Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2014-07-01

Unexpected Dramatic Change in Psychotherapy: Comparing Three Methods

Joseph Richard Horner
Brigham Young University - Provo

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

 Part of the Psychology Commons

BYU ScholarsArchive Citation

Horner, Joseph Richard, "Unexpected Dramatic Change in Psychotherapy: Comparing Three Methods" (2014). *All Theses and Dissertations*. 5253.
https://scholarsarchive.byu.edu/etd/5253

This Dissertation 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, ellen_amatangelo@byu.edu.

Unexpected Dramatic Change in Psychotherapy:

Comparing Three Methods

Joseph Horner

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

Doctor of Philosophy

Michael J. Lambert, Chair
Benjamin Ogles
Samuel Hardy
Bruce Carpenter
Patrick Steffen

Department of Clinical Psychology

Brigham Young University

June 2014

Copyright ©2014 Joseph Horner

All Rights Reserved

ABSTRACT

Unexpected Dramatic Change in Psychotherapy:
Comparing Three Methods

Joseph Horner
Department of Psychology, BYU
Doctor of Philosophy

Unexpected dramatic changes in psychotherapy have been observed historically and tied to high recovery rates. Many different methodologies that identify these changes are assumed to be capturing similar or identical phenomena. This study compared three methods – Sudden Gains (SG), Percentage Increase – 50% (PI-50), and Rapid Response (RR) - in a large database looking for similarities and differences. Results suggest that there are significant differences between SG, RR, and PI-50 as methods for operationally defining unexpected dramatic treatment response, and caution should be used when referring to SG, PI-50, and RR as the same phenomenon or interchangeable terms for unexpected dramatic treatment response. In particular, overlap in clients who experienced both a SG and RR was low. Experiencing any of the three phenomenon was associated with higher recovery rates, while differences abound in both which clients experience each of the phenomena and demographic characteristics of those clients. PI-50 identified inconsequential amounts of clients suggesting under its current methodological construction it would have limited useability. These results tying SG, RR, and PI-50 to significant rates of recovery and positive treatment change suggest possible future use as a predictive feedback tool for clients and clinicians alike to be better able to examine the effectiveness of treatment components during treatment.

Keywords*:* psychotherapy, psychotherapy outcome, sudden gains, rapid response, outcome measurement, treatment response, course of treatment

ACKNOWLEDGMENTS

I thank Dr. Michael Lambert and his mentorship and support throughout this process, the BYU Clinical Psychology program for the many opportunities it has afforded me, and my wife and family for their love and tolerance throughout this time consuming process.

Table of Contents

List of Tables

List of Figures

Unexpected Dramatic Change in Psychotherapy: Comparing Three Methods

For better and worse, the progress of clients in psychotherapy was and sometimes still has been assumed to be linear in nature and fairly similar among clients. In reality, it is likely that many different courses of progress are present, even among similar disorders, similar clients, and similar psychotherapies. Lutz and colleagues (Lutz et al., 2013; Lutz et al., 2005; Lutz, Stulz, & Köck, 2009) have found a subset of patients who change (a) positively and rapidly, (b) positively and steadily, (c) remain the same, and (d) who show a deteriorating course. These patterns of change have been used for clinical decision making and prediction of final outcome (Newnham & Page, 2010). In the case of deterioration, predictive methods have been used to identify treatment failure before it occurs and provide alarm signals that assist therapists in their attempts to improve psychotherapy outcomes (Shimokawa, Lambert, & Smart, 2010). On the other hand, observed dramatic, early, and rapid change patterns have implications for theories of change, because many clients improve before they have received a significant amount of treatment (Ilardi & Craighead, 1994) and a phenomena of unexpected dramatic treatment response patterns were beginning to be defined in the literature.

Early work using rudimentary measures of unexpected dramatic treatment response such as a $\geq$ 50% change in depressive symptomology from intake levels by session two (Renaud et al., 1998) and significant response within the first two weeks of treatment, predicted better outcomes at the end of treatment (Fennell & Teasdale, 1987) and even at up to 24 month follow-up (Renaud et al., 1998) when compared to patients not manifesting this phenomena. However, the mechanisms, definition, and even what to label these phenomena has varied from study to study. Labels used have included “Sudden Gains” (Tang & DeRubeis, 1999b), “Rapid Response” (Fennell & Teasdale, 1987), “Early Response” (Ilardi & Craighead, 1994), “Large

Sudden Improvements" (Gaynor et al., 2003), and "Early Sudden Gains" (Stiles et al., 2003). Consequently, clarification is needed to understand if the operational definitions used by various researchers are capturing the same phenomena. Parsimony in terms, however, can only be obtained if the phenomena being studied are truly similar. The following will review operational definitions of three phenomena and their criteria used to operationalize the stated phenomenon of dramatic unexpected change in psychotherapy. In addition to capturing unexpected dramatic change, each of the three phenomena may provide potential prediction tools for clinicians to identify clients who may recover at higher rates than the general client population.

Percentage Improvement

A common criterion used early in the study of this phenomenon is a greater than or equal to 50% reduction in score on an outcome measure. Used originally with the Beck Depression Inventory (BDI) (Beck & Steer, 1987; Beck, Steer, & Brown, 1996), this criterion called for examining cases where a "decline of ≥ 50% in the BDI score from pretreatment until the beginning of the 2nd session of psychotherapy." While psychopharmacology research often uses this method, it has been used sparingly in psychotherapy studies. In a classic study, adolescents who met this criterion showed better outcomes at the end of treatment and at 1 and 2 year follow-up periods (Renaud et al., 1998).

More recent iterations of this method have referred to these phenomenon as Percentage Improvement 50% (PI-50). The PI-50 method compares change based on a baseline rate of distress. In other words, if a 50% change in a measure of psychopathology is seen between the initial score and the measure's cutoff for clinically distressed or impaired, the client has experienced the PI-50 phenomenon. Only a few recent studies utilized the PI-50 method for change with psychotherapy (Dimidjian et al., 2006; Strauman et al., 2006). A recent study by

Hiller, Schindler, and Lambert (2012) demonstrated comparable accuracy in showing clinically significant change between the PI-50 method the Reliable Change Index (discussed below). However, the PI-50 method, with one notable exception (Renaud et al., 1998), was historically used with end of treatment scores and may require more research before being used as a predictive method during treatment. Nevertheless, due to its simplicity in measurement and wide use for other functions, the PI-50 method may show promise for use in a predictive function and identification of unexpected dramatic treatment response in clients.

Sudden Gains

Spurred on by the work of Ilardi and Craighead (1994), researchers Tang and DeRubeis (1999a) reconsidered the idea of dramatic "between session" changes (although in actuality the change could have occurred during the session) in depression symptomology. Their initial study (Tang & DeRubeis, 1999b) led to subsequent research, especially in the CBT treatment literature. In their study they first offered a precise conceptual definition of the Sudden Gains phenomena. They then operationalized this definition into the methodological form that is largely used today. In their results they found many aspects of the Sudden Gains phenomena which continued to be supported in subsequent research. The theoretical definition of Sudden Gains was originally split into three requirements. Sudden gains were originally defined as the intersession change needing to be "large (a) in absolute terms, (b) relative to depressive symptom severity before the gain, and (c) relative to symptom fluctuations preceding and following the gain" (Tang & DeRubeis, 1999b). The research was initially done using the Beck Depression Inventory (Beck & Steer, 1987) or BDI as the outcome measure. This was later translated into methodological terms stating that a Sudden Gain occurred when:

> the gain was (a) ≥ 7 BDI points (absolute magnitude), (b) represented at least 25% of the pregain session's BDI score (relative magnitude), and (c) the difference between the mean BDI score of the three sessions before the gain (n - 2, n - 1, and n) and the three sessions after the gain (n + 1, n + 2, and n + 3) was at least 2.78 times greater than the pooled standard deviations of these sessions' BDI scores (relative to symptom fluctuation (Tang, DeRubeis, Hollon, Amsterdam, & Shelton, 2007).

This definition, while specific and replicable, is not without controversy. The criteria are somewhat arbitrary in nature (e.g., 7 point change), but have been used to some degree in most of the succeeding studies to some degree. Problems include the need for consecutive data of at least 6 sessions. Such data is not always available, particularly in routine care. Measures other than the BDI have been used to track client change in patients with depression, and disorders other than depression are frequently studied where the Sudden Gains phenomenon is observed.

Frequency and outcomes. Sudden Gains based on the Tang et al., criteria were found in 39% of the clients in the aforementioned study (Tang & DeRubeis, 1999b). Those who experienced a Sudden Gain showed an average of 11.2 points of change on the BDI. This change accounted for 51% of the total change of these clients by the end of 12-14 weeks of a RCT cognitive therapy treatment. Of those who experienced Sudden Gains, 79% reached the status of "recovered" by the end of treatment compared to 41% in those who did not experience a Sudden Gain. At follow-up of 6 months and 18 months after the end of treatment, those who experienced a Sudden Gain exhibited significantly reduced amounts of depressive symptoms on the BDI, compared to those who had not experienced a Sudden Gain (Tang & DeRubeis, 1999b).

Follow-up studies were conducted using much of the same criteria as the original criteria for Sudden Gains (Tang & DeRubeis, 1999b). The studies available for review (Andrusyna, 2008; Andrusyna, Luborsky, Pham, & Tang, 2006; Cavallini & Spangler, 2013; Gaynor et al., 2003; Hardy et al., 2005; Hofmann, Schulz, Meuret, Moscovitch, & Suvak, 2006; Lutz et al., 2013; Stiles et al., 2003; Tang & DeRubeis, 1999b; Tang, DeRubeis, Beberman, & Pham, 2005; Tang et al., 2007; Tang, Luborsky, & Andrusyna, 2002; Vittengl, Clark, & Jarrett, 2005) exhibited results and outcomes similar to those in the original study.

Prediction of better outcomes. In all of the studies mentioned, the presence of Sudden Gains was associated with better outcomes at the end of treatment. Multiple studies showed significant maintenance of treatment gains at some period of follow-up (Gaynor et al., 2003; Hardy et al., 2005; Stiles et al., 2003; Tang & DeRubeis, 1999b; Tang et al., 2007). Some studies, however, did not show gains at follow-up relative to the rest of the sample (Hofmann et al., 2006; Tang et al., 2002; Vittengl et al., 2005). Overall, a large amount of consistency in results was seen across Sudden Gains studies. This suggests that this phenomenon can be largely replicated by the use of the original methodological definition, despite its inherent statistically arbitrary nature (Andrusyna et al., 2006; Tang & DeRubeis, 1999b; Tang et al., 2007).

A recent meta-analysis was published estimating the effects of Sudden Gains (Aderka, Nickerson, Bøe, & Hofmann, 2012). They selected 16 studies with a total 1,104 participants. In their results they found a robust moderate effect at post-treatment (Hedges' $g = 0.75$) and follow-up (Hedges' $g = 0.56$) for those who experienced a SG. As such, in this study SG are associated with moderate, short, and long term improvements in depression and anxiety symptoms. They also analyzed treatment orientation as a moderating effect, and found the effect size tended to be higher in CBT studies compared to other methods when treating depression and anxiety

(Hedges' $g = 0.75$ for CBT vs Hedges' $g = 0.23$ for other treatments). This may be explained by the high degree of control in many of the CBT studies when compared to their non-CBT counterparts. It may also be evidence of CBT's tendency to cause Sudden Gains to occur. Unfortunately when theorizing how to explain Sudden Gains, this article shows considerable bias towards cognitive methods in the studies included, but nevertheless serves to illustrate the accumulating interest in the topic of treatment components. Indeed, the fact that SG occur in one intercession period that can follow a specific treatment component lends natural interest in this area.

Rapid Response to Psychotherapy

Spurred on by research contrasting the placebo effect in psychotropic anti-depressant medications and psychotherapy and the Sudden Gains literature, Haas, Hill, Lambert, and Morrell (2002) investigated the timing of dramatic response in psychotherapy and its relation to reported psychological distress at the end of treatment and at follow-up. Often in studies with anti-depressant medication, rapid response to treatment was a negative indicator for long-term outcomes following administration of psychoactive medications. In psychotherapy research indications of the opposite have been found to be true. Individuals from a counseling center at a private university whose mental health had been tracked over the course of their psychotherapy treatment were sent the Outcome Questionnaire-45 OQ-45: (Lambert, Kahler, Harmon, Shimokowa, & Burlingame, 2011) one to two years after the discontinuation of treatment to measure treatment outcome at follow-up. Of the 240 returned surveys, 61% had attended at least one session of psychotherapy, and had intake OQ-45 scores in a distressed range.

Clients were selected as having experienced a Rapid Response to psychotherapy when their OQ-45 scores following their first three sessions of treatment were significantly lower (10^{th}

percentile) than predicted scores based on statistical modeling of treatment response from a large sample of outpatients whose response to treatment was tracked on a weekly basis (Finch, Lambert, & Schaalje, 2001). Expected treatment response is based on deviations from patients with identical (or nearly identical) level of initial disturbance. The use of scores averaged across the first three sessions was a judgment call of the researchers in an attempt to best capture response as "rapid" (i.e., early) and stable. Clients' were categorized at treatment termination and follow-up as "improved" if the change in OQ-45 score decreased and met criterion equal to, or larger than, the reliable change index of the OQ-45 or as "recovered" if their OQ-45 score changed reliably and dropped below clinical levels of dysfunction.

Analyzing the data, the researchers found that those with more dramatic rates of response had better outcomes at termination and follow-up (Haas et al., 2002). Of the rapid responders, 84% were categorized as improved or recovered by the end of treatment. This made up 82% of all reliably improved cases and 61% of all recovered cases in the sample were unexpected dramatic responders. Maintenance of gains at follow-up was also shown in clients classified as rapid responders to psychotherapy.

Subsequent research. A related follow-up study by Lutz et al. (2009) confirmed a correlation between Rapid Response and outcome in treatment. In their study, all participants who experienced Rapid Response had significant change at the end of treatment. Further, 96% of their sample maintained these gains at follow-up. Both at the end of treatment and follow-up, the Rapid Response group outperformed the rest of the sample (Lutz et al., 2009).

The Necessity of "Early"

As stated earlier, many have speculated that this phenomenon of dramatic unexpected change in psychotherapy often occurs early in treatment (Andrusyna, 2008; Flückiger, Holtforth,

Znoj, Caspar, & Wampold, 2013; Haas et al., 2002; Tang et al., 2007; Van et al., 2008). For example, in CBT literature, it has been suggested that as much as 60-70% of gains are made in the first four weeks (Ilardi & Craighead, 1994). Often unexpected dramatic change occurs before it is plausible to attribute the change to the supposed mechanisms of the treatment being used (Lambert, 2005). Many of the Sudden Gains studies reviewed above showed the median session of occurrence to be session five. Of note in one study (Tang et al., 2007), looking at a data set with two different Sudden Gains criteria, a slight modification disallowing identification of Sudden Gains before session five of treatment, among other changes, seemed to eliminate the presence of Sudden Gains association with outcome in that particular study. Likewise, in the reference study for Rapid Response the clients were only labeled as having experienced a Rapid Response to psychotherapy if it occurred in the first three sessions of treatment (Haas et al., 2002), a claim that has seen other support in naturalistic treatment settings (Delgadillo et al., 2014). A study by Busch, Kanter, Landes, and Kohlenberg (2006) found that of those who experienced Sudden Gains, experienced them in the first half of treatment and had significantly better outcomes at the end of treatment than those who did not have a Sudden Gain over this same time period. Those who had experienced Sudden Gains in the last half of treatment showed no such differences. In Stiles et al. (2003), no significant relationship to outcome was seen in those who experienced a Sudden Gain after session 16 in a naturalistic sample.

These studies offer evidence that the occurrence of Sudden Gains or Rapid Response to psychotherapy may be more meaningful when they occur early in the course of therapy and before many of the specific therapeutic techniques, such as the challenging automatic thoughts portion of many CBT protocols, had been offered (Busch et al., 2006). However, in the Sudden Gains literature there is little consensus of how many sessions constitutes an operationalization

of "early". A range of cutoff scores are observed that vary from session five (Tang et al., 2007) to session 16 (Stiles et al., 2003), to throughout therapy (Tang & DeRubeis, 1999b). Timing of Sudden Gains even vary significantly in the same study. Looking at CBT treatment for eating disorders, Cavallini and Spangler (2013) observed that 50% of SG were experienced between sessions 3 and 11 on an outcome measure for eating behaviors; and between sessions 5 and 15 on an outcome measure for body image concerns. Due to this variability, most studies still include and state all Sudden Gains that occur throughout treatment as it relates to ultimate outcome and other variables whereas Rapid Response studies historically examine RR much earlier in treatment.

RCT Versus Naturalistic ITT Studies

It should be noted that sudden gains have mainly been studied in randomized clinical trials (RCTs) with set doses of therapy of about 12-14 sessions rather than in routine care where doses are not set and treatments are of both longer and shorter duration.

In RCTs it is common that to participate in the study the patient has to agree to complete either all or a high percentage of the treatment. If this does not happen, it is likely that the patient and their results are excluded from the study. This data is excluded with no knowledge on why the patient decided to no longer participate and if it related to being a subject in the study. This is a logical part of a controlled study endeavoring to find if therapy has an effect, but may not represent how psychotherapy is conducted in a naturalistic practice setting. As such, one may conjecture if early dropouts are excluded in the RCT studies, a relatively large number of Sudden Gains or Rapid Response to psychotherapy phenomena may be excluded. Conflicting evidence to this effect is seen in the research literature. Haas et al. (2002) used a naturalistic university counseling center database for their work. The therapy did not have a fixed dosage

and generally did not last as long as a clinical trial. They observed a higher percentage of clients with dramatic change than the average Sudden Gains study. In contrast, Stiles et al. (2003), who also studied outcomes in a naturalistic practice, found a much lower percentage of patients experienced Sudden Gains than in the average RCT Sudden Gains study.

Another difference found between naturalistic or intent-to-treat and RCT studies is the control of the treatment administered. Whereas in most RCT studies treatments are tightly controlled, naturalistic studies often endeavor to have clinicians operate in the manner in which routine care is administered. As such, it is common for a whole host of different treatment theoretical orientations to be used in naturalistic studies. Further, the pace, implementation, and components of even similarly named treatments likely vary from practitioner to practitioner. This study examined Sudden Gains in a "real world" or routine care setting. Admittedly, it is not known whether the differences seen between RCT and ITT studies occur due to the inclusion of patients that would be excluded from RCT studies, the uncontrolled nature of the treatments implemented, or some other factor associated with studies in a naturalistic setting. Nevertheless, it is an important factor to be aware of and consider when making assertions in this area of research and especially relevant to the use of these different phenomena as feedback tools for clinicians.

Predictive Tools

Predictive tools have become more commonly pursued by researchers in an effort to give clinicians predictions of treatment outcome (Newnham & Page, 2010). Rapid Response, Percentage Increase – 50%, and Sudden Gains phenomena can be calculated and identified during the course of treatment easily by hand or by being built into the software of the measure being used to track treatment progress. The need for a predictive tool may be necessary due to

human error in identifying such clients by clinicians in real time. For example, in one study (Davies et al., 2006) clinicians in routine clinical practice were only able to retroactively identify half of the clients who had experienced a SG. Additionally, Flückiger et al. (2013) suggests many different variable may need to be monitored early in therapy in addition to unexpected dramatic response. As such, predictive tools may be helpful in this area.

Summary

As it turns out, psychotherapy does not appear to have incremental, session-by-session effects over its course for all patients, but instead, patients who show a substantial benefit and maintain these gains tend to respond rather unexpectedly and dramatically. Research has yet to show why this is the case, but it is obvious that such changes precede rather than follow many of the techniques thought necessary to bring about change.

There are also a variety of ways of operationally defining the dramatic, unexpected change in psychotherapy phenomenon, but little agreement on methods to do so. Given the research just reviewed, the general purpose of the current study was to explore three different *methods* (operational definitions) of defining the phenomena of unexpected dramatic change during psychotherapy using the same set of clients. The Sudden Gains method focuses on a relatively large and absolute (set) amount of change at a single intersession period (relative to initial disturbance and surrounding sessions). The Percentage Increase – 50% method relies on the arbitrary cutoff of half of the clinical scale being used to measure treatment change. The Rapid Response method depends on a positive change that is somewhat rare (experienced by only 10% of cases at any particular session) relative to the changes made by the typical patient at the same session of care who started with the same measured distress level. The identification of

unexpected dramatic treatment response before termination using one of these methods may benefit clinicians and clients alike in addition to having implications for theory.

Research Questions

Many of the outcomes related to Percentage Increase – 50%, Sudden Gains and Rapid Response phenomena are similar. This may suggest, as stated earlier, that the phenomena being studied may be in some part correlated or overlapping and that their operational differences may be able to be used interchangeably. However, to make this assumption, one would assume similarity between those identified as having experienced each of these three phenomena. If these phenomena are in actuality different we would see differences in who experiences them and in other variables related to treatment outcome. To gain a more complete understanding the following questions were addressed in this study.

1. When using the separate historically defined criteria for Percentage Increase – 50% (Renaud et al., 1998), Sudden Gains (Tang & DeRubeis, 1999b) and Rapid Response (Haas et al., 2002), how much overlap in clients are identified between methods?
2. Do descriptive characteristics such as outcomes at the end of treatment, dosage amounts, and other characteristics differ significantly between those in the sample who experienced a Percentage Increase – 50%, Sudden Gains or Rapid Response criteria?

Method

Participants

Clients were recruited from the BYU Counseling Center archival database. Those seeking only academic services were excluded. Clients provided at least two measurements (one inter-session period) where an OQ-45 measurement was taken. In addition, all clients included

began treatment in the clinically distressed range of the OQ-45. Those included in the database were university students who received individual psychotherapy between 1996 and 2013 and had completed treatment by March 2013.

Therapy at the BYU counseling center is offered free of charge and without session limits to full-time students of the university. Clients are referred or self-referred for a wide range of problems, the majority of which are adjustment, anxiety or depression related. Individual therapy generally consisted of the traditional 50 minute hour. All therapists were either PhD level psychologists, psychologist trainees, or licensed clinical social workers. Therapy theoretical system and implementation were used at the discretion of the therapist including cognitive-behavioral, psychodynamic, client-centered, existential, systems, and integrative modalities.

All clients used in the study completed treatment prior to data analysis. Prior approval from the research committee overseeing the database was obtained following approval of the Institutional Review Board. All clients routinely take the OQ-45 at each therapy session. The OQ-45 was administered in the current study either by paper or electronically to patients prior to the beginning of each therapy session, thus providing a measure of mental health functioning for the week preceding each session attended. Clients and therapists were aware that outcome scores were stored for research purposes, and informed consent was obtained at intake.

Outcome Measure

Psychological outcome was assessed during treatment using the Outcome Questionnaire-45 (OQ-45) by Lambert et al. (2011) which is a 45-item instrument designed to measure client distress and functioning over the last week and typically administered prior to each therapy session to track progress in therapy. Items are rated on a 5-point Likert scale. Total scores can

range from 0 to 180, with higher scores reflecting more severe distress and lower scores reflecting less distress. It contains items to assess risk for suicide, substance abuse, and violence at work. It has three domains: a depression and anxiety loaded symptom distress domain, a domain reporting interpersonal functioning of the individual, and a domain concerning the individual's reported social roles. Its main use in this study was to track change of general distress levels in these areas over time. It has a reported alpha level of .92. It has a correlation of .63 with the Beck Depression Inventory (Lambert et al., 1996; Lambert et al., 2011). Test-retest reliability was reported as .84 over a 3-week period. The OQ-45 is also reported, however, as sensitive to change, improving an average of 17.47 points in a sample of 40 patients receiving psychotherapy. Norms have been established for individuals between the ages of 18 and 80 within university, private practice, community mental health, outpatient, and inpatient settings. Scores on the OQ-45 greater than or equal to 64 are considered in the clinical range (Lambert et al., 2011).

Although OQ-45 scores can be explored as a continuous variable, the measure also offers cut-offs for a reliable change index (RC) and clinically significant change (CS), derived from the model of statistically operationalized clinically significant change proposed by Jacobson and Truax (1991). RC is defined as change in observed scores that exceeds the amount of variation expected within the standard error of measurement, which, in the case of the OQ-45, equals at least 14 points. CS is distinguished by two criteria: a) the change observed is equal to or exceeds the RC index, and b) the score leaves the clinical range of functioning (in the case of the OQ-45, scores > 63) and enters the normal range of functioning (OQ-45 $\leq$ 63). Both RC and CS were used in the current study to identify clinically significant change. Additionally, a change of 14 points in a negative direction (where the client is worse than when they began) defines reliable

deterioration, another category explored for significant differences between groups. These categories (i.e., RC, CS, and deterioration) are non-overlapping, and a significant portion of clients experienced neither reliable change nor reliable deterioration.

Identification

Clients from the database were identified for each of the phenomena based on the following criteria:

50% Percentage Increase (PI-50). The criteria for PI-50 phenomenon was taken from a previous study where the PI-50 phenomena was used (Hiller et al., 2012) in order to keep in line with literature on the topic and best represent historical operationalization of the phenomena. This resulted in two criteria as necessary for a client to have experienced a PI-50:

1) The measured amount of change between beginning and OQ-45 at the session examined was greater than 50% of the difference between the beginning OQ-45 score and the basal OQ-45 score to qualify as clinically distressed (OQ ≥ 64). In other words, if the change in a client's OQ-45 score was greater than or equal to a 50% change between the initial score and the measure's cutoff for clinically distressed or impaired. For example, if client A started treatment at an OQ-45 score of 104 and then by session two reduced his OQ-45 score more than half the distance between his initial score (104) and the cutoff for clinically distressed (64), or in this case 20 points reduction on the OQ-45, that client would meet this criterion.
2) The total amount of change was equal to or greater than 25% of the entire range of the instrument, which would be 45 points on the OQ-45. This criterion was proposed by Hiller et al. (2012) to eliminate clients who may qualify for the first criterion despite a low amount of change due to an initial OQ-45 score near the clinically

distressed cutoff of 64. For example, if Client *B* started treatment at an OQ-45 score of 70 and was measured as a score of 60 on session two that client would technically meet criterion 1 (≥ 50% reduction in OQ-45 score) but not meet criterion 2's requirement of a 25% reduction over the entire range of the OQ-45.

In addition, while PI-50 criteria is often used as a total change measurement at the end of treatment, in this study it is being used as a method of identifying unexpected dramatic treatment response during treatment. A historical study by Renaud et al. (1998) used PI-50 to identify those who changed early and in greater amounts than expected and its correlation to final outcome and follow-up maintenance. They operationalized "early" as PI-50 occurring in the first two sessions, an admittedly arbitrary number. It was decided to expand the criterion to the first three sessions to be in line with the more extensive literature on Rapid Response (discussed below). This decision is supported both by recent research suggesting three sessions as a cutoff point (Delgadillo et al., 2014) and an analysis we conducted using different possible cut-off points for the OQ-45. The analyses examined PI-50 using cutoffs of 2, 3, 4, and 5 sessions. There was limited added value beyond that observed at session 3. Specifically, little difference in mean number of weeks in therapy, initial distress levels, and ending distress levels was seen after that point. As such, to keep in line with cutoff scores for RR (see below), past literature, recent studies, as well as the current analysis, those who experienced a PI-50 at session 3 or sooner were examined in this analysis.

Sudden Gains (SG). The criteria for identifying Sudden Gains have changed over the years from the original Sudden Gains study (Tang & DeRubeis, 1999b). In this study, we reference multiple studies and progression of the criteria, drawing largely from the definition in the most recent SG study by Tang and DeRubeis (Tang et al., 2007), who were the original

identifiers of the Sudden Gains phenomenon, and that of Stiles et al. (2003) who adapted the criteria from its original use with the Beck Depression Inventory to a general distress measure similar to the OQ-45. The Tang et al. (2007) criteria were as follows:

(a) the gain was ≥ 7 BDI points (absolute magnitude),

(b) the gain represented at least 25% of the pregain session's BDI score (relative magnitude), and

(c) the difference between the mean BDI score of the three sessions before the gain (n - 2, n - 1, and n) and the three sessions after the gain (n + 1, n + 2, and n + 3) was at least 2.78 times greater than the pooled standard deviations of these sessions' BDI scores (relative to symptom fluctuation) (Tang et al., 2007).

Concerning criterion A, Tang and DeRubeis (1999b) noted that they had tried different amounts of BDI points, but settled on seven points due to the outcomes obtained at the end of therapy. Stiles et al. (2003) recognized that 7 points on the BDI was close to its Reliable Change Index. They then used the Reliable Change Index for their measure for this criterion. Seggar, Lambert, and Hansen (2002) have found the BDI to have a SD of 6.33 averaged across hundreds of studies of non-patient samples and 9.99 in clinical samples, with an RCI of 8.46. To keep in form with past procedures, we used the reliable change index from the OQ-45, which is 14 points (Lambert et al., 2011). Consequently, we will "consider a sudden gain to have occurred between Session N and Session N + 1 if:

a) The gain was ≥ 14 OQ-45 points

b) The gain represented at least 25% of the pre-gain sessions OQ-45 scores ($OQ_N - OQ_{N+1} \geq 0.25 \text{ X } OQ_{N)}$

c) The difference between the mean OQ-45 score of the three sessions before the gain (OQ_{N-2}, OQ_{N-1}, and OQ_N) and the three sessions after the gain (OQ_{n+1}, OQ_{n+2}, and OQ_{n+3}) was at least 2.78 times greater than the pooled standard deviations of these sessions' OQ-45 scores (relative to symptom fluctuation). Keeping in line with Tang (2007) when applied to session 2, OQ_{N-2} will not be used. Also OQ_{N+3} is not used if gains occurred on the third to the last session. "Thus, only changes following the first session and the second-to-last session were excluded. (Tang et al., 2007)."

This last criterion is admittedly arbitrary in nature (Stiles et al., 2003; Tang & DeRubeis, 1999b; Tang et al., 2007), but was used in the present study to be at least partially consistent with past research and, as such, in this study, the best way to compare the Sudden Gains phenomena with the other operational definitions.

Rapid Treatment Response (RR). In the software for the OQ-45, cases are labeled as rapid responders if their reported mental health functioning improves more than expected. This determination uses the initial OQ-45 score of the participant and compares it to at least 220 individuals from a database of individuals who had the same initial OQ-45 score. Hierarchal Linear Modeling curves were used to model expected progress (Finch et al., 2001). A score is considered a RR if the degree of improvement, relative to others with the same initial OQ-45 score, places it within the top 10^{th} percentile. In this study the OQ-45 software (OQ-Analyst) was used for identification of this criterion.

Haas et al. (2002) took the mean of the first 3 sessions of therapy and then compared the means to the expectancy curves. Some evidence has shown that using the first three sessions as a cutoff point when comparing treatment progress to initial OQ-45 scores as RR does, has some merit (Delgadillo et al., 2014). In this study, those who experienced a RR in the first three

sessions of therapy were simply labeled Rapid Response (RR) and no attempt was made to average across sessions as was done by Haas et al. (2002). It should be noted that the cutoff for inclusion was if the RR was observed by session three, not measurement three. For a more complete discussion of RR criteria and measuring unexpected dramatic response based on initial distress scores, the following references can be consulted (Finch et al., 2001; Lambert et al., 2011; Lambert, Whipple, Smart, Vermeersch, & Nielsen, 2001).

Recovery and reliable change. A client was considered "recovered" if they reliably improved (i.e., 14 or more points) and entered the ranks of normal functioning as defined by the Jacobson and Truax (1991) cut-off for normal functioning (i. e., were below 64 at the time of termination). A client was considered reliably improved if they improved by 14 or more points but did not end treatment in the non-distressed range (Lambert et al., 2011). Those with OQ-45 scores below a 64 were in the non-distressed range. A client was categorized as experiencing "no change" if no reliable change occurred, and "deteriorated" if their OQ-45 score at the end of treatment was reliably worse (i.e., 14 or more points) than at the beginning of treatment.

Results

Groups

Of the 14,689 clients in the archival data base, 12,060 (82.10%) did not meet any of the criterion for dramatic early change while 2629 (17.90%) experienced some form of unexpected dramatic treatment response: 1,768 clients (12.04%) experienced a RR, 967 clients (6.58%) experienced a SG, and 255 clients (1.74%) met PI-50 criteria (Table 1). It should be noted that the same client could experience one of more of the phenomenon and as such, the percentages do not add up to an even 100%. It is obvious from Table 1 that wide variation in the number of cases identified by each methods exists. This makes the likelihood of finding overlap in

identified cases between cases quite low. As can be seen in Table 1 few clients were classified as having a dramatic response, particularly when the PI-50 criteria were applied, with the RR criteria identifying many more clients.

Table 1

Amount of Sample Who Experienced Each Phenomenon (Overlapping Groups)

Group	*n*	Percentage of Total Sample
None	12,060	82.00%
PI-50	255	1.74%
SG	967	6.58%
RR	1,768	12.04%

Note. As a client could experience more than one phenomenon, percentages do not add up to 100%.

Occurrence Rates and Overlap Among Groups

Out of the 2,629 clients who experienced one or more of the phenomena 2295/2629 (89%) only experienced a single phenomenon exclusively. Accordingly, 112/2629 clients (0.76%) met PI-50 criteria only and not a SG or RR, while 1,469/2629 clients (10.00%) experienced only a RR and not a PI-50 or SG, and 714/2629 clients (4.86%) experienced only a SG and not a PI-50 or RR. As a vast majority of those who experienced some form of unexpected dramatic response did not experience more than one type, this supports the notion that despite theoretical similarities SG, RR, and PI-50 are markedly different phenomenon. Overlap that does exist is shown in Table 2 and Table 3 and explained in more detail below.

Table 2

Overlap Rates Among PI-50, RR, SG (Overlapping Groups)

Base Group	Overlap Group	*n*	Percentage Overlap of Base Group
PI-50	RR	108	42.35%
	SG	62	24.31%
RR	PI-50	108	6.11%
	SG	218	12.33%
SG	PI-50	62	6.41%
	RR	218	22.54%

Note. As a client could experience more than one phenomenon, percentages do not add up to 100%.

Table 3

Overlap Rates Among Independent Phenomena Groups

Group	*n*	Percentage of Total Sample	PI-50 Overlap	RR Overlap	SG Overlap
none	12060	82.10%	-	-	-
PI-50 Only	112	0.76%	43.92%	-	-
RR Only	1469	10.00%	-	83.09%	-
SG Only	714	4.86%	-	-	73.84%
PI-50 + RR	81	0.55%	31.76%	4.58%	-
PI-50 + SG	35	0.24%	13.73%	-	3.62%
RR + SG	191	1.30%	-	10.80%	19.75%
PI-50 + RR + SG	27	0.18%	10.59%	1.53%	2.79%

Overlap in PI-50. Comparing the 14,689 clients in the database and those who experienced a PI-50 (n = 255), 35/14,689 clients (0.24%) experienced both a PI-50 and SG, 81/14,689 clients (0.55%) experienced both PI-50 and RR, while 27/14,689 clients (0.18%) experienced all three phenomena. As such, those who experienced a PI-50 made up the smallest portion of the overall sample.

When looking more specifically at the 255 clients who met PI-50 criteria a high amount of overlap was observed: 43.92% experienced only a PI-50 and not a SG or RR, 13.73% both a PI-50 and a SG, 31.76% both a PI-50 and a RR, and 10.59% experienced all three phenomena (Table 3 above). As such, 42.35% of those who experienced a PI-50 also experienced a RR while 24.31% also experienced a SG (Table 2 above). Accordingly, we observe over 50% overlap between those who experience a PI-50 and either SG or RR. This suggests that while PI-50 may be the most restrictive criteria when related to the overall sample population, many clients who experience a PI-50 likely experience another phenomenon.

Overlap in RR. Comparing the 14,689 clients in the database and those who experienced a RR (n = 1768), 81/14,689 (0.55%) experienced both a RR and PI-50, 191/14,689 clients (1.30%) experienced both a RR and SG, while 27/14,689 clients (0.18%) experienced all three phenomena. As such, those who experienced a RR made up the larger portion than either SG or PI-50 but still minority portion of the overall sample.

When looking more specifically at the 1768 clients who experienced a RR, a relatively low amount of overlap with other phenomena was observed: 83.09% only experienced a RR and not a PI-50 or SG, 4.58% experienced both a RR and PI-50, 10.80% experienced both a RR and SG while 1.53% experienced all three phenomena (Table 3 above). As such, 6.11% of those who experienced a RR also experienced a PI-50 while 12.33% also experienced a SG (Table 2

above). Accordingly, a vast majority of those who experienced a RR did not experience an overlapping SG or PI-50 phenomenon. Despite having more similarities in methodology with the PI-50 phenomenon in terms of dosage, the highest amount of overlap for RR was with those who experienced a SG also. However, the low mount of clients who experienced both a RR and another phenomenon (16.91%) suggests that the RR phenomenon differs largely from others despite the theoretical and methodological similarities mentioned previously.

Overlap in SG. Comparing the 14,689 clients in the database and those who experienced a SG (n = 967), 35/14,689 (0.24%) experienced both a SG and PI-50, 191/14,689 (1.30%) experienced both a SG and RR, while 27/14,689 (0.18%) experienced all three phenomena. As such, those who experienced a RR made up the larger portion than either SG or PI-50 but still minority portion of the overall sample.

When looking more specifically at the 967 clients who experienced a SG a relatively low amount of overlap with other phenomena was observed: 73.84% only experienced a SG and not a PI-50 or RR, 3.62% experienced a SG and PI-50, 19.75% experienced a SG and RR, while 2.79% experienced all three phenomena (Table 3 above). As such, 6.41% of those who experienced a SG also experienced a PI-50 while 22.54% also experienced a RR (Table 2 above). As a vast majority of those who experienced a SG did not experiencing an overlapping RR or PI-50 phenomenon this suggest vast differences between SG and RR or PI-50 despite the theoretical and methodological similarities mentioned previously.

Comparing Descriptives

Each phenomenon was compared on different unique descriptive characteristics in an effort to analyze for differences present in the populations who met criteria for each phenomenon (see Table 4).

Table 4

Descriptive Statistics for Independent Groups

Group		Total Treatment Sessions	Number of Measure-ments	Total Weeks in Treatment	Initial OQ-45 Scores	Ending OQ-45 Scores	Total Change on OQ-45
None (n = 12060)	*M*	12.71	8.42	54.50	83.49	73.66	9.82
	Md	6.00	4.00	14.43	81.00	73.00	6.00
	Mode	2	1	0.00	68	69	0.00
	SD	20.18	11.79	84.34	14.62	21.17	19.06
	Min.	1.00	1.00	0.00	64.00	0.00	-102.00
	Max.	670.00	233.00	810.14	171.00	176.00	119.00
PI-50 Only (n = 112)	*M*	10.43	7.22	53.72	94.77	48.47	46.29
	Md	5.00	4.00	12.71	92.00	44.50	49.50
	Mode	3	3	2	106	71	51
	SD	19.38	9.95	82.06	17.61	23.78	23.34
	Min.	3.00	2.00	0.86	64.00	3.00	-35.00
	Max.	173.00	83.00	415.43	135.00	114.00	94.00
RR Only (n = 1469)	*M*	10.63	8.43	46.25	89.51	65.48	24.03
	Md	6.00	5.00	14.71	88.00	64.00	24.00
	Mode	3	3	1	73	74	21
	SD	16.79	10.51	63.79	15.98	21.03	18.64
	Min.	2.00	2.00	0.29	64.00	5.00	-55.00
	Max.	280.00	139.00	383.14	148.00	144.00	95.00
SG Only (n = 714)	*M*	26.47	20.69	112.60	83.44	55.09	28.36
	Md	16.00	14.00	74.29	81.00	53.50	28.00
	Mode	7	7	7	80	46	39
	SD	30.86	20.73	112.35	14.12	23.35	24.21
	Min.	4.00	4.00	1.71	64.00	0.00	-62.00
	Max.	417.00	216.00	663.71	136.00	141.00	104.00

(continued)

Table 4. *Descriptive Statistics for Independent Groups*

Group		Total Treatment Sessions	Number of Measure-ments	Total Weeks in Treatment	Initial OQ-45 Scores	Ending OQ-45 Scores	Total Change on OQ-45
PI-50 + RR (n = 81)	*M*	11.44	8.81	63.62	99.98	56.19	43.79
	Md	6.00	5.00	19.14	100.00	52.00	50.00
	Mode	3	3	4	86	30	45
	SD	15.62	10.29	83.37	17.20	28.48	25.66
	Min.	3.00	2.00	1.14	67.00	6.00	-26.00
	Max.	97.00	57.00	386.00	156.00	160.00	85.00
PI-50 + SG (n = 35)	*M*	12.00	9.99	72.84	105.80	53.17	52.63
	Md	8.00	7.00	25.86	105.00	50.00	59.00
	Mode	5	4	6.57	95	26	66
	SD	10.23	7.57	101.13	14.10	28.07	26.04
	Min.	4.00	4.00	2.86	77.00	8.00	7.00
	Max.	49.00	32.00	516.71	140.00	121.00	107.00
RR + SG (n = 191)	*M*	19.27	16.28	71.57	87.21	53.80	33.41
	Md	10.00	9.00	35.86	86.00	49.00	34.00
	Mode	4	4	5	72	43	32
	SD	27.68	20.80	88.62	14.98	23.83	21.81
	Min.	4.00	4.00	2.71	64.00	3.00	-48.00
	Max.	211.00	173.00	396.43	143.00	150.00	104.00
PI-50 + RR + SG (n = 27)	*M*	9.63	8.70	56.77	103.96	46.67	57.30
	Md	6.00	6.00	17.86	101.00	45.00	60.00
	Mode	4	4	3	81	41	61
	SD	11.78	8.63	74.63	16.36	16.03	17.34
	Min.	4.00	4.00	3.00	81.00	21.00	22.00
	Max.	59.00	39.00	241.14	144.00	88.00	106.00

Note. When multiple modes existed, the smallest value was used.

Beginning of treatment. In order to be analyzed for differences in distress levels at the beginning of treatment, clients were divided into non-overlapping groups based on the phenomena they had experienced. These included clients who met criteria for only one of the three phenomena (SG, RR, PI-50), all paired combinations (i.e., SG and PI-50, SG and RR, etc), those who met all three criteria, and those who met criteria for none of the groups. An ANOVA ($F(7, 14681) = 71.02$, $p < .000$) showed significant differences in distress at the beginning of treatment between the non-overlapping groups tested. Those who met criteria for PI-50 and SG had the highest initial severity *(M* =105.80), while those who met criteria for SG only had the lowest *(M* = 83.44). Using Tukey's *B* Post-hoc analysis, four populations were seen among the eight possible combinations, as seen in Table 5, which displays each group's mean distress scores at the beginning of treatment. Each of the independent groups of clients in the same population in Table 5 did not differ significantly from one another on mean OQ-45 scores at the start of treatment but did differ significantly from the other populations. As such, we see all groups that experienced a PI-50 had significantly higher beginning of treatment distress scores than the groups that did not experience a PI-50 but that there was a large amount of overlap seen throughout the groups.

In order to better understand the differences in beginning of treatment distress levels, a regression was run using initial severity as the dependent variable and entering whether or not a client met criteria for each phenomenon (i.e., SG, RR, or PI-50) into the model as a dichotomous variable. The regression showed that meeting criteria for experiencing each of the phenomena predicted beginning of treatment distress levels significantly. Examining the unstandardized beta coefficient reveals the nature of the effect of experiencing each of the phenomena. When

examining those who had experienced none of the phenomena, the regression model indicates that the mean initial severity score, as measured by the OQ-45, was approximately 83.50. Those who had experienced SG had a mean score that was .04 points lower than the overall group mean and was not statistically significant ($t(14681) = -0.84$, $p = .93$); those who had experienced a PI-50 had a mean score 13.02 points higher ($t(14681) = 13.77$, $p < .05$); and those who met criteria for RR had a mean score 5.67 points higher ($t(14681) = 14.97$, $p < .05$). This indicates that the PI-50 and RR methods tended to identify clients who were more distressed at treatment onset while the SG method identifies clients who are typical in disturbance of all those who entered therapy.

Table 5

Tukey Comparison for Beginning of Treatment Distress

Different Groups	*n*	Mean OQ-45 Scores			
		Population 1	Population 2	Population 3	Population 4
SG only	714	83.44			
None	12060	83.49			
RR + SG	191	87.21			
RR only	1469	89.51	89.51		
PI-50 only	112		94.77	94.77	
PI-50 + RR	81			99.98	99.98
PI-50 + RR + SG	27				103.96
PI-50 + SG	35				105.80

Note. Group means did not differ from one another in same population but did differ from other population. See text for more details.

These unstandardized coefficients can be understood additively. For example, those who had experienced both a RR and PI-50 would have a predicted mean on the OQ-45 at the

beginning of treatment approximately 18.69 points higher than the overall mean score of 83.50. Additionally, those who met criteria for both SG and RR would have a predicted initial severity approximately 5.63 points higher than the overall mean score. These findings are congruent with the Tukey's *B* reported earlier, where those who had experienced only SG showed significantly lower initial severity than the RR and PI-50 groups and those who had experienced PI-50 showed significantly higher initial distress levels, and that those who experienced both SG and RR along with those who experienced just a SG were not significantly different from those who experienced none of the phenomena. As such, it is likely that RR and PI-50 criteria are somewhat closely related to elevated initial OQ-45 scores while SG is not dependent on initial scores.

End of treatment. Comparing non-overlapping groups categorized by the phenomena they had experienced based on their end of treatment OQ-45 scores, an ANOVA ($F(7, 14681) = 146.21$, $p < .000$) showed significant differences in psychological distress at the end of treatment. Those who met criteria for none of the phenomena had the highest end severity *(M* = 73.66.), while those who met criteria for all three of the phenomena *(M* = 46.67) had the lowest end severity. Using Tukey's *B* Post-hoc analysis, two populations were observed were observed among the eight possible combinations, as seen in Table 6. Accordingly, clients in the independent groups that make up population one are alike in their end of treatment distress levels while all clients in the independent groups that make up population two were alike but each population differed significantly from one another in end of treatment distress scores. As such, clients who experienced none of the phenomena and clients who experienced just a RR made a single population (population 2 on the table below) and did not differ significantly from one

another in end of treatment distress but had significantly different end of treatment distress scores than the other population (population one on the table below).

With the exception of those who experienced only a RR and not a SG or PI-50, we see that those who experienced any combination of phenomena did not differ from one another in end of treatment distress scores but as a population together displayed significantly lower end of treatment distress scores than those who experience none of the phenomena.

Table 6

Tukey Comparison for End of Treatment Distress

Different Groups	*n*	Mean OQ-45 scores	
		Population 1	Population 2
PI-50 + RR + SG	27	46.67	
PI-50 only	112	48.47	
PI-50 + SG	35	53.17	
RR + SG	191	53.80	
SG only	714	55.09	
PI-50 + RR	81	56.19	
RR only	1469		65.48
None	12060		73.66

When analyzing end of treatment scores on an outcome measure the need arises to control for beginning treatment scores. An ANOVA ($F(8, 14680) = 689.09$, $p <$ .000) was run, controlling for beginning treatment scores on the OQ-45, and also showed significant differences in psychological distress at the end of treatment.

A regression showed that meeting criteria for experiencing each of the phenomena predicted end of treatment distress levels significantly. The regression indicated a mean ending severity score for all subjects on the OQ-45 of approximately 73.49. Those who had experienced SG had a mean score that was 16.40 points lower than the mean score for the entire population ($t(14685) = -22.84$, $p < .05$); those who had experienced a PI-50 had a mean score 15.14 points lower ($t(14685) = -11.06$, $p < .05$); and those who met criteria for RR had a mean score 7.03 points lower ($t(14685) = -12.80$, $p < .05$). These findings are somewhat congruent with the Tukey's *B* reported earlier, where only two groups, those who had experienced none of the phenomena and those who experienced only a RR and not a SG or PI-50, showed significantly higher ending severity than the group that experienced two or three of the phenomena. This is more evidence that in the area of end of treatment distress SG and PI-50, are similar to one another while RR seems to be somewhat closer to the population mean and perhaps in some cases not significantly different from clients who experienced no other phenomena.

A regression controlling for beginning OQ-45 scores showed that meeting criteria for experiencing each of the phenomena predicted end of treatment distress levels significantly. The regression indicated a mean ending severity score on the OQ-45 of approximately 60.19 after adjusting for beginning OQ-45 scores. Those who had experienced PI-50 had a mean score that was 24.02 points lower than the mean score for the entire population after adjusting for initial distress scores ($t(14684) = -19.76$, $p < .05$); those who had experienced a SG had a mean score 16.37 points lower ($t(14684) = -25.85$, $p < .05$); and those who met criteria for RR had a mean score 10.90 points lower ($t(14684) = -22.34$, $p < .05$). When correcting for beginning OQ-45 scores the distance from the overall mean end of treatment distress score increased for those who experienced any of the three phenomena.

Total change. To compare non-overlapping groups categorized by the phenomena they had experienced and based on their total change in OQ-45 scores, an ANOVA ($F(7, 14681) = 324.15$, $p < .000$) showed significant differences in distress at the end of treatment. Those who met criteria for none of the phenomena had the lowest amount of overall change during treatment *(M* = 9.82.), while those who met criteria for all three phenomena *(M* = 57.30) had the highest amount of change in distress levels over the course of treatment. Using Tukey's *B* Post-hoc analysis, six populations were seen among the eight possible combinations, as seen in Table 7. Those groups in each population do not differ significantly from one another but do differ from groups in other populations in mean total change over the course of treatment. These scores indicate a high amount of difference and little consensus among each of the independent groups in the amount of total change at the end of treatment.

Table 7

Tukey Comparison Total Change

Different Groups	*n*	Mean Q-45 Scores by Population					
		Population 1	Population 2	Population 3	Population 4	Population 5	Population 6
None	12060	9.82					
RR only	1469		24.03				
SG only	714		28.36	28.36			
RR + SG	191			33.41			
PI-50 + RR	81				43.79		
PI-50 only	112				46.29	46.29	
PI-50 + SG	35					52.63	52.63
PI-50 + RR + SG	27						57.30

However, it is of note that total change scores for those who experience none of the phenomenon *(M* = 9.82) are almost 2.5 times less total change on the OQ-45 than the next closest mean total change score, RR only *(M* = 24.03), indicating that experiencing any of these phenomena is associated with dramatically higher amounts of total change over the course of treatment (Figure 1). Also, those groups that include PI-50 seem to show no overlap with any group that does not include a PI-50, indicating significantly higher amounts of total change among those who experience a PI-50.

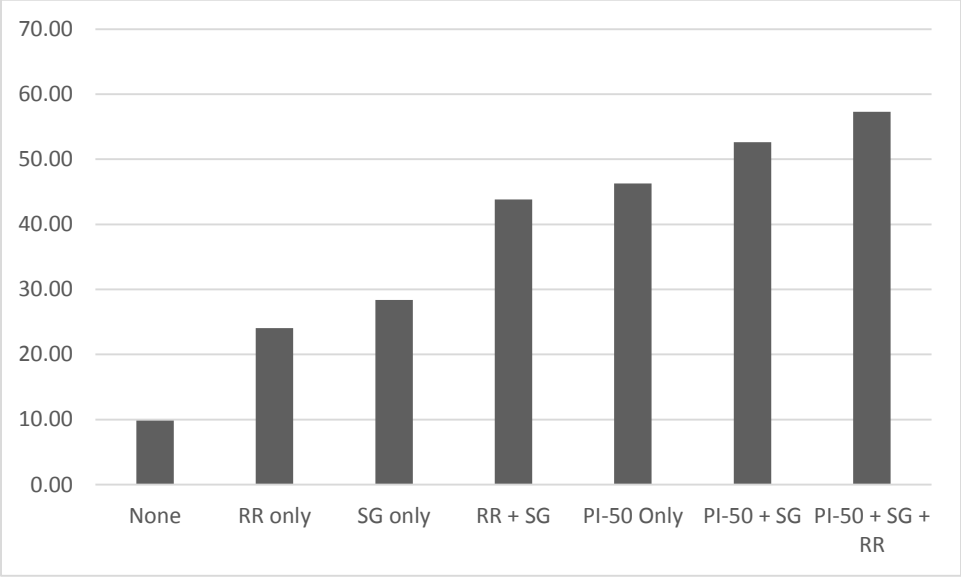

Figure 1. Total Change in Distress, over treatment by independent group

To determine whether or not total change over the course of treatment was still predicted by experiencing a SG, RR, or PI-50 when controlling for beginning distress levels an ANOVA ($F(8, 14680) = 415.50$, $p < .001$) was run controlling for initial scores on the OQ-45which also showed significant differences in total change of psychological distress as predicted by experiencing these phenomena.

A regression showed that meeting criteria for experiencing each of the phenomena significantly predicted the total amount of change in mental health functioning over the course of treatment. The regression indicated that the mean total change over the course of treatment for all subjects was approximately 10.01 points on the OQ-45. Those who had experienced a RR had a mean score that was 12.70 more points total change than the population mean ($t(14684) = 25.46$, $p < .05$); those who had experienced a SG had a mean score 16.36 points more ($t(14684) = 25.46$, $p < .05$); and those who met criteria for PI-50 had a mean score 28.16 points more ($t(14684) = 22.63$, $p < .05$). These findings are congruent with the Tukey's *B* reported earlier, where those who had experienced none of the phenomena showed significantly lower total change over the course of treatment than those that met criteria for any of the phenomena. It also coincided with the finding that experiencing a PI-50 in any combination with other phenomena is correlated with those who experienced the most change throughout treatment but that different combinations of experiencing each phenomenon seemed somewhat independent of one another.

A regression controlling for beginning OQ-45 scores showed that meeting criteria for experiencing each of the phenomena predicted total change in distress levels significantly. The regression indicated that the mean total change score of approximately 3.81 for all subjects after adjusting for beginning distress levels. Those who had experienced PI-50 had a mean total change score that was 24.02 points higher than the population mean ($t(14684) = 19.76$, $p < .05$); those who had experienced a SG had a mean score 16.37 points higher ($t(14684) = -25.85$, $p < .05$); and those who met criteria for RR had a mean score 10.90 points higher ($t(14684) = -22.34$, $p < .05$).

Number of sessions. Comparing non-overlapping groups categorized by the phenomena they had experienced based on the amount of sessions they attended, an ANOVA ($F(7, 14681) =$

49.53, $p < .000$) showed significant differences in total amount of sessions attended in treatment. Those who experienced all three phenomena attended the lowest amount of sessions during treatment *(M* = 9.63.), while those who experience only a SG met for the most sessions over the course of treatment *(M* = 26.47). Using Tukey's *B* Post-hoc analysis, three populations were seen among the eight possible combinations, as seen in Table 8. Those groups in each population did not differ significantly from other groups in the same population in total sessions attended but did significantly differ from groups in the other populations. There is a high amount of overlap among these independent populations as many groups occupy two populations. This may indicate a significant variance in total number of sessions across the population but a lack of separation between each independent group.

Table 8

Tukey Comparison for Total Number of Sessions Attended Throughout Treatment

Different Groups	*n*	Mean Sessions attended		
		Population 1	Population 2	Population 3
PI-50 + RR + SG	27	9.63		
PI-50 only	112	10.43	10.43	
RR only	1469	10.63	10.63	
PI-50 + RR	81	11.44	11.44	
PI-50 + SG	35	12.00	12.00	
None	12060	12.71	12.71	
RR + SG	191		19.27	19.27
SG only	714			26.47

A regression showed that meeting criteria for experiencing each of the phenomena significantly predicted the total amount of sessions attended over the course of treatment. This regression indicated that the mean total sessions attended for all subject was approximately 12.79. Those who had experienced SG had attended an average of 12.06 more sessions than the population mean ($t(14685) = 17.47$, $p < .05$); those who had experienced a PI-50 had attended 3.80 fewer sessions ($t(14685) = -2.89$, $p < .05$); and those who met criteria for RR had attended 2.46 fewer sessions ($t(14685) = -4.66$, $p < .05$). These findings are congruent with the Tukey's *B* post hoc analysis reported earlier, where those who had experienced a SG attended significantly more sessions over the course of treatment, while, overall, any combination of experienced phenomena were closely bunched together in number of sessions attended.

Weeks in treatment. Comparing non-overlapping groups based on the phenomena they had experienced, an ANOVA ($F(7, 14681) = 50.51$, $p < .000$) showed significant differences in weeks in treatment. Those who met criteria for only a RR had the lowest amount of weeks in treatment *(M* = 46.25), while those who met criteria for only a SG had the highest amount of weeks in treatment *(M* = 112.60). Using Tukey's *B* Post-hoc analysis, two populations were seen among the eight possible combinations, as seen in Table 9. Those groups in the same population did not differ significantly on total weeks in treatment from other groups in the same population but did differ significantly from those groups in the other populations. These populations indicate little discernible significant difference is seen between those who experience any of the phenomena in total number of weeks in treatment with the exception of those who experienced a SG and no other phenomena. This group spent significantly more weeks in therapy than all other independent groups studied.

Table 9

Tukey Comparison for Total Weeks of Treatment

Different Groups	*n*	Mean number weeks in treatment	
		Population 1	Population 2
RR only	1469	46.25	
PI-50 only	112	53.72	
None	12060	54.50	
PI-50 + RR + SG	27	56.77	
PI-50 + RR	81	63.62	
RR + SG	191	71.57	
PI-50 + SG	35	72.84	
SG only	714		112.60

A regression showed that meeting criteria for experiencing each of the phenomena predicted differences in the number of weeks between the first and last sessions. The regression model indicated the mean weeks in treatment for all subjects was approximately 54.95. Those who had experienced PI-50 had approximately of 2.46 fewer weeks in treatment more than the population mean ($t(14685) = -0.46$, $p = 0.65$) and was not significantly different from the constant. Clients who had experienced a SG had on average 49.18 more weeks in treatment ($t(14685) = 17.41$, $p < .05$); and those who met criteria for RR had on average 10.90 fewer weeks in treatment ($t(14685) = -5.05$, $p < .05$). These findings are somewhat congruent with the Tukey's *B* reported earlier, where those who had experienced a SG had significantly higher amounts of total weeks in treatment, while those who experienced a SG or RR did not differ from those who experienced none of the phenomena in weeks in treatment. Accordingly, we would expect those who experienced a SG and no other phenomenon to be more likely to be in treatment longer than the overall sample population, those who experienced a RR or PI-50, or even those who experienced a SG in combination with other phenomena.

Recovery. Of the 14,689 clients in this study, 4,278/14,689 (29.1%) recovered by the end of treatment, while an additional 2,013/14,689 (13.7%) showed reliable improvement but did not fit criteria for recovery, 7,499/14,689 (51.1%) experienced no significant change, and 899/14,689 (6.1%) had deteriorated. Individual rates of recovery, reliable change, no change, and deterioration are shown in table 10. Recovery rates for those in all eight possible combinations of the phenomena (no criteria, only PI-50, only SG, only RR, PI-50 + SG, PI-50 + RR, SG + RR, PI-50 + SG + RR) were significantly different ($\chi^2 = 1871.96$, $p < .001$). In seven of the eight independent groups (excluding those who experienced only a RR and no other phenomena), a majority of clients recovered by the end of treatment, all 8 independent groups had at least 70% reliable improvement or recovery, and none had over a 5% deterioration rate (Table 10).

Table 10

Recovery Rates Among RR, SG, PI-50 (Non-overlapping Groups)

		PI-50 only	RR only	SG only	PI-50 + RR	PI-50 + SG	RR + SG	PI-50 + RR + SG	None	Total
Recovered	*n*	80	676	431	55	24	131	24	2857	4278
	% recovered	71.4%	46.0%	60.4%	67.9%	68.6%	68.6%	88.9%	23.7%	29.1%
Reliable Improvement	*n*	21	465	92	12	6	31	3	1383	2013
	% recovered	18.8%	31.7%	12.9%	14.8%	17.1%	16.2%	11.1%	11.5%	13.7%
No Change	*n*	9	283	162	11	5	24	0	7005	7499
	% recovered	8.0%	19.3%	22.7%	13.6%	14.3%	12.6%	0.0%	58.1%	51.1%
Deteriorated	*n*	2	45	29	3	0	5	0	815	899
	% recovered	1.8%	3.1%	4.1%	3.7%	0.0%	2.6%	0.0%	6.8%	6.1%

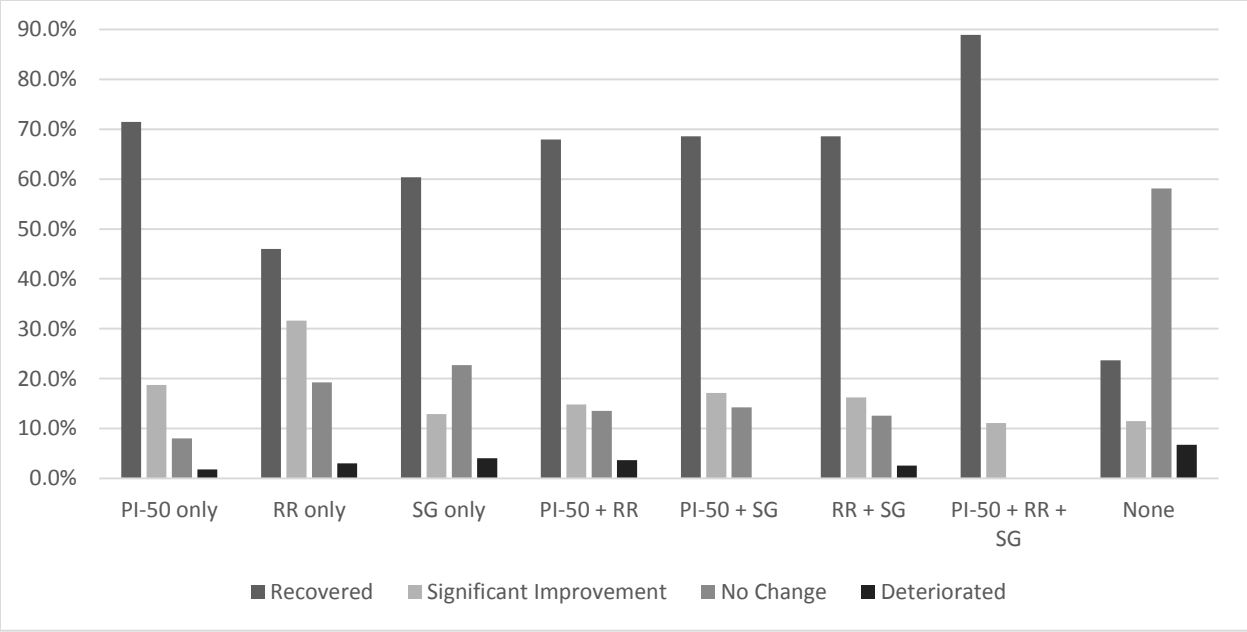

Figure 2. Percentage of Those who experienced each phenomena that recover, significantly improve, or deteriorate

Table 11

Recovery Rates Among PI-50, RR, SG (Overlapping Groups)

		PI-50	RR	SG	None
Recovered	*n*	183	886	610	2857
	% recovered	71.76%	50.11%	63.08%	23.69%
Reliable Improvement	*n*	42	511	132	1383
	% recovered	16.47%	28.90%	13.65%	11.47%
No Change	*n*	25	318	191	7005
	% recovered	9.80%	17.99%	19.75%	58.08%
Deteriorated	*n*	5	53	34	815
	% recovered	1.96%	3.0%	3.52%	6.76%

Those clients who experienced any of the phenomena accounted for a significant percentage of the overall group of those who were classified as "recovered" by the end of treatment (n = 4278) (Figure 2). Those who experienced a RR accounted for 20.71%, PI-50 for

4.28%, and SG for 14.26% of all clients in the sample classified as recovered at the end of treatment.

Testing RR that Occurred All Throughout Treatment

A secondary analysis was completed looking at clients who experienced a RR any time throughout their treatment instead of just the first three sessions. In this sample 14,694 clients were analyzed from the same database as the primary analysis and 3,549/14,694 (24.15%) experienced a RR at any time throughout treatment. Comparing all clients in the database and those who experienced a RR, 610 clients (17.18%) experienced both a RR and SG. Conversely these 610 clients who experienced both a SG and a RR throughout treatment accounted for almost half (48.28%) of all who experienced a SG in the sample. These overlap rates between those who experienced a SG who also experienced a RR were much higher than when examining RR that occurred in just the first the first three sessions. However, even at these higher rates of overlap, more than 50% of those who experienced a SG still did not experience a RR. This indicates that significant differences still exist between phenomena, even when the criteria for RR is expanded in timeframe.

A comparison in recovery rates was made using RR criteria that occurred any time in treatment rather than just the first three sessions. Of those who experienced a RR any time during treatment 1860 (52.41%) recovered. In addition, those who experienced a RR any time throughout treatment and were classified as “recovered” accounted for 43.47% of all the clients in the entire sample who were classified as “recovered” by the end of treatment. An examination of clients who experienced both a SG and RR and met criteria to be “recovered” at the end of treatment yielded 422 clients accounting for 22.69% of all those who experienced a RR any time throughout treatment and met criteria for recovery. Conversely, these 422 clients accounted for

52.95% of all those who experienced a SG and were classified as "recovered" at the end of treatment, a portion of the population much higher than when examining RR in only the first three sessions. This is further evidence that more similarities lie between RR and SG when RR criteria is expanded to any time throughout treatment but that there are still significant differences in those who experience each of the phenomena and are classified as "recovered" at the end of treatment. Further, when comparing RR that occur any time in treatment the percentage of clients who experience a RR that recover is larger. This somewhat challenges notion that "early" is necessary criteria for best outcomes with the RR method.

Discussion

The purpose of this study was to compare three methods for operationalizing unexpected dramatic treatment response and analyze the degree of overlap between methods. Similarities and differences between methods with regards to base rates of occurrence will first be discussed, followed by an examination of differences in outcomes of identified clients. Interpretations and implications of the results will be explored in the context of past research and clinical practice. Limitations of the study will then be examined while differences in particular of methodology between SG, PI-50, and RR will be highlighted followed by a conclusion section which attempts to summarize the study and suggest future research directions.

Similarities Between Phenomena

Phenomena are predictive. Despite differences between method base rates, clients who experienced any of the three phenomena had better treatment outcomes than those clients who did not meet criteria. This is in line with research literature on the topic. In these past studies, better outcomes were shown at the end of treatment and even up to two years after the end of treatment across all three phenomena. The primary RR study referenced (Haas et al., 2002)

demonstrated fewer psychological symptoms at termination of treatment and follow-up with a maintenance of the unexpected dramatic response seen during treatment. The reference PI-50 study (Renaud et al., 1998) demonstrated similar predictions of better outcomes at the end of treatment, one year, and two year follow-up periods. Lastly, a meta-analysis of 16 SG studies (Aderka et al., 2012) demonstrated moderate effects on outcome measured at post treatment and follow-up in CBT and non-CBT studies that used RCT and ITT analyses, specific symptom measures and general measures of distress, and across a variety of different diagnoses and treatment settings.

It might be said that "all roads lead to Rome" in the case of treatment recovery concerning the three methodologies used to identify unexpected dramatic treatment response. In this study, meeting criteria for a PI-50, RR, or SG gain was tied to higher recovery and reliable change at the end of treatment. Those clients not identified by any of the three methods had a recovery rate of 24%, with an additional 12% reliably improving but not recovering. In contrast, 73 to 90% improved or recovered if identified by any of the methods. This is evidence that despite differences in methodology base rate of identification, SG, PI-50, and RR may all be promising methodologies for use as predictors of recovery or improvement.

Interestingly, in our secondary analysis we observed that a similar rate of recovery was seen when comparing those who experienced a RR in the first three sessions and those who experienced a RR any time throughout therapy (50% vs 52%). However, differences were seen in the *number* of the total sample of clients classified as "recovered" at the end of treatment captured by the RR method. In the more restricted first three sessions method, 20.71% of those who were classified as "recovered" at the end of treatment experienced a RR.

In contrast, when expanded to allow RR any time throughout therapy, 43.47% of those who were classified as "recovered" at the end of treatment had experienced a RR sometime during treatment. In addition, more overlap with those who were classified as "recovered" and had experienced a SG was seen with this expanded criterion. When in the secondary analysis criteria was expanded to include RR any time in treatment, of those who experienced a SG just over half of the clients (52.95%) also experienced a RR and were classified as "recovered" at the end of treatment. This may be an indication that more overlap with SG exists between RR that occur throughout treatment than in those that occur in the first three sessions of treatment. Nevertheless, the overlap between SG and RR across all sessions is not total and a vast majority of those who experienced a RR any time throughout treatment did not experience a SG suggesting that even in this condition RR criteria is the least restrictive in identifying unexpected dramatic treatment response and that significant differences between SG and RR exist.

Overlap between PI-50 and other phenomena. This analysis looked at clients who experienced more than one phenomena to identify overlapping cases. These overlapping cases would demonstrate similarities between methodologies used to capture unexpected dramatic change in psychotherapy. While the overlap overall is low between those who experienced RR, PI-50, and SG respectively there seems to be some overlap within groups that is of interest. In those who experienced a PI-50, over 40% also experienced a RR (42.35%) and almost a quarter also experienced a SG (24.31%). Some of this overlap may be explained by PI-50 being experienced by a considerably smaller amount of clients when compared to SG and RR and as such, taking a narrower view of unexpected dramatic treatment response. Therefore, this narrower sample of those who experienced a PI-50 may somewhat overlap with RR's more liberal identification. In addition, RR and PI-50 share a common criterion of needing to see

unexpected and dramatic response by session three of treatment. Due to this shared methodology criteria and the possibility PI-50 may be examining a narrower range of clients, the over 40% overlap may be expected.

Differences Between Phenomena

Occurrence rate. In this study, vastly different numbers of clients were identified based on operational criteria for each method. Applying PI-50, RR, and SG methodology to the same database, a relatively larger sample of individuals met criteria for RR (n = 1768) than other methods: SG (n = 967) PI-50 (n = 255). It should be noted that the percentage of the population identified is much lower in this study than past RR and SG studies. For example, the meta-analysis by (Aderka et al., 2012) with past SG studies have shown the SG phenomenon present in anywhere from under 15% (Hofmann et al., 2006) to over 52% (Doane, Feeny, & Zoellner, 2010) of the samples examined with a large variance from study to study. Stiles et al. (2003) looked at a community sample similar to that of this study, that utilized multiple diagnoses and treatment modalities like this study, and used a similar outcome measure to the OQ-45, the CORE-SF. They identified SG in 17.8% of their sample. This is still over twice the percentage of the 967 (6.58%) clients who identified as experiencing a SG in this study. Further, in Haas et al. (2002) exactly 1/3 of the sample experienced a RR, which is significantly higher than the just over 12% identified in this study. When in our secondary analysis, we examine RR across all sessions we do get a larger sample (24.15%) that experience a RR that is somewhat closer to the 1/3 seen in Haas et al. (2002).

While the reasons for these dissimilar occurrence rates are unknown they do offer some points of comparison between unexpected dramatic treatment response methodologies. These dissimilar occurrence rates suggest that that RR was the most liberal indicator of unexpected

dramatic response even though it was based on just the first three sessions of care, while PI-50 was by far the most conservative with SG in the middle.

In particular, PI-50 only being experienced by less than 2% of the sample, may pose limited usefulness as a predictive measure or clinician feedback tool due to its unlikeliness of observation. This is likely correlated with the second requirement that 25% of the entire range of the instrument take place to experience a PI-50. In the OQ-45 this amounts to 45 points of change. Further, the clinical cutoff score from instrument to instrument falls at a different proportion of the scale. As such, our limited occurrence rate for the PI-50 may be due to this discrepancy and as such as Hiller et al. (2012) suggests occurrence rate of PI-50 should likely not be compared across measures.

One possible future area of exploration that may enhance the utility of PI-50 as a method of identification of unexpected dramatic treatment response early in treatment are the exploration of the changing of the second criteria requiring 25% of the entire instrument. This criteria may be change to be 25% of the entire clinical range or even different percentages examining if more clients are identified while still maintaining its high correlation with treatment outcomes.

Overlap rates. A main focus of this study was to examine the similarity between SG, PI-50, and RR by examining how many of the same clients would meet criteria for experiencing one or more of the presumably overlapping phenomena when applied to the same sample and using the same outcome measure. If a high percentage of the same clients met criteria for two or more of the phenomena, it would support uniformity and similarity in the practical application of these concepts, whereas a low percentage of overlap suggests the opposite.

It was surprising how few of those who experienced each of the unexpected dramatic response definitions overlapped with others. A large majority of those who experienced a RR

(83.09%) did not meet criteria for a SG or PI-50 while a large portion of those who experienced a SG (73.84%) did not experience a RR or PI-50 phenomena. When identifying clients who experienced all three of the phenomena, a tiny number of clients qualified (n = 27). This only accounts for 0.18% of the total sample and 1.03% of those who met criteria for having experienced one or more of the PI-50, RR, and SG phenomena. If PI-50, SG, and RR were all truly similar in identifying unexpected dramatic treatment response, we would likely observe many more clients who experienced all three phenomena. Accordingly, the lack of overlap suggests that RR, SG, and PI-50 are accounting for vastly different phenomenon despite their similar goal to attempt to identify unexpected dramatic treatment response.

Of particular note, a vast majority of those who experience a SG or RR did not experience any other phenomena illustrating a particularly large dissimilarity between these two methodologies. This lack of overlap between SG and RR is likely the most important finding of this study. Careful effort was taken to use the criteria on the same database of individuals with the same measure and similar, naturalistic treatment setting to find clients who experienced multiple phenomena rather than merely to compare incident rates across studies yet differences persist and relatively little overlap was seen among clients. This is evidence in practical application methodology demonstrating differences rather than only comparing similar theoretical underpinnings of unexpected dramatic treatment response.

Differences in criteria between methods may account for lack of overlap between the clients who experienced each phenomena. For example, both PI-50 and RR require unexpected dramatic change to be observed in the first three sessions of treatment. In contrast, SG methodology has a higher requirement for sustaining gains since they have to occur for at least two measurements post gain when compared to the two measurements pre gain. This pre-gain

requirement alone in the methodology suggests that it may be difficult for many who experience a SG to meet either the PI-50 or RR early treatment requirement.

The secondary analysis that examined RR occurring throughout the entire treatment and compared them with the RR that occurred in the first 3 sessions supports these observed differences. In the analysis using the less restrictive RR criteria that occurred any time in treatment, it was found that over 48% of those who experienced a SG also experienced a RR. While this is a higher amount of overlap with SG, a majority of those who experienced a SG still did not experience a RR even with these more loose criteria, while the expanded RR criteria also identified many more respondents that did not experience a SG than those that did. As such, even with the expanded criteria allowing a RR throughout treatment, SG and RR still show significant differences in the clients identified that experience them.

These findings indicate, therefore, that despite a lack of total synchronicity between all three phenomena, there is some overlap between two phenomena, such as PI-50 and RR that may be due to PI-50 methodology's more conservative approach. However, while there are definite similarities in outcome and in the way each of the phenomena is theoretically conceptualized as capturing unexpected dramatic treatment response, the lack of high overlap in clients identified using SG, RR, and PI-50 methodology suggests that each phenomena is significantly different from one another. Therefore, caution may be warranted when using SG, RR, and PI-50 as interchangeable operationalizations for those who experience unexpected dramatic treatment response.

Comparing Differences in Descriptive Statistics

Without exception, experiencing a RR, SG, or PI-50 predicted differences on every descriptive statistic examined and compared in our analysis. These differences were seen in

beginning of treatment distress level, end of treatment distress level, and total change of distress levels. As we have observed that there is at least some undoubtable differences between each of the three phenomena, some helpful illuminations might be taken from examining available descriptive information concerning the clients who experienced each of the phenomena.

Beginning of treatment distress. Significant differences were seen in beginning of treatment distress levels between phenomena. These differences were not only significant but varied widely. Interestingly, when looking at independent groups, any combination of those who experienced a SG or RR without a PI-50 were grouped in the same population of starting distress levels with those who experienced none of the three phenomena. Also, the regression model suggested that those who experienced a SG did not differ from the population mean in beginning OQ-45 distress. In contrast, experiencing a PI-50 had an additive effect over the overall population mean of over 2 times the OQ-45 points than that of RR in beginning distress levels. This data suggests that those who experienced a SG may differ little from those who experience none of the phenomena on beginning distress scores while some who experience a RR and most of those who experience a PI-50 start with significantly higher distress scores.

Some of these differences may even be due to the operational definition in the concepts behind each phenomenon. For example, RR is largely concerned with gains that exceed *expected* change with an emphasis on beginning distress scores (Haas et al., 2002). In contrast, SG creators defined their operational definition of the phenomenon as "sudden, substantial, and stable improvements that occur in one between-session interval" (Busch et al., 2006; Tang & DeRubeis, 1999b). Thus it is possible that clients who experienced a RR had significantly higher beginning of treatment OQ-45 scores in the regression model due in part to conceptual focus on beginning OQ-45 scores in defining its methodology. Accordingly, beginning of treatment

distress levels are a descriptive characteristic that may illuminate drastic differences between these supposedly similar phenomena despite their conceptual similarities.

The fact that those who experienced a PI-50 had higher initial scores on the OQ-45 is a more confusing finding. Hiller et al. (2012) purports that a benefit of the PI-50 criteria in identifying reliable change is its adjustment to initial severity level: "probably the most important advantage of the PI method that response requires more improvement (as compared to using the RCI) in severely ill persons than in those with only mild or moderate signs of psychopathology" (p. 8). Using this reasoning would suggest that those who have higher distress levels would be less likely to experience a PI-50, but the opposite is seen in this study.

However, one possible explanation is that of the number of clients who experienced a PI-50 was a more conservative sample of the unexpected dramatic treatment response phenomenon due to criteria 2 in the PI-50 methodology. This criterion requires a change of 25% of the entire range of the scale to qualify. It is strictly in place to discourage false positive identification of those clients near the cutoff line of being significantly distressed on the outcome measure to be able to improve a few points and qualify as experiencing a PI-50 by meeting the 50% reduction criteria between their initial score and score on the session measured. As a side-effect, this second PI-50 criteria may lead sampling of those who experienced a PI-50 to be higher in reported initial severity by excluding those whose initial OQ-45 were close to the cutoff line for being clinically distressed. Overall, this may be an indication that the PI-50 methodology seems to be much more sensitive to particular scaling of the outcome measure used in its calculation likely because of its use of percentages for calculating its criteria and the arbitrary choice to use "50%".

End of treatment distress. Differences were also seen in end of treatment distress scores between those who experienced each of the phenomena. Interestingly, experiencing any of the three phenomena predicted lower end of treatment distress levels in the regression analysis, even after controlling for initial OQ-45 scores. When comparing the eight different independent combination groups of SG, RR, and PI-50, only two populations emerged: population one contained clients who experienced none of the phenomenon and clients who experienced only a RR and not a SG or PI-50. Population two contained any clients who experienced any combination of RR, SG or PI-50. In addition, when examining the regression analysis, having experienced a PI-50 or SG was worth over twice as much in reduction of end of treatment distress as compared to RR. Those who experienced all three phenomena had the lowest mean end of treatment distress levels (at the mean of the non-dysfunctional normative sample) and all groups in population two did not differ from one another in this aspect.

One possible explanation of this may be the sample size of the two groups. Both the group of individuals who experienced none of the phenomenon and those who experienced only a RR were considerably larger than the other six independent groups. As the number of clients who experienced a RR was much higher than the other phenomena, it may be that the RR methodology is a more liberal assessment of unexpected dramatic treatment response, encompassing a wider swath of individuals than the other two phenomena, with the effects on end of treatment distress are much closer to the overall population mean.

Nevertheless, this discrepancy suggests that when using RR as a predictive tool, it is more likely for a client to experience a RR in the sample, but less reliable about predicting significant differences in end of treatment distress scores, especially when the RR does not co-occur with one of the other phenomena. This again offers more evidence that SG, PI-50, and RR

phenomena are significantly different from one another in important aspects that will be key to its future uses as a clinical feedback and research tool for unexpected dramatic treatment response.

Total change. Differences were seen in total change on the OQ-45 between those who experienced any combination of SG, RR, and PI-50 and the rest of the population. This difference was large and the distance between the mean of those who experienced none of the three phenomena (9.82) and those who experienced a RR only and not a SG or PI-50 (12.70), the next highest of the eight independent groups, was over 14 OQ-45 points. Further, when accounting for beginning OQ-45 scores, the baseline total change score is reduced to below 4 OQ-45 points (3.81) in the regression model, indicating that experiencing any of these phenomena has a significant additive effect on the total change that occurs during treatment even when accounting for beginning distress levels. In the analysis, true additive effects are seen in that generally the larger the amount of total change, the more of these phenomena that were experienced in combination with one another. Consequently, those who experienced a SG, RR, and PI-50 experienced on average over 57 points in change on the OQ-45 (see Figure 1 in results above) as opposed to the 9.82 points of change experienced by those who did not experience any of the phenomena, a difference of over 47 points.

It is notable that according to the regression model, experiencing a PI-50 was worth more than double additional points on the OQ-45 over the group mean for total change when compared to a RR (28.16 vs 12.71). This may challenge the notion that early occurrence of the change is required for prediction of the largest improvement during therapy as both RR and PI-50 in this analysis had to occur in the first three sessions of treatment to be included. However, it should be noted the PI-50 criteria again is the most exclusive of the three methodologies

studied, identified the least amount of clients, and started treatment with the highest distress scores. Further studies may need to examine if the restrictive nature of PI-50 methodology is correlated to higher amounts of change over the course of treatment in those who experience a PI-50.

Dosage statistics. With only a few exceptions, a high amount of overlap was found between those who experienced a SG, RR, and PI-50 and those experienced none of the phenomena on the number of weeks in treatment and in the number of sessions attended. This was particularly true of independent groups containing some combination of experiencing a PI-50 or RR phenomenon. This offers some further evidence that dropping out of therapy may not be an accurate indicator of treatment failure, as those who experienced RR and PI-50 had much higher recovery rates than the overall population mean, despite receiving a similar dosage of treatment.

However, the analysis models that were run did note that experiencing any of the phenomena study was associated with differences in weeks in treatment and sessions attended. This outcome likely has to do with those who experienced a SG and not a PI-50 or RR. This independent group attended far more weeks of treatment compared to clients who experienced some combination of a RR or PI-50. This may be indicative of the difference between SG and the other two phenomena's requirement to occur "early" in treatment. Further, in the regression model, experiencing a RR or PI-50 was associated with number of weeks in treatment and sessions attended that were similar or significantly less than the overall group mean. In contrast, experiencing a SG was worth an additional 49 weeks in treatment or attending over 12 additional sessions. Interestingly, those who experienced a SG in combination with either a RR or PI-50 did not differ on post hoc analysis from the other groups in weeks in treatment or sessions

attended; further evidence on the contrast between "early" criteria of RR and PI-50 with SG ability to occur throughout treatment.

Considering the session restriction in the PI-50 and RR phenomena criteria stating that they needed to occur in the first three sessions of treatment, experiencing either a PI-50 or RR may be associated with dropping out sooner from therapy than those who experience a SG. Indeed we know from other research that after experiencing clinically significant change, clients tend to drop out of treatment (Kadera, Lambert, & Andrews, 1996). Thus, while experiencing a SG in this study is still predictive of high recovery rates, many clients who experience them may be in treatment longer and for more sessions until they experience the SG, their distress levels lower considerably, and then treatment is discontinued.

Use of RR and SG in Examining Treatment Components

SG, RR, and PI-50 may be used to examine effectiveness of treatment components, but their effectiveness in doing so may not overlap as many differences were seen between phenomena. Of particular note, the SG method seem to be used in past research to attempt to fully examine effective or causal treatment components that lead to unexpected dramatic response. This may be due to unique identification tool of SG vs PI-50 or RR. Whereas the latter two phenomena's criteria are based upon initial distress score and deviation therefrom, the SG criteria focuses on unexpected dramatic change in a single intersession period. While this intersession period occurs more often early in therapy, it can and has been observed, including in this study, to occur throughout therapy.

Further, in SG literature SG have been repeatedly used in an effort to try and offer evidence of cognitive change or other treatment components positively effecting treatment components (Aderka et al., 2012). As such, Aderka et al. (2012), Tang and DeRubeis (1999b),

and others have theorized the single intercession period measured in a SG capture the beginning of an "upward spiral" positive feedback loop of behavioral changes in the client signifying improvement. Others have purported SG as part of gradual change pattern leading to positive outcomes regardless of the techniques that are offered (Thomas & Persons, 2013). In summation, that SG criteria focus on a single intercession period may make it more appropriate for use at studying effective treatment components.

Evidence of this flexibility may be seen in this study in that those who experience a SG are much more likely to be in treatment longer or attend more sessions but still have a similarly high recovery rate. Further, in our secondary analysis we saw higher but still only about 50% overlap between clients who experienced a SG and a RR any time in therapy. As such, this offers evidence of some utility of SG over RR in studying these specific periods of time where unexpected dramatic improvement occurs from one session to the next and is associated with high recovery rates.

In contrast, both RR and PI-50 occur so early in the treatment process that by definition they may precede any theoretical component that may contribute to change (Haas et al., 2002). Further, Lambert (2005) cites RR research among other research occurrences of "early" unexpected dramatic treatment response at challenging the need for specific components of therapy and suggests that RR may be indicative of the power of common factors or client factors in successful treatment process and outcome. The RR phenomon is based on outperforming other clients with identical initial OQ-45 scores. Thus, RR may allow comparison of these client characteristics that may contribute to high recovery rates which may be present in those who experience a RR when they begin treatment and before treatment components can take effect. Regardless, it is clear that PI-50 and RR, when observed in the first three sessions of therapy, are

predictive of success despite lack of theoretical underpinnings to explain their existence as a predictive phenomenon. Further research in this area may lead to increased knowledge of what makes a successful therapeutic process aside from treatment components.

Limitations

While many of the findings and associated characteristics of those who experience a PI-50, RR, and SG are similar to past studies, a main component of difference between this and historical studies was the low occurrence of the phenomena in the overall sample. Of the overall sample, 6.6% experienced a SG. This is significantly lower than historical SG studies that use an RCT format and even lower (6.6% vs 17.8%) than Stiles et al. (2003), a SG study that examined a non-RTC community sample and used a similar general distress measure. Likewise, comparing those who experienced a PI-50 in this study with our historical example (Renaud et al., 1998), we see that our study identified a dramatically smaller portion of the clients as having experienced a PI-50 (1.7% vs 31.0%). However, Renaud et al. (1998) is an example of an RCT study that only included clients who met depression criteria as opposed to our clinically distressed but diagnosis non-verified sample. Lastly, those who experienced a RR were the most identified out of any portion of the sample (12.0 %) but still less than when looking at the historical study the criteria was based upon that used a similar ITT outpatient university population (Haas et al., 2002).

These differences in occurrence rate may be associated with differences in populations of clients used for previous studies. Of the SG studies reviewed above, none looked at a college counseling center setting. Similarly the PI-50 studies reviewed were conducted in either ITT controlled outpatient research settings (Hiller et al., 2012) or an RCT for treatment of adolescents (Renaud et al., 1998). However, this study used a college counseling center data. Previous work

done by Lambert et al. (2011) demonstrated that once low scoring non-clinical clients are removed, as they were for this study, it represents the same level of disturbance found in outpatient therapy and clinical trials. As such, while the clients for this study may have unique, unknown characteristics of treatment that contribute to the considerably lower occurrence rates than historical studies, it most likely represents a typical large ITT outpatient sample.

One may suspect that another limitation of this study may be seen in the adaptation of these phenomena for use with the OQ-45. However, this limitation may not hold weight as the SG was originally tested for use with the BDI (Tang & DeRubeis, 1999b) and also shown use with the CORE inventory (Stiles et al., 2003); both measures which has been shown to have a high correlation with the OQ-45 (Lambert et al., 2011). Further the PI-50 was originally used with the Hamilton Depression scale but later adapted for use with the BDI (Hiller et al., 2012). So, while it is possible that some of the unique characteristics seen about each phenomenon in this study may be due to an adaptation of SG and PI-50 criteria to measures that they were not originally tested with the high correlation between measures likely eliminates most of this variance.

However, some concern with the PI-50 methodology exists due to its use of simple percentages of the OQ-45 distress scale are used in calculation. For comparison the BDI has a range with a top maximum at 60 with scores 14 and over in the clinical range offering a 46 point span of which the PI-50 percentages are taken. Conversely, the OQ-45 starts the clinical range at 64 and goes up to 180 offering a clinical range of 116 points. However, most OQ-45 scores fall below 100. As a result, Hiller et al. (2012) and others have noted, that the PI-50 method is likely more sensitive to the unique characteristics of the outcome measure used and that comparisons of PI-50 scores across measures should be interpreted with caution.

RR, as conceptualized in this study, was originally tested for use with the OQ-45. Therefore, we may expect that this is likely not a limitation of the RR method.

Some might also consider the use of a naturalistic sample to be a limitation of this study. It is true that some limitations of this dataset may be present, especially in the area of background information and diagnosis. However, caution is warranted against only studying any unexpected dramatic treatment responses in RTC or completer samples. When discussing the concept of those who respond to therapy quickly and dramatically, Haas et al. (2002) stated the following,

> Unfortunately studying *only* patients who have longer treatment durations could result in missing the phenomenon because in routine clinical practice (as opposed to clinical trials research) patients tend to drop out of treatment soon after reaching criteria for clinically significant change (Kadera et al., 1996). Thus, examining only longer treatment durations would mean studying a preponderance of patients who do not show an early treatment response.

Further, one future use of SG, RR, or PI-50 methodology may be to use the presence of the phenomenon as a predictive tool of good outcomes or a feedback tool for clients or clinicians. As such, restriction to only RTC settings may give unrealistic expectations of its functionality in most clinical settings.

Conclusion

In summation, this study suggests that there are significant differences between SG, RR, and PI-50 as methods for operationally defining unexpected dramatic treatment response, caution should be used when referring to SG, PI-50, and RR as the same phenomenon or interchangeable terms for unexpected dramatic treatment response. Experiencing any of the three phenomenon

was associated with higher recovery rates than the population norm despite their theoretical and methodological differences. While all three methods identify clients who are likely to have high recovery rates, differences abound in both which clients experience each of the phenomena and demographic characteristics of those clients.

Experiencing a PI-50 seemed to predict the highest recovery rates, highest starting distress score, and largest total change during treatment while also reflecting the lowest weeks in treatment, sessions attended, and end of treatment OQ-45 scores, but it identified only a small, even trivial, number of clients. In contrast, those who met criteria for a RR were associated with moderate recovery rates and end of treatment OQ-45 scores but identified substantially more clients who went on to recover. Lastly, SG demonstrated a strong tie to high recovery rates but included clients all throughout the treatment process rather than just in the first three sessions. Clients who experienced a SG were on average in treatment for significantly longer amounts of time, attended more sessions and did not differ from the overall group mean on starting distress scores. The SG and RR methods identified significantly more clients than the PI-50 method, which may limit the PI-50 methods usability as a prediction feedback tool. When comparing those who experience a RR any time throughout treatment instead of just the first three sessions, overlap in percentage of "recovered" clients identified is higher, but not 100%, with "recovered" clients who experience a SG signifying significant differences between RR and SG.

These results tying SG, RR, and PI-50 to significant rates of recovery and positive treatment change suggest possible future use as a predictive feedback tool for clients and clinicians alike to be better able to examine the effectiveness of treatment components during treatment. As such, clinicians looking for a predictive tool during treatment who observe either a SG, RR, or PI-50 in their clients may be able to use the presence of the phenomenon as an

indication of a higher likelihood of positive treatment outcomes. As predictive tools have become more commonly pursued by researchers in an effort to give clinicians predictions of treatment outcome (Newnham & Page, 2010) and evidence continues to mount suggesting that they can be effective in providing useful feedback in other contexts such as predicting negative treatment outcomes (Shimokawa et al., 2010), the exploration of use of RR, PI-50, and SG as predictive methods becomes more relevant. Further, RR, PI-50, and SG phenomena can be calculated and identified during the course of treatment and easily calculated by hand or by being built into the software of the measure being used to track treatment progress. This will likely help clinicians identify clients who have unexpected dramatic change in real time, while helping reduce human error that has been observed in past studies that ask clinician to identify clients who experience these phenomena (Davies et al., 2006). Evidence in this study offers preliminary hope that further research in use of SG, RR, and PI-50 in this predictive manner may be fruitful in increasing treatment effectiveness.

SG, RR, and PI-50 may be used to examine effectiveness of treatment components, but their effectiveness in doing so may not overlap as many differences were seen between phenomena. SG may be more appropriate for examining treatment components whereas RR challenges the notion that treatment components are necessary to experience an unexpected dramatic response and high recovery rates. As such, further research exploring the etiology of these phenomena and the validity of some of the more arbitrary criteria may help clinicians understand how to facilitate more unexpected dramatic treatment response to occur and as such improve treatment outcomes. Further, future research about what causes unexpected dramatic change to occur and what client characteristics may also contribute to future improvements in mental health treatment and the understanding of the mechanisms of change in psychotherapy.

References

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93-101. doi: 10.1037/a0026455

Andrusyna, T. P. (2008). *The mechanism and time course of sudden gains in behavioral activation treatment for depression*. Evanston, Illinois: Northwestern University.

Andrusyna, T. P., Luborsky, L., Pham, T., & Tang, T. Z. (2006). The mechanisms of sudden gains in Supportive-Expressive therapy for depression. *Psychotherapy Research, 16*(5), 526-535. doi: 10.1080/10503300600591379

Beck, A., & Steer, R. (1987). *Manual for the Revised Beck Depression Inventory.* San Antonio, TX: Psychological Corporation.

Beck, A., Steer, R., & Brown, G. (1996). *Manual for Beck Depression Inventory II (BDI-II).* San Antonio, TX: Psychology Corporation.

Busch, A. M., Kanter, J. W., Landes, S. J., & Kohlenberg, R. J. (2006). Sudden gains and outcome: A broader temporal analysis of Cognitive Therapy for depression. *Behavior Therapy, 37*(1), 61-68. doi: 10.1016/j.beth.2005.04.002

Cavallini, A. Q., & Spangler, D. L. (2013). Sudden gains in cognitive behavioral therapy for eating disorders. *International Journal of Cognitive Therapy, 6*(3), 292-310. doi: 10.1521/ijct.2013.6.3.292

Davies, L., Leach, C., Lucock, M., Stiles, W. B., Iveson, S., & Barkham, M. (2006). Therapists' recall of early sudden gains in routine clinical practice. *Psychology and Psychotherapy: Theory, Research and Practice, 79*(1), 107-114. doi: 10.1348/147608305x41209

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose–response in low intensity psychological interventions. *British Journal of Clinical Psychology, 53*(1), 114-130. doi: 10.1037/t06165-000

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658-670. doi: 10.1037/0022-006X.74.4.658

Doane, L. S., Feeny, N. C., & Zoellner, L. A. (2010). A preliminary investigation of sudden gains in exposure therapy for PTSD. *Behaviour Research and Therapy, 48*(6), 555-560. doi: 10.1016/j.brat.2010.02.002

Fennell, M. J., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research, 11*(2), 253-271. doi: 10.1007/bf01183269

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy, 8*(4), 231-242. doi: 10.1002/cpp.286

Flückiger, C., Holtforth, M. G., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research, 23*(1), 1-13. doi: 10.1080/10503307.2012.693773

Gaynor, S. T., Weersing, V. R., Kolko, D. J., Birmaher, B., Heo, J., & Brent, D. A. (2003). The prevalence and impact of large sudden improvements during adolescent therapy for

depression: A comparison across cognitive-behavioral, family, and supportive therapy. *Journal of Consulting and Clinical Psychology, 71*(2), 386. doi: 10.1037/0022-006X.71.2.386

Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains. *Journal of Clinical Psychology, 58*(9), 1157-1172. doi: 10.1002/jclp.10044

Hardy, G. E., Cahill, J., Stiles, W. B., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology, 73*(1), 59-67. doi: 10.1037/0022-006x.73.1.59

Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research, 22*(1), 1-11. doi: 10.1080/10503307.2011.616237

Hofmann, S. G., Schulz, S. M., Meuret, A. E., Moscovitch, D. A., & Suvak, M. (2006). Sudden gains during therapy of social phobia. *Journal of Consulting and Clinical Psychology, 74*(4), 687-697. doi: 10.1037/0022-006x.74.4.687

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice, 1*(2), 138-156.

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*(1), 12.

Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough?: A session-by-session analysis of the psychotherapy dose-effect relationship. *The Journal of Psychotherapy Practice and Research, 5*(2), 132.

Lambert, M. J. (2005). Early Response in Psychotherapy: Further evidence for the importance of Common Factors rather than 'Placebo Effects'. *Journal of Clinical Psychology, 61*(7), 855-869. doi: 10.1002/jclp.20130

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy, 3*(4), 249-258.

Lambert, M. J., Kahler, M., Harmon, C., Shimokowa, K., & Burlingame, G. (2011). *Administration and Scoring Manual for the Outcome Questionnaire-45.2*. OQ Measures, LLC.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., & Nielsen, S. L. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11*(1), 49-68.

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasz, C., . . . Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden Losses identified with session reports. *Psychotherapy Research, 23*(1), 14-24. doi: 10.1080/10503307.2012.693837

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., . . . Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology, 73*(5), 904-913. doi: 10.1037/0022-006X.73.5.904

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal Of Affective Disorders, 118*(1-3), 60-68. doi: 10.1016/j.jad.2009.01.019

Newnham, E. A., & Page, A. C. (2010). Bridging the gap between best evidence and best practice in mental health. *Clinical Psychology Review, 30*(1), 127-142. doi: 10.1016/j.cpr.2009.10.004

Renaud, J., Brent, D. A., Baugher, M., Birmaher, B., Kolko, D. J., & Bridge, J. (1998). Rapid response to psychosocial treatment for adolescent depression: A two-year follow-up. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(11), 1184-1190. doi: 10.1097/00004583-199811000-00019

Seggar, L. B., Lambert, M. J., & Hansen, N. B. (2002). Assessing clinical significance: Application to the Beck Depression Inventory. *Behavior Therapy, 33*(2), 253-269. doi: 10.1016/s0005-7894(02)80028-4

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*(3), 298-311. doi: 10.1037/a0019247

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., . . . Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology, 71*(1), 14-21. doi: 10.1037/0022-006x.71.1.14

Strauman, T. J., Vieth, A. Z., Merrill, K. A., Kolden, G. G., Woods, T. E., Klein, M. H., . . . Kwapil, L. (2006). Self-system therapy as an intervention for self-regulatory dysfunction

in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology, 74*(2), 367-376. doi: 10.1037/0022-006X.74.2.367

Tang, T. Z., & DeRubeis, R. J. (1999a). Reconsidering rapid early response in cognitive behavioral therapy for depression. *Clinical Psychology: Science and Practice, 6*(3), 283-288. doi: 10.1093/clipsy/6.3.283

Tang, T. Z., & DeRubeis, R. J. (1999b). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894-904. doi: 10.1037/0022-006x.67.6.894

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in Cognitive-Behavioral Therapy for Depression. *Journal of Consulting and Clinical Psychology, 73*(1), 168-172. doi: 10.1037/0022-006x.73.1.168

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting and Clinical Psychology, 75*(3), 404-408. doi: 10.1037/0022-006x.75.3.404

Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology, 70*(2), 444-447. doi: 10.1037/0022-006x.70.2.444

Thomas, C., & Persons, J. B. (2013). Sudden gains can occur in psychotherapy even when the pattern of change is gradual. *Clinical Psychology: Science and Practice, 20*(2), 127-142. doi: 10.1111/cpsp.12029

Van, H. L., Schoevers, R. A., Kool, S., Hendriksen, M., Peen, J., & Dekker, J. (2008). Does early response predict outcome in psychotherapy and combined therapy for major depression? *Journal Of Affective Disorders, 105*(1-3), 261-265. doi: 10.1016/j.jad.2007.04.016

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of Depression. *Journal of Consulting and Clinical Psychology, 73*(1), 173-182. doi: 10.1037/0022-006x.73.1.173