Journal of Consulting and Clinical Psychology
2017, Vol. 85, No. 5, 484–494

© 2017 American Psychological Association
0022-006X/17/$12.00 http://dx.doi.org/10.1037/ccp0000173

Feedback in Group Psychotherapy for Eating Disorders: A Randomized Clinical Trial

Annika Helgadóttir Davidsen
Stolpegaard Psychotherapy Centre, Capital Region of Denmark, and University of Copenhagen

Stig Poulsen
University of Copenhagen

Jane Lindschou and Per Winkel
Copenhagen University Hospital

Marjun Frígerð Tróndarson, Mette Waaddegaard, and Marianne Lau
Stolpegaard Psychotherapy Centre, Capital Region of Denmark

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Objective: To investigate the effect of client feedback in group psychotherapy on attendance and treatment outcome for patients with eating disorders. ***Method:*** We conducted a randomized clinical trial with central randomization stratified for diagnosis and treatment type according to a computer-generated allocation sequence concealed to the investigators. One-hundred and 59 adult participants, diagnosed with bulimia nervosa, binge eating disorder, or eating disorder not otherwise specified according to *DSM–IV*, were included. Eighty participants were allocated to the experimental group, and 79 participants to the control group. Both groups received 20–25 weekly group psychotherapy sessions. In the experimental group, participants gave and received feedback about therapy progress and alliance, measured before and after each session using the Outcome Rating Scale and the Group Session Rating Scale. The primary outcome was rate of attendance to treatment sessions; the secondary outcome was severity of eating disorder symptoms measured with the Eating Disorder Examination interview. Exploratory outcomes were psychological distress measured with the Symptom Checklist-90-R and the Outcome Rating Scale, social functioning measured with the Sheehan Disability Scale, and episodes of self-harm and suicide measured with a modified version of the Self-Harm Inventory. ***Results:*** Feedback compared with control did not affect the rate of attendance (0.59 vs. 0.58; $p = .96$), the severity of symptoms (2.03 vs. 2.02; $p = .46$), or any of the exploratory outcomes (p values from 0.06 to 0.67). ***Conclusions:*** Feedback neither increased attendance nor improved outcomes for outpatients in group psychotherapy for eating disorders. The results are discussed from different perspectives.

What is the public health significance of this article?
This trial highlights the importance of flexibility in the treatment setting, when implementing feedback-informed treatment and assessing the effect of it on psychotherapy outcome.

Keywords: group psychotherapy, feedback-informed psychotherapy, eating disorders, systemic and narrative psychotherapy

Eating disorders (EDs) are serious mental disorders affecting up to 10% of the population, primarily women (Smink, van Hoeken, & Hoek, 2013). One of the prevalent challenges is the high number of patients who prematurely withdraw from ED treatment. A recent comprehensive meta-analysis of psychotherapy dropout, analyzing 669 studies and 83,834 adult patients with a nonpsychotic disorder, found that 23.9% of patients in psychotherapeutic treatment dropped out of ED treatment; these patients often exhibit poorer treatment outcomes (Swift & Greenberg, 2012). In group therapy especially, dropout not only affects the patient and the therapists; it can have an adverse effect on the remaining group members, sometimes resulting in a "wave effect" leading to other dropouts (Yalom & Leszcz, 2005). Related to dropout is irregular attendance to treatment sessions. Poor psychotherapy attendance

This article was published Online First March 23, 2017.

Annika Helgadóttir Davidsen, Mental Health Services, Stolpegaard Psychotherapy Centre, Capital Region of Denmark, and Department of Psychology, University of Copenhagen; Stig Poulsen, Department of Psychology, University of Copenhagen; Jane Lindschou and Per Winkel, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Department 7812, Copenhagen University Hospital; Marjun Frígerð Tróndarson, Mette Waaddegaard, and Marianne Lau, Mental Health Services, Stolpegaard Psychotherapy Centre, Capital Region of Denmark.

We particularly acknowledge the participation of the patients at a difficult time in their lives. We are grateful to our clinical and research colleagues at Stolpegaard Psychotherapy Centre for their interest and engagement in making this trial possible. We would also like to thank Bruce Wampold and Scott D. Miller for helpful comments on the manuscript.

Correspondence concerning this article should be addressed to Annika Helgadóttir Davidsen. E-mail: annikahelgadottir@gmail.com

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

has consistently been related to a poorer treatment outcome (Montoya et al., 2005; Page & Hooke, 2009; Reardon, Cukrowicz, Reeves, & Joiner, 2002) and, accordingly, the development and testing of interventions to increase attendance and prevent dropout from group psychotherapy is highly relevant.

Monitoring in psychotherapy provides information about treatment progress and allows the therapist together with the patient to tailor the treatment and address potential problems leading to nonattendance and dropout. Different systems are available but most of the previous effectiveness research is based on the Outcome Questionnaire System (OQ System; Green & Latchford, 2012; Lambert, Hansen, & Finch, 2001; Lambert & Shimokawa, 2011; OQ-Measures, 2015) and the Partners for Change Outcome Management System (PCOMS; Bargmann & Robinson, 2011; Bertolino, Bargmann, & Miller, 2012; Duncan, 2012; Green & Latchford, 2012; Hubble, Duncan, & Miller, 1999). Although differences exist between the OQ System and the PCOMS, both systems have demonstrated their potential to enhance treatment outcomes (Duncan & Reese, 2015).

Results from meta-analyses of randomized clinical trials (RCTs) have suggested that providing client feedback has a significant positive effect on psychotherapy outcome, especially in individual and couple's therapy and for patients at risk for deterioration or achieving less than expected change (Knaup, Koesters, Schoefer, Becker, & Puschner, 2009; Lambert & Shimokawa, 2011; Sapyta, Riemer, & Bickman, 2005). More recent reviews have also suggested positive effects of client feedback on treatment outcome (Krageloh, Czuba, Billington, Kersten, & Siegert, 2015) although effect sizes vary and seem to weaken with more severe psychiatric populations (Davidson, Perry, & Bell, 2015).

Three RCTs have investigated the effect of different types of client feedback for patients with EDs. Schmidt et al. (2006) included 61 women with bulimia nervosa (BN) or eating disorders not otherwise specified (EDNOS). They were randomly allocated to 14 sessions of cognitive–behavioral guided self-care with or without added personalized feedback (not a previously published feedback system). Outcomes were measured with the Short Evaluation of Eating Disorders (Bauer, Winn, Schmidt, & Kordy, 2005), and treatment dropout. Client feedback had a significant effect on dietary restriction ($p = .03$), but no effect on dropout ($p = .67$). Truitt (2011) randomly assigned 51 women with anorexia nervosa (AN) or BN to treatment with client feedback (FB) or without (no feedback; NFB). Primary outcome was global psychological functioning measured with the OQ System. In this trial, client feedback significantly predicted positive change in individual global psychological dysfunction across the course of treatment ($p = .02$). Simon and colleagues (Simon et al., 2013) included 133 women diagnosed with AN, BN, or EDNOS in their trial. Patients were randomly assigned to treatment with or without client feedback using the OQ System. Primary outcome was global psychological functioning measured with the OQ System; secondary outcome was change in body mass index (BMI). The difference in global psychological functioning between the FB condition and NFB condition at end of treatment was statistically significant ($p < .05$). Feedback did not have a significant effect on BMI, which may have been because AN participants were encouraged to gain weight while patients with EDNOS or BED most likely were not.

Recently, two RCTs have evaluated the effect of client feedback in group psychotherapy (Schuman, Slone, Reese, & Duncan, 2015; Slone, Reese, Mathews-Duvall, & Kodet, 2015) using PCOMS. Schuman et al. (2015) included 263 soldiers and randomly assigned them to FB or NFB. Results indicated that clients in the FB group achieved significantly more improvement on the ORS ($p = .01$), and attended significantly more sessions compared with clients in the NFB group (4.16 vs. 3.55, $p < .01$). Slone et al. (2015) evaluated the efficacy of PCOMS in group therapy with 84 clients presenting with self-reported anxiety, stress, or depression. Results indicated that clients in the FB condition had significantly larger prepost group therapy gains measured on the ORS compared with clients in the NFB condition ($p < .001$). Clients in the FB condition attended more group sessions (8.0 vs. 6.6, $p < .05$). While these results indicate that client feedback may improve outcome of group psychotherapy, the implementation and use of feedback in group therapy still needs to be investigated further. In particular, the fact that (a) several individuals provide feedback in a group therapy session, and (b) feedback on the therapeutic alliance is more complex in groups than in individual therapy because of the multiple relationships present might constitute specific challenges to group therapists collecting feedback.

In conclusion, while the existing studies of client feedback in treatment for EDs and group therapy indicate various positive effects of feedback, the findings concerning the specific impact on attendance and dropout are conflicting. Group psychotherapy is widely used to treat individuals with EDs (Kalodner, Coughlin, & Seide, 2014) and dropout and nonattendance is common among patients in ED treatment. Because therapy attendance is a prerequisite for a good therapy outcome, it is important to find ways to increase attendance. Consequently, we planned the F-EAT trial ("feedback vs. no feedback in improving patient outcome in group psychotherapy for eating disorders") to assess the effect of client feedback on attendance and outcome for patients with ED in group therapy.

Method

Trial Objectives and Design

The objective of the F-EAT trial was to examine the effects of continuous client feedback on treatment attendance and outcomes in group therapy. Our primary hypothesis was that continuous feedback to patients and therapists, with subsequent adjustments of the treatment, would increase treatment attendance. Our secondary hypothesis was that feedback would improve treatment outcome, measured by ED symptoms. We furthermore wanted to explore whether client feedback would improve psychological and social functioning, as well as reduce the presence of suicidal or self-harming tendencies. The F-EAT trial is an investigator-initiated randomized, clinical superiority trial using a parallel group design with a 1:1 allocation ratio and blinded outcome assessments for outcomes, analyses, and drawing of conclusions. The design and methods have been described in detail in a previous publication (Davidsen, Poulsen, Waaddegaard, Lindschou, & Lau, 2014).

Participants

Patients. The participants were treatment-seeking adults, referred to outpatient treatment for EDs at Stolpegaard Psychother-

apy Centre (PCS), Mental Health Services in the Capital Region of Denmark. To be eligible, they were required to be aged 18 years or older, to be diagnosed with BN, BED, or EDNOS as their primary diagnosis, to have a BMI $\geq$20 kg/m^2, and to provide informed consent. Exclusion criteria were acute suicidal risk, psychosis, severe depression, severe or nonregulated physical comorbidity, pregnancy, abuse of alcohol or medicine, use of cannabis more than once monthly, concomitant psychotherapeutic or psychiatric treatment outside PCS, inability to understand Danish, previous participation in the current trial, or that the patient was considered unable to attend treatment sessions as planned.

Therapists. Fifteen therapists participated in the trial, two males and 13 females. Their mean age was 44.3 years (*SD* = 9.1). The mean years of experience with psychotherapy in general were 7.2 (*SD* = 6.6), and the mean years of experience with ED treatment were 3.8 (*SD* = 5.2). Six therapists were licensed social workers, three were licensed psychologists, four were psychiatrists/physicians in training, and two were licensed physiotherapists. The primary theoretical orientation for all therapists was systemic and narrative.

The therapists had two training sessions (3 hr each) in feedback-informed clinical work, led by an external certified PCOMS trainer and associate at the International Center for Clinical Excellence (ICCE, 2016). The therapists were introduced to the approach before patients were included in the trial, and were instructed in the use of the measures in clinical practice (Bargmann & Robinson, 2011). Only two of the therapists had worked with PCOMS in previous work positions. To control for therapist allegiance, each therapist was, whenever possible, placed in both a FB and a NFB group and thus provided both treatment modalities. To reinforce the implementation process and to ensure that the therapists followed the feedback approach, the training was supplemented with 1[1/2]-hr PCOMS case supervision monthly in the data collection period (18 months).

Immediately after the second training session, the therapists answered an attitude survey asking if they believed working with GSRS and ORS would make a positive difference in their therapeutic work (Davidsen, 2013). All therapists agreed that it would improve their clinical work, demonstrating the presence of therapist allegiance. In order to investigate if a "rush" influenced the first attitude survey after the training session, the therapists were asked again before they started using the measures. The answers were consistent with the first survey.

Procedure

Participants were included between August 2012 and February 2014 at PCS. All patients referred to group psychotherapy for BN, BED, or EDNOS diagnosed according to the *DSM–IV* (APA, 2000) were invited to participate in the trial. The trial is registered on www.clinicaltrials.gov (NCT01693237) and was approved by the regional ethics committee for the Capital Region of Denmark (journal number H-3–2011-151) and by the Danish Data Protection Agency (journal number 2007–58-0015).

Sample Size

We expected to find that the participants in the FB (experimental) group on average attended at least three more sessions than the participants in the NFB (control) group. Using a standard deviation of six sessions (based on unpublished data from routine care at PCS) and an alpha of 5% (Type I error), we needed to include 64 participants in each group (total 128) to be able to reject the null hypothesis that number of attended treatment sessions in the experimental and control group is equal with power of 80%. We also estimated the sample size using a power of 90%. This resulted in 170 participants (2 $\times$ 85 participants). We therefore planned to recruit a minimum of 128 participants, and in order to reduce the risk of Type II error, we aimed to recruit up to 170 participants, or as many as possible during our 18-months recruitment period. Calculations were made using the PS Power and Sample Size Calculations program version 3.0.14 (Dupont & Plummer, 1990, 1998).

Randomization and Blinding

Randomization was carried out centrally by the Copenhagen Trial Unit (CTU) according to a computer-generated allocation sequence with varying permuted block sizes of four, six, and eight. The allocation sequence and block sizes were kept concealed to the investigators. The allocation sequence was stratified for ED diagnosis (BN, EDNOS, or BED) and treatment type (basic or elaborate). After assessment of eligibility, the investigators telephoned the CTU to allocate the participant. The CTU staff then allocated the participant to one of the intervention groups according to data entered into a computer program while on the telephone with the investigator. The experimental groups contained patients randomized to the experimental condition only, while the control groups contained patients randomized to the control condition only.

Patients and therapists were naturally aware of the patient allocation; however, blinding was maintained by instructing the therapists to withhold patient information from the research team. The research team had no contact with the patients during group psychotherapy. Furthermore, the statistical analyses were conducted blinded with the two intervention groups coded as X and Y. Two conclusions were drawn, one assuming X was the experimental group and Y was the control group, and one assuming the opposite (Gotzsche, 1996; Jarvinen et al., 2014). After that, the code was broken.

Interventions

Standard treatment. Standard treatment was offered in both intervention groups, based on recommendations from the Danish National Board of Health (Danish Health Authorities, 2005) and on guidelines for treating EDs in the Danish Regions (Danish Regions, 2014). Based upon an assessment of ED severity, comorbidity, medical and/or social factors that are believed to complicate the treatment of the ED, the patient was offered one of two standard treatments: basic or elaborate. Both treatment programs included 20 group sessions for patients with BN or EDNOS, and 25 group sessions for patients with BED, provided on a weekly basis. Alongside group therapy, the patients were offered therapy sessions with a physician, dietician, physiotherapist, and social worker as well as sessions with relatives (as needed). The elaborated treatment was somewhat more extensive with treatment duration (from first assessment session to last follow-up session) of 12–14 months compared with 10 months for basic treatment.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Patients in the elaborate treatment were offered an extended medical assessment and more sessions with relatives and dietician (Danish Regions, 2014). Patient treatment status could be discussed at weekly team conferences. There were seven patients and two therapists in each group. The groups were rolling, that is, open to new patients as others ended treatment. Central to treatment was a food diary that patients were asked to keep and discuss in the group therapy sessions. Patients were weighed before each session and weight fluctuation was monitored and addressed in the case of, for example, rapid weight gain or weight loss. The patients were encouraged to set individual goals for the treatment, typically concerning food, body and appearance, relations, and future (Plambech, Lau, & Christensen, 2000).

Experimental intervention. In the experimental group, the therapists received feedback from the participants using the Outcome Rating Scale (ORS) and the Group Session Rating Scale (GSRS) from the PCOMS feedback system, which is defined as "a pantheoretical approach for evaluating and improving the quality and effectiveness of behavioral health services" (Bertolino, Bargmann, & Miller, 2012, p. 2). By using the ORS and GSRS, the patient continuously evaluates the outcome and psychotherapeutic alliance, and the feedback allows the therapists to monitor the progress from session to session and to tailor the treatment in continuous conversation with the patient (Bargmann & Robinson, 2011). We adjusted the guidelines for implementation of PCOMS in individual settings (Bertolino, Axsen, Maeschalck, Miller, & Babbins-Wagner, 2012) to suit the group setting, because official guidelines for groups were not available.

The patients marked their scores on the ORS before each group session, and on the GSRS after each group session, using a tablet computer with a computer-based application (FIT-Outcomes, 2012). The management system subsequently produced a graph illustrating the therapeutic progress on the ORS and GSRS, the cut-off scores and the expected treatment response. Based on the latest score on the ORS and GSRS, the management system provided immediate feedback to the therapists. The therapists were encouraged to discuss the ORS and GSRS scores with the patient in the present or the following group session. Adherence to treatment protocol was not monitored, but we regularly checked FIT-Outcomes to assure that the patients actually scored the measures before and after each group session.

Control intervention. In the control group, the patients filled out a paper version of the ORS before each group session. They did not, however, receive any feedback based on the ORS during therapy, nor did their therapists see the ORS forms, which were stored in sealed envelopes until data analyses.

Diagnosis, Outcomes, and Covariates

A detailed description of the outcome and assessment measures is published elsewhere (Davidsen et al., 2014). We will briefly summarize the measures in the following.

Clinical Assessment

Diagnosis was determined using the Eating Disorder Examination (EDE) interview, version 12, including the BED module from version 16 (Cooper & Fairburn, 1987; Fairburn & Cooper, 1993). Eligibility was assessed with the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). The possible presence of *DSM–IV* personality disorder was assessed with a researcher-rated version of the Standardized Assessment of Personality—Abbreviated Scale (SAPAS; Moran et al., 2003).

Treatment Attendance

The primary outcome was treatment attendance in the intervention period defined as a rate between 0 and 1.0. The rate (R) is the number of sessions attended (A) over the number of planned sessions (N), that is, ($R = A/N$), where N is 20, 25, or the number of sessions before a bilateral agreement to terminate. If the therapists and patient bilaterally agreed to discontinue treatment, for example, at Session 17 of 20, the number of planned sessions (N) was set to 17. If the patient unilaterally discontinued treatment, for example, at Session 17 of 20, N remained unchanged at 20.

Eating Disorder Symptoms

The secondary outcome was severity of ED symptoms measured by the EDE global score (Cooper & Fairburn, 1987; Fairburn & Cooper, 1993).

Psychological Symptoms, Functional Impairment, and Self-Harm

We analyzed several exploratory outcomes. The patient's experience of well-being in his or her individual, interpersonal, and social functioning was assessed with the ORS (Duncan & Miller, 2000). Psychological problems and symptoms of psychopathology were measured with the Global Severity Index (GSI) of the Symptom Check-List (SCL-90–R; Derogatis, Lipman, & Covi, 1973; Derogatis, Rickels, & Rock, 1976). Functional impairment (family, work, and social functioning) was measured with the Sheehan Disability Scale (SDS; Sheehan, Harnett-Sheehan, & Raj, 1996). Psychological well-being was measured with the WHO-Five Well-being Index (WHO-5; Bech, 2004). The presence of self-harm and suicidal tendencies was measured with a modified version of the Self-Harm Inventory (SHI), translated to Danish for use in the present trial (Sansone & Sansone, 2010). All measures were authorized Danish versions, besides the SHI. All outcomes were measured at the end of trial and all but the primary outcome, suicide and self-harm and the ORS at baseline. ORS measurements were sometimes first initiated after the intervention had started. Therefore, the ORS baseline values in the two groups are not compatible and were in consequence ignored. Treatment attendance, suicide and self-harm were only measured posttreatment.

Therapist Use of the PCOMS Measures

Because previous studies have pointed to the importance of the therapists' usage of the feedback measures (Amble, Gude, Ulvenes, Stubdal, & Wampold, 2016; De Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; Lutz, Rubel et al., 2015), the participating therapists ($N = 15$) were asked to answer a short survey about how often they used the ORS and GSRS, and how useful they found the PCOMS after the end of the trial (see

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4). Answers from 11 therapists (73%) were rated on a 5-point Likert scale (Questions 1–5 ranging from *never* to *every time*; Questions 6–14 ranging from *not at all* to *very helpful*, scored from 0 to 4).

Statistical Analyses

The analyses were intention to treat analyses using two-sided significance test at 5%. All outcome analyses were regression analyses including as covariates the binary intervention indicator (feedback vs. no feedback), the protocol specified stratification variables (ED diagnosis: BN, EDNOS, or BED; and treatment type: basic or elaborate) and the baseline value, when measured. In the analysis of the rate data, the model fit of a Poisson model and a negative binomial model, both with offset equal to log (number of planned follow-up sessions), were compared and the best fitting model was chosen, provided it fitted the data reasonably well. Otherwise, a nonparametric test was used (see below). For the analysis of the continuous outcomes, proc calis (SAS 9.3) was used. Proc calis uses a direct maximum likelihood estimation (i.e., if a value is missing, the maximum likelihood function is integrated over all possible values that the missing value may take on). Proc calis furthermore includes auxiliary variables, defined as variables not in the regression model but correlated ($r > .40$) with a variable (with missing values) included in the model. The inclusion of auxiliary variables improves the standard errors because they are correlated with the variable that has missing values. This is an analysis that is equivalent to a multiple imputation, which includes auxiliary variables. If data are missing at random (MAR), that is, the probability of missing values only depends on observed values, the results will be unbiased when proc calis—which is based on the structural equations principles—is used. Including auxiliary variables in the analysis should furthermore improve the efficiency. If the assumptions of a regression were not fulfilled, a nonparametric test (van Elteren's test) was used adjusted by ED diagnosis. The primary and secondary outcomes were tested in that order, each at the 5% level of significance. If the first test was not significant at the 5% level, the null hypothesis of the secondary outcome was accepted without test. This procedure keeps the family wise error rate $<.05$. The remaining variables were exploratory.

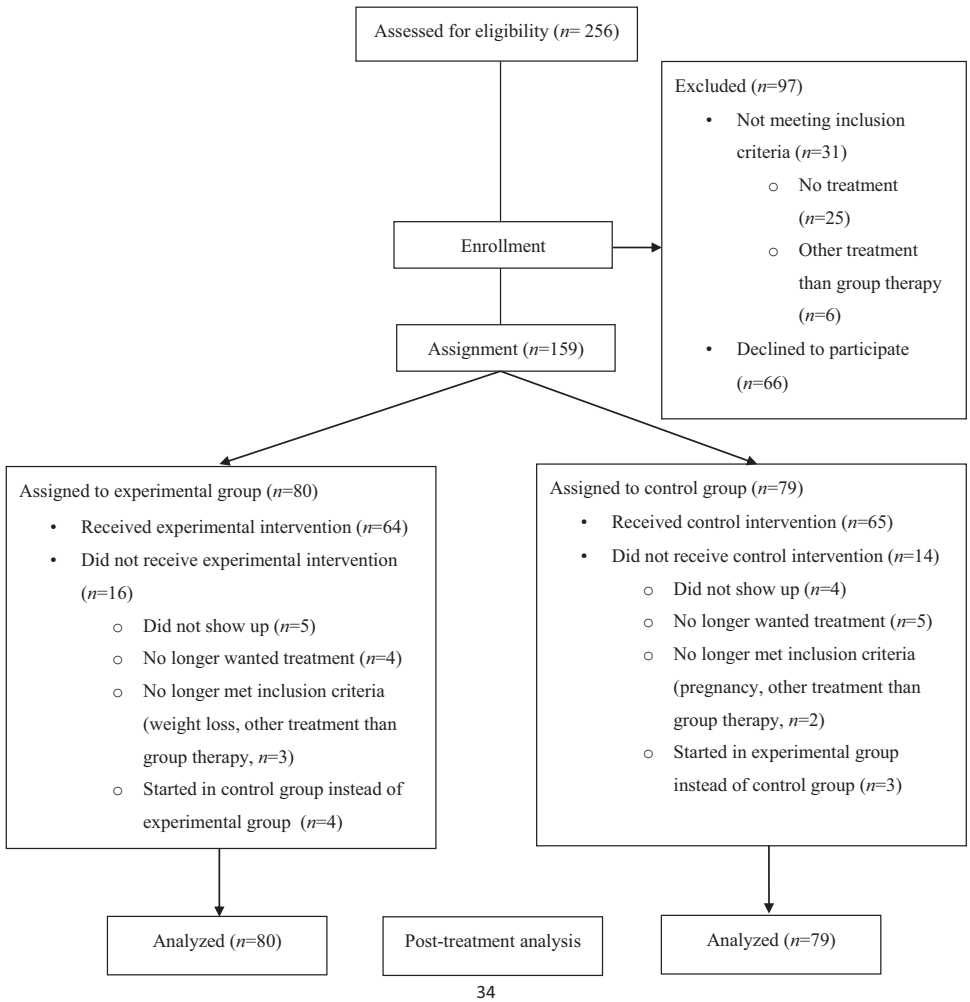

Figure 1. Flow of participants.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Results

Flow of Participants and Baseline Characteristics

A total of 256 patients were screened for eligibility. Of these patients, 97 were excluded; 31 patients (12.1%) did not meet the inclusion criteria, and 66 (25.8%) declined to participate in the trial. This left 159 participating patients (see Figure 1). A total of 30 patients did not receive the planned treatment (16 in the experimental group and 14 in the control group); most of these because they did not show up, no longer wanted treatment or no longer met the inclusion criteria. Seven of these patients started in the wrong intervention group; this was either due to misplacements by clinical staff ($n = 4$), or because it was not possible for patients to attend group therapy on the specific weekday ($n = 3$). The majority of patients in both intervention groups were single females, without children and undergoing education (see Table 1). The stratification variables at baseline are reported in Table 1. BN was the most frequent diagnosis in both intervention groups. Values for all outcomes at baseline and end of intervention are reported in Tables 2 and 3.

Treatment Attendance

The results of the analysis of the rate of attended sessions (number of sessions attended/number of planned sessions) showed no difference between FB and NFB groups. When the Poisson model was used, p of the effect of the intervention was .77. There were, however, clear signs of over dispersion with a deviance/*d.f.* of 7.50, which was remedied when the negative binomial model was used (deviance/*d.f.* = 1.23). The p value of the effect of intervention was .96; p of the nonparametric test was .98. To test whether the hierarchical structure of the data with patients nested within therapy groups unduly influenced the results, a post hoc mixed model analysis (using the negative binominal distribution and proc glimmix, SAS 9.3) with therapy group as a random variable and the binary intervention indicator (FB or NFB), treatment type (basic or elaborate) and diagnosis as covariates was conducted. The mixed model analysis uses the direct maximum-likelihood analysis, but discards any cases where one or more covariates are missing. Provided the covariates have no missing values (which was the case for our data), the results should be unbiased provided the data are MAR. In this analysis, the p value of the intervention was .14.

As p of the primary outcome was $>.05$, the null hypothesis of the secondary outcome was accepted without test (see statistical analysis plan). However, for exploratory purposes, we now present the results of the analysis of the secondary outcome as if it were an exploratory outcome.

Table 1
Baseline Characteristics for Patients in Feedback and No Feedback Groups

Characteristics/categories	Feedback ($N = 80$)	No feedback ($N = 79$)
Basic treatment type (vs elaborate), *n* (%)	31 (38.8)	34 (43)
Diagnosis, *n* (%)		
Eating disorder not otherwise specified	29 (36.3)	28 (35.4)
Bulimia nervosa	37 (46.3)	36 (45.6)
Binge eating disorder	14 (17.5)	15 (19.0)
Age in years, *M* (*SD*)	26.4 (8.4)	27.5 (8.9)
Female, *n* (%)	78 (97.5)	78 (98.7)
Duration of eating disorder >5 years, *n* (%)	55 (68.8)	49 (62.0)
Body mass index (*SD*)	26.3 (7.9)	26.3 (7.5)
Marital status: single, *n* (%)	55 (68.8)	57 (72.2)
Children under the age of 15: no, *n* (%)	68 (85%)	65 (82.3)
Education: ≥10 years of schooling, *n* (%)	59 (73.8)	66 (83.5)
Employment status: student, *n* (%)	34 (42.5)	37 (46.8)
Comorbidity[1]: ≥1 comorbid diagnosis, *n* (%)	24 (30.0)	24 (30.4)
SAPAS score ≥4, *n* (%)	30 (37.5)	35 (44.3)

Note. SAPAS = Standardized Assessment of Personality—Abbreviated Scale.
[1] Comorbidity was assessed with the Mini International Neuropsychiatric Interview (MINI).

Eating Disorder Symptoms, Psychological Symptoms, Functional Impairment, and Self-Harm

For the secondary outcome and each of the exploratory outcomes, the distributions of the outcome and the corresponding baseline values in each intervention group were all normal with reasonable approximation. After the auxiliary variables had been identified among all the outcomes and corresponding baseline variables, the structural equation regression models (*SEM* models) were identified. The results of the regression analyses using *SEM* are shown in Table 3. No differences between the FB and the NFB groups were found for neither the secondary outcome nor any of the exploratory outcomes; the p value was $>.05$ in all cases.

No auxiliary variables correlated with the binary variable "presence of suicidal tendencies or self-harm" were identified. However, because only the outcome had missing values, the results of the regression analysis should be unbiased provided the outcome is MAR (Carpenter & Kenward, 2013). The result of the logistic regression analysis showed $p = .48$.

As with the primary outcome, post hoc exploratory mixed model analyses with therapy group as a random variable and adjusting for the prespecified covariates were conducted. Again, all p values were $>.05$.

Ancillary Exploratory Analyses

To provide a better understanding of the results from the current trial, we conducted a number of post hoc exploratory analyses.

First, because the EDE global score does not reflect the core symptoms of a bulimic ED, we performed supplementary exploratory analyses with binging (binge eating episodes during the last month) and purging (sum of vomiting, laxative, and/or diuretic misuse episodes during the last month) as outcomes. These analyses showed no differences between the FB and NFB groups concerning binging (difference between Ms = −2.19, 95% CI [−5.83, 1.44], $p = .23$) or purging (difference between Ms = −2.90, 95% CI [−7.49, 1.67], $p = .21$). We also examined comorbidity (one or more comorbid diagnoses) as a possible moderator of the relationship between intervention group (FB or NFB) and rate of attendance and ED severity. The p values for these analyses were insignificant (both p values = .81).

Second, we calculated effect sizes (ES) using baseline and follow-up scores of ORS, primarily to be investigate the changes all participants have gone through during the period with therapy.

Table 2
Outcome Measures at Baseline and End of Intervention

	Baseline						End of intervention					
	Feedback[a]			No feedback[b]			Feedback			No feedback		
Measures	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Sessions planned	21.1	2.10	80	21.1	2.11	79	n.a.	n.a.		n.a.	n.a.	
Sessions attended	n.a.	n.a.		n.a.	n.a.		12.4	8.08	80	12.3	8.03	79
Rate of attendance[c]	n.a.	n.a.		n.a.	n.a.		.59	.37	80	.58	.37	79
EDE global score	3.88	.99	80	3.87	1.05	79	2.03	1.44	51	2.02	1.41	50
SDS global score	17.7	5.45	72	18.1	5.25	74	11.3	8.50	46	10.6	7.02	45
WHO-5 score	37.9	19.1	77	34.5	16.2	76	53.9	20.9	46	48.6	20.5	51
SCL-90 score	1.32	.54	74	1.38	.51	73	.78	.53	46	.96	.61	51
ORS score	n.a.	n.a.		n.a.	n.a.		26.2	11.1	64	23.9	9.37	65

Note. EDE = Eating Disorder Examination; SDS = Sheehan Disability Scale; WHO-5 = WHO-Five Well-Being Index; SCL = Symptom Check List; *M* = mean; *SD* = standard deviation. Self-harm was a dichotomous variable (yes/no) and is not illustrated in this table.
[a] $N = 79$. [b] $N = 80$. [c] (Number of sessions attended)/(Number of sessions planned).

ES was calculated by using the formula (Kazdin, 1994): $ES = (Mean_{pre} - Mean_{post})/s$ where s = pooled standard deviation ($\sqrt{(sd_1^2 - sd_2^2)}/2$). ES was .76 for the total sample, .86 for the FB condition and .65 for the NFB condition.

Third, we studied the therapist ratings of how often they used the ORS and GSRS, and how useful the measures were. Questions and answers (with median values) are shown in Table 4. The findings indicate that the therapists routinely looked at the ORS and GSRS scores (Q1 and Q2; *Mdn*s = 4), but the scores did not often result in a discussion with the cotherapist (Q4; *Mdn* = 2), and were rarely discussed at team conferences (Q5; *Mdn* = 1). The median score was 2 for the usefulness of discussions with patients (Q12), and the usefulness of PCOMS with regard to adjusting or ending the therapy course (Q11). These findings indicate that although the therapists reviewed their patients' scores regularly, they did not find the measures particularly useful. Accordingly, in order to clarify whether the client feedback led to additional clinical action (i.e., a more individualized approach to treatment) we compared the mean number of individual or network sessions with group therapist, physician, dietician, and body therapy sessions in the two trial groups. The results showed no significant differences (p values $>.05$), indicating that client feedback did not prompt therapists in the FB groups to offer more or less additional therapy sessions compared to patients in the NFB groups.

Fourth, as previous studies have indicated that client feedback has a stronger positive effect for patients who deteriorated in the course of therapy (Lambert & Shimokawa, 2011; Probst et al., 2013), we performed a subgroup analysis for patients who deteriorated ("not-on-track" [NOT]) and compared the outcomes of patients NOT in the FB and the NFB groups. We used the same primary, secondary and exploratory outcomes as for the whole sample. The patients were categorized post hoc using Slone et al.'s (2015) definitions: Clients who did not make at least a 5-point increase in the first three sessions, or clients who deteriorated by at least 5 points from their baseline measure at some point during treatment were considered to be deteriorating and classified as NOT (Slone et al., 2015). The number of patients defined as NOT in our sample was high: 52 of the 64 FB patients and 51 of the 65 NFB patients were considered NOT according to this definition. The subgroup analyses showed no significant differences in rate of attendance between NOT patients in the FB and NFB groups. The means (with SDs in parenthesis) for the FB and NFB groups were .72 (.27) and .73 (.27), respectively. The p value of the difference between the FB and NFB groups in rate of attendance was .68. The p values of the difference between the FB and NFB groups on the secondary and exploratory outcomes ranged from .17 to .90.

Finally, because we did not have a pilot phase, the therapists could have improved their performance in using the PCOMS

Table 3
Secondary and Exploratory Outcomes: Difference in Estimated Means Between the Feedback Group and the Control Group

Outcome (type)	Difference between estimated means	95% Confidence interval (CI)	*p* value
EDE global score (secondary)	.026	[−.43, .48]	.46
ORS score (exploratory)	2.67	[−.76, 6.10]	.06
SDS global score (exploratory)	1.08	[−1.46, 3.62]	.21
WHO-5 score (exploratory)[a]	−.56	[−6.79, 14.7]	.32
SCL-90 score (exploratory)	−.035	[−.14, .14]	.67

Note. EDE = Eating Disorder Examination; ORS = Outcome Rating Scale; SDS = Sheehan Disability Scale; WHO = WHO Well-Being Index; SCL = Symptom Checklist.
[a] Using auxiliary variables convergence was not obtained. Instead, only the covariates of the analytical model were used. However, this analysis should still produce unbiased parameter estimates as long as the values are only missing at random.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
Therapist Survey (N = 11)

Questions	Median[a]
1. How often did you look at the patients′ ORS scores?	4
2. How often did you look at the patients′ GSRS scores?	4
3. How often did you (alone) reflect upon the patients′ progress, based on the ORS and GSRS?	3
4. How often did you discuss with your co-therapist the patients′ progress, based on the ORS and GSRS?	2
5. How often were the ORS/GSRS-scores discussed at team conferences?	1
6. How useful was the ORS graph?	2
7. How useful was the distribution of ORS scores into green and red areas?	3
8. How useful was the expected treatment response (ETR) graph?	2
9. How useful was the GSRS graph?	3
10. How useful was the GSRS cut-off score?	1
11. How useful were the ORS and GSRS with regards to adjusting or ending the treatment course?	2
12. How useful was discussing the ORS and GSRS scores with the patients?	2
13. How useful was discussing the ORS and GSRS scores with your co-therapist?	2
14. How useful was the FIT-supervision?	3

Note. ORS = Outcome Rating Scale; GSRS = Group Session Rating Scale; FIT = Feedback-informed treatment. Answers were rated on a 5-point Likert scale and scored from 0 to 4.
[a] *Mdn* = 0–4.

system after the first months of implementation. We therefore explored possible differences between early and late treatment courses by performing a *t* test comparing the ORS post treatment scores for FB patients included in the first 9 months ($n = 29$) versus the last 9 months ($n = 35$) of the patient inclusion period. There was not a significant difference in the post ORS scores for the FB patients included in Phase 1 ($M = 26.29$, $SD = 10.88$) and Phase 2 ($M = 25.23$, $SD = 11.75$); $t(62) = 0.37$, $p = .71$.

Discussion

The current trial is the first RCT examining the effect of client feedback on treatment attendance and outcomes for patients in group therapy for an ED. The primary null hypothesis was not rejected: Continuous client feedback neither resulted in increased treatment attendance nor improved treatment outcome, measured in ED symptoms. The results are consistent with the Schmidt et al. (2006) trial that found no effect of FIT on dropout. They are also consistent with Davidson et al. (2015) results indicating that effect sizes tend to diminish with more severe psychiatric populations. The results differ from the other trials reviewed in this article, which all found that FIT has a positive effect on attendance and outcome (Schuman et al., 2015; Simon et al., 2013; Slone et al., 2015; Truitt, 2011). In the following, we will discuss the results from various perspectives.

The results from the survey of the participating therapists in our trial indicate that although the therapists routinely viewed their patients' PCOMS scores, they rated the usefulness of the measures relatively low. It was surprising that the usefulness of the ORS and GSRS to adjust or end treatments, and of discussing the ORS and GSRS scores with the patients were not rated higher as these questions capture the essence of FIT: To alter the treatment based on a discussion with the patients about their scores. The therapists' view of the usefulness was especially surprising since all therapists initially believed that implementing the measures would improve their therapeutic efforts in the groups. Furthermore, 81% of the patients in the FB groups were classified as NOT at some point during treatment, indicating that some kind of clinical action was relevant. However, the post hoc analysis revealed no differences in the number of additional sessions between patients in the FB groups and patients in the NFB groups. These results indicate that the therapists did not use the PCOMS feedback as intended, that is, to individualize treatment by adjusting or altering treatment length or actions according to client feedback.

The organizational context of the F-EAT trial might have had an impact on the limited treatment adjustment used and is therefore relevant to the interpretation of the main results. The trial took place at PCS, which is situated in the Mental Health Services in the Capital Region of Denmark. To ensure quality and equal treatment for all patients, psychiatric treatment in the Danish Regions since 2012 has been standardized so that the number of hours allocated to each patient and clinical actions such as assessment, psychoeducation or group sessions are preplanned and largely fixed (Danish Regions, 2014). The standardization of the psychotherapeutic treatment is in contrast with the individualized and flexible approach that FIT encourages by prompting the therapists to adjust the therapy according to their patients′ feedback. Organizational factors such as limited flexibility and lack of time have previously been proposed as potential barriers for client feedback to have an effect (De Jong et al., 2012), and during the trial and the implementation of the FIT measures, the therapists did actually repeatedly express a wish for more time to use the feedback actively in therapy. This was, however, not possible due to organizational factors, even though it would probably have enhanced the therapists' feeling of the usefulness of the measures. The lack of extra time to address the issues raised in the client feedback may be particularly problematic in group therapy where the feedback is exceedingly complex since several clients provide feedback and the feedback on the therapeutic alliance concerns multiple relationships. On the other hand, other possible obstacles to productive implementation of feedback were avoided because the feedback was not accessible to team leaders or managers and there were no consequences for therapists whose patients reported negative feedback. The therapists were aware of this before they started using the measures.

The implementation process did not seem to affect the results. FIT was implemented in the clinic simultaneously with the start of data collection, and none of the therapists participating at the onset of the trial had previous experience with the method. This meant that the initial months of data collection were also a period where the therapists had to acquaint themselves with using the measures therapeutically, as well as learning the technical aspects in the management system. Therefore, the results might have been different if the effect was measured after a pilot phase. There was, however, no significant difference in the post ORS scores for the FB patients in the first and the second phases of the study.

Therapist characteristics have previously been found to affect outcome in psychotherapy research (De Jong & De Goede, 2015; De Jong et al., 2012). De Jong et al. (2012) found that internal feedback propensity (higher likelihood to trust their own opinion

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

than feedback from external sources), self-efficacy, and commitment to use the feedback moderated the effects of feedback in a large RCT including 413 patients. Other studies have come to the same conclusion: that client feedback may not be effective for all therapists (Lutz, De Jong, & Rubel, 2015; Simon, Lambert, Harris, Busath, & Vazquez, 2012). Because the therapists in our trial worked in pairs, it was difficult to plan any statistical analyses isolating a possible individual therapist effect. We therefore considered to perform post hoc subgroup analyses to test if effectiveness differed between the therapy groups but the small number of patients in each of the five FB groups (number of patients ranged from five to 22), resulting in decreased statistical power to detect subgroup differences (Assmann, Pocock, Enos, & Kasten, 2000), prevented us from performing the analyses. Other therapist effects that might have influenced the results are "carryover effects." As the therapists served as their own controls, that is, worked in a FB and NFB group whenever possible, it is plausible that the therapists in the NFB groups might have been extra attentive to patient feedback because of their experience and practice in the FB groups.

Another possible interpretation of the results may be that because the standard treatment was intensive, we observed large improvements in all participants during the intervention period, which means that a ceiling effect might have occurred where the addition of feedback could not provide a further increase in outcome.

The F-EAT trial has several strengths. The randomization was conducted through central, stratified allocation by a computer-generated sequence unknown to the investigators (Gluud et al., 2008; Kjaergard, Villumsen, & Gluud, 2001; Savovic et al., 2012; Wood et al., 2008). We stratified for type of ED and treatment intensity, our intervention groups seemed well randomized, and we considered stratification in our analyses (Kahan & Morris, 2012; Kernan, Viscoli, Makuch, Brass, & Horwitz, 1999; Pocock, 1983). The outcome assessors were blinded to the treatment allocation of the patients and we observed no breaking of the intervention code (Gluud et al., 2008; Kjaergard et al., 2001; Savovic et al., 2012; Wood et al., 2008). We conducted our analyses blinded to intervention and drew our conclusions blinded to intervention group (Gotzsche, 1996; Jarvinen et al., 2014). We performed intention-to-treat analyses, and the missing data were handled by direct maximum likelihood estimation which assures that bias will not occur as long as the mechanism causing data to be missing only depends on observed the observed data (Carpenter & Kenward, 2013). We are thus able to exclude with reasonable likelihood that the experimental intervention should be substantially superior to the control intervention, at least in the form implemented in the present trial. Many previous feedback trials reporting findings in favor of feedback have been performed by a limited number of research groups, most of them advocates of the approach or developers of the systems. The presence of researcher allegiance could possibly inflate the hitherto positive results of client feedback. This is not the case in the F-EAT trial. Moreover, while most of the previous trials use the progress measure as their primary outcome, we used attendance and ED symptoms as primary and secondary outcomes, which provide a more reasonable test of client feedback. We also find it particularly important to publish null findings to prevent publication bias, that is, that the published trials are unrepresentative of the population of completed studies (Rothstein, Sutton, & Borenstein, 2006). Until now, mostly confirmatory feedback studies and trials have been published.

There are, however, also limitations to this trial. The homogeneity of the sample, for example, concerning level of distress, diagnosis, and demographics, limits the generalizability of the findings. A relatively high number of the randomized patients did not begin the planned treatment due to reasons usually seen in naturalistic settings, such as moving away, not wanting treatment after all, or being unable to attend on a specific weekday. These circumstances are difficult to avoid. We did not monitor the therapy sessions through audio or video recordings, which prevented us from observing the less visible and measurable therapy adjustments that occur in therapy sessions. The organization of our data presented multiple levels of nesting: The individual patient score in our dataset was nested within the group scores, which again was nested within the therapist pairs leading the group. The fact that the data are nested may result in nonindependence in the group treatment data and inflate Type I error (Tasca, Illing, Joyce, & Ogrudniczuk, 2009), and multilevel linear modeling analyses would be relevant for data organized at more than one level (Tabachnick & Fidell, 2013). We refrained from choosing this analytic approach in the prespecified statistical analysis plan, because (a) the groups were rolling, meaning that the composition of the group continuously changed over time; and (b) the therapist pairs were unstable during the trial, that is, the therapists occasionally changed groups and teamed up with other therapists, which made it impossible to establish a consistent set of values of the therapist level. However, to ensure that the structure of the data did not unduly influence the outcomes, we conducted supplementary post hoc mixed model analyses for the primary and secondary variables with therapy group as a random variable. None of these analyses showed statistically significant differences between the interventions (p values $>.05$).

Because the F-EAT trial is the first RCT performed in group therapy including patients with EDs, the trial must be replicated before conclusions that are more firm can be drawn. The results seem to indicate that flexibility in the treatment setting is a prerequisite when using client feedback, that is, that the therapists are able to adjust treatment according to the feedback provided by the PCOMS measures. Future studies should use client feedback to illuminate the complex process of therapeutic change in groups, for example, how each patient's change relates to the other group members' change, both individually and overall. Video and audio recordings of the group sessions would further inform these processes, help to reveal in detail how patients and therapists use the measures, and would be a way to monitor therapist adherence.

References

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

APA. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet, 355,* 1064–1069. http://dx.doi.org/10.1016/S0140-6736(00)02039-0

Bargmann, S., & Robinson, B. (2011). Feedback-informed clinical work: The basics [Manual]. *The ICCE Manuals on Feedback-Informed Treatment (FIT)*.

Bauer, S., Winn, S., Schmidt, U., & Kordy, H. (2005). Construction, scoring and validation of the Short Evaluation of Eating Disorders (SEED). *European Eating Disorders Review, 13,* 191–200. http://dx.doi.org/10.1002/erv.637

Bech, P. (2004). Measuring the dimensions of psychological general well-being by the WHO-5. *Quality of life newsletter, 32,* 15–16.

Bertolino, B., Axsen, R., Maeschalck, C., Miller, S. D., & Babbins-Wagner, R. (2012). Manual 6: Implementing Feedback-informed Work in Agencies and Systems of Care. *The ICCE Manuals on Feedback-Informed Treatment*.

Bertolino, B., Bargmann, S., & Miller, S. D. (2012). What works in therapy: A primer [Manual]. *The ICCE Manuals on Feedback-Informed Treatment (FIT)*.

Carpenter, J. R., & Kenward, M. G. (2013). Multiple imputation of quantitative data. In *Multiple imputation and its application* (pp. 77–88). Chichester, UK: Wiley. http://dx.doi.org/10.1002/9781119942283

Cooper, Z., & Fairburn, C. G. (1987). The eating disorder examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6,* 1–8. http://dx.doi.org/10.1002/1098-108X(198701)6:1<1::AID-EAT2260060102>3.0.CO;2-9

Danish Health Authorities. (2005). *Spiseforstyrrelser. Anbefalinger for organisation og behandling* [Eating disorders. Recommendations for organization and treatment]. Retrieved from http://www.sst.dk/~/media/A297287DE0F548A88DC8F8C44441A8A9.ashx

Danish Regions. (2014). *Pakkeforløb for spiseforstyrrelser* [Standardized treatment for eating disorders]. Retrieved from http://regioner.dk/media/1252/spiseforstyrrelser.pdf

Davidsen, A. H., Poulsen, S., Waaddegaard, M., Lindschou, J., & Lau, M. (2014). Feedback versus no feedback in improving patient outcome in group psychotherapy for eating disorders (F-EAT): Protocol for a randomized clinical trial. *Trials, 15,* 138. http://dx.doi.org/10.1186/1745-6215-15-138

Davidsen, A. H. (2013). F-EAT allegiance scale (unpublished).

Davidson, K., Perry, A., & Bell, L. (2015). Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychology and Psychotherapy: Theory, Research and Practice, 88,* 21–37. http://dx.doi.org/10.1111/papt.12032

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

de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22,* 464–474. http://dx.doi.org/10.1080/10503307.2012.673023

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale—Preliminary report. *Psychopharmacology Bulletin, 9,* 13–28.

Derogatis, L. R., Rickels, K., & Rock, A. F. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *The British Journal of Psychiatry, 128,* 280–289. http://dx.doi.org/10.1192/bjp.128.3.280

Duncan, B. L. (2012). The partners for change outcome management system (PCOMS): The heart and soul of change project. *Canadian Psychology, 53,* 93–104. http://dx.doi.org/10.1037/a0027762

Duncan, B., & Miller, S. D. (2000). The client's theory of change. *Journal of Psychotherapy Integration, 10,* 169–187. http://dx.doi.org/10.1023/A:1009448200244

Duncan, B. L., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client's frame of reference. *Psychotherapy: Theory, Research, & Practice, 52,* 391–401. http://dx.doi.org/10.1037/pst0000026

Dupont, W. D., & Plummer, W. D., Jr. (1990). Power and sample size calculations. A review and computer program. *Controlled Clinical Trials, 11,* 116–128. http://dx.doi.org/10.1016/0197-2456(90)90005-M

Dupont, W. D., & Plummer, W. D., Jr. (1998). Power and sample size calculations for studies involving linear regression. *Controlled Clinical Trials, 19,* 589–601. http://dx.doi.org/10.1016/S0197-2456(98)00037-3

Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination. In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment and treatment* (12th ed., pp. 317–360). New York, NY: Guilford Press.

FIT-Outcomes. (2012). *Feedback informed treatment*. Retrieved from https://www.fit-outcomes.com

Gluud, L. L., Thorlund, K., Gluud, C., Woods, L., Harris, R., & Sterne, J. A. (2008). Correction: Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine, 149,* 219. http://dx.doi.org/10.7326/0003-4819-149-3-200808050-00023

Gotzsche, P. C. (1996). Blinding during data analysis and writing of manuscripts. *Controlled Clinical Trials, 17,* 285–290.

Green, D., & Latchford, G. (2012). *Maximising the benefits of psychotherapy*. Chichester, UK: Wiley-Blackwell. http://dx.doi.org/10.1002/9781119967590

Hubble, M. A., Duncan, B. L., & Miller, S. D. (1999). *The heart and soul of change: What works in therapy* (1st ed.). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11132-000

ICCE. (2016). *International Center for Clinical Excellence*. Retrieved from https://www.centerforclinicalexcellence.com/

Järvinen, T. L., Sihvonen, R., Bhandari, M., Sprague, S., Malmivaara, A., Paavola, M., . . . Guyatt, G. H. (2014). Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *Journal of Clinical Epidemiology, 67,* 769–772. http://dx.doi.org/10.1016/j.jclinepi.2013.11.011

Kahan, B. C., & Morris, T. P. (2012). Reporting and analysis of trials using stratified randomisation in leading medical journals: Review and reanalysis. *British Medical Journal, 345,* e5840. http://dx.doi.org/10.1136/bmj.e5840

Kalodner, C. R., Coughlin, J. W., & Seide, M. (2014). Psychoeducational and counseling groups to prevent and treat eating disorders and disturbances. In J. L. DeLucia-Waack, C. R. Kalodner, & M. T. Riva (Eds.), *Handbook of group counseling and psychotherapy* (2nd ed., pp. 484–494). Thousand Oaks, CA: Sage.

Kazdin, A. E. (1994). Methodology, design and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 19–71). Oxford, UK: Wiley.

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology, 52,* 19–26. http://dx.doi.org/10.1016/S0895-4356(98)00138-3

Kjaergard, L. L., Villumsen, J., & Gluud, C. (2001). Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine, 135,* 982–989. http://dx.doi.org/10.7326/0003-4819-135-11-200112040-00010

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

Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

mental health services: A scoping study and typology. *Psychiatric Services, 66,* 224–241. http://dx.doi.org/10.1176/appi.ps.201400141

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172. http://dx.doi.org/10.1037/0022-006X.69.2.159

Lambert, M., & Shimokawa, K. (2011). Collecting client feedback. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (Vol. 2, pp. 203–223). New York, NY: Oxford University press. http://dx.doi.org/10.1093/acprof:oso/9780199737208.003.0010

Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research, 25,* 625–632. http://dx.doi.org/10.1080/10503307.2015.1079661

Lutz, W., Rubel, J., Schiefele, A. K., Zimmermann, D., Böhnke, J. R., & Wittmann, W. W. (2015). Feedback and therapist effects in the context of treatment outcome and treatment length. *Psychotherapy Research, 25,* 647–660. http://dx.doi.org/10.1080/10503307.2015.1053553

Montoya, I. D., Schroeder, J. R., Preston, K. L., Covi, L., Umbricht, A., Contoreggi, C., . . . Gorelick, D. A. (2005). Influence of psychotherapy attendance on buprenorphine treatment outcome. *Journal of Substance Abuse Treatment, 28,* 247–254. http://dx.doi.org/10.1016/j.jsat.2005.01.004

Moran, P., Leese, M., Lee, T., Walters, P., Thornicroft, G., & Mann, A. (2003). Standardised assessment of personality—abbreviated scale (SAPAS): Preliminary validation of a brief screen for personality disorder. *The British Journal of Psychiatry, 183,* 228–232. http://dx.doi.org/10.1192/bjp.183.3.228

OQ-Measures. (2015). *OQ measures*. Retrieved from http://www.oqmeasures.com/

Page, A. C., & Hooke, G. R. (2009). Best practices: Increased attendance in inpatient group psychotherapy improves patient outcomes. *Psychiatric Services, 60,* 426–428. http://dx.doi.org/10.1176/ps.2009.60.4.426

Plambech, B., Lau, M., & Christensen, G. (2000). Stolpegårdmodellen - et behandlingstilbud [The Stolpegaard model—A treatment option]. In J. Andreasen & E. Christiansen (Eds.), *Spiseforstyrrelser. En tværfaglig antologi* [Eating disorders: An interdisciplinary anthology] (pp. 343–365). København, Denmark: Frydenlund.

Pocock, S. J. (1983). Methods of randomization. In S. J. Pocock (Ed.), *Clinical trials—A practical approach* (pp. 66–89). Chichester, UK: Wiley.

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

Reardon, M. L., Cukrowicz, K. C., Reeves, M. D., & Joiner, T. E. (2002). Duration and regularity of therapy attendance as predictors of treatment outcome in an adult outpatient population. *Psychotherapy Research, 12,* 273–285. http://dx.doi.org/10.1080/713664390

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. New York, NY: Wiley.

Sansone, R. A., & Sansone, L. A. (2010). Measuring self-harm behavior with the self-harm inventory. *Psychiatry, 7,* 16–20.

Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology, 61,* 145–153. http://dx.doi.org/10.1002/jclp.20107

Savović, J., Jones, H. E., Altman, D. G., Harris, R. J., Jüni, P., Pildal, J., . . . Sterne, J. A. C. (2012). Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Annals of Internal Medicine, 157,* 429–438. http://dx.doi.org/10.7326/0003-4819-157-6-201209180-00537

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

Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research, 25,* 396–407. http://dx.doi.org/10.1080/10503307.2014.900875

Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. A. (1996). The measurement of disability. *International Clinical Psychopharmacology, 11,* 89–95. http://dx.doi.org/10.1097/00004850-199606003-00015

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for *DSM–IV* and ICD-10. *The Journal of Clinical Psychiatry, 59,* 22–33.

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

Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research, 22,* 638–647. http://dx.doi.org/10.1080/10503307.2012.698918

Slone, N. C., Reese, R. J., Mathews-Duvall, S., & Kodet, J. (2015). Evaluation the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice*.

Smink, F. R., van Hoeken, D., & Hoek, H. W. (2013). Epidemiology, course, and outcome of eating disorders. *Current Opinion in Psychiatry, 26,* 543–548. http://dx.doi.org/10.1097/YCO.0b013e328365a24f

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80,* 547–559. http://dx.doi.org/10.1037/a0028226

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tasca, G. A., Illing, V., Joyce, A. S., & Ogrodniczuk, J. S. (2009). Three-level multilevel growth models for nested change data: A guide for group treatment researchers. *Psychotherapy Research, 19,* 453–461. http://dx.doi.org/10.1080/10503300902933188

Truitt, K. G. (2011). *Modeling treatment outcomes in eating disorders* (PhD Dissertation), Loma Linda University, Loma Linda, CA.

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., . . . Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *British Medical Journal, 336,* 601–605. http://dx.doi.org/10.1136/bmj.39465.451748.AD

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy* (5th ed.). New York, NY: Basic Books.

Received May 23, 2016
Revision received November 15, 2016
Accepted November 22, 2016 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.