cogent
psychology

CLINICAL PSYCHOLOGY & NEUROPSYCHOLOGY | RESEARCH ARTICLE

Effects of feedback-assisted treatment on post-treatment outcome for eating disordered inpatients: A follow-up study

Megan M. Bowen[1], Michael J. Lambert[1]*, Arjan Berkeljon[1], Tyler E. Orr[1], Michael Berrett[2] and Witold Simon[1,3]

*Corresponding author: Michael J. Lambert, Psychology Department, Brigham Young University, Provo, UT 84602 USA
E-mail: michael_lambert@byu.edu

Reviewing editor:
Peter Walla, University of Newcastle, Australia

Additional information is available at the end of the article

Abstract: Only one randomized clinical trial (RCT) has examined feedback-assisted (Fb) treatment in an inpatient eating disordered population. Results from this study suggested that those who received Fb treatment were more likely to recover than participants in the treatment-as-usual condition; however, long-term effects of this treatment have not been investigated. This is especially pertinent in eating disordered populations, where outcomes tend to be poor and course of illness chronic. In the current study, fifty-three women from the aforementioned RCT were contacted three to four years after leaving inpatient care to assess their current distress level and psychological functioning. Results suggested no significant difference between treatment conditions. The vast majority of women sought out multiple forms of treatment over the follow-up period, regardless of treatment condition. This is consistent with past research suggesting that women with more severe pathology (i.e. those requiring inpatient treatment) tend to experience a more chronic pattern of symptoms even after intensive treatment. Overall, the superiority of feedback-assisted treatment found at discharge diminished over time and could not be detected at follow-up. Suggestions for further research are delineated.

ABOUT THE AUTHOR

This research came out of the lab of Michael J. Lambert who directed the Doctoral Dissertation of Megan M. Bowen. Michael J. Lambert, PhD is a professor of psychology at Brigham Young University. He has more than four decades of experience studying the effects of psychotherapy on patients who suffer from psychological disorders. His research primarily concerns improving outcomes in routine clinical care through the use of measuring, monitoring, and feedback to providers and patients. This paper summarizes a 4-year follow-up study of a clinical trial conducted within an inpatient setting for eating disordered patients in which half the patients' therapists received feedback on clinical progress and half received treatment as usual without the benefits of monitoring and feedback. The original trial documented improved mental health outcomes for eating disordered patients whose therapist received progress feedback and problem-solving tools.

PUBLIC INTEREST STATEMENT

While a majority of individuals who enter psychotherapy show a benefit, many do not respond, and a small number (8%) worsen while they in treatment. Worsening and non-response can be reduced substantially if mental health functioning is monitored throughout the course of treatment using brief measurement of common complaints such as anxiety, depression, interpersonal problems, and poor role functioning. When sophisticated statistical modeling is used to identify cases at risk for poor ultimate outcomes while that treatment is ongoing this information can be provided to practitioners and patients in order to alter treatments. Given the heavy burden of psychological illness, effective methods of preventing treatment failure are an important innovation in routine care that is not only effective but also cost effective.

cogent · oa

Subjects: Health and Social Care; Medicine, Dentistry; Nursing & Allied Health; Social Sciences

Keywords: eating disorders; outcome research; mental health services research; progress feedback

1. Introduction

In this era of accountability, clinicians have been placed under increasing pressure to demonstrate the effectiveness of their treatment in bringing about positive patient outcomes. While much of this effort has been invested in the use of empirically supported treatments, alternative evidence-based practices are becoming increasingly popular. One of these alternative methods is the use of patient-focused research in which a therapist is provided feedback about their patients' progress. More specifically, this approach seeks to enhance positive improvement and reduce deterioration in clients by identifying patients at risk for a negative outcome (not-on-track; NOT), providing this predictive information to clinicians and occasionally directly to clients, and assisting clinicians with decision-making tools to interrupt the course of deterioration (Lambert, Hansen, & Bauer, 2008).

The theoretical justification for the use of feedback in treatment has been elaborated upon by Riemer, Rosof-Williams, and Bickman (2005) in the form of contextual feedback theory, which posits that clinician performance improves when information about their performance, especially their errors, is provided in a timely manner. In the current line of research this refers to informing therapists that patients are predicted to have a negative treatment outcome, a phenomenon that they often fail to recognize and, therefore, an instance where feedback is likely to help (Hannan et al., 2005). It has been repeatedly demonstrated in clinical research that predictions relying on statistical or actuarial methods generally tend to fare better than clinical judgment alone (Grove, 2005). Contextual feedback theory also suggests that feedback is more effective at improving patient outcome if it includes suggestions for problem solving. In this case, clinical decision-making tools (clinical support tools: CST) have been provided within the course of treatment with the intent of further interrupting the course of deterioration and changing it toward a positive outcome with clients at risk for treatment failure (Harmon, Lambert, Smart, & Hawkins, 2007; Simon, Lambert, Harris, Busath, & Vazquez, 2012; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Whipple et al., 2003). The CST intervention relies on assessment of therapeutic alliance, patient motivation, social support, and untoward life events with corresponding recommendations for effective actions to improve outcome (Lambert, Bailey, White, Tingey, & Stephens, 2015).

Interventions that incorporate feedback consistent with contextualized feedback theory appear to substantially reduce deterioration rates in patients predicted to be treatment failures. Among other findings, the Shimokawa, Lambert, and Smart (2010) meta/mega-analysis of the effects of applying the OQ®-Analyst system (an electronic form of providing feedback) found that deterioration rates could be reduced from the baseline of 20% in NOT cases to 13% when therapists were alerted to negative patient progress status. The use of the CST intervention further reduced deterioration rates to about 5.5%. A possible limitation to these findings is that the majority of the meta-analyzed studies from the Shimokawa et al. (2010) analysis were conducted at the same university counseling center (Harmon et al., 2007; Lambert et al., 2001, 2002; Slade et al., 2008; Whipple et al., 2003). Counseling center clients typically have a limited range of complaints with relatively low symptom severity. Many of these clients do not meet formal diagnostic criteria, are young, and are likely experiencing their first episode of illness.

In contrast, one of the six studies was conducted in a hospital-based outpatient clinic with a far more disturbed population, all of whom met criteria for a disorder, had long-standing symptoms, often comorbid diagnoses, and complicated psychopharmacology (Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004). The effects of feedback in this study reached statistical significance in some comparisons but were smaller than those found in college counseling center samples. Simon et al. (2012) replicated the Hawkins study within the same outpatient clinic with essentially the same

findings. In addition, Crits-Christoph et al. (2012) reported a study focused on seriously disturbed samples of substance and alcohol abusing patients drawn from sites in New York, Philadelphia, and Salt Lake City. This study found the feedback methods improved patient outcomes. There were no significant differences between the feedback and no-feedback groups at the session at which patients went off-track ($t(115) = .98$, $p = .33$). However, statistically significant differences between the feedback and no feedback conditions were found on drug and alcohol use from the off-track point onwards ($t = 2.56$, $p = .011$). Clinically significant change was not reported in this study, but the effects of the intervention were smaller than those found in the college counseling center samples. None of the aforementioned studies included a follow-up to examine the lasting effects of the feedback interventions.

In the most recent study to come out of this lab, Simon et al. (2013) examined the degree to which delivering patient progress feedback with alarm signals and clinical support tools to psychotherapists would enhance the mental health outcomes of inpatient psychotherapy provided for individuals diagnosed with eating disorders (ED) including anorexia, bulimia, eating disorder not otherwise specified, and obesity. One hundred thirty-three individuals were randomly assigned to a treatment-as-usual (TAU) condition or the experimental feedback (Fb) condition. In the experimental Fb condition, therapists were provided with the OQ-based progress feedback, which indicated if the patients were responding as expected or appeared to be at risk for treatment failure. Clinicians could then discuss the patient's progress with her during the next session and use the Clinical Support Tools to pinpoint which interventions might be most useful in improving outcome.

Results of this study suggested that the feedback intervention based on the OQ-Analyst was effective in enhancing mental health functioning as measured by the OQ-45. There was a consistent pattern of findings, many of them statistically significant, favoring the group that received feedback over the TAU condition. The magnitude of these differences were often large, and they appeared to be clinically meaningful, with 53% of patients in the Fb condition compared to 29% of TAU patients achieving Jacobson and Truax's (1991) criteria for recovery.

As previously mentioned, none of the feedback studies conducted a follow-up study to assess the lasting effects of feedback-assisted treatment when compared to TAU. Women with EDs are an ideal group to assess superiority of feedback-assisted treatment at follow-up, simply because they represent a more disturbed population with higher relapse rates post-treatment (Berkman, Lohr, & Bulik, 2007; Lowe, Zipfel, Buchholz, Dupont, & Reas, 2001). For example, in a large 21-year follow-up study, fifty-one percent of patients with anorexia nervosa were found to be fully recovered at follow-up, while the other half experienced a chronic or lethal course; at follow-up, 21% were partially recovered, 10% still met full diagnostic criteria, and 16% were deceased due to causes related to anorexia nervosa (Lowe et al., 2001). Found similar findings in that over half of patients with anorexia nervosa did not improve in symptomatology over a 6-year follow-up study. If feedback-assisted treatment can reduce relapse rates and improve long-term outcome in a highly disturbed population, such results would be not only impressive but also clinically meaningful. It is possible that those effects could be extended to other severely disturbed populations with poor long-term trajectories.

The current study is considered an extension of the program of research on feedback conducted by Simon et al. (2013). The purpose of the present study was to assess if delivering patient progress feedback and clinical support tools in inpatient ED treatment improves long-term functioning and outcome in this population over TAU at follow-up.

2. Research objectives

The primary research objective for this study was to determine if the use of progress feedback during inpatient treatment impacted the long-term mental health outcome of women with eating disorders. In particular we hoped to determine:

cogent ·· psychology

(a) Is there still a significant difference between Fb and TAU conditions at the point of follow-up (relative to pre-treatment, relative to post-treatment) on mental health functioning?

(b) Are relative differences between conditions on ratios of recovered, improved, unchanged or deteriorated present at follow-up? Particularly, is there a significant difference between the TAU and Fb conditions in the number of women who meet full recovery criteria on the OQ-45 at follow-up?

(c) Have the Fb and TAU conditions maintained their weight within the normal range since leaving treatment?

In addition to the above hypotheses, we were interested in exploring supplementary research questions that were not addressed in the previous study such as:

(a) Have the women received additional treatment since their inpatient stay? If so, what types of treatment have they received and what was the duration of that treatment?

(b) Did the rate of women who underwent additional treatment differ between the Fb and TAU conditions?

(c) What does the women's eating symptomatology look like at follow-up?

Finally, our central hypothesis (in the null form) is that there will be no difference in mental health outcomes at the follow-up point for Fb and the TAU conditions.

3. Method

3.1. Participants

A total of 133 adult female patients (aged 18 or above) who completed treatment (discharged from the hospital by mutual agreement) in the Simon et al. (2013) study were identified using the center for change (CFC) records/database and were invited to participate in the current study. Table 1 presents demographic information on the original study participants in terms of their age, gender, ethnicity, marital status, and employment, as well as the duration of their disorder. The trial methodology, details about the treatment program, and the quality of randomization in the original study are presented in the original report. Of the 133 original participants, 56 of the women responded to our contacts. Three women from this collected sample declined to participate leaving a total of 53 (40%) of the original participants. Twenty-six of the women belonged to the Fb condition and 27 belonged to the TAU condition. Table 2 presents demographic information on the current study's participants in terms of their age, gender, ethnicity, marital status, and employment, as well as the duration of their disorder.

All potential participants were presented with the informed consent prior to their completion of the provided questionnaires (approved by the Human Subjects Institutional Review Boards of Brigham Young University and Center for Change), which described the goal of the study and risks and benefits to participation. Because a large number of the women were contacted via telephone a description of the study was provided and participants gave a verbal consent to participate.

3.2. Procedures

Participants were invited to participate in the current study by postal service mail, email, or by telephone. The questionnaires took approximately 15–20 min for the women to complete and contained a total of 72 items. Survey procedures consisted of several sections: a follow-up satisfaction telephone survey, which assessed current eating behaviors, ratings of recent relationship quality, satisfaction with life and treatment, emotional functioning, and ability to fulfill daily social roles (12 items); an assessment of clients' current height and weight, via self-report, in order to determine body mass index (BMI) (2 items); questions assessing whether participants found the use of the OQ-45 and the clinical support tools to be helpful during their original treatment (4 items); the OQ-45, a

cogent ·· psychology

Table 1. Simon et al. (2013) sample demographics			
	All participants	**TAU**	**Fb**
N	133	64	69
Age			
M	25.51	25.17	25.83
SD	7.67	6.79	8.44
Range	17–54	18–54	17–48
Gender			
Females	133 (100%)	64 (100%)	69 (100%)
Ethnicity			
White	123 (92.48%)	58 (90.63%)	65 (94.2%)
Black	2 (1.5%)	1 (1.56%)	1 (1.45%)
Latino	1 (.75%)	1 (1.56%)	0 (0%)
Asian	4 (3.01%)	2 (3.12%)	2 (2.9%)
Mixed	3 (2.26%)	2 (3.12%)	1 (1.45%)
Marital status			
Single	94 (70.68%)	44 (68.75%)	50 (72.46%)
Engaged	1 (.75%)	0 (0.0%)	1 (1.45%)
Married	28 (21.05%)	17 (26.56%)	11 (15.94%)
Divorced	8 (6.0%)	3 (4.69%)	5 (7.25%)
Other	2 (1.5%)	0 (0.0%)	2 (2.90%)
Employment			
Employed	54 (40.6%)	28 (43.75%)	26 (37.69%)
Unemployed	79 (59.4%)	36 (56.25%)	43 (62.31%)
Duration of disorder (in years)			
Mean	10.93	10.44	11.38
SD	7.5	6.63	8.24
Range	1–35	1–35	1–33

measure of current psychological distress (45 items); and Section III of the LIFE-EAT-II, which assessed additional treatment sought since leaving treatment (14 items).

3.3. Measures

3.3.1. Outcome questionnaire
The Outcome Questionnaire-45 (OQ-45; Lambert, et al., 2013) is a well-established instrument that has been validated across a broad range of normal and client populations. The measure was designed to assess three aspects of the client's life: subjective discomfort/symptoms, problems in interpersonal relationships, and problems in social role performance. The items also measure personally and socially relevant characteristics that affect the individual's quality of life, attempting to quantify both positive and negative functioning. Each item is scored on a 5-point scale with the total score yielding a range of possible scores of 0–180; higher values indicate the endorsement of pathology. Completion of the OQ-45 takes approximately 5–7 min. The OQ-45 has adequate internal consistency ($r = 0.93$) and 3-week test–retest reliability ($r = 0.84$). Concurrent validity is moderate to high ($r = 0.50$–0.85) when correlated with measures most often used to assess psychotherapy outcome in clinical trials. Concurrent validity of the OQ-45 total score has been examined by correlating it with the Symptom Checklist-90 (SCL-90; Derogatis, 1977), Beck Depression Inventory (BDI; Beck, Steer, & Carbin, 1988), Zung Depression Scale (Zung, 1965), and the State-Trait Anxiety Inventory (STAI; Spielberger, 1983), and additional measures. Evidence supporting the factor structure of the

cogent ·· psychology

Table 2. Current sample demographics			
	All participants	**TAU**	**Fb**
N	53	27	26
Age			
M	30	28.92	32.77
SD	7.71	3.85	10.02
Range	21–53	24–39	21–53
Gender			
Females	53 (100%)	27 (100%)	26 (100%)
Ethnicity			
White	48 (90.57%)	24 (88.88%)	24 (92.30%)
Black	1 (1.89%)	1 (3.70%)	0 (0.0%)
Latino	1 (1.89%)	1 (3.70%)	0 (0%)
Asian	3 (5.66%)	1 (3.70%)	2 (7.69%)
Mixed	0 (0.0%)	0 (0.0%)	0 (0.0%)
Marital Status			
Single	40 (75.47%)	21 (77.77%)	19 (73.08%)
Engaged	0 (0.0%)	0 (0.0%)	0 (0.0%)
Married	3 (5.66%)	1 (3.70%)	2 (7.69%)
Divorced	1 (1.78%)	0 (0.0%)	1 (1.78%)
Other	9 (16.98%)	5 (18.52%)	4 (15.38%)
Employment			
Employed	20 (37.74%)	6 (22.22%)	12 (46.15%)
Unemployed	33 (62.26%)	21 (77.77%)	14 (53.84%)
Duration of disorder (in years)			
Mean	17.06	15.15	18.96
SD	7.87	4.98	9.69
Range	6–39	7–29	6–39

OQ-45 has been reported by Bludworth, Tracey, and Glidden-Tracey (2010), de Jong et al. (2007), and Lo Coco et al. (2008). The OQ-45 has been shown to be sensitive to change in clients over short periods while remaining stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000).

In the original study, the OQ-45 was administered on a once-a-week basis prior to and during their inpatient stay. The range of length of stay for participants was one week to 39 weeks. The average length of stay for participants was 13.1 weeks in the hospital (with a standard deviation of 6.39). In the current study, the client only completed the OQ-45 once. The OQ-45 was used specifically for the purpose of assessing patient mental health functioning three to four years after the original in-patient treatment was received. In addition to the total score, the OQ-45's three additional subscales were helpful in assessing patients' quality of interpersonal relations, social role functioning, and symptom distress.

3.3.2. Perceived helpfulness of the OQ-45
Four additional questions were added to the study survey to explore the clients' perceptions and attitudes about the use of the OQ-45 and receiving feedback from their therapists (if they received it) during their inpatient treatment.

3.3.3. Body mass index

Clients were asked to provide their current height and weight in order to calculate their BMI. The previous study weighed each of the women on a weekly basis and calculated their BMI as a means to assess overall symptom improvement; however, it should be noted that the body weight of clients with eating disorders can fall within the normal range despite experiencing clinical levels of symptomatology. The previous study and the current study calculated BMI by using the equation BMI = (weight (lbs.)/height (in.) × height (in.) × 703; www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html). A portion of the women indicated they did not know their current weight, and therefore, either did not provide it or simply "guessed" for the purposes of the study. This is to be expected given that the Center encouraged them not to weigh themselves regularly once out of the hospital.

3.3.4. Phone survey

The CFC research team developed a satisfaction survey that could be conducted by phone or on a written survey form, and they use it regularly in their routine follow-up/aftercare procedures. This survey is administered at long-term intervals to assess former clients' perceptions of their recovery and general well-being. The survey consists of 12 questions, which include eating patterns and frequency of bulimic or anorexic behaviors in the past month, ratings of recent relationship quality, satisfaction with life, spirituality, emotional functioning, and ability to fulfill daily social roles. A total long-term outcome score can be computed based on patients' responses with higher scores indicating more disturbances; however, a total score was not computed for each of the women in this study. The survey was primarily used for qualitative data about the women's current eating patterns; we simply did not ask questions about relationships or spirituality for the sake of brevity, as very similar questions are addressed on the OQ-45. While typically researchers would use a much more extensive measure looking at eating disorder behaviors, such as the Eating Attitudes Test or the Eating Disorder Inventory 3, out of necessity for brevity we decided to forgo these types of questionnaires. Our aim in this study was to shorten the amount of questions the women needed to answer given that our main questionnaire (OQ-45) was lengthy and we simply wanted to minimize our intrusion into their lives. Also, we needed our survey to be short enough to motivate the women to participate.

The Center for Change has conducted an in-house study assessing the psychometric properties of this measure that suggested it correlated well with the EAT test and the OQ-45 and provided evidence that the survey was a valid long-term outcome measure for eating disorder patients, regardless of the eating disorder diagnosis; however, because the Center's study was never published and was not available for review at the time of this study, these psychometric properties cannot be substantiated. We also have no ability to examine how our participants compare to the previous women at the Center who completed the phone survey. We recognize that our choice to use this measure is a limiting factor.

3.3.5. The longitudinal interval follow-up evaluation

Clients' treatment history since leaving the inpatient facility was gathered using The Longitudinal Interval Follow-up Evaluation (LIFE-EAT-II), a modified version of the LIFE II (Keller et al., 1987). The LIFE-EAT-II is an instrument designed for collecting longitudinal data on eating disorders, comorbid psychopathology, treatment participation, and psychosocial functioning. The LIFE-EAT-II is a very comprehensive measure that would take a significant amount of time to administer to each participant. Because we wanted to minimize our intrusion into the women's lives and motivate them to participate in a *brief* survey, only section III of the Longitudinal Course of Treatment was used from the instrument. The addition of this section of the survey was to determine the extent to which clients sought additional treatment since leaving the inpatient facility and whether additional treatments may explain current levels of functioning and/or any improvement outside of the treatment they received during their inpatient stay. This portion of the survey was used to assess: (a) treatment received since leaving the in-patient setting; (b) types of treatment received; (c) duration of specific treatment received; (d) assess the circumstances under which additional treatment was sought after leaving the inpatient facility.

3.4. Data collection

The participants in the original study were admitted to and discharged from the Center for Change between January 2009 and January 2011. This means that by the time we contacted the participants for the current follow-up study in 2014, approximately three or four years had elapsed since their original stay at the hospital. The sheer length between the participants' original stay and our attempts to contact them several years later raised concerns about our ability to actually contact the women. It was anticipated that many had moved away from the address they resided at during the time of their inpatient stay or that many had changed their email addresses/phone numbers since that time. For this reason the research team concluded that it would be important to utilize multiple methods of contact to maximize participation. In order to maximize response rate, four elements were utilized in designing the survey and associated methodology: (1) we developed a participant-friendly questionnaire, (2) we initiated three types of contact, (3) our procedures allowed for ease of electronic submission, (4) participants were sent two reminders via postcard and email after the initial packet was sent, and (5) the researchers offered a monetary compensation in the form of a $10–$20 Amazon or Target gift card.

The lead researcher and the lead Aftercare employee first hand-addressed envelopes that contained a brief letter signed by the Aftercare employee (a person whom the women were familiar with) explaining the purpose of the study, a consent form, and the written survey materials. We then mailed the packets out on 25 February 2014 to the mailing addresses on file for each participant. In an effort to maximize response rate we then sent the survey materials (introductory letter, OQ-45, phone survey, and follow-up questions) through a Qualtrics generated email link on 28 February 2014 to the email addresses we had on file for each of the women. After this first attempt to contact the women we only received three returned mailed packets and only 18 of the women completed the survey through Qualtrics. For this reason, it was decided that we would send a reminder to the women through mail/email. The first mailed reminder was sent on 6 March 2014 approximately a week after the first packet was sent (to account for the time it would take to get to the addresses) and the Qualtrics reminder link was sent on 14 March 2014. After these first reminders were sent, five additional women completed the survey via Qualtrics and one additional woman mailed a completed packet back to us. Again, because we were not receiving the response we had hoped for we sent out a second reminder via mail/email to the women. This second reminder advertised our need for each individual woman's response in order to complete the study and we advertised that the compensation was now increased from $10 to $20 to further elicit. The second mailed reminder was sent on 20 March 2014 and the second emailed reminder was sent on 25 March 2014. After these reminders were sent we received seven additional completed surveys in the mail and one additional participant completed the survey via Qualtrics.

A limiting factor to this point was that our ability to contact participants was dependent on the contact information that was available through the Center, as not all of the women maintained the email addresses or postal addresses they provided at discharge. For this reason, the lead researcher and lead Aftercare employee began to directly contact participants between the dates of 1 April 2014 and 1 June 2014 via the telephone numbers provided in their file. Our ability to contact participants via telephone was limited by confidentiality concerns, in that we left no voicemails about our purpose for calling and we did not disclose our purpose for calling to other individuals answering the telephone. Eventually, our calling efforts led to contact with 18 additional women who completed the survey orally over the phone (after receiving their consent). As participants completed the phone, web-based, or mail-in survey, their data was entered and compiled into the CFC database by an employee until the survey participation was closed.

3.4.1. Compensation

Participants were offered a choice of a $10–$20 Amazon or Target gift card upon their completion of the survey. The gift card was sent using the email address or mailing address used to contact the participants for the study. Gift cards were sent to participants approximately within one week after the full data-set had been collected.

3.4.2. Confidentiality

Survey data collected electronically from participants was initially stored on the inpatient facility's account on Qualtrics.com. All survey forms completed online were safeguarded to the degree permitted by Qualtrics.com; the Qualtrics system maintains data behind a firewall. Upon completion of the study, responses were downloaded from Qualtrics.com to the facility's existing data-set. Survey data collected from the telephone and from postal mail were manually entered into the facility's electronic database by an Aftercare employee. In the case of mail-in surveys, the only identifying information contained in the mailed packet was an assigned number given to each participant. Based on participants' responses, summary assessments were generated and sorted into individual participant charts for clinical use and data analysis. All confidential files and charts remained on site and the necessary information for this study was entered into a separate computer data file. After receiving permission from the facility's partners and the facility's human subjects review committee, the investigator was given access to the computerized data file with all participants identified by an assigned number. Thus, all personally identifying information was kept separate from participant responses to measures. The data file included data relevant only to this proposed study.

3.5. Data analysis

First, several preliminary linear regression models were conducted using the collected data to determine study variables that might covary with the dependent variable and hence, predict follow-up outcome. Several of the covariates examined in the models included pre-treatment scores on the OQ-45, post-treatment scores on the OQ-45, length of treatment at the center, and treatment received after inpatient stay. The variables that most strongly predicted follow-up outcome were used as the covariate in a one-way analysis of covariance (ANCOVA) with the post-treatment-to-follow-up change (as measured by the OQ-45) as the dependent variable. ANCOVA allowed precise comparisons between groups on post-treatment outcome as a function of experimental group while accounting for a covariate. The standardized mean difference statistic, d effect size (Cohen, 1988), was applied.

The ratio of recovered, improved, unchanged, or deteriorated patients was derived from Jacobson and Truax (1991) criteria for clinical significance (see also Lambert et al., 2008). In the case of the OQ-45, patients are considered as having made "reliable change" when their scores change in a positive or negative direction by at least 14 points. Based on the reliability of the OQ-45 this degree of change exceeds measurement error. The cut-off score for the OQ-45 that determines whether a person's score is more likely to come from a dysfunctional population than a functional population is estimated to be 64/63 (Lambert et al., 2013). Therefore, in order to have made clinically significant change and to be considered as functioning similarly to a non-patient population, a person's score must fall at or below 63. In order to be considered recovered clients must improve their OQ score by 14 or more points and pass the clinical cut-off score of 64. Support for the validity of the OQ-45's reliable change and clinical significance cut-off score have been reported by Lunnen and Ogles (1998) and Beckstead et al. (2003). Chi-square analyses were used to assess if there was a significant difference between the two groups on the proportions of those meeting recovered, improved, or deteriorated criteria. Descriptive statistics were used to characterize the demographics of the patients and relationships between the various measures used in the study.

4. Results

4.1. Preliminary analyses

Figure 1 illustrates the OQ-45 mean scores for each treatment condition across time. As can be seen, both treatment conditions follow a similar trajectory in terms of distress (as measured by the OQ-45). Observation of the graph alone would suggest that while the feedback condition had a slight advantage over TAU at post-treatment, at follow-up both treatment conditions appear to have approached one another. These findings will be more closely examined later in the results section.

Figure 1. Mean OQ-45 scores for each treatment group across time excluding outliers.
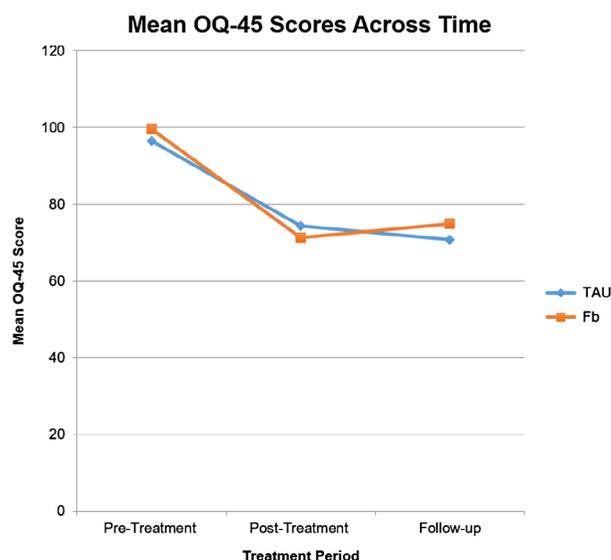
Prior to conducting any outcome analyses, we examined how the individuals who participated in the follow-up study differed in terms of distress levels from those individuals we were unable to reach for follow-up. Specifically, we compared individuals in the TAU and feedback conditions who were followed with individuals in each condition that we were unable to observe at follow-up. Sample means were compared at the pre-test and post-test points using t-tests for independent samples. First, we ran a t-test comparing the *pre-treatment* OQ-45 scores of our follow-up sample to the non-follow-up sample across all treatment conditions. At pre-treatment the follow-up sample did not reliably differ from those we did not contact ($t(131) = 1.0$; $p = .32$; Non-sample $M = 94.23$ $SD = 20.13$; Follow-up $M = 97.91$ $SD = 21.87$). Thus, those included in the follow-up sample appeared to be representative of the whole sample in level of disturbance before inpatient treatment. Second, the *post-treatment* OQ-45 scores of our follow-up sample were compared with the non-follow-up sample across all treatment conditions. We found that there was no significant difference between the entire follow-up sample's OQ-45 scores and the entire non-follow-up sample's OQ-45 scores at post-treatment ($t(131) = .87$; $p = .39$; Non-sample $M = 63.66$ $SD = 25.00$; Follow-up $M = 67.74$ $SD = 28.53$). Thus, it was concluded that the follow-up sample as a whole had improved to an equivalent degree as the non-follow-up sample and may be considered equivalent at post-treatment.

Perhaps, more important was to test for differences between the follow-up feedback sample and their non-follow-up feedback counterpart. Results showed no reliable difference between these groups at pre-treatment ($t(67) = .34$; $p = .73$; non-sample $M = 97.65$ $SD = 19.32$; follow-up $M = 99.5$ $SD = 24.92$) or post-treatment ($t(67) = .44$; $p = .66$; non-sample $M = 60.56$ $SD = 28.05$; follow-up $M = 63.81$ $SD = 32.26$). At the beginning and end of treatment our feedback follow-up sample was essentially equivalent to the un-followed feedback participants.

With regard to our follow-up TAU sample and the non-follow-up TAU sample, neither of these groups significantly differed at pre-treatment on their OQ-45 scores ($t(62) = 1.22$; $p = .23$; non-sample $M = 90.24$ $SD = 20.59$; follow-up $M = 96.37$ $SD = 18.82$). Our follow-up TAU sample and our non-follow-up TAU sample also did not significantly differ on their OQ-45 scores at post-treatment ($t(62) = .75$; $p = .46$; non-sample $M = 67.27$ $SD = 20.71$; follow-up $M = 71.52$ $SD = 24.44$). We found this data to be particularly encouraging, as it appears that the sample we were able to obtain at follow-up closely resembled the remaining treated women who were not able to be reached at follow-up. This has important implications for our ability to generalize what we found with our sample to the remaining women who were lost at follow-up.

4.2. Long-term feedback results

Preliminary regression models were used to determine covariates that predict the difference between the post-treatment OQ-45 score and the follow-up OQ-45 score. Several moderate covariates were found including: restricting behavior (p = .01); group therapy attendance (p = .03); and pre-treatment OQ-45 score (p = .37). While the initial level of disturbance (pre-treatment OQ-45 score) did not meet the .05 level of confidence to be considered a covariate, we decided to include it because it was a significant covariate in the original study. These covariates were then used in the ANCOVA model.

Prior to conducting the ANCOVA, the data was examined to assess the number of outlier OQ-45 scores that were in each treatment condition. It was determined that there was one outlier score in the TAU group and three outlier scores in the feedback group. While there is no empirical evidence for what constitutes an "outlier" OQ-45 score, for the purposes of this study, we considered a score below 20 to constitute an outlier. An OQ-45 score below 20 (about a standard deviation healthier than community non-patient adults) would be feasible if the scores were produced by a person in a non-patient normative sample; however, the outlier scores were obtained while the women were on an inpatient unit. For this reason, it was assumed that those scores were a result of disingenuous reporting. Logic would suggest that if someone has severe enough symptoms to be on an inpatient unit, their level of distress would exceed that of an outpatient population average score. Therefore, these outlier scores were excluded while conducting the One-Way ANCOVA.

The ANCOVA yielded a non-significant p-value (p = .62), which indicates that the feedback group and TAU group could not be distinguished at the end of the follow-up ((F (1, 9) = .25, p = .62); see Table 3). In addition, neither group showed improvement over time from post-treatment to follow-up.

It should be noted that when examining effects using small sample sizes, significance testing can be misleading because statistical significance is difficult to obtain. However, the mean change over time and the difference between groups at follow-up were quite small, suggesting that even with a larger sample they would have likely failed to meet the .05 level of confidence. Nevertheless, Cohen's d was calculated, as this statistic is independent of sample size and can estimate the size of effects. The calculated effect size Cohen's d, between the post-treatment OQ-45 scores and the follow-up OQ-45 scores for the feedback group was d = −0.12 (see Table 4). The effect size between the post-treatment OQ-45 scores and the follow-up OQ-45 scores for the TAU group was d = .16 (see Table 4) and was in a positive direction. The calculated effect size comparing both groups at follow-up was d = −0.15 (see Table 4). Cohen's criteria for small effects suggest the d-value should exceed 0.20 for it to be considered even a small effect. By this standard it must be concluded that patients did not change much during the follow-up period, and even though the raw OQ-45 mean score indicated the feedback group regressed on average and the TAU group slightly improved, the effect sizes did not meet criteria for a small change. The TAU condition improved on average 4.19 OQ-45 points from post-treatment to follow-up (after adjusting for covariates). The TAU group ended treatment in the 91.9th percentile of the normal population as measured by the OQ-45 and at follow-up they fell in the 90th percentile of the normal population. The feedback condition actually deteriorated by an average of 3.56 OQ-45 points from post-treatment to follow-up. The feedback group ended treatment at the 84th percentile of the normal population on the OQ-45 but fell at the 93.3rd percentile of the normal population at follow-up. Even though the feedback condition was better off at termination than the TAU condition, the TAU condition appears to have closed this margin and was slightly better off at follow-up; however, keep in mind that the inferential statistic indicates that no reliable change in the groups across time could be found.

4.3. Clinical significance

To further examine the impact of feedback on long-term outcome, i.e. to examine the outcome for each individual patient and determine whether it was clinically meaningful, final individual patient outcomes were categorized according to the number of patients that responded to treatment (i.e. met either reliable or clinically significant change criteria) and those that did not respond to

Table 3. One-way ANCOVA[a]				
Treatment conditions	**M^b**	**SD**	**F**	**P**
Treatment as usual	−4.19	32.34	.25	.62
Feedback	+3.56	25.25		

[a]Dependent variable is change between post-treatment and follow-up OQ-45 scores and covariates used were pre-treatment OQ-45 score, restricting behaviors, and group therapy.

[b]Mean change in OQ-45 points.

Table 4. Outcome results by treatment group measured by OQ-45 with outliers omitted					
	Post	**Follow-up**	**Change**	**Post to follow-up**	**Follow-up to follow-up**
Feedback group (n = 23)					
M	71.35	74.91	+3.56[a]		
SD	25.79	32.34			
d				−0.12[b]	−0.15[c]
Treatment as usual (n = 26)					
M	74.27	70.73	−4.19		
SD	20.21	25.25			
d				0.16	−0.15[c]

[a]This number is positive because the feedback group increased their total score.

[b]This effect size is reflected as a negative because it means the Fb group's OQ-45 scores experienced change in a negative direction i.e. they became worse.

[c]Comparison between TAU and feedback change scores at follow-up; this effect size comparing both groups at follow-up is negative, which signifies that the TAU group was functioning slightly better at follow-up than Fb group.

treatment (reliably worsened or did not change). Looking from pre-treatment to follow-up, clients exposed to the feedback condition met criteria for clinically significant change at a similar frequency to those participating in the TAU condition (30% vs. 35%). Clients from the feedback condition also met the lesser standard of reliable change at a similar frequency to those participating in the TAU condition (35% vs. 33%). An equal proportion of feedback and TAU patients remained unchanged from pretest to follow-up (Fb 22% vs. TAU 22%), as well as an equal proportion of patients from each group reliably worsened (12% vs. 7%). A chi-square test for homogeneity was conducted to assess if outcome classification had the same probability across treatment groups. The results suggested that there was no difference in the proportions of deteriorated, unchanged, improved, or recovered across the two treatment groups (χ^2 (3, N = 53) = 0.4035, p = .94; see Table 5). It should be noted that every patient in the study could improve by 14 points but not every patient in the study could cross the threshold from dysfunctional to functional. In other words, if patients were in the functional range at termination they could not pass the "functional" threshold to be considered recovered at follow-up. An explanation for the similar category membership across treatment groups may be the fact that our sample size was not sufficient to find a significant difference in category membership. In addition to small sample size, using categorical data instead of means and standard deviations is less sensitive to actual differences, if they do exist.

When examining clinically significant change in our sample from only post-treatment to follow-up we found that a larger proportion of those in the feedback condition had reliably worsened (TAU 27% vs. Fb 39%; see Table 6). However, clients exposed to the feedback condition met criteria for clinically significant change at a similar frequency to those participating in the TAU condition (26% vs. 31%). Clients from the feedback condition also met the lesser standard of reliable change at a similar frequency to those participating in the TAU condition (13% vs. 8%). A larger proportion of TAU patients remained unchanged from post-treatment to follow-up when compared to Fb patients (35% vs. 22%). These ratios are consistent with the effect size findings, and it can be observed that the Fb group has a larger percentage of women who had reliably worsened since post-treatment.

Table 5. Percentage of patients meeting clinically significant change, reliable change, reliably worsened, or unchanged criteria on the OQ-45 from pre-treatment to follow-up

Outcome classification	Treatment as usual *n* (%)	Feedback *n* (%)
Reliably worsened[a]	2 (7)	3 (12)
Unchanged[b]	6 (22)	5 (22)
Reliably improved[c]	9 (33)	8 (35)
Clinically significant change (recovered)[d]	9 (35)	7 (30)

[a]Worsened by at least 14 points on the OQ-45 from pre-treatment to follow-up.

[b]Improved less than 14 points or worsened by less than 14 points on the OQ-45.

[c]Improved by at least 14 points on the OQ-45 but did not pass the cut-off between dysfunctional and functional populations.

[d]Improved by at least 14 points on the OQ-45 and passed the cut-off between dysfunctional and functional populations; TAU N = 26; Feedback N = 23 (excluded outliers).

4.4. Diagnoses

Although the original study assessed differential outcome across diagnostic categories, given the small sample size within diagnostic groupings in our study it made little sense to analyze data within diagnostic categories.

4.5. BMI

Changes in weight, as measured via the body mass index, were analyzed to see if mental health functioning feedback had any effect on each client's body mass index (BMI). In the field of eating disorder treatment, BMI is recognized as an important marker of the effectiveness of psychotherapy for individuals with ED, especially those with Anorexia Nervosa and obesity. In this study we followed standard BMI range categories (in kg/m^2) to classify the participants as: starvation <16.00; severely underweight 16.00–16.99; underweight 17.00–18.50; normal 18.51–25.00; overweight from 25.01–30.00; obese class I 30.01–35.00; obese class II 35.01–40.00; obese class III >40.00.

The individuals that were contacted for this follow-up study were a combination of those who began treatment outside of a normal weight category and those who did not. Thirty-one out of the 53 individuals (59%) who were contacted for follow-up began treatment outside of the normal BMI category (22/53 or 41% began treatment within the normal range). Of particular interest were those 31 individuals who could improve their BMI between pre-treatment and post-treatment and how many individuals reached the normal weight category (by post-treatment and by follow-up) if they started outside normal limits. It is implied here that "moving toward" a normal BMI category means moving from a less favorable category to a more favorable one, for example, from obese class II to obese class I, or from starvation to severely underweight.

Of the 31 individuals in our sample that began outside of the normal BMI category 14 individuals 45% (14/31) belonged to the TAU group and 55% (17/31) belonged to the feedback group. Simply looking at admission to post-treatment, 27 of the 31 traceable individuals reached a normal BMI category (see Table 7). Specifically, of the 14 individuals in the TAU condition, 13 (92.85%) improved with regard to weight classification, and 13 (92.85%) reached normal weight. Of the 17 individuals who participated in the feedback condition, 15 (88.23%) improved their weight classification and 14 (82.35%) reached normal BMI. We can conclude that the Center was relatively successful in improving BMI during the women's stay.

Assessing movement toward normal BMI category from *admission to follow-up* was significantly more difficult because of the number of women who did not provide their BMI at follow-up; eleven women in the sample did not provide their current BMI at follow-up. Many of the participants simply did not provide their BMI on their written materials and others admitted that as part of their recovery they had not weighed themselves in quite some time, and as such, their estimations would likely be incorrect. Of the 12 women who began outside of normal at pre-treatment and who also provided a

cogent ··psychology

Table 6. Percentage of patients meeting clinically significant change, reliable change, reliably worsened, or unchanged criteria on the OQ-45 from post-treatment to follow-up		
Outcome classification	Treatment as usual *n* (%)	Feedback *n* (%)
Reliably worsened[a]	7 (27)	9 (39)
Unchanged[b]	9 (35)	5 (22)
Reliably improved[c]	2 (8)	3 (13)
Clinically significant change (recovered)[d]	8 (31)	6 (26)

[a]Worsened by at least 14 points on the OQ-45 from pre-treatment to follow-up.

[b]Improved less than 14 points or worsened by less than 14 points on the OQ-45.

[c]Improved by at least 14 points on the OQ-45 but did not pass the cut-off between dysfunctional and functional populations.

[d]Improved by at least 14 points on the OQ-45 and passed the cut-off between dysfunctional and functional populations; TAU N = 26; Feedback N = 23 (excluded outliers).

BMI at follow-up and received TAU, 7 (58.33%) improved with regard to their weight classification and 7 (50.00%) reached normal BMI at follow-up. Of the 13 individuals who began outside of normal at pre-treatment, provided a BMI at follow-up, and received feedback, 10 (76.92%) improved with regard to their weight classification but only 5 (29.41%) reached normal BMI by follow-up (see Table 8). It appears from these percentages that recidivism may be occurring, as only a small percent of the traceable women who fell into the normal BMI category at post-treatment fell in the normal BMI weight category at follow-up (87% (27/31) post-treatment vs. 39% (12/31)) at follow-up. Again, keep in mind that the numbers are reduced because some participants did not know their weight at follow-up.

We wanted to assess movement toward and movement away from the normal BMI category from post-treatment to follow-up, as any deterioration that would take place would be expected during this time period (see Table 9). As mentioned previously, 11 of our follow-up sample did not report their BMI; this left us with 42 women across both treatment conditions who could be tracked from post-treatment to follow-up. Of these 42 women we found that 36 of them or 85.7% ended treatment in the normal range. At follow-up, 26 of these 36 women or 61.9% had maintained the BMI that they left the hospital with. Of the TAU condition, 66.6% were able to maintain their BMI level after leaving the hospital. Of the feedback condition 57.1% maintained their post-treatment BMI.

Improvement from post-treatment to follow-up was minimal, as there were only two individuals (one from Fb and one from TAU) or 4.8% who improved their BMI or moved closer toward a normal weight category. Unfortunately, a number of the women's BMIs deteriorated since they left the hospital. Specifically, 14 of the 42 women or 33% of the sample moved further away from a normal weight category since leaving the hospital; 28.6% (6/21) of individuals in the TAU group deteriorated in terms of BMI; and 38.1% (8/21) of individuals in the feedback condition deteriorated in terms of BMI (See Table 9). Within the anorexic group, where the largest number of individuals was outside normal BMI to begin with and where it could be argued that extreme thinness presents the greatest health risk, the original study did not find an effect for feedback. Nevertheless, it was hoped that such an effect might emerge at follow-up. Specifically, we were hoping to find that the women in the feedback condition with anorexia would have maintained their normal BMI or improved during the follow-up time period. This is not what was found; those with anorexia, especially in the feedback group, experienced the most difficulty in maintaining their weight compared to other diagnostic categories.

4.6. Further treatment
It was anticipated that many of the women in the study would seek out additional treatment after leaving the Center to either maintain gains or prevent a relapse, as it is well known that eating disorders tend to have a chronic course. It is also standard practice in this setting to arrange outpatient care before discharge, thus ensuring some additional outpatient care. Therefore, participating in

Table 7. Movement toward normal BMI category from pre-treatment to post-treatment				
	ALL[a]	**AN[b]**	**BN[c]**	**EDNOS[d]**
Started outside normal				
TAU & Fb	31/53 (58.5)	22/31 (70.9)	4/31 (12.9)	5/31 (16.1)
TAU	14/31 (45.2)	10/14 (71.4)	2/14 (14.3)	2/14 (14.3)
Fb	17/31 (54.8)	12/17 (70.6)	2/17 (11.8)	3/17 (17.6)
Moved toward normal				
TAU & Fb	28/31 (90.3)	22/22 (100)	4/4 (100)	2/5 (40)
TAU	13/14 (92.9)	10/10 (100)	2/2 (100)	1/2 (50)
Fb	15/17 (88.2)	12/12 (100)	2/2 (100)	1/3 (33.3)
Reached normal				
TAU & Fb	27/31 (87.1)	21/22 (95.5)	4/4 (100)	2/5 (40)
TAU	13/14 (92.9)	10/10 (100)	2/2 (100)	1/2 (50)
Fb	14/17 (82.4)	11/12 (91.7)	2/2 (100)	1/3 (33.3)

[a]ALL, all participants.
[b]AN, anorexia nervosa.
[c]BN, bulimia.
[d]EDNOS, eating disorder not otherwise specified.

Table 8. Movement toward normal BMI category from pre-treatment to follow-up				
	ALL[a]	**AN[b]**	**BN[c]**	**EDNOS[d]**
Started outside normal				
TAU & Fb	31/53 (58.5)	22/31 (70.9)	4/31 (12.9)	5/31 (16.1)
TAU	14/31 (45.2)	10/14 (71.4)	2/14 (14.3)	2/14 (14.3)
Fb	17/31 (54.8)	12/17 (70.6)	2/17 (11.8)	3/17 (17.6)
Moved toward normal				
TAU & Fb	17/25 (68)	15/18 (83.3)	1/3 (33.3)	1/4 (25)
TAU	7/12 (58.3)	6/8 (75)	1/2 (50)	0/2 (0)
Fb	10/13 (76.9)	9/10 (90)	0/1 (0)	1/2 (50)
Reached normal				
TAU & Fb	12/31 (38.7)	10/22 (45.5)	1/4 (25)	1/5 (20)
TAU	7/14 (50)	6/10 (60)	1/2 (50)	0/2 (0)
Fb	5/17 (29.4)	4/12 (33.3)	0/2 (0)	1/3 (33.3)

[a]ALL, all participants.
[b]AN, anorexia nervosa.
[c]BN, bulimia.
[d]EDNOS, eating disorder not otherwise specified.

treatment during the follow-up period was not necessarily an indication of treatment failure. In order to assess the types of treatment the women had received since they left, we asked if they had participated in individual therapy, group therapy, psychoeducation, family therapy, nutritional management, medical care, or hospital/residential services since leaving the center. We wanted to explore the percentage of participants who received further treatment and whether those who sought additional treatment belonged to the feedback or TAU group. We were also interested in conducting a correlation to assess the relationship between receiving or not receiving further treatment after discharge and maintenance/loss of the treatment gains made in the hospital.

Table 9. Movement toward normal BMI category from post-treatment to follow-up				
	ALL[a]	**AN[b]**	**BN[c]**	**EDNOS[d]**
Normal at post-treatment				
TAU & Fb	36/42 (85.7)	17/18 (94.4)	14/15 (93.3)	5/9 (55.5)
TAU	19/21 (90.5)	8/8 (100)	9/9 (100)	2/4 (50)
Fb	17/21 (80.9)	9/10 (90)	5/6 (83.3)	3/5 (60)
Deteriorated at follow-up				
TAU & Fb	14/42 (33.3)	7/18 (38.9)	3/15 (20)	4/9 (44.4)
TAU	6/21 (28.6)	2/8 (25)	1/9 (11.1)	3/4 (75)
Fb	8/21 (38.1)	5/10 (50)	2/6 (33.3)	1/5 (20)
Improved at follow-up				
TAU & Fb	2/42 (4.8)	0/18 (0)	0/15 (0)	2/9 (22.2)
TAU	1/21 (4.8)	0/8 (0)	0/9 (0)	1/4 (25)
Fb	1/21 (4.8)	0/10 (0)	0/6 (0)	1/5 (20)
Maintained BMI				
TAU & Fb	26/42 (61.9)	11/18 (61.1)	12/15 (80)	3/9 (33.3)
TAU	14/21 (66.7)	6/8 (75)	8/9 (88.9)	0/4 (0)
Fb	12/21 (57.1)	5/10 (50)	4/6 (66.7)	3/5 (60)

[a]ALL, all participants.

[b]AN, anorexia nervosa.

[c]BN, bulimia.

[d]EDNOS, eating disorder not otherwise specified.

Upon closer examination it became clear that essentially everyone who participated in the study had sought out additional treatment after leaving the center, all except for one person in the feedback group. Therefore, it is clear that we do not have enough people in the sample who did *not* receive additional treatment in order to make a statement about receiving extra treatment and maintaining gains vs. *not* receiving extra treatment and maintaining gains.

For descriptive purposes the data can be summarized to list the types of additional treatment that was received by the sample. First, everyone in the sample, both TAU and feedback, sought out *multiple* forms of additional treatment, meaning they often combined different forms of treatment such as individual therapy and medication. None of the participants had tried only one form of additional treatment (this is of course excluding the one participant who sought no additional therapy). The majority of participants from both groups sought out at least three types of additional treatment or more, except for two participants who only sought out two additional forms of treatment. The majority of our sample had received nutritional management at some point since leaving the hospital, with only four individuals not receiving this additional treatment. The majority of the sample had also received additional individual therapy; only two individuals did not receive individual therapy after leaving the Center. The majority of the sample had also received medication during their hospital stay and after leaving, except for six individuals who began medication after leaving the hospital.

Several implications can be made based on the above data. Despite receiving additional treatment, gains were not obvious in either group. Or in other words, further treatment after hospitalization did not bring individuals any closer to the normative population. So, at best, the additional treatment sought may have prevented relapse or helped individuals to maintain their gains rather than increase their gains after leaving the hospital. It can likely be concluded that the follow-up sample, and perhaps all of the treated patients, despite the large gains they made in the inpatient unit, were likely not well enough to go without additional treatment after discharge. More

importantly it appears that participation in the feedback condition of this study did not decrease the need for further treatment after discharge.

4.7. Eating disordered behaviors

Another area we were interested in assessing and describing in our sample was whether there were any eating disordered behaviors that the women engaged in at follow-up. The behaviors we wanted to assess were the frequency of bingeing, purging, restricting/starvation, and laxative use in each group. We felt that exploring this information would be crucial in understanding how the sample is fairing three to four years after treatment. After all, one can have a normal BMI and be functioning well while engaging in subclinical eating disordered behaviors.

Table 10 presents the respective count and percentages of women engaging in eating disordered behaviors in both the feedback condition and TAU condition. Because of how small the samples were and the failure in our early analysis to find a feedback effect on any of the studied variables, we concluded that the most reliable data would come from combining the behaviors across the groups. We debated whether or not to describe the average or provide a mean of each behavior for each group, but ultimately felt that portraying the data that way could be misleading because the women's individual endorsement levels and their symptom severity would be lost in an average score; Table 8 is more representative of the women's levels of endorsement. Based on the women's self-report, the vast majority of treated individuals (77.3%) never or rarely binged at the follow-up time point, while 22.6% continued to binge. Laxative use was similar in that most treated individuals (90.5%) never or rarely used laxatives as a means to control their weight, while 9.4% reported using laxatives regularly. Most treated individuals (71.6%) reported that they never or rarely purged, while 28.2% admitted that they continued to struggle with regular or frequent purging. The sample appeared to be engaging most in restrictive eating behaviors across both groups. Approximately half of the sample (45.2%) denied that they restricted their eating; however, 56.5% of the treated sample continued to struggle with restricting their intake on a regular basis. Indeed, this was the most common disordered eating behavior participants endorsed, regardless of their diagnosis.

4.8. Correlation of OQ-45 distress score with bmi and disordered eating

A major criticism of the original Simon et al. (2013) study was the fact that it relied solely on the OQ-45 as both a progress and outcome measure with no inclusion of an eating disorder specific standardized measure. It was assumed that the OQ-45 was capable of detecting increased severity of disordered eating behavior indirectly via a total distress score. We wanted to assess this assumption and determine if indeed there was a detectable relationship between the OQ-45 distress score and the severity of disordered eating the women were engaging in. It should be noted again that the measure we used to assess eating behaviors was a measure the Center for Change created for the purpose of assessing eating patterns in previously treated patients (it has not been validated or determined to be a reliable measure).

A chi-square test for independence demonstrated that, indeed, the relation between severity of disordered eating and OQ-45 distress score was significant (χ^2 (12, $N = 42$) = 24.1, $p = .02$; see Tables 11 and 12). Patients with more severe eating disordered behaviors at follow-up were more likely to experience a reliable worsening in their OQ-45 distress scores. Endorsement of increasingly severe disordered eating and percentage of patients experiencing worsening distress is illustrative of this relationship. For example, of the individuals who denied engaging in disordered eating practices only 30% (3/10) experienced a worsening of their distress on the OQ-45. Of those individuals who endorsed engaging in disordered eating "rarely" only 27% (3/11) experienced a worsening of their distress on the OQ-45 over the follow-up time frame. Interestingly, of the individuals who endorsed engaging in disordered eating approximately once per week, none (0/8) experienced a worsening of their distress over the follow-up time period. However, when individuals endorsed engaging in disordered eating several times per week the percentage of those experiencing a worsening on their OQ-45 distress score increased to 73% (8/11). And lastly, those who endorsed engaging in disordered

cogent ··psychology

Table 10. Percentage of eating disordered behavior across treatment conditions				
Endorsement rating	**Bingeing (%)**	**Purging (%)**	**Restricting (%)**	**Laxatives (%)**
0 (Never)	32 (60.4)	29 (54.7)	12 (22.6)	47 (88.7)
1 (Rarely)	9 (16.9)	9 (16.9)	12 (22.6)	1 (1.8)
2 (Once per week)	6 (11.3)	4 (7.5)	8 (15.1)	0
3 (Several times per week)	4 (7.5)	6 (11.3)	13 (24.5)	2 (3.7)
4 (Several times per day)	2 (3.8)	5 (9.4)	9 (16.9)	3 (5.7)

eating several times a day, 46% (6/13) experienced a worsening of their distress levels across the follow-up time frame.

Another critical aspect of recovery for these patients that would seem to be intricately connected with psychological distress is that of BMI. It is assumed that BMI would be closely tied to one's level of distress, since most women with eating disorders engage in more disordered eating the more distressed they become; therefore, we would expect that their weight would reflect that distress. However, in the Simon et al. (2013) study both conditions made significant and equivalent improvements in BMI, despite the changes in mental health functioning favoring the Fb group. Interestingly enough, when a chi-square test for independence was conducted between absolute OQ-45 score at follow-up and severity of BMI classification (not change in these scores) we found that they actually were *not* independent of one another. At follow-up we found that there was a significant correlation between one's OQ-45 score and their severity of BMI classification (χ^2 (9, N = 42) = 20.69, p = .01; see Tables 13 and 14). Those who fell in the abnormal categories of BMI including starvation, severely underweight, obese class III, obese class II, obese class I, underweight, and overweight fell almost exclusively (13/16 of individuals, 81%) in the clinical range of the OQ-45 (clinical range on the OQ-45 was considered to be a score of 64 or higher). Of those who were in the normal BMI category, 42% (11/26) fell in the clinical range on the OQ-45, and 58% (15/26) fell in the non-clinical range on the OQ-45 (non-clinical on the OQ-45 included scores of 63 and below). This variation in severity of OQ-45 scores of those in the normal BMI category suggests that those who are experiencing a significant amount of psychological distress can still fall within a normal weight category. However, overall those who fall within an abnormal BMI category appear more likely to be experiencing higher amounts of psychological distress than those who fall in the normal BMI category.

A related question was the relationship between negative *change* on the OQ-45 and negative *change* in BMI category membership (negative BMI change is any movement away from the normal BMI category). A chi-square test for independence revealed that there was no significant correlation between worsening on the OQ-45 and deterioration in BMI category (χ^2 (4, N = 42) = 6.0, p = .19; see Tables 15 and 16).

4.9. Severity and number of contacts

As mentioned in the data collection section, it took several attempts of contact in order to reach our sample participants. We were curious if there was a difference in distress level (as measured by the OQ-45) in the participants who responded to our first outreach attempt, second outreach attempt, and third outreach attempt. Any systematic difference could reveal if those who did not participate in the follow-up study might be expected to reliably differ from those who did participate (e.g. be more disturbed). We ran an analysis of variance (ANOVA) to analyze the differences between the number-of-contact group means and it suggested that there was not a statistically significant difference in the distress levels of each group (p = .06; see Table 17). However, because this p-value could be considered on the borderline of statistical significance we ran a series of post hoc *t*-tests to look more closely at number-of-contact group mean differences.

The post hoc *t*-test comparing the distress levels of the group who responded after one contact and the group who responded after two contacts suggested that there was not a significant

cogent ·· psychology

Table 11. OQ-45 and eating behavior frequencies

Outcome classification	Frequency of eating behavior					
	0[a]	1[b]	2[c]	3[d]	4[e]	Total
Reliably worsened[f]	3 (6%)	3 (6%)	0 (0%)	8 (15%)	6 (11%)	20 (38%)
Unchanged[g]	2 (4%)	2 (4%)	4 (8%)	2 (4%)	6 (11%)	16 (30%)
Reliably improved[h]	0 (0%)	1 (2%)	0 (0%)	1 (2%)	1 (2%)	3 (6%)
Recovered[i]	5 (9%)	5 (9%)	4 (8%)	0 (0%)	0 (0%)	14 (26%)
Total	10 (19%)	11 (21%)	8 (15%)	11 (21%)	13 (25%)	53 (100%)

[a]Never.

[b]Rarely.

[c]Once per week.

[d]Several times per week.

[e]Several times per day.

[f]Worsened by at least 14 points on the OQ-45 from pre-treatment to follow-up.

[g]Improved less than 14 points or worsened by less than 14 points on the OQ-45.

[h]Improved by at least 14 points on the OQ-45 but did not pass the cut-off between dysfunctional and functional populations.

[i]Improved by at least 14 points on the OQ-45 and passed the cut-off between dysfunctional and functional populations.

Table 12. OQ-45 and eating behavior chi-square analysis

Statistic	DF	Value	Probability
Chi-square	12	24.05	0.02
Likelihood ratio chi-square	12	32.44	0.00
Mantel-Haenszel Chi-Square	1	5.92	0.02
Phi Coefficient		0.67	
Contingency Coefficient		0.55	
Cramer's V		0.39	

Table 13. OQ-45 and BMI classification frequency

BMI classification	OQ-45 classification				
	Clinical 1[a]	Clinical 2[b]	Normal 1[c]	Normal 2[d]	Total
Extreme[e]	0 (0%)	4 (10%)	1 (2%)	0 (0%)	5 (12%)
Moderate[f]	2 (5%)	4 (10%)	1 (2%)	1 (2%)	8 (19%)
Normal[g]	3 (7%)	8 (19%)	7 (17%)	8 (19%)	26 (62%)
Severe[h]	3 (7%)	0 (0%)	0 (0%)	0 (0%)	3 (7%)
Total	8 (19%)	16 (38%)	9 (21%)	9 (21%)	42 (100%)

[a]OQ-45 scores from 64 to 77.

[b]OQ-45 scores of 78 and above.

[c]OQ-45 scores from 50 to 63.

[d]OQ-45 scores of 49 or below.

[e]Extreme classification includes BMI categories of starvation, obese class III, and obese class II.

[f]Moderate classification includes BMI categories of underweight and overweight.

[g]Normal classification includes the BMI category of normal.

[h]Severe classification includes BMI classifications of severely underweight and obese class I.

cogent • • psychology

Table 14. OQ-45 and BMI classification chi-square analysis

Statistic	DF	Value	Probability
Chi-square	9	20.69	0.01
Likelihood ratio chi-square	9	19.41	0.02
Mantel-Haenszel chi-square	1	0.02	0.88
Phi coefficient		0.70	
Contingency coefficient		0.57	
Cramer's V		0.40	

Table 15. OQ-45 and BMI change frequency

OQ-45 change	BMI change			
	Improved	No change	Worsened	Total
Unchanged	1 (2%)	8 (19%)	3 (7%)	12 (29%)
Reliable worsening	1 (2%)	6 (14%)	8 (19%)	15 (36%)
Reliable improvement	0 (0%)	12 (29%)	3 (7%)	15 (36%)
Total	2 (5%)	26 (62%)	14 (33%)	42 (100%)

Table 16. OQ-45 and BMI change chi-square analysis

Statistic	DF	Value	Probability
Chi-square	4	6.0	0.19
Likelihood ratio chi-square	4	6.62	0.16
Mantel-Haenszel chi-square	1	0.002	0.96
Phi coefficient		0.38	
Contingency coefficient		0.35	
Cramer's V		0.27	

difference in distress levels (p = .85). The post hoc t-test comparing distress levels of the group who responded after two contacts and the group who responded after three contacts also was not statistically significant (p = .12). For a point of reference, the difference in mean distress level between the two contact group and three contact group was 26.7 OQ-45 raw score points, which we would expect is a large enough gap to be statistically significant. This lack of statistical significance is likely a product of our low power. Although the difference is not statistically significant, there is a practical difference between the two contact and three contact group. It is possible that with a larger sample size the difference may become statistically significant.

The post hoc t-test comparing distress levels of the one contact group with the three contact group was also not statistically significant (p = .08); however, the point difference between the two groups means was 20.7, which is sizeable and would suggest a statistically significant difference. Again, it is assumed that we do not have enough power to detect a statistically significant difference between these two groups if it exists. This data suggests the possibility that those who finally consented to participate were more disturbed than those who initially did so. By implication, it is possible that further attempts to recruit participants may have produced a sample with more seriously disturbed individuals.

cogent ·· psychology

Table 17. Mean distress levels and standard errors based on number of contacts			
Number of contacts	***n* (%)**	**Mean follow-up OQ-45 score**	**Standard error**
1	38 (71.7)	67.9	4.1
2	6 (11.3)	62.0	10.2
3	9 (16.9)	88.7	8.3

5. Discussion

5.1. Current findings and implications

The current study followed a portion of women who were treated in an intensive inpatient study reported by Simon et al. (2013). That study sought to assess the effects of progress feedback with alarm signals and clinical support tools on the mental health of females with eating disorders. Of 141 females recruited into the study, four declined participation leaving 137 who were randomly assigned to either treatment as usual (*n* = 68) or feedback-assisted treatment (*n* = 69); four patients withdrew from TAU in the first week, leaving 133 patients that could be followed in the original study (64 TAU & 69 whose therapists received feedback). Both the feedback group and treatment-as-usual group received routine hospital care that included a wide variety of treatment programs including nutrition counseling, daily group therapy, weight monitoring, twice weekly individual psychotherapy, and other treatments. Ninety percent of the patients were on psychoactive medications. Randomization of patients to the experimental and control groups was done within therapists, meaning that the 16 therapists who delivered individual psychotherapy treated half their cases while receiving feedback. The other half of their caseload was monitored but feedback was withheld.

Simon et al. (2013) found that the feedback intervention reliably enhanced mental health functioning when compared to TAU. Patients of therapists who received feedback, had similar rates of deterioration but higher rates of recovery (TAU = 28.6% vs. feedback = 52.9%). Despite these positive effects for feedback on mental health functioning, this did not extend to a superior outcome for positive changes in weight, where both groups showed similar improvements. The findings of Simon et al. (2013) extended the positive effects of feedback found in a variety of outpatient settings (Shimokawa et al., 2010) and inpatient care of psychosomatic patients in Germany (Probst et al., 2013).

Despite consistent evidence that tracking patient progress and supplying therapists with feedback about their patients mental health functioning (e.g. that a patient is off-track for a positive outcome, and supplying clinicians with problem-solving tools for off-track cases), no study has examined long-term patient outcome to see if the positive effects of feedback are lasting. This study examined patient status three to four years after completion of hospital services as reported by Simon et al. (2013) to see if the improvements in mental health functioning due to feedback were still observable.

Multiple attempts to recruit the 133 patients studied by Simon et al. (2013) resulted in only capturing information from 26 individuals who were in the feedback group and 27 who received TAU. Thus, we could only follow 53 individuals from the original study sample of 133 (47%). Despite this disappointing response rate, analysis of OQ-45 scores at various time points comparing the follow-up sample with the original sample suggested that the follow-up sample as a whole was quite similar in level of disturbance at pre-treatment. In respect to post-treatment functioning, the follow-up groups were quite similar to the larger sample from which they came, i.e. the feedback patients that could be followed had better mental health functioning at post-treatment than the TAU patients that could be followed. This finding suggests that the follow-up samples were representative of the larger samples in the original report by Simon et al. (2013). An exception to this conclusion was the comparison of scores in relation to how many contacts it took to get patients to respond to the follow-up questionnaires. The data suggested that additional attempts to recruit patients may have

resulted in studying more individuals who had an especially high level of disturbance. However, we could not draw firm conclusions, perhaps because of the small N's.

The central purpose of this follow-up study was to examine the long-term outcome of individuals who were assigned to either TAU or an experimental feedback group to see if the advantage of feedback on mental health functioning could still be detected years later. Our first objective was to find out if there was still a significant difference between Fb and TAU conditions at the point of follow-up (relative to pre-treatment, relative to post-treatment) on mental health functioning. The ANCOVA yielded a non-significant p-value (p = .62), which indicates that the feedback group and TAU group could not be distinguished at the end of the follow-up period. Effect size calculations also yielded no sizable effect/difference between the two groups at follow-up. The calculated effect size Cohen's d, between the post-treatment OQ-45 scores and the follow-up OQ-45 scores for the feedback group was very small and in a negative direction suggesting that the feedback condition had deteriorated slightly. The effect size between the post-treatment OQ-45 scores and the follow-up OQ-45 scores for the TAU group was also very small but in a positive direction, suggesting that the TAU condition improved slightly; however, these effect sizes were too low to meet the Cohen's criteria for a "small" effect. Overall, the data suggests that the feedback group relinquished the superior outcomes they had at post-treatment and more closely approached the TAU group, while the TAU slightly improved during the follow-up time period. We can likely conclude that the superior effects of feedback observed at post-treatment in the original study were not observed three to four years later at follow-up and both groups were almost equivalent to one another in terms of psychological distress.

In further support of the idea that both treatment groups closely resembled one another at follow-up, the chi-square analysis demonstrated no difference between groups in reliable change proportions of recovered, improved, reliably worsened, and unchanged when looking from pre-treatment to follow-up. This significantly differs from the original study's results that suggested that more Fb participants met criteria for recovery than those in the TAU condition. It of course should be noted that *because* more of the Fb patients were in the normal range at post-treatment, they had more opportunity to actually deteriorate compared to those in the TAU condition, and they were less likely to cross the threshold from dysfunctional to functional to be considered "recovered" during the follow-up period.

Overall the effects of the feedback intervention found in the original Simon et al. (2013) study were not maintained and could not be detected three to four years after the original treatment. That is to say that treatment utilizing feedback and clinical support tools does not show a lasting effect in our sample, a sample who we have reason to believe is fairly representative of the original sample. While the longevity of feedback-assisted treatment is doubtful, important clinical information about the women's BMIs, eating behaviors, and further treatment should still be considered, as it sheds light on the long-term outcome trajectory for women with more severe forms of eating pathology.

At the end of the Simon et al. (2013) study both conditions made significant and equivalent improvements in BMI, despite the changes in mental health functioning favoring the Fb group. Therefore, the original researchers concluded that a replicated study would suggest that progress feedback and CSTs do not facilitate movement toward normal BMI. What we found at follow-up supports this original conclusion. In fact, both treatment groups had similar percentages of women who maintained their post-treatment BMI (TAU 66.7%; Fb 57.1%), improved (TAU 4.8%; Fb 4.8%), and deteriorated (TAU 28.6%; Fb 38.1%); however, the Fb condition appears to have a few more women whose BMIs deteriorated since post-treatment when compared to the TAU condition. As would be expected, particularly those with anorexia nervosa were more likely to experience deterioration in BMI compared to other diagnostic categories. This finding is strongly supported in the literature, which suggests that those with AN are less likely than a non-disease comparison group to be a normal body weight and they generally remain more symptomatic than other eating disorders on ED specific measures (Berkman et al., 2007). This finding also confirms what the original study

hypothesized, which is that those with AN may experience BMI deterioration over time due to this disorder's higher relapse rates. It remains to be seen whether mental health functioning feedback can effect needed changes in BMI or if BMI itself needs to be monitored with an alert system that identifies patients at risk for failure in this area. Perhaps, the specific identification of BMI failure might allow clinicians to provide direct treatment for this issue and prevent future deterioration in BMI.

On a related note to BMI, we ran several chi-square tests of independence to explore the relationship between psychological distress as measured by the OQ-45 and BMI category. First, we found that there was no significant correlation between change on one's OQ-45 distress levels and change in one's BMI category. However, when examining the relationship between one's follow-up OQ-45 score (in terms of normal or abnormal) and one's BMI classification at follow-up, we found a significant correlation. Those who fell into abnormal BMI categories almost exclusively fell in a clinical/dysfunctional range on the OQ-45. So, while the OQ-45 may not be an effective measure to track or predict BMI changes over time, it does appear to correlate well with BMI at absolute points in time.

In addition to BMI, it was important to explore the women's eating patterns, as many of the women could not provide an accurate BMI at follow-up and those with normal BMIs could still exhibit disordered eating. Our hope was that understanding the women's eating patterns might compensate for the missing BMI data. We found that almost half of our entire sample engaged in some sort of disordered eating pattern at follow-up. Most commonly, the women endorsed restricting caloric intake on a regular basis, regardless of their original diagnosis. We also found that changes on the OQ-45 did have a significant relationship with changes in severity of disordered eating. Patients with more severe eating disordered behaviors were more likely to experience a reliable worsening in their OQ-45 distress scores across time. Endorsement of increasingly severe disordered eating and percentage of patients experiencing worsening distress was highly illustrative of this relationship. For example, when the women reported engaging in disordered eating patterns several times a week, over 70% of these same individuals' distress levels worsened as well. We felt that while the OQ-45 should probably not be used to accurately predict severity of eating behaviors, we are at least optimistic that the OQ-45 is capable of reflecting some of the pathology-specific behavior that characterizes this population.

Another interesting finding discovered was the sheer amount of treatment the women needed to maintain their gains after their inpatient stay. Overall, we discovered that this is a heavy therapy-using population. While we saw no major improvements in the women's distress scores over time, the vast majority of them were actively participating in multiple forms of treatment, most for two years or more. This seems to suggest that the women needed additional treatment in order to maintain the gains they made at the hospital, not improve upon the gains they had already made. Our most important conclusion is that these women participated in a lot of additional treatment after their stay and we did not find the Fb group to be utilizing less treatment than the TAU group. This equivalency in sought-after treatment seems to support the idea that the Fb group began to relinquish their superior gains over the follow-up period. If they continued to maintain their superior gains made at the hospital we would expect to see them utilizing less additional treatment or at least see a continued decrease in distress with this additional treatment. Logic would also suggest that if these women needed additional treatment then the initial treatment must have failed. However, it should be noted that the hospital actively encourages the women to seek out additional treatment after they leave and in many cases the hospital will set up outpatient therapy for the women once they do leave. Therefore, the mere fact that they sought out additional treatment does not necessarily mean that the initial treatment failed.

5.2. Consistency with the literature

Our research suggests that a large portion (half) of the women in our sample were engaging in disordered eating and were experiencing a significant amount of psychological distress at follow-up despite receiving intensive inpatient services (in addition to other treatment modalities including

cogent ·· psychology

psychoactive medication). This is consistent with the large Lowe et al. (2001) study and the study, both of which found over half of the women in the study still met partial criteria for an eating disorder 6–21 years after an initial inpatient stay. The women in our sample also engaged in multiple forms of treatment consistently since they left the hospital resulting in minimal gains. This fact alone is consistent with other research findings, which suggest that those who have severe forms of ED (and thus require inpatient treatment) have a poorer long-term outcome trajectory than those who primarily seek outpatient treatment. They tend to have more complicating illness factors and comorbidities, they tend to utilize more treatment, and they tend to experience higher rates of relapses. Overall, these women tend to have a rougher road to recovery than those in outpatient samples (Berkman et al., 2007; Nilsson & Hägglöf, 2005; Wentz, Gillberg, Anckarsater, Gillberg, & Rastam, 2009). Part of this has to do with the fact that those who are inpatient clients often have more Axis I/Axis II comorbidities, substance misuse, social problems, etc. that complicate the long-term success of treatment (Berkman et al., 2007). Perhaps, the efficacy of existing interventions is questionable for this level of pathological severity. The field needs more work in developing treatments to help the most severely impacted patients with eating disorders.

It is difficult to compare our findings with other research conducted on the OQ-45-based feedback system, simply because this is the first follow-up study looking at the long-term effectiveness of the system. While the previous studies by Simon et al. (2013), Hawkins et al. (2004), and Simon et al. (2012) demonstrated that the OQ-Analyst system led to statistically significant gains in more disturbed populations, our study is not demonstrating the enhanced mental health outcome (that was observed post-treatment) at the time of follow-up. Indeed, the large effects of feedback-assisted treatment may have declined over time and more closely approached the outcomes of those who received treatment-as-usual.

5.3. Limitations and further research
Several limitations of the study ought to be taken into consideration. In fact, a major limitation of this study was the low-response rate of participants. Despite our best efforts, we were unable to follow-up with the entire original sample. We were limited by the fact that our study took place several years after their initial inpatient stay; therefore, the addresses, telephone numbers, and email addresses we had for the women had likely changed by the time we contacted them. Many of them had moved away from home, gone to college, or moved out of state leaving us with no way to contact them. Because our response rate was so low we recognize that our study is likely underpowered. However, statistical analyses used that were independent of sample size indicated that there was not even a small effect to be found. Even if a larger sample could be procured, our effect size calculations suggest that we still would not find a meaningful difference between treatment groups. Overall, results of this study indicate that the advantages of feedback found at post-treatment were not sustained in this sample, as the majority sought out additional treatment regardless of treatment condition. It should be noted that our sample was representative of the entire original sample at both pre-treatment and post-treatment; therefore, our findings could very well be generalized to include the entire original sample.

It is difficult to know just how generalizable the results from this study are to other settings that focus on eating disorders. While we know these women had definite eating problems with relatively chronic histories, high comorbidities, and high level of disturbance, most of them participated in a private (even "elite") eating disorder hospital setting. Most of these women desired treatment and had high internal motivation to succeed as well as substantial social support and it is not clear if this is typical of most women seeking treatment for eating disorders. Of course, the generalization of our results is also limited by our aforementioned power problem when examining subgroups of patients. Future research would need to include a larger sample size in order to definitively outline the effects of feedback-assisted treatment on long-term trajectory in different eating disorders.

cogent •• psychology

We also must recognize that our data collection about further treatment the women received was not stellar. For example, we gave the women forced-choice questions about the amount/length of further treatment they received. For this reason, we do not know if most of the women needed continuous therapy, if they did two treatment modalities at once, or if they simply needed booster sessions sporadically across the follow-up period. It is recommended that future researchers conducting a follow-up study try to attain more precise data about additional treatment sought. A better way to assess this might have been to contact the women every six months after the hospital stay to assess what additional treatment they had sought out to that point.

Another serious limitation we encountered in our study was the fact that many of the women had not weighed themselves recently and could not provide an accurate BMI. This was not entirely unexpected given that the hospital encourages the women not to weigh themselves as part of their recovery process. This did leave a significant gap in our data and limited our interpretation about what has occurred with the women's weight over time.

We must note another limitation of the current study; in contrast to most studies conducted with eating disordered patients, we relied on a single general measure of mental health rather than the usual practice of focusing on measures of eating disorder psychopathology. This was a result of the fact that the original Simon et al. (2013) did not utilize a standardized eating disorder measure or battery of assessments, nor did they assess disordered eating behaviors in general. To maintain consistency with the data of the original study we also used the OQ-45 as a measure of progress and a measure of outcome. For this reason, a secondary objective of this study was to assess the correlation between change in the OQ-45 with change in BMI and with severity of disordered eating. The goal was to ascertain if the OQ-45 was able to detect change in core pathological symptomatology via the participants' level of distress. Unfortunately, there was not a correlation between increasing distress levels on the OQ-45 and BMI change; however, an absolute OQ-45 score at follow-up was significantly correlated with static BMI classification at follow-up. While we were able to correlate severity of disordered eating with reliable deterioration on the OQ-45, the measure we used to ascertain eating disordered behaviors was the Center's self-generated measure that had no established reliability or validity. We also do not have data from the original study about change scores on eating disordered behavior; therefore, we only have the follow-up eating behaviors to conjecture upon.

In future inpatient studies utilizing feedback-assisted treatment it will be important to include additional pathology-specific outcome measures from the beginning of treatment. It may also be worthwhile in the future to examine the value of creating a symptom-specific long-term outcome measure for eating disorders. This might yield larger effect sizes than can be achieved by relying on one general mental health functioning measure. Or, at the very least, examine how the OQ-45 correlates with other eating behavior measures. Perhaps, we might find that the OQ-45 questionnaire is comprehensive enough to capture the core pathology in this population (or in other words the fact that they are not well) without including specific eating disorder behavior items or alternative ED measures.

It should be noted that it is still unclear whether or not the OQ-45 Analyst System (and feedback in general) is useful in therapeutic situations where clients and therapists are not in collaboration or clients are not motivated to receive treatment. In this case, it is not uncommon for women with eating disorders to have been persuaded into treatment by concerned family members and many are reluctant to seek help for behaviors they do not find problematic. This lack of motivation for treatment is not unique to eating disorders and can be found in a number of populations including substance users. The OQ-45 is a fairly face-valid measure and a client who is not invested in the treatment process or motivated to reduce distress could purposefully report disingenuous distress levels (or lack thereof). Thus, such disingenuous reporting would appease therapists and family members with regard to progress and the patient would make very little actual change. One has to wonder if this type of phenomenon took place in our sample and may have skewed the results. We

suspect it may have, given the amount of outlier OQ-45 scores that were removed during the analyses. We recommend that further research should be aimed at better understanding the helpfulness of the OQ-45 system and feedback-assisted treatment in settings where clients are not in full collaboration with their therapists.

One last issue we came across during this study was the problematic verbiage Jacobson and Truax (1991) use to delineate who is recovered and who is not. It seems contradictory that a large portion of the women would be considered to be "recovered" by Jacobson and Truax's (1991) criteria while still engaging in disordered eating. This calls into question what recovery for eating disorders should hypothetically look like. Should clinicians be aiming for total extinction of disordered eating in treatment, or is that even feasible for this population? For this reason, we, as well as many other clinicians, are critical of Jacobson and Truax's (1991) criteria for recovery, as clearly many women in this study who are considered recovered are still engaging in pathological behaviors. Might we recommend that in the case of eating disorders (and other specific kinds of psychopathology) one must meet Jacobson and Truax's (1991) statistical "recovery" criteria in addition to cessation of disorder-specific symptomatology? Perhaps, a new definition of recovery for eating disorders might include falling in the normal range of mental health functioning on the OQ-45 *and* a BMI that falls within normal limits. While more stringent criteria for recovery would no doubt decrease the amount of significant positive treatment outcomes we would observe, such criteria would more accurately reflect the difficulty in treating this population and take into account the very real challenge these women face in both attaining and maintaining non-pathological functioning and behavior across time.

Competing Interests
The authors declare no competing interest.

Author details
Megan M. Bowen[1]
E-mail: mbowen4688@gmail.com
Michael J. Lambert[1]
E-mail: michael_lambert@byu.edu
Arjan Berkeljon[1]
E-mail: arjan@berkeljon.com
Tyler E. Orr[1]
E-mail: tylorr18@hotmail.com
Michael Berrett[2]
E-mail: michaelB@cfc.com
Witold Simon[1,3]
E-mail: witold.simon@wp.pl
[1] Psychology Department, Brigham Young University, Provo, UT 84602, USA.
[2] Center for Change, Orem, UT, USA.
[3] Department of Psychiatry, Institute of Psychiatry and Neurology, Warsaw, Poland.

Citation information
Cite this article as: Effects of feedback-assisted treatment on post-treatment outcome for eating disordered inpatients: A follow-up study, Megan M. Bowen, Michael J. Lambert, Arjan Berkeljon, Tyler E. Orr, Michael Berrett & Witold Simon, *Cogent Psychology* (2016), 3: 1191119.

References
Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100. http://dx.doi.org/10.1016/0272-7358(88)90050-5

Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K., & Vermeersch, D. A. (2003). Clinical significance of the outcome questionnaire (OQ 45.2). *The Behavior Analyst Today, 4,* 79–90. doi:10.1037/h0100015

Berkman, N. D., Lohr, K. N., & Bulik, C. M. (2007). Outcomes of eating disorders: A systematic review of the literature. *International Journal of Eating Disorders, 40,* 293–309. doi:10.1002/eat.20369

Bludworth, J. L., Tracey, T. J. G., & Glidden-Tracey, C. (2010). The bi-level structure of the Outcome Questionnaire-45. *Psychological Assessment, 22,* 350–355.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (pp. 51–84). Hillsdale, NJ: Erlbaum.

Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J. L., Lambert, M. J., Gallop, R., McClure, B., … Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of Substance Abuse Treatment, 42,* 301–309. doi:10.1016/j.jsat.2011.09.003

Derogatis, L. R. (1977). *The SCL-90 manual: Scoring, administration and procedures for the SCL-90.* Baltimore, MD: Johns-Hopkins University School of Medicine, Clinical Psychometrics Unit.

de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The outcome questionnaire-45 in a Dutch population: A cross cultural validation. *Clinical Psychology & Psychotherapy, 14,* 288–301.

Eisler, I., Simic, M., Russell, G. F., & Dare, C. (2007). A randomised controlled treatment trial of two forms of family therapy in adolescent anorexia nervosa: A five-year follow-up. *Journal of Child Psychology and Psychiatry, 48,* 552–560. doi:10.1111/j.1469-7610.2007.01726.x

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl.. *Journal of Clinical Psychology, 61,* 1233–1243. doi:10.1002/jclp.20179

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61,* 155–163. doi:10.1002/jclp.20108

Harmon, S. C., Lambert, M. J., Smart, D. M., & Hawkins, E. (2007). Enhancing outcome for potential treatment failures: Therapist–client feedback and clinical support tools. *Psychotherapy Research, 17*, 379–392. doi:10.1080/10503300600702331

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 14*, 308–327. doi:10.1093/ptr/kph027

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. doi:10.1037/0022-006x.59.1.12

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The longitudinal interval follow-up evaluation. *Archives of General Psychiatry, 44*, 540–548. doi:10.1001/archpsyc.1987.01800180050009

Lambert, M. J., Bailey, R. J., White, M., Tingey, K. M., & Stevens, E. (2015). *Clinical Support Tool Manual—Brief Version-40*. Salt Lake City, UT: OQMeasures.

Lambert, M. J., Hansen, N. B., & Bauer, S. (2008). Assessing the clinical significance of outcome results. In A. M. Nezu & C. Maguth-Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychological interventions* (pp. 359–378). New York, NY: Oxford University Press.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., ... Burlingame, G. M. (2013). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.2)*. Orem, UT: American Professional Credentialing Services.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11*, 49–68. doi:10.1080/713663852

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. K. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy, 9*, 91–103. doi:10.1002/cpp.324

Lo Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, M. J. (2008). The factorial structure of the outcome questionnaire-45: A study with an Italian sample. *Clinical Psychology & Psychotherapy, 15*, 418–423.

Lowe, B., Zipfel, S., Buchholz, C., Dupont, Y., & Reas, D. L. (2001). Long-term outcome of anorexia nervosa in a prospective 21-year follow-up study. *Psychological Medicine, 31*, 881–890. doi:10.1017/S003329170100407X

Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology, 66*, 400–410. doi:10.1080/10503300902849483

Nilsson, K., & Hägglöf, B. (2005). Long-term follow-up of adolescent onset anorexia nervosa in northern Sweden. *European Eating Disorders Review, 13*, 89–100. doi:10.1002/erv.631

Probst, T. A., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R. E., & Tritt, K. A. (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. doi:10.1016/j.jpsychores.2013.07.003

Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child & Adolescent Psychiatric Clinics of North America, 14*, 241–254. doi:10.1016/j.chc.2004.05.002

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*, 298–311. doi:10.1037/a0019247

Simon, W., Lambert, M. J., Busath, G., Vazquez, A., Berkeljon, A., Hyer, K., ... 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. doi:10.1080/10503307.2013.787497

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

Slade, K. L., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology & Psychotherapy, 15*, 287–303. doi:10.1002/cpp.594

Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory: STAI (formY)*. Palo Alto, CA: Consulting Psychologists Press.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74*, 242–261. http://dx.doi.org/10.1207/S15327752JPA7402_6

Wentz, E., Gillberg, I., Anckarsater, H., Gillberg, C., & Rastam, M. (2009). Adolescent-onset anorexia nervosa: 18-year outcome. *The British Journal of Psychiatry, 194*, 168–174. doi:10.1192/bjp.bp.107.048686

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*, 59–68. doi:10.1037/0022-0167.50.1.59

Zung, W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12*, 63–70. http://dx.doi.org/10.1001/archpsyc.1965.01720310065008

cogent ⊷ psychology

Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com