The Effect of Patient Progress Feedback on Psychotherapy Outcome

Henning Schöttke Michael Unrath Christina Uhlmann

Poliklinische Psychotherapieambulanz, Institut für Psychologie, Klinische Psychologie und Psychotherapie, Universität Osnabrück, Osnabrück, Germany

Keywords

Clinical psychology · Psychotherapy · Feedback · Treatment · Outcome

Abstract

Background: There is evidence that outpatient psychotherapy is effective in treating mental disorders. However, 5–10% of the patients get worse during treatment. **Patients and Methods:** This randomised, controlled trial investigates the effect of systematic feedback of patients' progress to therapists on the outcome of psychotherapy. A total of 230 adult patients were randomly assigned to three conditions: progress monitoring with feedback, progress monitoring without feedback, no progress monitoring. Subsequently, they were treated with cognitive behavioral therapy in an outpatient clinic. Feedback was based on three self-report questionnaires: Questionnaire for the Evaluation of Psychotherapeutic Change (FEP-2), Outcome Questionnaire (OQ-30), and Assessment for Signal Cases (ASC). Variance analyses with repeated measurements were conducted to verify whether feedback enhances treatment outcomes. **Results:** There were no statistical differences between the three conditions with regard to sociodemographic and disease-relevant variables. Patients' impairment improved significantly in all three groups, but there were no significant differences between groups. Moreover, feedback had no differential effect on patients without initial improvement ($n = 49$). **Discussion and Conclusion:** In our study, collective patients' progress feedback to therapists had neither a global nor a differential effect on treatment outcomes. In order to better demon-

strate a potential feedback effect, future studies should aim for a higher frequency of feedback, measure therapists' attitude towards such feedback systems, and use "clinical support tools."

© 2019 S. Karger AG, Basel

Einfluss von Verlaufsfeedback auf die Behandlungseffektivität ambulanter Psychotherapien

Schlüsselwörter

Klinische Psychologie · Psychotherapie · Feedback · Behandlung · Effektivität

Zusammenfassung

Hintergrund: Obwohl die Wirksamkeit ambulanter Psychotherapie bei psychischen Erkrankungen hinreichend belegt wurde, zeigen 5–10% der Patienten eine Verschlechterung im Therapieverlauf. **Patienten und Methoden:** In dieser randomisierten kontrollierten Studie wurde untersucht, welchen Effekt eine formalisierte therapiebegleitende Rückmeldung an den Therapeuten über Veränderungen der psychischen Belastung, Beeinträchtigung und Therapiemotivation des Patienten auf den Therapieverlauf hat. In einer poliklinischen Psychotherapieambulanz wurden 230 Patienten auf drei Studiengruppen randomisiert verteilt: Verlaufsmessung mit Feedback (EG), Verlaufsmessung ohne Feedback (KG1),

keine Verlaufsmessung (KG2). Während der anschließenden verhaltenstherapeutischen Behandlung wurden die Verlaufsmessungen zur Erhebung des Therapiefortschritts regelmäßig mithilfe des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP-2), des Outcome Questionnaires (OQ-30) und des Assessment for Signal Cases (ASC) durchgeführt. In der EG wurden die Ergebnisse der Verlaufsmessungen den Therapeuten zurückgemeldet. Mittels Varianzanalysen mit Messwiederholungen wurde der Einfluss des Feedbacks auf die Wirksamkeit der Therapien überprüft. **Ergebnisse:** Die Patienten der drei Studiengruppen unterschieden sich nicht signifikant hinsichtlich soziodemografischer und erkrankungsrelevanter Variablen. Die psychische Beeinträchtigung aller Gruppen ging im Therapieverlauf zurück. Zwischen den Gruppen zeigten sich keine signifikanten Unterschiede. Bei Therapien, die bis zur ersten Verlaufsmessung keine Verbesserung aufwiesen ($n = 49$), zeigte sich ebenfalls kein differentieller Effekt des Feedbacks. **Diskussion und Schlussfolgerung:** Feedback brachte in diesem Studienkollektiv einer Ausbildungsambulanz weder einen globalen noch einen differentiellen Vorteil auf den Therapieverlauf. Um einen potentiellen Feedbackeffekt besser nachweisen zu können, sollten zukünftige Studien eine höhere Frequenz des Feedbacks anstreben, die Einstellung der Therapeuten gegenüber derartigen Feedbacksystemen erheben sowie "Clinical Support Tools" einsetzen.

© 2019 S. Karger AG, Basel

Introduction

The effectiveness of psychotherapy has been adequately proven scientifically for a broad spectrum of mental disorders, yet 5–10% of patients get worse during the course of treatment [Lambert and Ogles, 2004; Mohr, 1995]. An estimated one-third of patients do not respond to treatment or get worse [Rubel et al., 2017]. Patient-oriented psychotherapy research, therefore, seeks to improve the quality of therapeutic care on an individual basis [Howard et al., 1996; Lutz, 2002]. Therapist assessments rarely accurately predict negative outcomes of treatment [Hannan et al., 2005]. Formalized feedback to the therapist about changes in the patient's psychological burden or other relevant factors could enhance treatment success [Harmon et al., 2005; Slade et al., 2008]. Feedback systems based on psychometric data have been used in both inpatient and outpatient settings [Berking et al., 2006; De Jong et al., 2014; Delgadillo et al., 2018; Hawkins et al., 2004; Lambert et al., 2001, 2002b; Strauss et al., 2015; Whipple et al., 2003]. In the Outcome Questionnaire (OQ) system [Lambert et al., 2002a, 2004a], patients regularly fill in standardized questionnaires as their therapy proceeds, which cover, among other things, their mental health on the symptom level,

social functioning, and quality of interpersonal relationships. In German-speaking countries, the Fragebogen zur Evaluation von Psychotherapieverläufen (Questionnaire for the Evaluation of Psychotherapeutic Change, FEP-2) [Lutz and Böhnke, 2008] takes a similar approach. The therapist receives feedback in the form of a psychometric evaluation of the recorded items. In the OQ system, a problematic course of therapy is defined on the basis of the changes between therapeutic sessions, defined by established algorithms ("not-on-track" therapies). Specific suggestions have been made on how the therapy could be made more successful with a so-called "Clinical Support Tool" [Lambert et al., 2007; Lambert, 2013], based on evaluation of potential reasons for the progression (e.g., lack of therapeutic motivation, poor therapeutic alliance, inadequate social support, or critical life events). Research findings on the effects of psychometric feedback are inconsistent. Some studies have found that feedback has a global positive effect on various measures of efficacy [Amble et al., 2015; Berking et al., 2006; Hawkins et al., 2004]. Other studies in the USA and Europe have failed to show this effect [Probst et al., 2014; Strauss et al., 2015] or found a smaller, more differential effect, such that patients whose therapy was not progressing well could benefit especially from formalized feedback [De Jong et al., 2014; Lambert et al., 2001, 2002b; Simon et al., 2012; Whipple et al., 2003]. A recent English study [Delgadillo et al., 2018], using a cluster-randomized design with $n = 77$ psychotherapists and $n = 2,233$ outpatients with anxiety and depressive symptoms, also showed a significant beneficial effect on outcomes (effect sizes $d = 0.23$ or $d = 0.19$) from psychometric feedback for "not-on-track" therapies. Another recent systematic review concluded that a global beneficial effect for routinely applied feedback systems cannot currently be proven, but also established their differential favorable effect for "not-on-track" therapies at similar effect sizes (effect size $d = 0.22$) [Kendrick et al., 2016]. An even more recent systematic review likewise estimated the differential effect to be small (effect size $d = 0.33$) [Lambert et al., 2018].

There has so far been only one large cluster-randomized study for outpatient care in Germany, with $n = 403$ therapists; it was unable to prove a global effect of feedback on the course of therapy [Strauss et al., 2015; Wittmann et al., 2011]. They used various instruments to ascertain the symptom burden in routine care, cooperating with a large statutory health insurance ("TK Study") to provide feedback to the therapist on the progression of the therapy, at specific points in time. In this and almost all previous studies, the effect of the explicit feedback to the therapist could not be separated from the effect of repeated progress monitoring – i.e., repeated questioning of the patients about their health and factors relevant to their therapy. However, since it seems plausible that re-

peated questioning of patients might cause changes in and of itself [Durham et al., 2002], in our study, we introduced an additional control group, with repeated progress monitoring but no feedback.

The goal of the present study was to investigate whether feedback to the therapist has a global positive effect on the outcome of outpatient psychotherapy, or whether it is primarily a differential effect that applies exclusively to “problem therapies.” Further research should explore whether feedback actually has a positive effect beyond that of repeated progress monitoring.

Method

Design and Context

The effect of feedback on the progress of outpatient psychotherapy patients was studied in a randomized controlled trial, by comparing the feedback as an intervention condition (experimental group, EG) with 2 different control conditions (control group, CG1: progress monitoring without feedback; CG2: no progress monitoring). The study was performed at the Outpatient Psychotherapy Clinic of the University of Osnabrück, with a total of $n = 20$ study therapists. Of these, $n = 15$ study therapists (75%) were psychologists in training to become psychological psychotherapists, with a focus on behavioral therapy ($n = 13$ female, $n = 2$ male). Their treatment of the patients in the context of this training program took place under the statutory conditions prescribed by the Ausbildungs- und Prüfungsverordnung für Psychologische Psychotherapeuten (Training and Qualifying Examinations for Psychological Psychotherapists, PsychTh-AP) and the Psychotherapy Guidelines. At the time of the study, the therapists in training had already completed at least half of the course, including Practical Activities I and II. The other treatments of patients that were included in the study were performed by $n = 5$ licensed psychotherapists in the Outpatient Research Department of the university ($n = 4$ female and $n = 1$ male). On average, the training therapies were supervised by an expert after every fourth therapy session.

In the indication phase before the probational sessions begin, the patients are given an admission interview, a structured clinical interview (SCID-I), and psychometric assessments. Exclusion criteria for treatment in the Outpatient Psychotherapy Clinic and thus for participation in the study were: acute suicidality, acute (physical) violence, an acute organic mental disorder (ICD-10 F00–F09), an acute psychotropic substance disorder (F10–F19), or an acute psychotic disorder (F20–F29). After the admission interview and before the first probational session, research associates at the clinic collected the psychometric data and conducted the structured clinical interview. To participate in the study, the patients were consecutively referred to the therapists from a waiting list and were randomly distributed to the study groups (EG, CG1, CG2) (randomization at the individual patient level). A ratio of the group sizes of 2 (EG):1 (CG1):1 (CG2) was sought, so that when the 2 control groups were combined in the statistical analyses, they would be equivalent to the size of the experimental group. All patients gave their written consent to participate in the study.

Data Collection, Instruments, and Measurements

The patients of all the study conditions filled out the following questionnaires before the start of therapy (measurement point T1, before the first therapy session) and at the end of therapy (measurement point T3, e.g., within the last 3 sessions before the end of

therapy): FEP-2 [Lutz and Böhnke, 2008], OQ-30 [Lambert et al., 2002a, 2004a], Beck Depression Inventory (BDI-2) [Hautzinger et al., 2006], Symptom Checklist-90 (SCL-90-R) [Franke, 2002], and Patient Health Questionnaire (PHQ-D) [Löwe et al., 2002]. In the groups where there was progress monitoring (EG, CG1), data were also collected at the end of each calendar quarter by means of the FEP-2, OQ-30, and the German version of the Assessment for Signal Cases (ASC) [Lambert et al., 2007; Lambert, 2013]. Only the EG received the quarterly Questionnaire on Use of Feedback [Rubel et al., 2017].

The FEP-2 is a multidimensional, change-sensitive instrument that ascertains the progress of psychotherapy in detail. It consists of 40 items that are grouped into 4 scales (well-being, symptoms, interpersonal relationships, goal congruence). A total score of mental impairment can also be determined, which in the present study is the main target criterion for assessing the efficacy of the therapy. The retest reliability (non-clinical sample) for the total score of mental impairment is $r = 0.69$. The internal consistency is given as between $\alpha = 0.74$ and 0.94 [Lutz et al., 2009].

The OQ-30 records the general life impairment, with 30 items. It is structured in several subscales, and a total score can also be calculated. Scores of 0.59 – 0.93 were reported for internal consistency and scores of 0.71 – 0.87 for retest reliability [Lambert et al., 2002a].

The BDI-2 is a self-description tool with 21 items to ascertain the severity of depressive symptoms. Internal consistency, according to Cronbach's α , is above 0.89 . Retest reliability is $r = 0.78$ [Hautzinger et al., 2006].

The SCL-90-R is a self-assessment instrument for ascertaining the subjective impairment by physical and psychological symptoms, using 90 items. In addition to scoring the individual scale values for 9 symptom scales (e.g., depression, somatization), total scores for psychological burden can be determined. Internal consistency, according to Cronbach's α , for the 9 scales is between 0.78 and 0.91 , and that of the Global Severity Index (GSI), which measures the average psychological burden, is 0.98 . Retest reliability is between $r = 0.74$ and $r = 0.88$ [Franke, 2002].

The PHQ-D is a self-assessment questionnaire that is based on the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [Sass et al., 2003] and is used in the form of a screening. It includes questions about somatoform disorders, depressive disorders, eating disorders, panic and other anxiety disorders, as well as alcohol abuse. The scale sum scores can be used as an estimate of the severity of the disorder. Internal consistency, according to Cronbach's α , was 0.88 for the depression scale and 0.79 for the somatization scale [Gräfe et al., 2004].

The ASC identifies the area of the patient's problems that may prevent therapeutic success [Lambert, 2013]. It comprises 40 items, which can be assigned to the 4 subscales therapeutic alliance, therapy motivation, social support, and stressful life events. There are no quality criteria for the German translation of this instrument yet.

The Questionnaire on Use of Feedback records the therapists' information about how they used the psychometric feedback during the course of therapy. For this purpose, therapists were presented with a list of 11 options, beginning every calendar quarter. These items describe possible adaptations of therapeutic behavior based on the feedback (sample item: I have tried to increase the patient's motivation because of the feedback). A more detailed description of the questionnaire is to be found in Rubel et al. [2017]. In this study, the percentage frequency of use of each of the 11 intervention adjustments was estimated by assessing for the EG per therapy, whether the opportunity was used in at least one quarter or was not used. The frequencies of use per intervention were then divided by the number of therapies performed in the EG after the probational sessions.

Fig. 1. An example of graphical feedback from the FEP-2 total score. The therapist can obtain the following information from this feedback: the absolute change in the FEP-2 score over time, information on the statistical significance of the changes and the clinical relevance of the remaining symptoms. Feedback from the OQ-30 was treated similarly.

Fig. 2. An example of feedback from the ASC at 2 measurement points, consisting of a descriptive progression graph of the scale means plus verbal information. Higher scores on the ASC represent an improvement. The therapist can obtain the following information from this feedback: The figure shows that the patient is improving between the 2 points of ASC data collection (1st and 2nd quarter, 2011), e.g., on the scale "social support," but has gotten worse on the scale "therapeutic motivation." In addition, the items are included in the feedback that the patient evaluated extremely negatively at the points of data collection.

Feedback

At the beginning of each calendar quarter, all therapists in the EG received written feedback in the form of an evaluation of the previous quarter's progress monitoring from the FEP-2, OQ-30, and ASC on possible causes of the lack of therapeutic success (Fig. 1, 2). The evaluation comprised both text and graphics. Figure 1 shows an example of feedback from the FEP-2 total score. The therapist can use this feedback to determine the absolute change in the FEP-2 over time, as well as to gain information on the statistical significance of the changes and the clinical relevance of the remaining symptoms. Feedback was similarly handled by the OQ-30 and the ASC (Fig. 2). The changes were rated according to the methods of clinical significance [Jacobson and Truax, 1991] and written up for the therapists in a patient-oriented individual case evaluation. There was no feedback in the other 2 study groups (CG1, CG2).

Statistical Analyses

A power analysis was conducted a priori with the program G*Power version 3.1.9.2. This resulted in an assumed smallest detectable effect size of $d = 0.20$ [Poston and Hanson, 2010], a sample size of $n = 195$ for a comparison by ANOVA (3 groups, 2 measurement points), with a statistical power of $1 - \beta = 0.95$ and an alpha error of $\alpha = 0.05$.

The statistical analyses were done with IBM® SPSS® Statistics [IBM Corp., 2016] and in the form of intention-to-treat analyses. This means that the analyses were based on randomized grouping, regardless of whether the patient dropped out during the probational session or the therapy, which would have meant that no feedback could be used at all. This conservative approach was deliberately chosen to prevent any distortions due to systematically occurring dropouts and to avoid overestimating the effect of the intervention [Weiss, 2008]. In addition to descriptive statistics, inferential statistical methods were used, such as χ^2 tests (or Fisher's exact test) and univariate analyses of variance.

In order to best account for the assumed hierarchical data structure (each therapist conducted multiple therapies), a linear multilevel model was tested, which included the FEP-2 total score at T3 as the target variable. The fact that in some cases very few observation units existed per therapist was problematic for a reliable model evaluation, so 5 therapists each conducted a maximum of 3 therapies. The analyses showed that only a very small proportion of variance, not significantly different from 0, could be explained by differences among the therapists ($ICC = 0.0219$; $p = 0.482$). Also, the model quality did not improve significantly with the introduction of a random intercept (Likelihood Ratio Test with

Fig. 3. Result of random assignment of the patients to the 3 study groups EG, CG1, and CG2, taking into account the lacking indications in the probational phase and the dropouts from therapy.

$df = 1$; $\chi^2 = 0.735$; $p = 0.391$). Therefore, we gave up using hierarchical linear models. Instead, we used one-way repeated measure analysis of variance, with the FEP-2 total score before start of therapy (T1) as the baseline value, to examine the impact of feedback on the efficacy of the therapy. The main target criterion was the FEP-2 total score at the end of therapy (T3).

In addition to analysis of the total sample, equivalent repeated measure analyses of variance were performed in a defined subgroup of therapies to test the differential efficacy of the intervention for difficult therapeutic progressions. This subgroup of “not-on-track” therapies was defined based on the difference between the OQ-30 total score in the first quarter after start of therapy and the OQ-30 total score before the start of therapy. A stable or higher OQ-30 total score at the end of the first quarter was classified as a difficult course of therapy, in that there was no early improvement. The chosen definition reflects research findings that early changes in the course of therapies significantly predict the later course of therapy [Flückinger et al., 2013]. The method of clinical significance was not used in these more exploratory subgroup analyses in order to ensure a sufficient group size.

Missing values in the main target criterion (total score of FEP-2) were replaced in all 3 groups by the last observation carried forward (LOCF) or, in the case of a single measured value, the baseline observation carried forward (BOCF) to counteract a loss of power on account of dropouts. Simply replacing missing values with the LOCF poses the risk of distorted effect estimators and overestimated measurement accuracy [Pedersen et al., 2017], so additional sensitivity analyses were performed (complete case analysis [CoCa], i.e., cases without missing values in the main target criterion, and multiple imputation [MI]), in order to estimate possible distortions of the results due to the handling of missing values.

Results

There were $n = 230$ patients included in the study (EG: $n = 103$, CG1: $n = 58$, CG2: $n = 69$), and $n = 39$ patients dropped out during the probation sessions and $n = 39$

during therapy (without or with final measurement, according to the main target criterion). Figure 3 illustrates the result of randomization. The patients did not differ significantly when they were first included in the study with respect to sociodemographic variables, number of diagnoses per patient, and severity of initial distress. The number of therapy sessions required and the proportion of patients treated by therapists in training versus licensed psychotherapists also did not differ among the 3 study conditions (Table 1). In total, $n = 188$ therapies (81.7%) were conducted by therapists in training and $n = 42$ (18.3%) by licensed psychotherapists. 47.8% of patients ($n = 110$) had a comorbidity (≥ 2 diagnoses) with an average of 1.6 diagnoses ($SD = 0.75$) per patient. The main diagnosis made in the samples was as follows: 44.3% ($n = 102$) affective disorders (ICD-10 F3x.x), 43.5% ($n = 100$) neurotic, stress-related, and somatoform disorders (ICD-10 F4x.x), 6.1% ($n = 14$) personality disorders (F6x.x), 4.8% ($n = 11$) behavioral disorders with physical disorders (F5x.x), 0.9% ($n = 2$) psychotic disorder (F2x.x), 0.4% ($n = 1$) organic disorders (ICD-10 F0x.x). The distribution of the main diagnoses did not differ significantly among the 3 study groups ($df = 10$, $\chi^2 = 15.042$; $p = 0.131$). On average, each patient received about 30 sessions of behavioral therapy. In total, $n = 72$ patients (31.3%) were missing a value in the main target criterion (FEP-2 total score at T3). The proportion of missing values was similarly distributed among the groups (EG: $n = 36$ [35.0%], CG1: $n = 16$ [27.6%], CG2: $n = 20$ [29.0%]). There were no significant correlations between the lack of a value and participation in the study, the level of initial distress according to FEP-2, or the other measures of distress. For the study participants who did not have missing values in

Table 1. Sociodemographic variables, treatment variables, and psychological burden before the start of therapy and at the end of therapy

Characteristics*	Categories	EG (<i>n</i> = 103)	CG1 (<i>n</i> = 58)	CG2 (<i>n</i> = 69)	Total (<i>n</i> = 230)
Gender, %	male	25.2 (<i>n</i> = 26)	29.3 (<i>n</i> = 17)	31.9 (<i>n</i> = 22)	28.3 (<i>n</i> = 65)
	female	74.8 (<i>n</i> = 77)	70.7 (<i>n</i> = 41)	68.1 (<i>n</i> = 47)	71.7 (<i>n</i> = 165)
Marital status, %	married	35.0 (<i>n</i> = 36)	29.3 (<i>n</i> = 17)	26.1 (<i>n</i> = 18)	30.9 (<i>n</i> = 71)
	divorced/separated	12.6 (<i>n</i> = 13)	20.7 (<i>n</i> = 12)	14.5 (<i>n</i> = 10)	15.2 (<i>n</i> = 35)
	single	48.5 (<i>n</i> = 50)	48.3 (<i>n</i> = 28)	58.0 (<i>n</i> = 40)	51.3 (<i>n</i> = 118)
	widowed	3.9 (<i>n</i> = 4)	1.7 (<i>n</i> = 1)	1.4 (<i>n</i> = 1)	2.6 (<i>n</i> = 6)
Partnership, %	stable life partner	58.3 (<i>n</i> = 60)	63.8 (<i>n</i> = 37)	65.2 (<i>n</i> = 45)	61.7 (<i>n</i> = 142)
	no stable life partner	38.8 (<i>n</i> = 40)	34.5 (<i>n</i> = 20)	31.9 (<i>n</i> = 22)	35.7 (<i>n</i> = 82)
High school diploma, %	(vocational) high school diploma	41.7 (<i>n</i> = 43)	46.6 (<i>n</i> = 27)	42.0 (<i>n</i> = 29)	43.0 (<i>n</i> = 99)
	intermediate secondary school	18.4 (<i>n</i> = 19)	13.8 (<i>n</i> = 8)	13.0 (<i>n</i> = 9)	15.7 (<i>n</i> = 36)
	secondary general school	1.0 (<i>n</i> = 1)	–	–	0.4 (<i>n</i> = 1)
	no secondary school diploma	1.0 (<i>n</i> = 1)	1.7 (<i>n</i> = 1)	2.9 (<i>n</i> = 2)	1.7 (<i>n</i> = 4)
	still in school	3.9 (<i>n</i> = 4)	–	–	1.7 (<i>n</i> = 4)
	other	34.0 (<i>n</i> = 35)	37.9 (<i>n</i> = 22)	42.0 (<i>n</i> = 29)	37.4 (<i>n</i> = 86)
Age, years, M; SD	–	35.6; 13.3	36.4; 12.3	36.2; 12.2	36.0; 12.7
Training therapy, %	yes	86.4 (<i>n</i> = 89)	82.8 (<i>n</i> = 48)	84.1 (<i>n</i> = 58)	84.8 (<i>n</i> = 195)
	no	13.6 (<i>n</i> = 14)	17.2 (<i>n</i> = 10)	15.9 (<i>n</i> = 11)	15.2 (<i>n</i> = 35)
Diagnoses per patient at start of therapy, M; SD	–	1.6; 0.8	1.6; 0.7	1.7; 0.8	1.6; 0.8
Therapy sessions, M, SD	–	29.3; 22.8	32.2; 23.8	30.0; 20.6	30.2; 22.4
FEP-2, M; SD	–	2.8; 0.7	2.8; 0.8	2.7; 0.7	2.8; 0.7
OQ-30, M; SD	–	1.7; 0.6	1.7; 0.7	1.7; 0.6	1.7; 0.6
GSI/SCL-90-R, M; SD	–	1.0; 0.7	1.1; 0.7	1.0; 0.6	1.0; 0.7
BDI-2, M; SD	–	22.3; 11.2	22.5; 11.6	21.3; 11.9	22.0; 11.5
FEP-2 at T3, M; SD	–	2.0; 0.6	2.0; 0.7	2.2; 0.6	2.1; 0.6
OQ-30 at T3, M; SD	–	1.0; 0.6	1.0; 0.7	1.1; 0.6	1.0; 0.6

M, mean; SD, standard deviation; GSI, Global Severity Index of the SCL-90-R. * Missing: *n* = 1 for variables OQ-30, GSI, BDI-2 at T1; *n* = 6 for variable “partnership” at T1; *n* = 72 for variable FEP-2 at T3; *n* = 73 for variable OQ-30 at T3.

the main target criterion (*n* = 158), there was a change between T1 and T3, with the effect size $d = 0.96$ [Morris and DeShon, 2002]. In the EG, *n* = 19 participants dropped out during the probational period, so feedback could not be used at all. In the other *n* = 84 therapies, feedback of *n* = 62 therapists (73.8%) was used at least once during the course of therapy to change the therapeutic care. The most frequent forms of feedback use were discussion of the questionnaire evaluation with the patients (*n* = 45, 53.6%), the adaptation of therapeutic interventions (*n* = 42, 50.0%), and encouragement of the patient's competencies (*n* = 36, 42.9%). Figure 4 illustrates the proportion of therapists who used the feedback in various ways, as well as the results of a feedback study in the Outpatient Training Department of the University of Trier [Rubel et al., 2017]. Although the frequency and type of feedback differ significantly in the 2 studies, a similar pattern of feedback usage can be described. Frequency differences between the 2 studies appear to lie “in the at-

tempt to improve the therapeutic relationship,” “to encourage the patient's resources,” and “to discuss the results of the questionnaires with the patient.”

Global Effect of Feedback

A one-way repeated measure analysis of variance (2 measurement points: T1, T3) examined the effect of feedback on the total score for psychological impairment using FEP-2. If all therapies were included in the analyses (*n* = 230; imputation using LOCF/BOCF), a significant effect for the time factor was found ($F[1.000; 227.000] = 122.317$; $p < 0.001$; effect size $\eta^2 = 0.350$). Psychological impairment decreased across all groups as therapy proceeded (Fig. 5). There were no significant differences, however, between the groups ($F[2.000; 227.000] = 0.390$; $p = 0.677$; effect size $\eta^2 = 0.003$), and there was also no significantly different progression of the FEP-2 total score in the groups over time, i.e., no time $\times$ group interaction ($F[2.000; 227.000] = 0.342$; $p = 0.711$; effect size $\eta^2 = 0.003$).

Fig. 4. Relative frequency of different types of feedback in the EG over the course of therapy, presented for the therapies that were actually initiated after the probational phase ($n = 84$), compared to another German study.

In the 2 groups that received progress monitoring (EG, CG1), additional repeated measure analyses of variance were performed that included more measurement time points (LOFC/BOCF) (Fig. 6). Here too, there were significant effects only for the time factor, with effect sizes between $\eta^2 = 0.189$ (5 measurement points) and $\eta^2 = 0.349$ (2 measurement points).

Sensitivity analyses included only therapies that had values for the main target criterion (FEP-2 total score at T3) (CoCa, $n = 158$). This led to very similar results, both in the comparison of all 3 groups with respect to T1 and T3, as well as with inclusion of up to 5 measurement points for groups EG and CG1 ($n = 109$ to $n = 97$, depending on the number of measurement points). Although missing values for the main target criterion were replaced by MI ($n = 230$), the same pattern of results was found in all analyses of variance as was found for imputation using LOCF/BOCF. Table 2 gives an overview of the effect estimators for the comparison between the 3 study groups, taking into account the different methods for dealing with missing values.

Differential Effect of Feedback for Therapies with Initial Non-Improvement

In $n = 49$ of the EG and CG1 therapies (30.4% of all therapies in these groups), there was no improvement in the OQ-30 total score in the first quarter after start of therapy, so that these were defined as “not-on-track” therapies. This accounted for $n = 26$ for the EG and $n = 23$ for the CG1. Independently of the number of measurement points considered ($n = 2$ to $n = 5$), the repeated measure analyses of variance performed within this

Fig. 5. Change in the FEP-2 score from T1 to T3 in the 3 comparison groups, EG, CG1, and CG2.

subgroup had a significant effect only on the change in FEP-2 total score over time, but no significant differences between groups and no significantly different progression over time for the feedback group. Figure 7 illustrates the progression of the FEP-2 total score with inclusion of 5 measurement points (LOCF/BOCF; $n = 49$; $F[2.940; 138.175] = 12.089$; $p < 0.001$; effect size $\eta^2 = 0.205$).

Sensitivity analyses were performed only for therapies with complete diagnostics (CoCa) or if the missing main target criterion was replaced with MI. Both approaches yielded only marginally different results for the LOCF/BOCF analysis.

Fig. 6. The FEP-2 total score in EG and CG1 over time, with up to 5 measurement points. PM, progress monitoring.

Table 2. Results of one-way repeated measure analyses of variance for comparison of the 3 study groups, broken down by method for dealing with missing values

Effect	df, numerator	df, denominator	F value	Effect estimator η^2	p value
LOCF/BOCF ($n = 230$)					
Time	1	227	122.317	0.350	0.000
Group $\times$ time	2	227	0.342	0.003	0.711
Group	2	227	0.390	0.003	0.677
CoCa ($n = 158$)					
Time	1	155	151.087	0.494	0.000
Group $\times$ time	2	155	0.232	0.003	0.794
Group	2	155	0.006	0.006	0.645
MI ($n = 230$)					
Time	1	227	221.978	0.494	0.000
Group $\times$ time	2	227	0.912	0.008	0.403
Group	2	227	0.317	0.003	0.729
df, degrees of freedom.					

Discussion

We found neither a global positive effect of progression feedback on the efficacy of the therapies nor a differential positive effect, especially in the group of patients with initial non-improvement (“not-on-track” therapies). There were neither significant group differences nor significant effect modifications over time. Thus, progress monitoring without feedback had no advantage over a control group without progress monitoring.

Other studies using the OQ system with outpatient and inpatient psychotherapy patients in the USA and Eu-

rope have so far mostly shown small effects, especially or exclusively for patients with problematic courses of therapy [De Jong et al., 2014; Delgadillo et al., 2018; Probst et al., 2014; Simon et al., 2012; Strauss et al., 2015]. The results of 2 recent systematic reviews [Kendrick et al., 2016; Lambert et al., 2018] likewise point to at most a very small, clinically questionable relevant global feedback effect, but a small but distinct differential feedback effect. One explanation for the lack of a positive effect from feedback in the present study may be in its quite different framework conditions compared to previous studies. The first feedback studies were conducted at an American university counseling center [Shimokawa et al., 2010]. The patients treated there were young, educated, and received only a small average number of therapy sessions (5–7), suggesting a relatively low degree of impairment. In studies outside the German-speaking world, not only the professions of the treating therapists were heterogeneous [Amble et al., 2015; De Jong et al., 2014; Lambert et al., 2001, 2002b; Simon et al., 2012], but also the therapeutic methods [Amble et al., 2015; De Jong et al., 2014; Hawkins et al., 2004; Lambert et al., 2001, 2002b; Simon et al., 2012].

The therapies in the present study, however, were predominantly long-term outpatient therapies performed in the context of in-service training of psychological psychotherapists with a focus on behavioral therapy. The legal requirements for training therapies in Germany, such as regular supervision by experienced psychotherapists and the possibility of using videos of the therapy sessions for preparation and debriefing, are among the essential therapy-facilitating measures of these training therapies. This distinctive feature could have had an effect in our study, by aligning the treatment effects across the groups. The generalizability of the results may be limited. On the other hand, the findings on therapists’ usage of feedback

Fig. 7. FEP-2 total score in the group of patients with initial non-improvement (“not-on-track” therapies) at 5 measurement points. PM, progress monitoring.

show that the therapists are trying to use it in the therapeutic process and that this is done according to a similar pattern of the use of therapy adjustments as at another university training center, with evidence of feedback effects.

Another explanation for the lack of feedback effects in our study, especially in comparison to the study by Delgadillo et al. [2018] in which there was also regular collegial supervision, might be the low frequency of the surveys of progress and of the feedback. Progress monitoring occurred once per quarter in the present study, whereas other studies conducted weekly assessments [Delgadillo et al., 2018]. Sapyta et al. [2005] and Lambert [2015] also emphasize the impact of regular and direct feedback on therapeutic success. While Rubel et al. [2017] gave immediate feedback after each therapy session, the present study only reports the effect of multiple therapy sessions in the quarter.

Our study also differs from previous research with respect to the definition of at-risk patients, i.e., patients who have not progressed in the course of psychotherapy, which makes comparison difficult. We identified at-risk patients on the basis of their initial development in the psychotherapy, because changes in the early phase of therapy in particular contribute decisively to clinical outcome [Howard et al., 1993; Flückinger et al., 2013]. While the recent study by Delgadillo et al. [2018] and other studies defined clear statistical cut-offs for significant deterioration, in our study, patients were also defined as “not-on-track” who displayed no statistically significant deterioration. The reason for this was the insufficient number of patients with statistically significant failure to improve. But despite the broader defini-

tion of “not-on-track” therapies, the size of this subgroup ($n = 49$) is still too small to statistically verify small feedback effects.

The strengths of the present study are its design as a randomized clinical trial; the use of standardized, valid, and reliable measuring instruments; and appropriate methods of statistical analysis, including sensitivity analyses. Additional strengths are the sample size based on power calculations and the aims of the study, to replicate previous findings and to check the effect of repeated progress monitoring (2 control groups). This is an important contribution [Lambert, 2015], especially in light of the work’s critical examination of the stability of the research and treatment results and the failure to replicate empirical findings.

A limitation of the data is the relatively high proportion of missing values in the primary outcome ($n = 72$; 31.3%). The very similar and stable results obtained by the different imputation methods and the CoCa, however, suggest that the missing values did not cause any significant distortions of the results. There is also no evidence of a systematic pattern of missing values. Furthermore, the intention-to-treat analyses were conducted very conservatively, in that patients were included who had no indication for outpatient psychotherapy in the probational phase. If only the patients with an indication for outpatient psychotherapy are considered, the dropout rate is 17% ($n = 39$), which does not deviate from common dropout rates [Lutz et al., 2015b; Strauss et al., 2015].

Another limitation relates to the lack of information about the participating therapists’ attitude toward using feedback. Several authors have come to the conclusion, based on previous studies, that feedback is not always helpful, but the attitude or personal commitment of the therapist is a significant modulating factor [Ioannidis et al., 2014; Lutz et al., 2015a, 2015b; Miller et al., 2015], which may also be an explanation for the lack of effects. The proportion of therapists who indicated in our study that they used the feedback in various ways, however, does not appear to be lower than in the only large German study on the use of feedback in outpatient psychotherapy (TK Study) [Lutz et al., 2011].

Future studies should assess the therapist’s attitude toward the use of feedback systems in more detail and amend their a priori reservations about feedback systems. Feedback should also be given as frequently and promptly as possible, and sample sizes should be large enough to perform subgroup analyses (for “not-on-track” therapies) with sufficient power. So-called “clinical tools” [Lambert et al., 2007] could also be used to increase the positive effects of feedback, by giving therapists concrete instructions for changing the therapeutic procedure, in addition to the feedback itself. These scientifically sub-

stantiated recommendations are derived from the psychometric data of the progress monitoring. Several studies have already shown increased effectiveness with the use of “clinical tools” [Lambert, 2013; Shimokawa et al., 2010]. In summary, important insights are still lacking that would help us understand and evaluate the effectiveness of therapeutic feedback.

Translated by Susan Welsh

Statement of Ethics

All patients gave their written consent to participate in the study.

Disclosure Statement

The authors declare that they have no conflicts of interest.

References

- Amble I, Gude T, Stubdal S, Andersen BJ, Wampold BE. The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. *Psychother Res*. 2015;25(6):669–77.
- Berking M, Orth U, Lutz W. Wie effektiv sind systematische Rückmeldungen des Therapieverlaufs an den Therapeuten? Eine empirische Studie in einem stationär-verhaltenstherapeutischen Setting. *Z Klin Psychol Psychother*. 2006;35(1):21–9.
- De Jong K, Timman R, Hakkaart-Van Roijen L, Vermeulen P, Kooiman K, Passchier J, et al. The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: a randomized controlled trial. *Psychother Res*. 2014;24(6):629–39.
- Delgadillo J, de Jong K, Lucock M, Lutz W, Rubel J, Gilbody S, et al. Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *Lancet Psychiatry*. 2018 Jul;5(7):564–72.
- Durham CJ, McGrath LD, Burlingame GM, Schaalje GB, Lambert MJ, Davies DR. The effects of repeated administrations on self-report and parent-report scales. *J Psychoed Assess*. 2002;20(3):240–57.
- Flückiger C, Grosse Holtforth M, Znoj HJ, Caspar F, Wampold BE. Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychother Res*. 2013;23(1):1–13.
- Franke GH. SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version - Manual. Göttingen: Beltz; 2002.
- Gräfe K, Zipfel S, Herzog W, Löwe B. Screening psychischer Störungen mit dem Gesundheitsfragebogen für Patienten (PHQ-D). *Diagnostica*. 2004;50(4):171–81.
- Hannan C, Lambert MJ, Harmon C, Nielsen SL, Smart DW, Shimokawa K, et al. A lab test and algorithms for identifying clients at risk for treatment failure. *J Clin Psychol*. 2005 Feb;61(2):155–63.
- Harmon C, Hawkins EJ, Lambert MJ, Slade K, Whipple JS. Improving outcomes for poorly responding clients: the use of clinical support tools and feedback to clients. *J Clin Psychol*. 2005 Feb;61(2):175–85.
- Hautzinger M, Keller F, Kühner C. Beck Depressions-Inventar (BDI-II). Revision. Frankfurt/Main: Harcourt Test Services; 2006.
- Hawkins EJ, Lambert MJ, Vermeersch DA, Slade KL, Tuttle KC. The therapeutic effects of providing patient progress information to therapists and patients. *Psychother Res*. 2004;14(3):308–27.
- Howard KI, Lueger RJ, Maling MS, Martinovich Z. A phase model of psychotherapy outcome: causal mediation of change. *J Consult Clin Psychol*. 1993 Aug;61(4):678–85.
- Howard KI, Moras K, Brill PL, Martinovich Z, Lutz W. Evaluation of psychotherapy. Efficacy, effectiveness, and patient progress. *Am Psychol*. 1996 Oct;51(10):1059–64.
- IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.
- Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014 Jan 11;383(9912):166–75.
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991 Feb;59(1):12–9.
- Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database Syst Rev*. 2016 Jul;7(7):CD011119.
- Lambert MJ, Bailey RJ, Kimball K, Shimokawa K, Harmon SC, Slade K. Clinical Support Tools Manual – Brief Version-40. Salt Lake City: OQMeasures LLC; 2007.
- Lambert MJ, Hannöwer W, Nisslmüller K, Richard M, Kordy H. Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). *Z Klin Psychol Psychother*. 2002a;3(1):40–7.
- Lambert MJ, Ogles BM. The efficacy and effectiveness of psychotherapy. In: Lambert MJ, editor. Bergin and Garfield's handbook of psychotherapy and behavior change. New York: Wiley; 2004a. pp. 139–93.
- Lambert MJ, Vermeersch D, Brown GC, Burlingame G. Administration and Scoring Manual for the OQ-30.2. OQ Measures [Internet]. 2004b. Available from: <http://www.oqmeasures.com/measures/adult-measures/oq-30/>
- Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy (Chic)*. 2018 Dec;55(4):520–37.
- Lambert MJ, Whipple JL, Smart DW, Vermeersch DA, Nielsen SL, Hawkins EJ. The effects of providing therapists with feedback on patient progress during psychotherapy: are outcomes enhanced? *Psychother Res*. 2001 Mar;11(1):49–68.
- Lambert MJ, Whipple JL, Vermeersch DA, Smart DW, Hawkins EJ, Nielsen SL, et al. Enhancing psychotherapy outcomes via providing feedback on client progress: a replication. *Clin Psychol Psychother*. 2002b;9(2):91–103.
- Lambert MJ. Outcome in psychotherapy: the past and important advances. *Psychotherapy (Chic)*. 2013 Mar;50(1):42–51.
- Lambert MJ. Progress feedback and the OQ-system: the past and the future. *Psychotherapy (Chic)*. 2015 Dec;52(4):381–90.
- Löwe B, Spitzer RL, Zipfel S, Herzog W. Gesundheitsfragebogen für Patienten (PHQ-D). Komplette Version und Kurzform. Testmappe mit Manual, Fragebögen, Schablonen. Karlsruhe: Pfizer; 2002.
- Lutz W, Böhnke JR, Köck K, Bittermann A. Diagnostik und psychometrische Verlaufsrückmeldung im Rahmen eines Modellprojektes zur Qualitätssicherung in der ambulanten Psychotherapie. *Z Klin Psychol Psychother*. 2011;40(4):283–97.
- Lutz W, Böhnke JR. Der „Fragebogen zur Evaluation von Psychotherapieverläufen“ (FEP-2): Validierungen und Manual. Trierer Psychologische Berichte. 2008;35(3):1–21.
- Lutz W, De Jong K, Rubel J. Patient-focused and feedback research in psychotherapy: where are we and where do we want to go? *Psychother Res*. 2015a;25(6):625–32.
- Lutz W, Rubel J, Schiefele AK, Zimmermann D, Böhnke JR, Wittmann WW. Feedback and therapist effects in the context of treatment outcome and treatment length. *Psychother Res*. 2015b;25(6):647–60.
- Lutz W, Schürch E, Stulz N, Böhnke JR, Schöttke H, Rogner J, et al. Entwicklung und psychometrische Kennwerte des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP). *Diagnostica*. 2009;55(2):106–16.
- Lutz W. Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychother Res*. 2002;12(3):251–72.
- Miller SD, Hubble MA, Chow D, Seidel J. Beyond measures and monitoring: realizing the potential of feedback-informed treatment. *Psychotherapy (Chic)*. 2015 Dec;52(4):449–57.
- Mohr DC. Negative outcome in psychotherapy: A critical review. *Clin Psychol Sci Pract*. 1995;2(1):1–27.

- Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods*. 2002 Mar;7(1):105–25.
- Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017 Mar;9:157–66.
- Poston JM, Hanson WE. Meta-analysis of psychological assessment as a therapeutic intervention. *Psychol Assess*. 2010 Jun;22(2):203–12.
- Probst T, Lambert MJ, Dahlbender RW, Loew TH, Tritt K. Providing patient progress feedback and clinical support tools to therapists: is the therapeutic process of patients on-track to recovery enhanced in psychosomatic inpatient therapy under the conditions of routine practice? *J Psychosom Res*. 2014 Jun;76(6):477–84.
- Rubel JA, Zimmermann D, Deisenhofer AK, Müller V, Lutz W. Nutzung von psychometrischem Feedback als empirische Unterstützung des Supervisionsprozesses bei Ausbildungstherapeuten. *Z Klin Psychol Psychother*. 2017;46(2):83–95.
- Sapyta J, Riemer M, Bickman L. Feedback to clinicians: theory, research, and practice. *J Clin Psychol*. 2005 Feb;61(2):145–53.
- Sass H, Wittchen H-U, Zaudig M. Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Göttingen: Hogrefe; 2003.
- Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and meta-analytic review of a psychotherapy quality assurance system. *J Consult Clin Psychol*. 2010 Jun;78(3):298–311.
- Simon W, Lambert MJ, Harris MW, Busath G, Vazquez A. Providing patient progress information and clinical support tools to therapists: effects on patients at risk of treatment failure. *Psychother Res*. 2012;22(6):638–47.
- Slade K, Lambert MJ, Harmon SC, Smart DW, Bailey R. Improving psychotherapy outcome: the use of immediate electronic feedback and revised clinical support tools. *Clin Psychol Psychother*. 2008 Sep-Oct;15(5):287–303.
- Strauss BM, Lutz W, Steffanowski A, Wittmann WW, Boehnke JR, Rubel J, et al. Benefits and challenges in practice-oriented psychotherapy research in Germany: the TK and the QSPSY-BAY projects of quality assurance in outpatient psychotherapy. *Psychother Res*. 2015;25(1):32–51.
- Weiss NS. Clinical Epidemiology. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*. Philadelphia: Wolters Kluwer Health; 2008. pp. 641–51.
- Whipple JL, Lambert MJ, Vermeersch DA, Smart DW, Nielsen SL, Hawkins EJ. Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *J Couns Psychol*. 2003;50(1):59–68.
- Wittmann WW, Steffanowski A. Qualitätsmonitoring in der ambulanten Psychotherapie: Ergebnisse des TK-Modellprojektes. *Psychotherapie Aktuell*. 2011;3(3):6–12.