



2019-05-01

# Treatment Failure Rates in Group Versus Individual Treatment Using the OQ-45: An Archival Replication

Cameron Todd Alldredge  
*Brigham Young University*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Social and Behavioral Sciences Commons](#)

---

## BYU ScholarsArchive Citation

Alldredge, Cameron Todd, "Treatment Failure Rates in Group Versus Individual Treatment Using the OQ-45: An Archival Replication" (2019). *All Theses and Dissertations*. 7471.  
<https://scholarsarchive.byu.edu/etd/7471>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

Treatment Failure Rates in Group Versus Individual

Treatment Using the OQ-45:

An Archival Replication

Cameron Todd Alldredge

A thesis submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of

Master of Science

Gary M. Burlingame, Chair  
Joseph Olsen  
Chad D. Jensen

Department of Psychology

Brigham Young University

Copyright © 2019 Cameron Todd Alldredge

All Rights Reserved

## ABSTRACT

### Treatment Failure Rates in Group Versus Individual Treatment Using the OQ-45: An Archival Replication

Cameron Todd Alldredge  
Department of Psychology, BYU  
Master of Science

In routine outcome monitoring (ROM), the rate and timing of treatment failure alerts has been related to the success of feedback in past randomized clinical trials. In a recent OQ-45 feedback study, Burlingame and colleagues (2018) found that the rates and timing of not-on-track (NOT) progress alerts in group treatment were different than those reported for individual treatment. Using data from 58 different therapy groups and 374 patients, NOT progress alerts occurred at 186% of the rate reported by Shimokawa et al. (2010) when they examined over 6,000 patients receiving individual therapy at the same clinics. Another significant difference was found on the timing of the first NOT progress alerts with group treatment's first alerts occurring two sessions later than individual treatment. The goal of the current study was to use de-identified archival OQ-45 data from patients receiving group and individual treatment at a comparable clinic to determine if these rate and timing differences were replicable. Data from individual therapy ( $N = 5,493$ ) and group therapy ( $N = 146$ ) patients' OQ-45 scores show that the present study duplicated the significant difference found in the rate of NOT alerts between these formats. Relative risk of alerting as NOT at least once in group therapy was calculated to be 1.43 compared to individual therapy (group patients are 143% more likely to alert than individual patients). On the other hand, the present study did not find a significant difference in the timing of first alerts between formats. The implication of these results are significant when considering ROM in group therapy. Patients participating in group therapy are much more likely to flag as not-on-track later during the course of treatment when compared to patients in individual therapy.

Keywords: routine outcome monitoring, group therapy, individual therapy, differential efficacy

## TABLE OF CONTENTS

Abstract .....	ii
List of Tables .....	iv
List of Figures .....	v
Treatment Failure Rates in Group Versus Individual Treatment Using the OQ-45: An Archival Replication .....	1
Introduction.....	1
Rationale and Hypothesis .....	7
Method .....	8
Measure.....	8
Participants from Burlingame et al. (2018).....	9
Groups from Burlingame et al. (2018).....	9
Participants from Shimokawa et al. (2010) as Used in Burlingame et al. (2018).....	10
Procedure .....	10
Results.....	12
Evaluating Rate of NOT Alerts.....	12
Evaluating Timing of First Alerts .....	13
Additional Temporal Analyses .....	14
Discussion.....	15
References.....	20

## LIST OF TABLES

Table 1 <i>Not-on-track Alert Rates by Treatment Format</i> .....	13
Table 2 <i>One Way ANOVA of First Alert Timing between Treatment Formats</i> .....	14

## LIST OF FIGURES

<i>Figure 1.</i> Using Burlingame et al.'s (2018) data to compare the percentage of first not-on-track (NOT) alerts by session between individual and group treatment.....	4
<i>Figure 2.</i> Comparing the percent of the NOT alerts frequency by session between individual and group treatment from Burlingame et al. 2018.....	5
<i>Figure 3.</i> Comparing the percentage of first not-on-track (NOT) alerts by session between individual and group treatment for the present study. ....	14
<i>Figure 4.</i> Comparing the percent of the NOT alert frequency by session between individual and group treatment in the archival data.....	15

Treatment Failure Rates in Group Versus Individual  
Treatment Using the OQ-45:  
An Archival Replication

**Introduction**

Routine outcome monitoring (ROM) is an evidence based practice that relies upon repeated assessment of mental health patients with an outcome instrument. Typically, patients are assessed at the beginning of treatment and at each significant clinical contact. ROM systems provide session by session progress feedback to clinicians regarding the clinical status of their clients. In 2006, the American Psychological Association (APA) Presidential Task Force on Evidence-Based Practice Progress endorsed the continuous monitoring of patient progress and adjusting mental health treatment accordingly (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 280).

The oldest ROM system with the greatest empirical support uses the 45 item-long Outcome Questionnaire (OQ-45) to compare patient progress using empirically derived algorithms (Lambert, 2015). These algorithms provide two types of feedback alerts for mental health clinicians. *Change alerts* categorize patient change from the beginning of treatment into four categories: recovered, improved, no change, or deterioration. These alerts are calculated using the reliable change index (RCI); a change metric which identifies the number of points a score must change from initial distress to be statistically significant. This type of alert is useful in determining whether or not the therapy is effectively helping the client to experience reliable change and whether or not they are feeling better or feeling worse

The OQ-45 also uses *progress alerts* which compare an individual patient's pattern of change with normative change trajectories that are associated with successful treatment outcome. When a patient's pattern of change deviates from these normative change trajectories, it is an indication that they are not having a positive response to treatment and are at risk for treatment

failure. The importance of these alerts is illustrated by Probst et al. (2013) as they found the trajectories of not-on-track (NOT) patients improved significantly more when therapists were informed of these NOT alerts compared to when they were not given this information.

Krageloh and colleagues (2015) conducted a scoping study where they categorized different methods of outcome data usage among clinicians and found that the most effective method included a formal structure in delivering the feedback. An example from a study included in Krageloh et al.'s scoping study where formal, structured feedback was delivered includes clients completing the OQ-45 at intake and each subsequent session and these scores being available to them before each session presented in a graph and color-coded system categorizing progress. In addition, some clients received also written feedback messages and those that flagged as NOT were encouraged to discuss their concerns about the treatment progress (Hawkins et al., 2004). An example of a feedback style on the opposite end of the spectrum included clients completing the OQ-45 before each session and the results were suspended from the therapists up until three weeks after the administration (Hannan et al., 2005). During that time, therapists were left to estimate clients' progress on their own.

Lambert (2015) summarized a dozen randomized clinical trials testing the effect of providing mental health clinicians with feedback using the OQ-A ROM system. All found a significant effect for progress feedback; specifically, there was a reduction in the rate of treatment failure and greater improvement for therapists using feedback on the OQ-45. However, there has been limited research suggesting that the effect of feedback may be moderated by at least two variables. Shimokawa, Lambert, and Smart (2010) found greater effects for feedback when there were at least 5 sessions of therapy after a progress alert. Thus, alerts occurring in early sessions of treatment may lead to greater effects for feedback because the therapist has

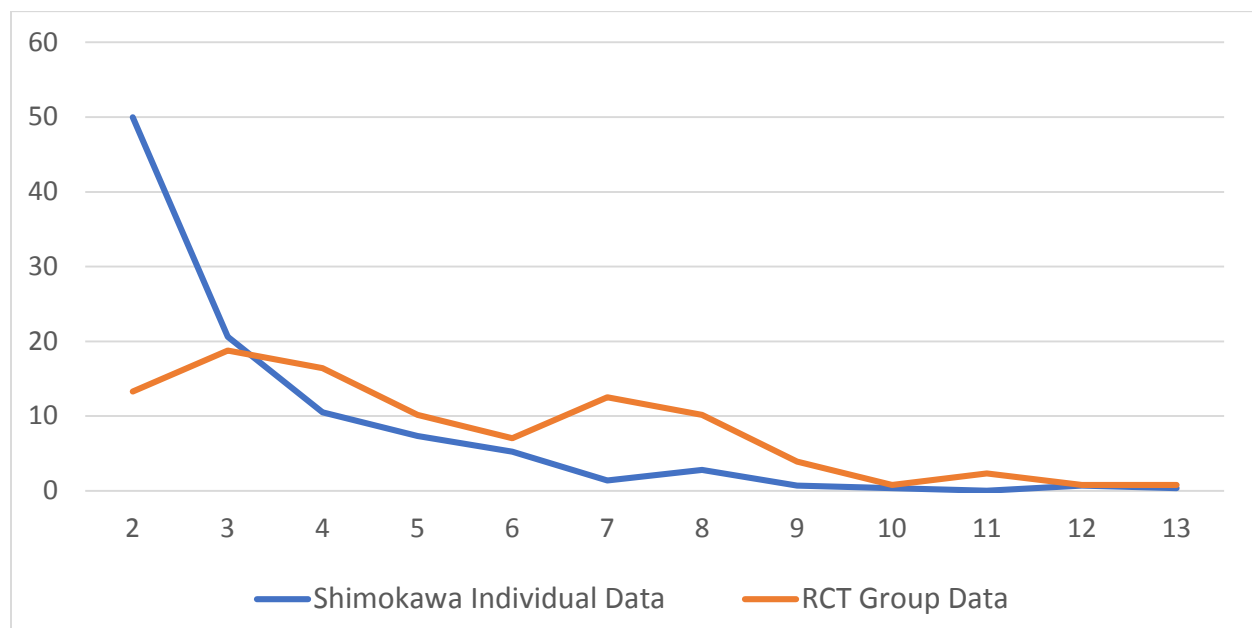


sufficient time to address states of symptom deterioration. In addition, the absolute rate of alerts has been shown to be differentially related to the effects of feedback. More specifically, one study reporting a lower base rate of progress alerts found smaller effects for progress feedback (Schuman, Slone, Reese, & Duncan, 2015). The attenuated effects of feedback may reflect lower levels of distress or an under-powered test of feedback due to an insufficient number of alerts. Interestingly, higher rates of alerts that reflect a more disturbed population have been associated with ineffectual progress feedback (Davidsen et al., 2017).

Past research testing ROM systems has been conducted primarily with clients treated in individual therapy. The average rate of progress alerts for studies using the OQ-45 and the OQ-A ROM system has been shown to be 18% (Shimokawa et al., 2010). Stated differently, 18% of individual therapy patients produce at least one treatment failure alert during the course of therapy. Of these alerts, about 60% occur at session 2 with a gradual decrease over the course of therapy to about 40% by session 13. This temporal pattern of progress alerts is of particular importance given Shimokawa et al's (2010) findings that early session alerts lead to larger effects for progress feedback systems.

Until recently, the effectiveness of ROM has not been tested in group therapy although past research from naturalistic settings has shown identical patterns of change on the OQ-45 for patients receiving either group or individual therapy (Burlingame, Seedback, et al., 2016). Thus, the literature has assumed that the rates of progress alerts—or cases that are not-on-track (NOT) for a successful outcome should be the same for group and individual therapy. However, in the first group therapy study testing the effects of OQ-45 feedback using the OQ-A, Burlingame and colleagues (2018) found evidence suggesting that the rates and timing of progress alerts may be quite different. They studied 430 patients who received treatment in 58 therapy groups at three

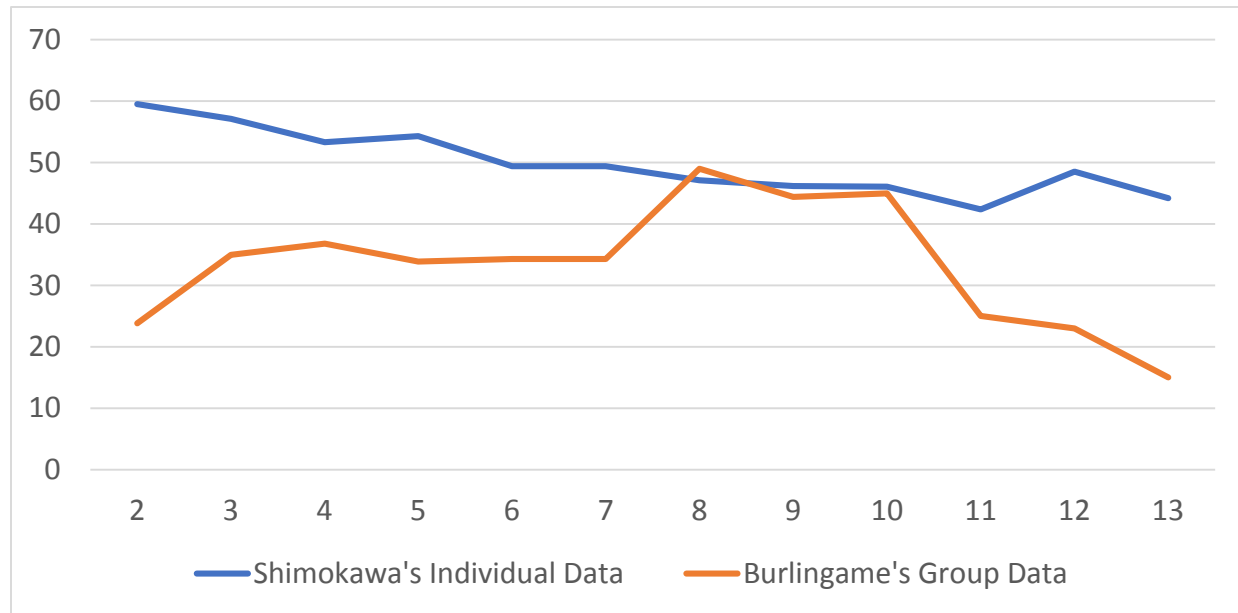
university outpatient clinics and found that 33.5% of group therapy clients' cases produce at least one alert over the course of therapy. This rate is 186% more than the rate reported by Shimokawa et al. (2010) on 6000+ patients receiving individual therapy at the same university outpatient clinics. The timing of when alerts first occur in group therapy compared to individual therapy was also different. First alerts of treatment failure in individual therapy appeared to occur earlier on in treatment when compared to group therapy. In addition, the overall pattern of when the alerts occurred is also different between the two formats.



*Figure 1.* Using Burlingame et al.'s (2018) data to compare the percentage of first not-on-track (NOT) alerts by session between individual and group treatment.

The pattern depicted in Figure 1 is problematic. Burlingame et al. (2018) reported that the average patient receiving group treatment in the university outpatient clinics attended 9.7 sessions. As evidenced in Figure 1, sessions 7 and 8 still have 10% or more of at-risk group therapy patients alerting for the first time. This leaves less time for therapists to therapeutically address the reasons for these patients being at-risk for treatment failure. The later on in treatment

the alerts occur, the more difficult it is for therapists to address these concerns before treatment ends which can be seen as a disadvantage. On the other hand, approximately 50% of the first NOT alerts occur in session two for individual therapy. This implies that the therapists have more time to address the issues contributing to the potential treatment failure before therapy ends.



*Figure 2.* Comparing the percent of the NOT alerts frequency by session between individual and group treatment from Burlingame et al. 2018.

The pattern of the frequency of all NOT alerts is shown in Figure 2 which demonstrates similar issues. The highest percentage of NOT alerts happens at the second session of individual therapy and gradually decreases overtime. The percent of group treatment's NOT alerts is very different as it seems to increase until about session 8 where it is equivalent with individual therapy until it drops off again at session 11. This also illustrates the problem of group therapists having less time to address the concerns of group members who may be alerting as not-on-track. Without sufficient time to facilitate resolution for these patients could lead to treatment failure.

Burlingame et al. (2018) found that NOT alert cases in group therapy produced worse outcomes than those in individual therapy. However, this finding could be due to the nonequivalence of the number (18% vs. 33.5%) and timing (early vs. late) of first alert cases in individual vs. group, respectively. The increasing application of ROM in clinical setting creates some urgency in determining if progress feedback is differentially effective across formats.

Burlingame et al. (2018) study's findings were based upon patients receiving treatment at three university counseling centers along the Wasatch front. If this finding is replicated in other samples of clients receiving group therapy, it has significant implications. For instance, the average group therapy client attends 9.7 sessions leaving only one session (on average) for the therapist to address the alert. This could mean that more group therapy clients are ending treatment worse off than when they started because of a different temporal pattern of treatment failure alerts compared to individual therapy. Shimokawa and colleagues (2010) warned that even though causation is difficult to demonstrate between psychotherapy and negative outcomes, just the quantity of patients who, despite receiving treatment, experience an increased level of distress should be alarming to professionals who seek to alleviate the suffering of their patients.

These findings are interesting in light of the fair amount of evidence that treatment outcomes don't differ between group and individual therapy. A 2016 meta-analysis by Burlingame, Seedback, and colleagues demonstrated evidence that the two treatment formats do not differ in acceptance, dropout, remission, and improvement rates. This is consistent with the well-established "no difference conclusion" in the group versus individual literature which suggests that, in most cases, neither treatment format is superior to the other (Lambert, 2013). With this information, it is puzzling as to why group therapy may yield significantly more NOT

alerts compared to individual therapy in addition to an apparent “disadvantage” in the timing of these alerts if the two formats supposedly do not differ in overall improvement of the patients.

Since the rate of progress alerts and the timing of first NOT alerts have been shown to moderate the effects of progress feedback in the individual therapy literature, and since the only study comparing rates and timing of feedback between formats has produced nonequivalence, Burlingame and colleagues (2018) called for a replication of the rates and timing of NOT alerts in group treatment using an identical clinical setting; a university outpatient clinic. If the rates and timing of NOT cases are replicated in a new data set, the clinical implications for ROM being applied to group therapy are significant.

### **Rationale and Hypothesis**

*Aim:* To draw upon a new data source—the University of Utah’s University Counseling Center (UCC) data—that uses the OQ-45 to investigate if the NOT frequency and first alert temporal patterns found in Burlingame et al. (2018) are common across all group therapies, or if the three counseling centers used in previous research were unusual.

*Hypotheses:* Without sufficient previous research to give evidence whether the NOT frequency and temporal patterns found by Burlingame et al. (2018) are common across all group therapies, we hypothesize that the new data source will show no difference between rate and timing of NOT alerts across group and individual therapy. In addition, when comparing UCC’s NOT alert frequency patterns to those reported by Burlingame et al. (2018) and Shimokawa et al. (2010), we hypothesize that, again, there will be no difference between group and individual treatment.

## Method

### Measure

**The Outcome Questionnaire-45-** The Outcome Questionnaire (OQ-45) is a 45-item self-report measure designed to assess clients' general level of distress and track the effectiveness of mental health services (Lambert et al., 2004). It uses a 5-point Likert scale from 0 ("never") to 4 ("almost always"; Lambert et al., 2013). The OQ-45 total score has a range of 0 to 180 where higher scores indicate more distress and poorer functioning. The OQ-45 also has three subscales intended to measure quality of interpersonal relationships, social role functioning, and symptom distress. The cut-off score for clinical distress is an OQ total score at or above 64.

According to the OQ-45's reliable change index (RCI), total scores fluctuating 14 points or more indicate that a patient's change is probably due to something other than error or random mood fluctuations (change alert). The total OQ score is used over the course of therapy to calculate and identify cases that are either on track (OT) or not-on-track (NOT) for a successful outcome. The five color-coded progress alerts are derived from empirically supported trajectories of successful outcome scores respective to given baseline scores. More specifically, should a total score be above the upper 68% of the tolerance interval surrounding the empirically supported trajectory, the alert status is yellow. If the total score is above the upper 80% of the tolerance interval, the alert status is red. These two color-coded progress alerts indicate that the patient is not-on-track for a successful outcome and actions should be taken to address these concerns to avoid treatment failure. The accuracy of detecting NOT individuals who are at-risk for treatment failure is well established (Lambert, 2015). Test-retest reliability over the period of three weeks for non-treatment samples has been estimated to be .84. Additionally, internal

consistency reliability is estimated at .93 (Lambert et al., 2004), with high concurrent validity with other outcome measures (.78–.88 and all significant at the  $p < .05$  level).

### **Participants from Burlingame et al. (2018)**

**Clients in Group Treatment-** participants from the study by Burlingame et al. (2018) consisted of individuals in group therapy in counseling centers at Brigham Young University, Southern Utah University, and Utah State University. 430 participants enrolled in the study, 55.8% were female, 88.9% were Caucasian, and 65.8% reported previous counseling. A total of 374 clients were included in the study and met the inclusion and exclusion criteria as described below.

*Inclusion Criteria:* Clients had to agree to the following criteria for inclusion in the current study: (1) willingness to participate in at least four sessions of group therapy, (2) willingness to complete the GQ and OQ weekly, and (3) willingness to have group psychotherapy as their primary mode of psychosocial treatment, meaning they do not receive no more than one session of individual psychotherapy for every three sessions of group psychotherapy.

*Exclusion Criterion:* If clients did not have an email address, they were excluded from the study.

### **Groups from Burlingame et al. (2018)**

A total of 58 groups were included in the study, averaging about 7.7 members per group. Each group was led by at least one licensed psychologist usually along with a co-leading trainee or intern. The group leaders were asked to include the feedback from the Group Questionnaire (GQ) and the OQ-45 into their daily practices (Burlingame, Gleave, et al., 2016). With the leaders, group members had the opportunity to decide how they wanted their outcome feedback

integrated into their treatment. The majority (67%) of these groups were interpersonal process groups. The other one-third of the groups were focused on specific clinical concerns and were often guided by an evidence-based protocol such as sexual concerns, eating disorders, and trauma. Most of the 58 groups lasted for only one semester (85%), eight continued for two semesters, and one lasted four semesters.

### **Participants from Shimokawa et al. (2010) as Used in Burlingame et al. (2018)**

Burlingame et al. (2018) used archival, individual therapy data from Shimokawa et al.'s (2010) meta-analysis in order to conduct the analyses duplicated in the present study. In the original meta-analysis, a total of 6,151 patients across six different studies were included. These patients received individual therapy at outpatient university clinics and outpatient hospital settings with a mean age of 24.18 (SD 4.17) and 63.2% females. In these patients' course of treatment, their OQ-45 results were used to monitor progress and to help identify those who were at risk for treatment failure via the progress alerts. The average initial OQ-45 total score was 80.3 and patients attended an average of 8.32 sessions. Approximately 18% of these participants receiving individual therapy had at least one progress alert of being not-on-track during the course of treatment.

### **Procedure**

We obtained and analyzed de-identified archival OQ-45 data that were similar to the timeframe of Shimokawa et al.'s date range (years 1999-2013) from group and individual patients at a comparable university outpatient clinic (The University of Utah's University Counseling Center [UCC]) located in Salt Lake City, Utah. All valid OQ-45 administrations with more than two data points for a single patient were eligible if categorized as *individual therapy* or *group therapy*. All other treatment types (i.e., couple, crisis, and career counseling) were



excluded. The eligible OQ-45 administrations were first organized into treatment episodes. The treatment episodes were considered a series of sessions occurring within 90 days of one another. If more than a 90 day gap was present between OQ-45 administrations, it was counted as a new episode of treatment. In some cases, there were as many as five treatment episodes but for the purpose of the present study and to preserve uniformity, we only analyzed the first treatment episode among patients.

Treatment episodes were then organized by whether they were exclusively individual therapy, exclusively group therapy, or included both individual and group therapy. If a patient was concurrently in group and individual therapy during a single treatment episode, their OQ-45 administrations were excluded. Because it is impossible to have an alert in the first administration of the OQ-45, we only looked at sessions 2-15 for the main analyses. However, the initial OQ-45 administrations and sessions beyond number 16 were included in other, more general analyses. The reason we determined the cut-off to be at session 15 was due to the issue of on-track trajectories becoming more complicated and less accurate after session 15 (Lambert, 2013). All eligible OQ-45 administrations within their respective treatment format, episode, and session number were entered into the OQ-Analyst.

The OQ-Analyst's algorithms and change trajectories produced the yellow and red alerts used to determine the frequency of patients who flagged as NOT and the timing of when the NOT alerts occurred for the first time. The data received from the OQ-Analyst was in long format and we reshaped it into wide format in order to conduct the necessary analyses. Before doing so, we created a dummy variable to indicate whether or not an individual had a yellow or red NOT alert any time during sessions 2-15 of treatment. A variable was also created for the first session at which a client had a yellow or red alert.

## Results

Following our inclusion and exclusion criteria, the UCC data left us with an  $N$  of 4,953 first-episode individual therapy patients and an  $N$  of 146 first-episode group patients. 35% of these patients were male ( $N = 1,793$ ), 54% were female ( $N = 2,730$ ), and the remaining 11% did not indicate their gender ( $N = 576$ ). The dates for when these therapy sessions were initially held ranged from September 1<sup>st</sup>, 1999 to October 30<sup>th</sup>, 2013.

Because greater posttreatment improvement has been associated with higher initial severity of OQ-45 scores (Lambert et al., 2013), an examination of intake scores among patients that flagged as NOT at least once during treatment revealed a nonsignificant trend,  $F(1, 1124) = 1.82, p = 0.18$ . The number of sessions an individual attends has also been linked to overall improvement (Lambert et al., 2013), and between the two treatment formats, a difference of just one session was found which was not significant,  $F(1, 1302) = 1.33, p = 0.25$ .

### Evaluating Rate of NOT Alerts

A chi-square test was conducted to assess whether patients in group or individual treatment produced more NOT alerts when compared to the total sample of cases  $N$  in each format. There was a significant difference between group and individual treatment NOT alerts  $\chi^2 = (2, 5639) = 7.25, p = 0.007$ . As seen in Table 1, the proportion of patients who flagged as NOT at least once in group therapy (31.5%) was greater than the proportion of patients in individual therapy (22.1%). The relative risk of alerting at least once as a group patient compared to individual patients was calculated to be  $RR = 1.43, p = .004, 95\% CI [1.12, 1.82]$ . This means that group patients are 143% more likely to have at least one NOT alert compared to individual patients. These results are consistent with what Burlingame et al. (2018) found in their RCT

where group therapy yielded significantly more NOT alerts (33.5%) compared to individual therapy (18%).

Table 1  
*Not-on-track Alert Rates by Treatment Format*

Format	Never Alerted	At least one NOT alert	Total
Individual therapy	4,279 (77.9%)	1,214 (22.1%)	5,493 (100%)
Group therapy	100 (68.5%)	46 (31.5%)	146 (100%)

$$\chi^2 (1) = 7.25, p = .007$$

### **Evaluating Timing of First Alerts**

A one-way analysis of variance (ANOVA) was conducted to compare the effect of treatment format on the timing of first alerts. As seen in Table 2, the UCC data set did not yield a significant effect of when the first alerts typically occurred between group and individual treatment  $F(1, 1258) = 3.55, p < .060$ . A post-hoc power analysis to detect differences was calculated to be at .27 for an effect size of .20 as low power in the present study is attributable to having 46 group-only patients alert as NOT. In individual therapy, the average session number where the first alert occurred was  $M = 4.70$  ( $SD = 2.88$ ) where first alerts seemed to occur just under one session later in group therapy ( $M = 5.52, SD = 3.49$ ). The temporal pattern of first alerts is shown in Figure 3 and is similar to those seen in Figure 1. However, these findings do not replicate the findings of Burlingame et al. (2018). In their original analyses, they found a significant effect of treatment format and when the first NOT alert occurred  $F(1, 443) = 61.84, p < .001$ . More specifically, they found that first alerts in group therapy tended to occur about two sessions later than individual therapy (3.21 vs. 5.05, respectively).

Table 2  
*One Way ANOVA of First Alert Timing between Treatment Formats*

	Sum of Squares	<i>df</i>	Mean Square	F
Between Groups	29.98	1	29.98	3.55
Within Groups	10622.73	1258	8.44	
Total	10652.71	1259		

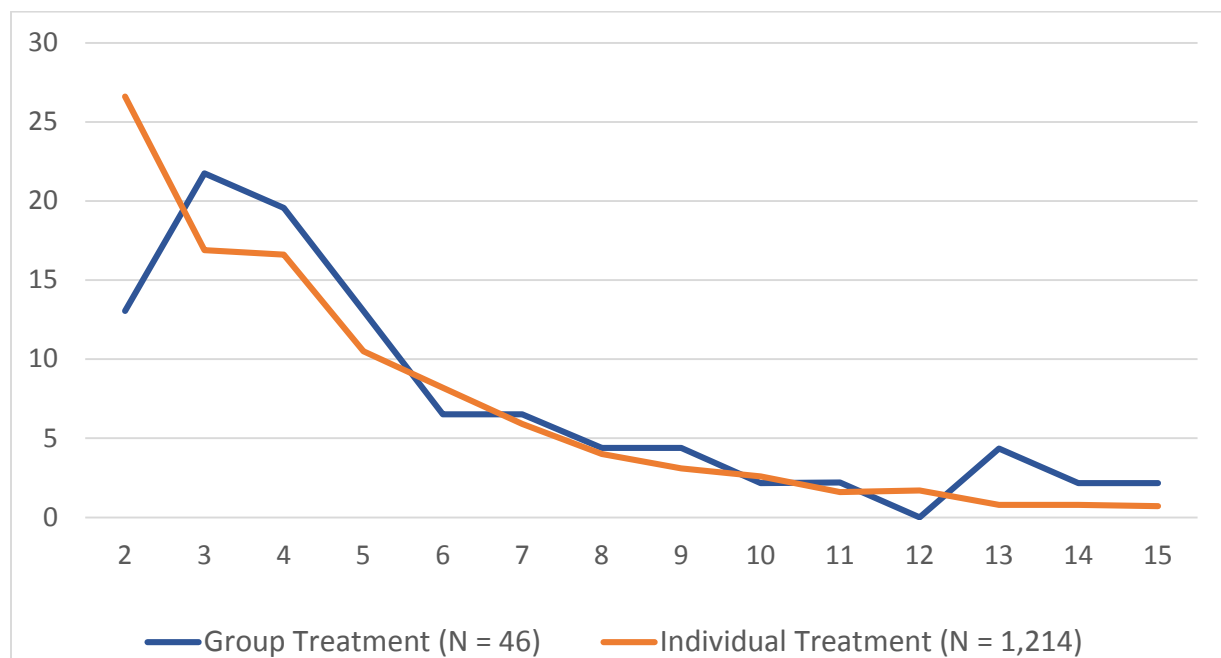


Figure 3. Comparing the percentage of first not-on-track (NOT) alerts by session between individual and group treatment for the present study.

**Additional Temporal Analyses**

Burlingame et al. (2018) reported a graph (Figure 2) demonstrating NOT alert frequency among the cases that alerted at least once during treatment. We created a similar graph with our archival data (Figure 4) and, contrary to our hypothesis, found a similar trend in the sense that group therapy had a dramatic increase in NOT alerts after session 9. It is important to note that these effects should not be attributed to attrition because *N* was the same (18) for sessions 9 and

10. It is unclear as to why there is a sudden increase in NOT alerts so late in the course of treatment and we believe this issue warrants further research. The clinical significance of the graphs' similarity may demonstrate that group therapy is at a disadvantage with the number of patients alerting as NOT increases later in treatment.



*Figure 4.* Comparing the percent of the NOT alert frequency by session between individual and group treatment in the archival data.

### Discussion

The major question addressed in this study was whether Burlingame et al.'s (2018) findings that patients receiving group treatment are more likely to receive NOT warnings and that these warnings occur later in treatment were replicable or unique. Due to the lack of research in this specific area, we hypothesized that there would not be a difference in NOT rate and first alert timing between group and individual treatment. In other words, we assumed that the

findings of Burlingame et al. were unique to their sample population and would not generalize into a similar university counseling center. Our results produced mixed findings with respect to our hypotheses.

First, the rate of NOT cases found within the comparable clinic's OQ-45 data appeared to replicate those of Burlingame et al.'s (2018) closely. Using archival individual data (Shimokawa et al., 2010), Burlingame et al. initially found a significant difference between the number of patients that flag as NOT at least once during treatment and the present study found a similar difference in rate between group and individual therapy. More specifically, group therapy yielded significantly more NOT alerts when compared to individual therapy. This can be problematic as it indicates that patients receiving group therapy are more likely to experience treatment failure. This means there is a higher risk of patients ending treatment feeling worse off in group versus individual treatment.

This replication is clinically important for a number of reasons. First, it indicates that approximately 1/3 patients in group therapy with flag as NOT at least once during the course of treatment. Thus, according to past research, clinicians running group psychotherapy should attend to outcome measures in order to know which of the members may be at-risk for treatment failure so they can address these issues which, in turn, may reduce the risk. If clinicians understand beforehand that group patients are at significantly greater risk for flagging as NOT during treatment they may be able to prepare themselves and take necessary steps to improve treatment outcomes for their group members. It may be useful to let group members know (when appropriate) that it is common for individuals in group therapy to feel worse than when they started while they are attending group sessions.

Second, consistent with our original hypothesis, we found that the timing of first-occurrence NOT alerts was not significantly different between group and individual treatment. This finding diverges from Burlingame et al.'s (2018) initial comparison which revealed a significant difference between treatment formats concerning when these NOT alerts first appeared. Although the present study also found that first NOT alerts occur earlier on in individual treatment compared to group treatment, the effect was not significant by the conventional standard of  $p < .05$ . However, there does appear to be a trend and power to detect differences was calculated to be at .27 for an effect size of .20. The effect size of .20 originates from Burlingame, Seedback, et al.'s (2016) meta-analysis looking at individual versus group treatment studies. While the current study's results may suggest that the groups included Burlingame et al.'s (2018) original study may be unique and their first alert temporal patterns are not generalizable to other, similar clinics, it is possible that the 46 group-only patients who alerted as NOT at least once included in the present study did not provide us with a large enough sample size to effectively detect an effect. Further research concerning the timing of first-occurring NOT alerts is warranted to investigate whether or not there is a significant difference between treatment format.

When looking at the frequency of NOT alerts session by session between treatment formats, there appears to be common pattern among patients in group therapy and individual therapy respectively. In both the present study and Burlingame et al.'s (2018) RCT, there appears to be a sharp increase in NOT alerts around session 8 whereas the archival individual treatment data typically showed a gradual decrease or minimally-changing rate of NOT alerts over the sessions. For there to be a dramatic increase of group patients flagging as NOT around session 8 is particularly interesting due to the average number of group sessions attended being 9.5. This is

problematic and can be interpreted various ways. Because at-risk group patients have been seen to attend more sessions than the average, one possible explanation is that group patients become aware of the soon-approaching end to the group and it causes them to report feeling worse on the OQ-45.

Various limitations of this study need to be addressed. First, beyond gender, additional demographic information in the deidentified data was extremely limited (i.e., no information on age, presenting concern, year in school, ethnicity, etc.). Because of this, we could only assume that the sites were fairly comparable in these realms. Had we been given more complete demographic information, we could have made additional efforts to ensure the data sets resembled each other more closely. Similarly, we also did not have information as to what kind of groups these patients were taking part in which could have influenced individual member outcomes.

Another limitation is the present study's group therapy sample size. During the data cleaning, we found that most individuals that attended group therapy were concurrently involved in individual therapy. For the sake of replication, we had already determined that we would only include those that were exclusively participating in either group or individual therapy. We were only left with 146 individuals whose sole form of treatment within their first episode was group therapy. It is also unclear how clinicians in the new data set implemented client results from the OQ-45. According to OQ Measures' records, the clinic did not purchase the OQ-Analyst software until March of 2007. Before owning the OQ-Analyst, the clinicians in the clinic would not have known whether their patients were on-track or not. Even after purchasing the software, therapist compliance and usage is unknown.



Other limitations may arise due to those that were inherent in the original studies included in this replication. For example, the homogeneity of the sample (all of the patients were receiving treatment in university counseling centers) and incentives may have influenced participant compliance with the measures.

Because some of our results were contrary to those of the original study, further research should be aimed at exploring the timing of first alerts across formats in order to reach a clearer conclusion. Future research should also investigate the reason why so many more individuals in group therapy flag as NOT when compared to those in individual therapy. It should also work to test plausible causes for the sudden increase in NOT alerts occurring at session 8 in group therapy. Lastly, attrition rates between treatment formats in congruence with NOT alerts should be more thoroughly investigated to see if attrition has a role in the varying frequency of alerts by session.

## References

- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285.
- Burlingame, G., Gleave, R., Beecher, M., Griner, D., Hansen, K., & Jensen, J. (2016). *Administration and scoring manual for the Group Questionnaire—GQ*. Salt Lake City, UT: OQ Measures.
- Burlingame, G., Seedback, J., Janis, R., Griner, D., Whitcomb, K., Barkowski, S., Rosendahl, J., & Bernhard, S. (2016). Outcome differences between individual and group formats when identical and nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy, 53*(4), 446-461.
- Burlingame, G. M., Whitcomb, K. E., Woodland, S. C., Olsen, J. A., Beecher, M., & Gleave, R. (2018). The effects of relationship and progress feedback in group psychotherapy using the group questionnaire and outcome questionnaire-45: A randomized clinical trial. *Psychotherapy, 55*(2), 116-131.
- Davidson, A. H., Poulsen, S., Lindschou, J., Winkel, P., Tróndarson, M. F., Waadegaard, M., & Lau, M. (2017). Feedback in group psychotherapy for eating disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 85*, 484–494.  
<http://dx.doi.org/10.1037/ccp0000173>
- Hannan, C., Lambert, M. J., Harmon, C., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155–163.

- Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., et al. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 3*, 294-298.
- Krageloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome in mental health services: A scoping study and typology. *Psychiatric Services, 66*(3), 224-241.
- Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change, 6<sup>th</sup> edition*. New York, NY: Wiley.
- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy, 52*(4), 381-390. <http://dx.doi.org/10.1037/pst0000027>
- Lambert, M., Morton, H., Hatfield, D., Harmon, C., Hamilton, S. Reid, R., ... Burlingame, G. (2004). OQ©-45 administration and scoring manual. Salt Lake City, UT: OQ Measures.
- Lambert, M. J., Kahler, M., Harmon, C., Burlingame, G., Shimokawa, K., & White, M. (2013). *Administration and scoring manual outcome questionnaire-45*. Salt Lake City, UT: OQ Measures.
- Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75*(3), 255–261.
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research, 25*, 396 – 407. <http://dx.doi.org/10.1080/10503307.2014.900875>

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. <http://dx.doi.org/10.1037/a0019247>