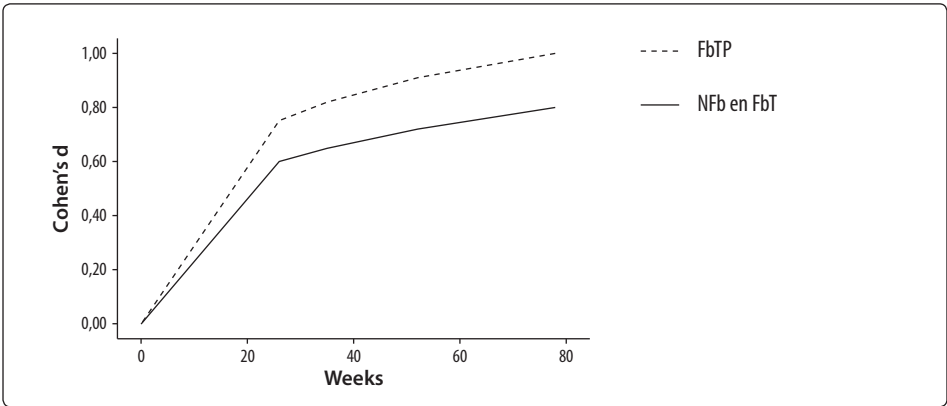
Note: NFb = No Feedback; FbT = Feedback to Therapist; FbTP = Feedback to Therapist and Patient; NOT = Not On Track; Time is the natural log of weeks + 1. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**Table 4** Estimated effect sizes (Cohen's *d*) of difference between groups

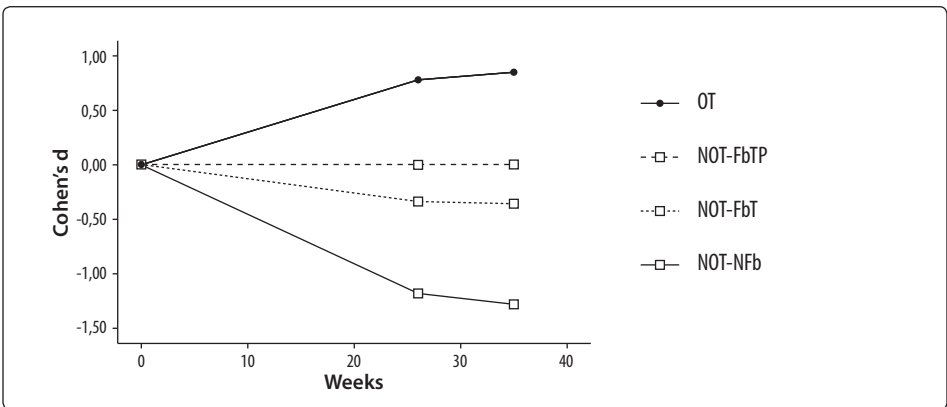
	All cases	Not On Track	
	NFb-FbT vs. FbTP	NFb vs. FbT	NFb vs. FbTP
All therapy lengths			
26 weeks	0.15		
35 weeks	0.16		
52 weeks	0.18		
78 weeks	0.20		
Short-term therapies			
26 weeks		0.84	1.18
35 weeks		0.91	1.28
Long-term therapies			
26 weeks	0.22		
35 weeks	0.24		
52 weeks	0.27		
78 weeks	0.29		

Note: NFb = No Feedback; FbT = Feedback to Therapist; FbTP = Feedback to Therapist and Patient

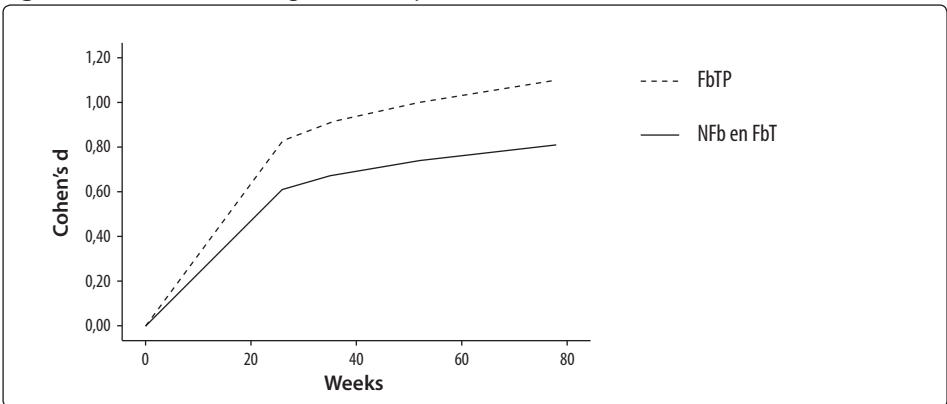
**Figure 2a** Effect sizes per group for all therapy lengths



**Figure 2b** Effect sizes for short-term therapies



**Figure 2c** Effect sizes for long-term therapies



between feedback (either FbT or FbTP) and the patient being not on track.

Next, two additional models were analyzed, for short and long-term therapies 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. NOT cases have a negative change over time. In the FbT and FbTP conditions receiving feedback had respectively a large and a very large effect 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 OT cases (see Table 3). Feedback to patient and therapists had a small advantage over feedback to therapists alone in the NOT cases, the additional effect sizes are 0.34 at 26 weeks and 0.37 at 35 weeks.

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, equally of OT and NOT cases (see Table 4).

### End state functioning

Table 5 shows the OQ-45 scores at the end of treatment. Although there were no overall significant differences between conditions ( $\chi^2(6) = 8.01, p = 0.24$ ), there was a trend for the FbTP condition to have the best results: the lowest rate of deteriorated patients ( $Z = -1.3; p = 0.097$ ) were 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.

**Table 5** Clinical significant and reliable 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

## Discussion

### Summary of results

In this study we aimed to demonstrate the effect of feedback about patient progress to therapists and patients. As anticipated, feedback to both therapists and patients was most effective. The benefits were strongest for cases who were not progressing well in short-term therapies. Feedback provided to the therapist alone was effective of Not On Track (NOT) patients in short-term therapies. No significant effect of being NOT was found in the full sample. In long-term therapies only feedback to therapist and patient was effective. Feedback influenced that rate of change, but did not significantly improve end state functioning.

### Short and long term effects

The effects in the short-term 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 in the effect of feedback, compared to others. Their studies typically demonstrate that feedback is most effective for NOT cases (Lambert, 2007; Shimokawa, Lambert, & Smart, 2010b). Our overall effect of feedback was 0.15 after 26 weeks and 0.20 after 78 weeks, which is considered a small effect. However, if we look at the effect of feedback for NOT cases in the short-term therapy group – which most resembles Lambert's samples – the effects are very similar to what they found. That is, feedback to therapists and patients had a very large effect in this subgroup of patients and feedback to therapists had a large (after 26 weeks) to very large (after 35 weeks) effect. For these cases the feedback reduced the number of negative outcomes.

For feedback in long-term therapies, Knaup *et al.* (Knaup, et al., 2009) found no significant effect in their meta-analysis. In contrast, we did find a small but significant effect ( $d = 0.22$  after 26 weeks and  $d = 0.29$  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 includes 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 group among whom not much progress might be expected.

These findings raise the question of why feedback seems most effective in NOT cases in short-term therapies but not in long-term therapies. One explanation might be that in longer therapies therapists have more opportunities to identify and correct the negative track. The results show that more positive change trajectories are found in the NOT cases for the long-term therapies. Another issue is that patients who receive long-term therapy are possibly not the same type as patients who receive short-term therapies. Although we did not find relevant differences between patients in long and short-term therapy, patients may differ on unmeasured constructs that are related with the complexity of their complaints. For instance, patients that are not progressing well might have dropped out before 35 weeks in therapy and could be underrepresented in the long-term therapy group. An alternative interpretation is that the OQ-45 might be better at measuring domains that are likely to change in short-term therapies, such as symptoms, than domains that are targeted in long-term therapies, such as character changes.

### **Therapist versus patient feedback**

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.* (Hawkins, et al., 2004) study feedback to patients and therapists outperformed feedback to therapists alone, but the studies by Slade *et al.* (Slade, Lambert, Harmon, Smart, & Bailey, 2008b) and Harmon *et al.* (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 (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 center, whereas the studies of both Harmon *et al.* and Slade *et al.* were done in a university counselling center 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 thus, probably resembles our group more than the counselling center sample does.

One could wonder why feedback to therapist and patient shows a more pronounced effect than feedback to therapists alone. Since the therapist is the one providing the therapy, the added effect of providing feedback to patients could be low. There are a couple of explanations that could be viable. For instance, it could be a matter of implementation. The therapist knows that the patient sees the feedback too and this might encourage the therapist to look at the feedback as well. A recent study amongst therapists showed that two major barriers for therapists to look at the feedback were other tasks that demanded attention and not having enough time to look at

the feedback (De Jong, 2012). If the therapist knows that the feedback is not seen by the patient, looking at the feedback might be assigned a lower priority than other administrative tasks. Therapists may also experience resistance to being evaluated (Riemer & Bickman, 2011) and might avoid looking at the feedback as a result. When the patient receives feedback as well, they are not in a position to disregard the feedback, since they know the patient might ask about it.

Another explanation might be that patients are more empowered when they receive feedback about their own progress in therapy. Some of the therapists in our study indicated that providing feedback to patients may have resulted in an increased sense of ownership of their own change process. By receiving the feedback, patients might be more alarmed if there is a lack of progress and might actively discuss this with their therapist and manage their own lack of progress. In that way it 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 and the effect of feedback to therapists and patients might actually be due to a response shift. 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.

## Limitations

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 less 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.

A factor that might complicate the generalization of our results is that it is unclear to what extent our sample is selective. Therapists could have made self-selections of patients they approached to participate in the study and we had no way of checking this possibility. Similarly, not all therapists may have reported all patients that refused to participate in the study. The fact that we do not have sufficient insight in the selectiveness of the sample is mainly related to partnering with private practices rather than a single department in a mental health care institution. It was particularly complicated to get information from them on patients for whom the therapists provided treatment in the same time period, but who were not enrolled in the study.

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., 2008b), 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 in two groups post hoc, which resulted in equal group sizes and thus optimal power to detect an effect for both groups, but may have 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 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 for being NOT by having more sessions. Also, since we used treatment duration in weeks, this division does not tell us much about dosage. Short-term treatments might have had a higher density (e.g. weekly sessions rather than bi-weekly).

## **Implications for practice and research**

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. It should be noted though that although feedback did seem to prevent a negative treatment course, the effect was not strong enough to result in a positive treatment course for these patients (see Figure 2b). However, previous studies have showed that NOT cases have an increased risk of achieving negative treatment outcomes (Lutz, et al., 2006) and if feedback can help prevent that, it should be considered. The use of an expected recovery curve in the feedback model and using clinical support tools to help the therapist may improve the effect of the feedback, since NOT cases could be identified sooner, based on the deviation of their expected progress (Shimokawa et al., 2010).

Although more studies are emerging on the topic of feedback, there is still much we do not know about the subject yet. 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.* (Newnham, Hooke, & Page, 2010) showed for instance that in an acute clinic, feedback was only effective for depressed patients. A recent study by Bickman *et al.* (Bickman, Douglas Kelley, Breda, De Andrade, & Riemer, 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 therapist 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.

## Acknowledgements

*This study was supported by a grant from The Netherlands Organisation for Health Research and Development (ZonMW), grant number 94506414. The authors would like to thank Andrew Page for his comments on the first draft of this article.*

