TRAJECTORIES OF CHANGE IN PSYCHOLOGICAL TREATMENT: ANALYSIS OF
CLIENT-RESPONSE TO TREATMENT IN A UNIVERSITY CLINIC

A Thesis in
Psychology
by
Samuel S. Nordberg

© 2008 Samuel S. Nordberg

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Master of Science

December, 2008
The thesis of Samuel S. Nordberg was reviewed and approved* by the following:

Louis G. Castonguay  
Professor of Clinical Psychology  
Thesis Adviser

Kenneth N. Levy  
Assistant Professor of Clinical Psychology

Melvin M. Mark  
Professor of Psychology  
Department Head

*Signatures are on file in the Graduate School.
ABSTRACT

This study explored treatment response trajectories in a naturalistic population of clients served at a university psychotherapy training clinic. Using data collected through repeated administrations of a treatment outcome, and methods new to the field of psychotherapy research (growth mixture models), this study attempted to model the shape of change of clients through the course of therapy. Specifically, this study explored the possibility that treatment response is better captured when groups or clusters of responders are modeled. In addition, this study examined the way in which pre-treatment symptom severity relates to trajectories of change. Three separate groups of treatment responders were found. Pre-treatment symptom severity was found to differentiate one low-symptom group from two higher-symptom groups but failed to differentiate between the two higher-symptom groups. These results indicate that Growth Mixture Models may be a useful tool for examining treatment outcome data and that pre-treatment symptom severity has limited usefulness in predicting treatment outcome.
# TABLE OF CONTENTS

List of Tables..............................................................................................................................................iv
List of Figures......................................................................................................................................................v

Chapter 1. INTRODUCTION.................................................................................................................................1
  Latent Group Analysis........................................................................................................................................3
  Initial Symptom Severity..................................................................................................................................4
  Hypotheses......................................................................................................................................................6

Chapter 2. METHODS.......................................................................................................................................7
  Participants....................................................................................................................................................7
  Measures.....................................................................................................................................................8
  Procedure....................................................................................................................................................9
  Plan for Analysis........................................................................................................................................9

Chapter 3. RESULTS.........................................................................................................................................13
  Clinically Significant Change......................................................................................................................16

Chapter 4. DISCUSSION.....................................................................................................................................17

References......................................................................................................................................................21
LIST OF TABLES

Table 1. Model Comparison: Goodness-of-fit statistics for one-, two-, and three-group models..........................................................................................................................10

Table 2. Summaries: Slope, Intercept and Membership by Group.................................................15
LIST OF FIGURES

Figure 1. Trajectories: Plotted Intercepts and Slopes for Each Responder Group……………..14
Chapter 1

Introduction

Traditional RCTs and, to a lesser extent, effectiveness studies allow for a causal link to be drawn between average patient improvement and the particular treatment manipulation in the study. Such studies have resulted in robust findings over the past two decades, demonstrating the effectiveness of many different treatments in the lab and in the field. Research seems to have clearly established that therapy works (Lambert and Ogles, 2004). What remains unclear is what the active agents of change are in psychotherapy and how they work. Process research, the study of during-treatment mechanisms of change, has long been overshadowed by efficacy and effectiveness research. Recently, a call has been made by division 12 (Society for Clinical Psychology) of the American Psychological Association for increased attention to aspects of the therapeutic process other than the aggregate impact of a particular therapy at the conclusion of treatment. Of particular relevance to this study is their call for the evaluation of new and different research paradigms capable of “advancing knowledge on the basis of modeling large numbers of realized cases” (Weisz, Hawley, Pilkonis, Woody & Follette, 2000).

One aspect of such research is the study of the way in which clients change during the course of treatment. Such research may offer additional insights beyond those of traditional therapy studies in that, by studying the ways in which clients change during treatment, we may better understand the way in which different treatments impact clients (exploring multiple paths to recovery), as well as predicting a client’s expected rate of change based on membership in one or another group of responders. It may be possible to identify critical interventions by their
session-to-session impact. Clinicians may also be able to determine which type of therapy will work best for a particular type of client based on their past or expected change trajectory. Research has shown that analytical methods used to investigate an individual’s change during treatment have more predictive validity than clinical judgment alone, offering the possibility that actual therapy practice could be improved as it progresses (Hannan, Lambert, Harmon, Nielsen, Smart, Shimokawa, et al., 2005). Indeed, some researchers have already called for methodological modifications to traditional pre-post design in order to further explore during-treatment change; in particular, calling for repeated measurements of outcome variables throughout the treatment (Hayes, Laurenceau, Feldman, Strauss & Cardaciotto, 2007). Research that repeatedly measures an individual’s distress and attempts to answer the question, "is this treatment working for this client?" has been termed "patient focused research" (Howard, Moras, Brill, Martinovitch & Lutz, 1996). Such research has traditionally taken place in naturalistic settings (Stultz & Lutz, 2007, although a recent call has been made for new RCTs with such a methodology, Hayes, Laurenceau, & Cardaciotto, 2007) and has as its hallmark a methodology based on frequent and repeated measurements of symptomatic distress over time.

One way of exploring the way in which clients change during treatment is by allowing a single sample to converge to multiple trajectories of recovery. Most efficacy and effectiveness studies to date have assumed that all of the subjects in their sample come from the same population, and that all subjects vary around a single, aggregate recovery trajectory. Some researchers have called for a departure from this assumption, whereby analyses of outcome data discard the assumption of a single aggregate slope and explore the possibility that the data is better described by multiple slopes, each representing a unique population (Laurenceau, Hayes, and Feldman, 2007). Researchers have found clusters of individuals with distinct patterns of
change that are missed by aggregate models and may better describe and predict an individual's change in therapy (Stultz & Lutz, 2007, Illardi & Craighead, 1994, Tang & DeRubeis, 1999). These preliminary results indicate a need for further exploration of change in therapy; specifically, the need to allow models of change that can capture clusters of treatment responders. The results from investigations using these tools are likely to generate new hypotheses, and eventually theories, of change that could be tested in more controlled settings.

**Latent Group Analysis**

Recent research by Stultz and Lutz (2007) has taken repeated-measures data generated from a managed care company using the COMPASS tracking system (Howard et. al., 1996), and explored the likelihood that treatment-responders cluster into groups. Using growth mixture models (GMM), the authors found that their data was better defined by three separate recovery trajectories than by one, average, trajectory. The majority of clients in their study conformed to a trajectory they termed as *phase model consistent*, a quadratic trajectory with diminishing returns for additional sessions that conformed to the phase model of psychotherapeutic outcome (for a summary of the phase model, see Howard, Lueger, Maling, & Martinovitch, 1993). They also found two additional clusters of responders, *partial rapid responders* who achieved greater change, more rapidly than other clients, and *symptomatically high impaired* clients who began treatment with more elevated symptomology and responded more slowly, although with the same quadratic change. These findings are consistent, at least in part, with a previous study using different methods. Illardi and Craighead (1994) laid out two discrete response patterns based on the rapidity of response to treatment. *Rapid responders*, those who's symptoms fell dramatically
in the first few sessions, seemed to achieve greater post-treatment symptom reduction than *non-rapid responders*, who had a flatter trajectory overall. In addition to demonstrating the need for analyses that allow for more than one group of treatment responder, these findings demonstrate the importance of examining the relationship of initial symptom severity to a client’s change trajectory during therapy.

In the current study, we attempted to replicate Stulz and Lutz’s use of Growth Mixture Models with a clinical population from a different treatment setting; a university-based training clinic for PhD candidates in clinical psychology. In addition, and following the recommendations of Stulz and Lutz, we attempted to use a less-global measure of well-being, and to focus on a particular patient population, namely those suffering from depression. Also following on the suggestions of Stulz and Lutz, we explored the relationship between clients’ initial symptom severity and their trajectory during therapy.

*Initial Symptom Severity*

Data collected in both naturalistic and controlled settings shows that clients respond differently to treatment depending on the severity of their symptoms (for a summary of these findings see Clarkin and Levy, 2004 and Garfield, 1994). Some studies show that clients with higher symptom severity prior to treatment exhibit steeper recovery trajectories (e.g. Sotsky et al., 1991). However, other studies have found the opposite effect; that higher symptom severity is associated with slower response to therapy (see Newman, Crits-Christoph, Gibbons, & Erickson, 2006). Those who have found that higher symptom severity predicted steeper recovery
have argued that their findings represent a regression to the mean of outlying clients, a theory based on a robust and well-supported statistical phenomenon found in many forms of analysis. Those who have found the opposite posit that increased symptom severity correlates with reduced client-efficacy in therapy and that more severely symptomatic clients will have fewer resources to recruit and will respond more poorly to treatment than those with fewer, or less significant, symptoms.

It is possible that these findings contradict each other because they are based on a single trajectory of change for subjects that may be better described by multiple trajectories such as in those found by Stultz and Lutz, 2007 (described above). The subjects from whom these findings are drawn may well be more representative of one or another population. For example, it may be that there are two distinct response curves for highly symptomatic subjects: one in which clients receive tremendous relief of their symptoms through therapy and another in which they do not. By allowing a model to converge to several different pathways to recovery (or non-recovery) it may be possible to better understand the relationship between initial symptom severity and change in therapy. This would be an important extension of the research summarized above as previous studies have not assessed participants before they began treatment to empirically test whether their symptom severity pre-treatment predicted membership in one or another population of treatment responders. As a further extension of previous research, we also believe that it is important to determine whether clients’ improvement captured by GMMs meets the criteria for clinically significant change (Jacobson and Truax, 1991). Over the past two decades, this measure has become the gold-standard for determining change that would be considered significant to a client or clinician.
Thus, the goal of this study was to explore multiple paths to recovery using GMMs to analyze data collected at a university training clinic. Specifically, we were interested in exploring whether multiple groups of treatment responders (and non-responders) existed, and whether they could be differentiated by their baseline symptom severity. We formulated two testable hypotheses:

1) That analyses performed using GMMs on psychotherapy outcome data collected at a university training clinic would result in a best fit model with more than one group.

2) That baseline symptom severity would differentially determine membership in groups of treatment responders.
Chapter 2
Methods

Participants

Data was collected from clients attending weekly individual therapy sessions at the Penn State University Psychological Clinic (PSU clinic) between July of 2002 and July of 2005. The PSU clinic is an outpatient facility serving clients from Penn State University and the Centre County region. Currently it staffs 36 graduate student therapists, 14 licensed clinical psychologist supervisors, one psychiatrist and two nurse practitioners. The PSU clinic treated over 400 clients between August of 2007 and August of 2008.

Of these clients, 158 exhibited symptoms of depression at intake that were one or more standard deviations above the population mean\(^1\) (as measured by the Treatment Outcome Package, detailed below) and therefore met inclusion criteria for the current study. This inclusion criterion was chosen in order to focus our exploration to a particular population, those with elevated symptoms of depression. As the most commonly seen symptom-picture at the PSU Clinic, depressed clients offered the largest single-symptom sample size with which to conduct our analyses. As an additional criterion, clients included in the study attended a minimum of fifteen sessions, to ensure enough data for meaningful analyses of change.

Subjects were mostly female (68.5%), with at least 2 years of college (59%). 81% of subjects were Caucasian, which is typical of the population served by the Penn State Clinic.

\(^1\) We replicated results using .5 and 1.5 standard deviations as our inclusion criteria with no significant differences.
Measure

The Treatment Outcome Package. The Treatment Outcome Package or TOP (Kraus, Seligman, & Jordan, 2005), is a measure of global well-being designed for use in clinical settings and available through Behavioral Health Laboratories (BHL). The TOP is a 58-question core outcome battery used in 33 states, with hundreds of thousands of clients. The TOP meets all of the recommendations of the Core Battery Conference (Horowitz, Lambert, & Strupp, 1997). The measure captures 12 symptom and functional domains: work functioning, sexual functioning, social conflict, depression, panic, psychosis, suicidal ideation, violence, mania, sleep, substance abuse, and quality of life. The TOP reports symptom severity on each of its 12 subscales in terms of standard deviations above or below the population mean (as normed on over 100,000 subjects).

Additionally, the TOP assesses demographics, health, substance use, stressful life events, treatment goals and satisfaction with treatment. It has good test-retest reliability (.76 to .94 for the 12 subscales), sensitivity to change, and high levels of convergent validity with scales such as the Beck Depression Inventory, the Brief Symptom Inventory and the Minnesota Multiphasic Personality Inventory-2. The TOP requires about 5 minutes to complete. Employed by managed health care companies such as Massachusetts Blue Cross Blue Shield, the TOP has also been used at the PSU Clinic as a measure of client improvement, since 2002. As such, it provides an excellent opportunity to examine the shape of change for clients treated in a university training setting.
Procedure

Clients were admitted to the PSU Clinic from multiple sources (e.g. physician’s referral, county referral, self referral). Once admitted, clients were scheduled for an intake interview with a graduate student-in-training. These interviews consisted of a broad, information-gathering discussion about the client, his problems, family, work and significant others, followed by an oral administration of the Structured Clinical Interview for the DSM IV (First, Spitzer & Williams, 1993.) Before their interview, clients filled out the TOP in a waiting area. The measure was then faxed to BHL for scoring and archiving, and then returned to the PSU Clinic and given to the intake-therapist for inclusion in the client’s paper-file.

After intake, clients were assigned to therapists based on the standard practices at the clinic, which account for therapist availability and the appropriateness of the client as a training case. When clients began therapy, they were asked to fill out the TOP before their first and fifteenth session. After the fifteenth session, clients continued to fill out the TOP every fifteen sessions. For this study, clients were seen in the clinic for a minimum of 30 sessions and a maximum of 60 sessions. The data was archived at BHL until requested by researchers at Penn State.

Plan for Analysis

In order to explore the possibility that subjects fell into distinct groups of treatment responders and non-responders, we used Growth-Mixture Modeling (GMM). GMMs allow for the identification of unobserved (latent) distinct groups or classes of individuals with similar
slopes and intercepts. As in mixed-effect models, GMMs also model each individual’s variation around their group’s slope.

This form of analysis results in one or more clusters of individuals, each with a mean slope and intercept, as well as individual variation around each slope and intercept. GMMs have been used with similar data to identify latent classes of treatment responders in recent studies (Stulz & Lutz, 2007). In addition, because class membership is an unobserved variable, estimated by the model, the proportion of subjects in each class is estimated. The slopes, intercepts and class memberships are estimated using maximum likelihood estimation.

It is also possible, using GMMs, to regress an observed variable onto the emergent latent classes in order to examine whether the observed variable reliably determines class membership. We regressed symptom severity at intake onto the latent class variable to explore the possibility that baseline severity determines membership in one or more of the classes.

Our dependent variable was symptom severity as recorded by self-report on the depression sub-scale of the TOP at the first, 15th, 30th, 45th and 60th sessions, entered into our model in months from the start of treatment (i.e. 15th session entered as 3.25 months). Symptom severity on the depression subscale at intake (n=158) was regressed on to the latent classes. Due to treatment dropout, data was missing in the fourth and fifth assessment periods (45th and 60th sessions). Of 158 clients, there were 158 observations in the first, second and third assessment period, 148 observations in the fourth assessment period and 90 in the fifth. As is standard practice in treatment studies with dropout, the last observation was carried forward for all missing data in order to include those who had terminated treatment before the end of the study.

Data was analyzed using mPlus version 3.12 (Muthen & Muthen, 2005). The best-fitting model was determined by starting with a model with one latent class, adding an additional class
and testing for significantly improved fit using three fit statistics: the Lo-Mendell-Rubin likelihood ratio test and the Bayesian Information Criterion (BIC) statistic and the entropy test. The Lo-Mendell-Rubin likelihood ratio test tests the null hypothesis that the data is equivalently explained by a model with one less class than the current model. The BIC accounts for the number of parameters in the model, favoring more parsimonious models; a lower BIC indicates a model that more accurately reproduces the data, accounting for the parameters used. Entropy indicates how well-separated the latent classes are by examining the posterior probabilities that each subject belongs to one group and not the others. Entropy ranges from 0 to 1. An entropy close to 1 reflects good class separation, and indicates that subjects are clearly differentiated into their separate groups.

As mentioned above, we also decided to assess whether the change demonstrated by each of the groups to emerge from our GMM analysis was significant. We calculated clinically significant change using both cutoff points and a Reliable Change Index (RCI), as suggested by Jacobson and Truax’s guidelines.

Cutoff points are markers that delineate healthy populations from patient populations. These cutoffs are based on the normative data on both populations. Scores above the cutoff point are considered representative of a patient population. Scores below the cutoff, are considered healthy. In the case of the depression sub-scale on the TOP, the cutoff is at .78 standard deviations above 0 (the healthy population mean). The RCI is a statistic that accounts for measurement error when determining whether a pre-post change can be considered clinically significant. The RCI for the depression sub-scale on the TOP is 1.23 standard deviations. In order to meet the gold standard for clinically significant change, clients must move from being
above the cutoff to below the cutoff and clients must change enough such that the data cannot be an artifact of measurement error, therefore meeting the criteria of the RCI.
Chapter 3
Results

In support of our first hypothesis, the results revealed that a linear model with three
groups was the best fitting model for the outcome data analyzed. We first compared a model
with one group to a model with two groups. The two-group model was a significant
improvement over the one-group model according to the Lo-Mendal-Rubin test, which resulted
in a p-value of <.0001, and the BIC, which was smaller for the two-group model. The three-
group model was a significant improvement over a 2-group model according to the Lo-Mendal-
Rubin test, which resulted in a p value of .0128 and the BIC, which was, again, lower for the
three-group model. A four-group model had a non-significant Lo-Mendal-Rubin value,
indicating that additional groups added no explanatory power. Lastly, the entropy for the three-
group model was higher than for the two-group model, indicating better separation between
groups (see table 1 for a summary of these statistics).

Table 1. Model Comparison: Goodness-of-fit statistics for one-, two-, and three-group models.

<table>
<thead>
<tr>
<th></th>
<th>1 Group**</th>
<th>2 Groups</th>
<th>3 Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>2562.605</td>
<td>2495.454</td>
<td>2480.835</td>
</tr>
<tr>
<td>Lo-Mendal Rubin*</td>
<td>NA</td>
<td>0.0001</td>
<td>0.0151</td>
</tr>
<tr>
<td>Entropy</td>
<td>NA</td>
<td>0.829</td>
<td>0.859</td>
</tr>
</tbody>
</table>

* NB the Lo-Mendal Rubin statistic reports an exact p value

** Entropy and Lo-Mendal Rubin tests cannot be calculated for a one-group model.
The three specific groups that emerged from the analysis and their recovery trajectories are depicted in figure 1. One group consisted of the majority (111 subjects) of the sample and had an intercept of 1.24 and a slope of -.036, achieving about ½ standard deviation pre-post change. Both the intercept and slope were significant at the .001 level. We term this group *low-symptom responders*.

**Figure 1.** Group Trajectories: Plotted Intercepts and Slopes for Each Responder Group.

The second group had 31 members, an intercept of 3.429 (p<.001) and a non-significant slope of -.018. We label this group as *high-symptom non-responders* and, in our sample, they
appear to represent a significant sub-population of clients who do not get better during treatment. The group’s proportional size (20%) roughly approximates the proportion of clients generally assumed to not improve during treatment (Lambert and Ogles, 2004).

The last group had 16 members, an intercept of 3.859 and a slope of -.260. Both the intercept and slope for group 3 were significant at the .001 level. This group has a highly significant and steep trajectory of change such that clients in this group have an average pre-post change of over 3.5 standard deviations, ending at the healthy population mean of 0. We term this group high-symptom rapid responders.

In partial support of our second hypothesis, our analyses demonstrated that baseline severity reliably differentiated membership in group 1 from groups 2 and 3 (p<.001); however, baseline severity did not reliably differentiate between groups two and three. Table 2 shows the slopes, intercepts and differential group memberships of each of the three latent classes.

**Table 2.** Trajectory Summaries: Slope, Intercept and Membership by Group.

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.237*</td>
<td>3.429*</td>
<td>3.859*</td>
</tr>
<tr>
<td>Slope</td>
<td>-0.036*</td>
<td>-0.018</td>
<td>-0.26*</td>
</tr>
<tr>
<td>Membership</td>
<td>0.972</td>
<td>0.873</td>
<td>0.86</td>
</tr>
</tbody>
</table>

* p<.001
Clinically Significant Change

In our study, the only group that met both of the criteria for clinically significant change was the rapid responders group. The high-symptom non-responders began treatment well above the cutoff between patient and healthy populations but did not change enough to move below that cutoff, or to dismiss the possibility that their change was simply due to measurement error. The low-symptom responder group began treatment in the patient range and ended in the healthy range but did not change enough to satisfy the requirements for Reliable Change.
Chapter 4

Discussion

The goal of this study was to apply growth-mixture modeling to treatment outcome data collected over time in a university training clinic, in order to explore whether allowing for multiple groups of treatment responders (and non-responders) would result in a better-fitting model than allowing for only a single group of responders. In addition, we sought to explore the relationship between pre-treatment symptom severity and change during treatment, in order to help clarify an existing debate in the field.

Overall, the results of our GMM analyses seem consistent with the findings of Stulz and Lutz (2007) discussed above. Both analyses converged to three groups. Our largest group begins treatment at a moderate level of impairment; progresses with a gradual reduction in symptoms and represents 70% of our total sample, consistent with the size (63%) and trajectory found by Stulz and Lutz for their phase model consistent group. The group of high-symptom non-responders appears similar to Stulz and Lutz’s symptomatically high impaired group in that both begin with high impairment, change little throughout treatment, and represent roughly 20% of the sample. The last group, high-symptom rapid responders, appears to be a departure from the findings by Stulz and Lutz in that, while both exhibit steep recovery trajectories, our group appears to begin treatment at a high degree of impairment, while theirs begins treatment at a level similar to the phase model consistent group.

Above and beyond providing a better fit to the data than a single-slope model would, the resulting model seems to offer more information as well. Specifically, this model confirms that a substantial portion of clients in this sample did not change during treatment. For those who improved, the model also clearly indicates that groups of clients changed at different rates during
treatment. The model lastly shows that clients in our sample seem to form two dichotomous initial symptom severity groups, one high-symptom group, composed of the high-symptom rapid-responder and high-symptom non-responder groups, and one lower-symptom group composed of the low-symptom responders. This last observation leads well into a discussion of our second hypothesis, related to initial symptom severity.

Our second hypothesis, that baseline symptom severity would differentiate between members of different groups, was partially supported. When baseline symptom severity was used to determine group membership, we found that it significantly predicted differential membership between the low-symptom responder group and the other two groups such that higher baseline severity predicted a higher probability of membership in either of the high-symptom groups, but did not differentially predict membership between the high-symptom non-responder and high-symptom rapid responder groups. It is notable that the findings with regard to baseline symptom severity seem to support both contradictory findings in the field (1. that high-symptom clients respond poorly to treatment and 2. that high-symptom clients respond more rapidly to treatment).

Importantly, this finding indicates that groups of clients with high initial symptom severity can have different trajectories of recovery. This suggests that the numerous studies showing contradictory findings with regard to the relationship between symptom severity and outcome may have involved different subgroups of clients or different proportions of those subgroups. Assuming that both groups of treatment responders exist, the relative size of one versus the other would influence the results of any aggregate analysis. Indeed, in a mixed-effects model run on these data, with only a single recovery trajectory, baseline symptom severity is negatively correlated with linear slope. One might conclude, then, that our data supports the findings that
increased severity results in more rapid recovery; however that would miss the entire 31-member group of *high-symptom non-responders*.

The criteria of clinically significant change add clarity to our results. The RCI and cutoff statistics reinforce the finding that the *high-symptom non-responder* group did not change meaningfully, as it met neither of those criteria. While the *low-symptom responder* group did not meet the criteria for RCI, it did pass the cutoff between healthy and clinical populations, achieving a “healthy” average depression score by the end of treatment. The results lastly show that the *rapid responders* group demonstrated a level of improvement that met the full criteria for clinical significance, both passing the cutoff into the healthy range and exceeding the requirements for RCI. This reinforces a picture of the three groups where two recover (*rapid responders and low-symptom responders*) and one does not (*high-symptom non-responders*).

This study has a number of limitations. TOP scales have a floor of -1.44 standard deviations. Put simply, this means that clients at lower initial symptom severity have less room for change. Therefore, a client who starts at 1 standard deviation and bottoms out at -1.44 standard deviations can only have a limited impact on the slope of change (-2.44 total change), whereas a client who begins at 4 standard deviations, can have a much larger impact (-5.44 total change). This floor effect may have influenced the trajectory of the *low-symptom responder* group in particular, reducing the negative slope and overall change over time.

As in all studies that use data collected in natural settings, this study traded high external validity for low internal validity. Unmeasured and uncontrolled variables could have accounted for significant differences in our sample. We lacked systematically collected diagnostic information on both therapists and clients and, although we used the depression sub-scale from the TOP in our analysis in order to focus on a particular domain, not all of our subjects may have
met a diagnosis of clinical depression and not all of them may have been primarily treated for such a clinical problem. Therefore, it is impossible to assume homogeneity of treatment across subjects. In addition, the study experienced significant dropout throughout the course of treatment. By carrying forward the last observation, this study may have biased the findings depending on the condition clients were in when they left treatment.

Despite these limitations, using Growth Mixture Modeling to describe treatment outcome data seems to offer several advantages over single-slope models. Allowing multiple groups of responders to emerge from the analysis seems to offer additional insight into how clients change during treatment. Interestingly, while our results replicate (in large part) and extend previous findings, they also raise questions to be explored in future studies. To date, the few models that researchers have generated seem to tell a similar story: the majority of clients appear to be moderately impaired and to gradually improve throughout treatment; some highly-impaired clients do not get better and some get better very quickly. Since symptom severity alone appears insufficient to reliably differentiate between the two high-severity groups, future research should harness the power of GMMs to test the predictive validity of additional variables, using the vast existing literature on variables predictive of outcome as a guide.

In addition, the current body of research using GMMs has almost entirely been conducted using global assessments of a client’s well being. Future research should seek to explore whether assessments with more specificity (such as anxiety, mixed anxiety and depression, or personality disorders) replicate the existing research or whether there are different models for different assessments of clients’ distress.
References


