

Utah State University

DigitalCommons@USU

---

Undergraduate Honors Capstone Projects

Honors Program

---

5-2018

## Modeling Change Trajectories for Mental Health Symptoms and Functioning During Psychotherapy

Elizabeth Wynn  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/honors>



Part of the [Psychology Commons](#)

---

### Recommended Citation

Wynn, Elizabeth, "Modeling Change Trajectories for Mental Health Symptoms and Functioning During Psychotherapy" (2018). *Undergraduate Honors Capstone Projects*. 451.

<https://digitalcommons.usu.edu/honors/451>

This Thesis is brought to you for free and open access by the Honors Program at DigitalCommons@USU. It has been accepted for inclusion in Undergraduate Honors Capstone Projects by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



MODELING CHANGE TRAJECTORIES FOR MENTAL HEALTH SYMPTOMS AND  
FUNCTIONING DURING PSYCHOTHERAPY

by

Elizabeth Wynn

Capstone submitted in partial fulfillment  
of the requirements for graduation with

UNIVERSITY HONORS

with a major in

Psychology  
in the department of Psychology

Approved:

---

Capstone Mentor  
Dr. Rick Cruz

---

Departmental Honors Advisor  
Dr. Scott Bates

---

University Honors Program Director  
Dr. Kristine Miller

UTAH STATE UNIVERSITY  
Logan, Utah  
Spring, 2018

Copyright © ELIZABETH WYNN 2018

ALL RIGHTS RESERVED

### Abstract

Psychological symptoms are routinely measured in clinic settings using self-report surveys to help researchers understand the nature of client progress. Past studies have generally used metrics that compare client scores at two time points (beginning and end of treatment) to classify progress by whether there has been significant improvement or deterioration in their symptom levels. However, contemporary practice often uses more frequent (e.g., weekly) assessment. Thus, methodologies incorporating data from every assessment, such as multilevel modeling, are used to provide more nuanced information about change trajectories. Though there is research on the uses of both methodological frameworks, little research has examined how these two methods can be used in conjunction with one another. In this study, I used secondary data to investigate if and how these two analytic methods can be used to complement one another. Deidentified data from 42 clients at a clinical psychology doctoral training clinic in Virginia were used to evaluate the study question. Assessment measures included the Brief Adjustment Scale (BASE-6), Generalized Anxiety Disorder Scale (GAD-7), and the Patient Health Questionnaire (PHQ-9). For each measure, RCI metrics and clinical significance thresholds were obtained from existing research and clients were grouped according to their pre-to-post treatment RCI and whether they had passed the clinical significance threshold during treatment. Multilevel models were constructed to describe change trajectories for each of these groups. From these models, descriptive and visual output was produced providing a foundation by which to compare results for each group of clients. This study will provide information concerning the nature of client progress across different analytic methods, and will advance a framework for future research in this area of study.

### Acknowledgments

I would like to thank my capstone mentor, Dr. Rick Cruz for his unfailing support and invaluable mentoring throughout the three years we have worked together. I am so grateful for his patience and guidance in my research as well as my educational and career goals. I would also like to thank the Clinical Psychology Doctoral Training Clinic at Virginia Tech University, particularly Haley Murphy and Dr. Lee Cooper, for providing the data for this study and for lending support throughout the analysis process. I am thankful for the Undergraduate Research Fellowship Program and Dr. Scott Bates for introducing me to research early on and stimulating my growth throughout in my undergraduate career. Additionally, I am grateful to many research mentors that have taught me and pushed me to grow intellectually. Finally, I would like to thank my family. To Mom, Dad, Camille, Olivia, Victoria, and Harrison, I do not know where I would be without you!

## CONTENTS

	Page
Abstract . . . . .	ii
Acknowledgments . . . . .	iii
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	vii
Introduction . . . . .	1
Methods . . . . .	6
Measures . . . . .	6
BASE-6 . . . . .	6
The PHQ-9 . . . . .	7
The GAD . . . . .	7
Participants . . . . .	8
Procedure . . . . .	8
Analysis . . . . .	9
Jacobson-Truax Framework . . . . .	9
Multilevel Modeling Across Full Treatment . . . . .	10
Early Detection of Treatment Outcome . . . . .	11
Results . . . . .	11
BASE-6 Analysis . . . . .	11
Jacobson-Truax Classification . . . . .	11
Multilevel Modeling Across Full Treatment . . . . .	13
Early Detection of Treatment Outcome . . . . .	14
GAD Analysis . . . . .	16
Jacobson-Truax Classifications . . . . .	16
Multilevel Modeling Across Full Treatment . . . . .	17
PHQ-9 Analysis . . . . .	19
Jacobson-Traux Classifications . . . . .	19
Multilevel Modeling Across Full Treatment . . . . .	19
Discussion . . . . .	21
Reflection . . . . .	24
References . . . . .	27
Author Biography . . . . .	30

Appendix A . . . . .	31
----------------------	----

## LIST OF TABLES

Table		Page
1	Items from the PHQ-9 . . . . .	7
2	Items from the GAD-7 . . . . .	8
3	The Reliable Change Indexes for different subsets of the BASE-6 data . . . .	9
4	The Clinical Cutoff Scores and Reliable Change Indexes for each measure. .	10
5	Descriptive statistics for Jacobson-Truax classifications . . . . .	12
6	Model fit statistics for multilevel model estimating total score. All classifica- tion variables are figured in relation to the Deteriorated class variable. . . .	13
7	Model fit statistics for multilevel model estimating total score. All classifica- tion variables are figured in relation to the Deteriorated class variable. . . .	15
8	Descriptive statistics for Jacobson-Truax classifications . . . . .	17
9	Model fit statistics for multilevel model estimating total score. All classifica- tion variables are figured in relation to the Deteriorated class variable. . . .	18
10	Descriptive statistics for Jacobson-Truax classifications . . . . .	20
11	Model fit statistics for multilevel model estimating total score. All classifica- tion variables are figured in relation to the Deteriorated class variable. . . .	20



## LIST OF FIGURES

Figure		Page
1	Individual trajectories for each client for the BASE-6. Different colors represent each client. . . . .	12
2	Average trajectories for each Jacobson-Truax classification. . . . .	14
3	Predicted trajectories for the first three timepoints across Jacobson-Truax classifications. . . . .	16
4	Average trajectories for each Jacobson-Truax classification. . . . .	19
5	Average trajectories for each Jacobson-Truax classification. . . . .	21

### Modeling Change Trajectories for Mental Health Symptoms and Functioning During Psychotherapy

In psychotherapy with mental health clients, clinical assessment with standardized questionnaires is used to understand a client's psychological symptoms at the initial visit (i.e., severity of mental health symptoms such as depression), and also to track changes in symptoms to measure treatment success. This practice is often referred to as routine outcome monitoring (ROM). A developing body of research has revealed the importance of basing clinical care and decision making on the data collected through ROM, a convention referred to as measurement based care (MBC). There are various benefits to using MBC in psychotherapy. It has demonstrated flexibility in that clinicians have successfully implemented it regardless of their training or the diagnosis of their clients (Scott and Lewis, 2015). MBC has also been shown to give clients a greater sense of involvement in their treatment (Eisen, Dickey, and Sederer, 2000). Additionally, the practice of MBC appears to be particularly helpful to clinicians in identifying clients who are at risk of treatment failure or dropout (Lambert, Harmon, Slade, Whipple, and Hawkins, 2005). Without ROM data, clinicians are inclined to highly overestimate the likelihood of treatment success, while underestimating treatment failure (Hannan et al., 2005). However, when clinicians utilize ROM data from measures assessing psychological symptoms, their ability to identify clients at risk of treatment failure greatly increases (Shimokawa, Lambert, and Smart, 2010).

Because of these findings, the use of ROM in psychotherapy clinics has seen an increase in past years. However, there is often a gap in routinely collecting data from clients (ROM) and actually using that data to inform clinical decision making (MBC). One of the barriers preventing the use of ROM data was the complexity of ROM when using paper-and-pencil measures. To alleviate the computational stress this process brings, there is a growing trend in the field to use computerized or web-based systems for ROM known as measurement feedback systems (MFS's). An MFS is a software tool that gathers routine

outcome data from clients and then gives the clinician feedback about their progress, often in the form of automatically generated graphs of clients' symptoms across time. There are a variety of these systems available, each with different formats and functions (Lyon, Lewis, Boyd, Hendrix, and Liu, 2016).

Viewing a client's scores on psychological symptom measures and related graphs within an MFS is not sufficient, however, unless clinicians are able to accurately and efficiently interpret the data and identify clients who are in danger of treatment failure. Because of this, some MFS developers have worked to create systems where the clinician is alerted if their client is off track from what would be labeled normal progress (Bickman, Kelley, and Athay, 2012; Cannon, Warren, Nelson, and Burlingame, 2010; Hannan et al., 2005; Lambert et al., 2005; Youn, Kraus, and Castonguay, 2012). In a meta-analysis of four different studies, Lambert et al. (2005) found that when clinicians received a notification that their client was off track, the clients' symptoms levels were far less severe at the end of treatment than when clinicians were not given these warnings. These findings demonstrate the utility of ROM using an MFS to measure clients' treatment response and make informed treatment decisions. Still, the procedures and algorithms used as the foundation of these MFS's are an area that deserves further attention to make these systems as informative as possible for clinical use.

Jacobson and Truax (1991) were instrumental in devising methods to describe the nature and magnitude of change in clinical symptoms and functioning. They suggested a twofold criterion for classifying clinical change in psychotherapy: First, whether or not the degree of change is significant, that is, reliably different from zero or no change, and second, whether or not clients reach a level of psychological functioning comparable to that of the normal functioning population (Jacobson, Roberts, Berns, and McGlinchey, 1999). To address the first criterion, a formula called the reliable change index (RCI) is used to measure if clients have made significant change that is either positive (i.e., symptom improvement) or negative (i.e., symptom deterioration). The RCI refers to the number of points a

client must change on a measure for the change to be statistically “reliable” (i.e., not due to measurement error). The RCI is calculated using the following formula:

$$RCI = \frac{X_2 - X_1}{\sqrt{2(S_E)^2}}$$

where  $X_1$  is a client’s baseline score, and  $X_2$  is that client’s post-treatment (or post-baseline) score.  $S_E$  represents the standard error of measurement for that particular clinical assessment measure and is computed using:

$$S_E = s_1 \sqrt{1 - r_{xx}}$$

where  $s_1$  is the standard deviation of the control group or pretreatment group and  $r_{xx}$  is the test-retest reliability of the measure.

To address the second criterion, a clinical cutoff point is established for each measure, with scores above the threshold representing the dysfunctional population and scores below representing normal psychological functioning (Jacobson, Follette, and Revenstorf, 1984). To find this clinical cutoff for each measure, researchers typically test the measure on both individuals with clinical symptoms and those functioning normally as measured by another validated measure or clinical interview. Researchers then test different cutoffs and compare statistics such as the sensitivity and specificity to see which cutoff is most successful at separating the clinical population from the typical population (Spitzer, Kroenke, Williams, and Löwe, 2006). Combining these two tools provides a framework by which each client can be classified by the level and extent of the change in their psychological symptoms. Clients are deemed *recovered* if they pass the clinical cutoff in the direction of clinical to non-clinical symptoms and their change is significant as measured by the RCI. If clients show significant change in the positive direction, but do not pass the clinical cutoff score, they are classified as *improved, but not recovered*, while if there is no significant change they are classified as *unchanged*. Finally, clients who experience significant change

in then negative direction are classified as *deteriorating*.

The Jacobson-Truax method has been used frequently to track client progress from session to session. While this is useful in that it is relatively simple to implement, there are also has several pitfalls to this measure. This framework of classifying clients uses two data points to communicate how the client is progressing. However, contemporary ROM is often administered more frequently than at intake and termination, and in many cases weekly. The Jacobson-Truax method also ignores the possibility of a non-linear rate of change in symptom levels throughout treatment and does not incorporate the fact that client trajectories may vary in treatment. For example, one client's symptoms may decrease rapidly in the beginning of treatment and then slow down as treatment progresses (i.e., deceleration of progress), while the opposite may be true of another client (i.e., acceleration of progress). This variation in the rate of change over time is often referred to as the *shape of change* (Laurenceau, Hayes, and Feldman, 2007).

Understanding the shape of change can help us understand the typical nature of change throughout therapy as well as enable us to assess how a client is progressing with respect to these norms. For example, Tang and Roberts (1999) found that clients had much better treatment outcomes when they experienced *sudden gains*. Sudden gains are defined as large symptom improvements relative to typical change in the overall population as well as with respect to the client's typical symptom fluctuations. Clients who have experienced this type of change at some point, usually showed long-lasting improvement as their therapy continued. Thus, clinicians can use this fluctuation in their clients' rate of symptom change to predict treatment success. This is just one example that illustrates why understanding the shape of change is critical.

One way to capture information concerning the shape of change throughout treatment is by modeling trajectories using all available data from treatment. As described above, it is becoming much more common to routinely track outcomes, and therefore individuals have more than just two data points. Trajectories incorporating all of these time

points can be created using multilevel modeling. Multilevel modeling is a statistical framework in which a response variable (i.e. outcome measure score) is a function of fixed effects (i.e. time in treatment), similar to a linear regression model. However, in the context of ROM, each client has multiple scores for an outcome measure, and thus the assumption of independence necessary for a linear regression model is not met. To resolve this issue, multilevel models allow for random effects as well. This enables us to assume random intercepts and slopes in our model for each client, with the intercept representing the client's beginning score and the slope representing their rate of change throughout treatment (Winter, 2013). Thus, the fixed effects provide information on the average effect of a variable on the response, and the random effects account for the variability around these averages.

One of the benefits of using multilevel modeling is that this method allows us to compute average trajectories for groups of clients and thus can give us information about the typical nature of change during therapy. We can use this information to assess progress of current clients and predict their future change based on the shape of their trajectories and how well they conform to the normal trajectory. Cannon et al. (2010) successfully used multilevel modeling to construct typical change trajectories for the Youth Outcome Questionnaire and then built a warning system to alert clinicians if clients did not follow the expected change trajectory. This warning system was able to predict treatment failure with as high as 73% accuracy.

The prior study shows the utility that multilevel modeling can have in measuring and predicting treatment outcomes. Multilevel modeling has the ability to give us information about the shape of change that the Jacobson-Truax framework is unable to provide. However, the Jacobson-Truax framework is a valuable tool in that it is simple and also gives concrete categories by which client progress can be classified. Thus both of these methods have their strengths in measuring, summarizing, and predicting future change in psychotherapy client progress. Despite this, there has been little research on how the

Jacobson-Truax framework and multilevel modeling can be used in conjunction with one another. Because of this, the primary aim of the current study was to investigate how multilevel modeling can be used to build upon and expand information obtained from the Jacobson-Truax Classification Framework. Furthermore, I wanted to explore the possibility of using multilevel modelling to predict within the first few sessions which Jacobson-Truax Classification clients would fall into at the end of treatment.

In addition to these research goals, one of the major purposes of this project was to develop reproducible code and practices that can be applied to new, larger dataset. The purpose of this was two-fold. First, there is the potential that more data will become available within the research group I have been working with and so I wanted to ensure that this data could be analyzed in the future. Second, an vital part of research in statistics and data science is assuring that your research is reproducible and accessible (Stodden, 2010). This enables other researchers to validate your findings and is also convenient because researchers can use already created code instead of developing their own. Thus, reproducibility was key in this study because of the importance of accelerated future research.

## Methods

### Measures

I examined three different measures for this study: the Brief Adjustment Scale (BASE-6), the Generalized Anxiety Disorder Scale (GAD-7), and the Patient Health Questionnaire (PHQ-9).

**The BASE-6.** This measure has six item measure with each item on a 7-point Likert scale (1 = not at all, 4 = somewhat, 7 = extremely). It is used to measure general psychological adjustment and is generally used as a shorter alternative to similar longer measures (Peterson, 2015). Some of the items in this measure include “To what extent have you felt unhappy, discouraged, and/or depressed this week?” and “How much has emotional distress interfered with feeling good about yourself this week?” This measure has

strong concurrent validity when correlated with another common outcome measure, the OQ-45 ( $r = 0.66$  to  $r = 0.81$ ). Additionally, the BASE-6 also exhibited excellent internal consistency with Cronbach's alpha scores ranging from  $\alpha = 0.87$  to  $\alpha = 0.92$ .

**The PHQ-9.** This is a measure typically used to track the severity of depression symptoms. Each of the nine items in this measure is evaluated on a scale of 0 (not at all) to 3 (nearly everyday). Items from the PHQ-9 can be seen in Table 1 The PHQ-9 exhibited excellent internal validity with Cronbach's alpha scores ranging from  $\alpha = 0.86$  to  $\alpha = 0.89$  (Kroenke, Spitzer, and Williams, 2001). It also showed strong construct validity shown by the association between PHQ-9 scores and scores on the SF-20, a general health questionnaire.

**Table 1:** Items from the PHQ-9

Over the last 2 weeks, how often have you been bothered by the following problems?
1. Little interest or pleasure in doing things
2. Feeling down, depressed, or hopeless
3. Trouble falling or staying asleep, or sleeping too much
4. Feeling tired or having little energy
5. Poor appetite or overeating
6. Feeling bad about yourself, or that you are a failure or have let yourself or your family down
7. Trouble concentrating on things, such as reading the newspaper or watching television
8. Moving or speaking so slowly that other people could have noticed. Or the opposite- being so fidgety or restless that you have been moving around a lot more than usual
9. Thoughts that you would be better off dead or of hurting yourself in some way

**The GAD.** Similar to the PHQ-9, the seven items on the GAD-7 are measured on a scale from 0 (not at all) to 3 (nearly everyday). This measure, however, is used to identify symptoms of Generalized Anxiety Disorder. The items of the measure can be viewed in Table 2. Like the other measures, the GAD-7 showed excellent internal validity, exhibited by a Cronbach's alpha of  $\alpha = 0.92$  (Spitzer et al., 2006). There was also a strong association between increasing GAD-7 scores and worsening function scores on the SF-20, indicating strong construct validity.

I chose to use these measures because of their widespread use within the OwlOut-



**Table 2:** Items from the GAD-7

Over the last 2 weeks, how often have you been bothered by the following problems?
1. Feeling nervous, anxious or on edge
2. Not being able to stop or control worrying
3. Worrying too much about different things
4. Trouble relaxing
5. Being so restless that it is hard to sit still
6. Becoming easily annoyed or irritable
7. Feeling afraid as if something awful might happen

comes library. Additionally, norms for longitudinal change in treatment for these measures have not been developed, so investigating these measures using multilevel modeling is needed.

### Participants

Deidentified data used in this study came from 42 clients at a Clinical Psychology doctoral training clinic at Virginia Polytechnic Institute and State University (Virginia Tech). All clients in the study used an electronic MFS called OwlOutcomes to complete at least one of the psychological measures in question (Base-6:  $n = 40$ ; GAD-7:  $n = 8$ ; PHQ-9:  $n = 10$ ). All 42 clients completed one of these measures at a minimum of two separate time points.

### Procedure

Data was collected from clients before psychotherapy sessions using OwlOutcomes software. This data was then deidentified and formatted in password protected Excel files and sent to researchers at Utah State University. After these files were obtained, two different time variables were computed. The first time variable that was computed was *session number* in which every therapy session was counted sequentially. For therapy sessions where the client did not complete a symptom questionnaire, the score for that entry was marked NA. Likewise, if a symptom questionnaire was taken, but no therapy session occurred (i.e. client completed multiple questionnaires between sessions) *session number* was marked as NA. The second measure of time was *days after intake*. This measure represented the number of days in between a client's first therapy session and the current

timepoint. The *days after intake* variable was later converted to *months after intake* to increase the interpretability and utility of the model.

### Analysis

**Jacobson-Truax Framework.** The first step in the analysis was to categorize each patient by whether or not they passed the clinical threshold during their treatment and if their change was reliable. Clinical cutoff scores and Reliable Change Indexes (RCI) for the PHQ-9 and GAD-7 were obtained from past research (Delgadillo, 2012; C. A. Griffiths and Griffiths, 2015; Kroenke et al., 2001; Spitzer et al., 2006). Clinical cutoff scores for the BASE-6 were obtained from unpublished pilot research and corresponded to the clinical cutoff of the commonly used OQ-45 measure.

The RCI for the BASE-6 was computed using data from two sources: the pretreatment scores from the current study data set, and BASE-6 scores from a sample of individuals recruited through Amazon Mechanical Turk (Mturk), a crowd-sourced online participant pool. The Mturk data was obtained in a previous study and included 812 participants, 192 of which had previously participated in counseling (Peterson, 2015). Jacobson and Truax's (1991) equation was used in the computation of the RCI. Several potential RCI scores were computed using various subsets of the participants. These subsets were: subjects who had received counselling, subjects who had not received counseling, subjects from the MTurk dataset, and pretreatment scores from the study data set. Table 3 shows the RCI score for each of these subsets. All values indicated that it was appropriate to use an RCI of six for the BASE-6.

**Table 3:** The Reliable Change Indexes for different subsets of the BASE-6 data

Label	SD	RCI Cutoffs
All Observations	8.89	6.52
No Counseling	8.71	6.39
Counseling	8.51	6.24
Vtech Pre-Treatment	9.20	6.75
M-Turk Counseling	8.37	6.14

Resulting clinical cutoff scores and RCI's for the BASE-6 and all other measures

are shown in Table 4

**Table 4:** The Clinical Cutoff Scores and Reliable Change Indexes for each measure.

Measure	Clinical Cutoff	Reliable Change Index
GAD-7	10	5
PHQ-9	10	8
BASE-6	19	6

After clinical cutoffs and RCI's were established for each measure, each research subject was classified according to the categories prescribed by Jacobson and colleagues: 1) Recovered (passing the clinical cutoff and exhibiting reliable change in the positive direction); 2) Improved but not recovered (exhibiting reliable change in the positive direction without passing the clinical cutoff); 3) No reliable change 4) Deterioration (reliable change in the negative direction).

**Multilevel Modeling Across Full Treatment.** After each participant was classified according to the Jacobson-Truax Framework, a multilevel model was formulated for each of the three measures using the *lmer* function in the *lme4* R package (Bates, Mächler, Bolker, and Walker, 2014). An initial model was created using the score of the measure in question as the response variable. Fixed effects terms in this initial model include linear and quadratic effects for times in therapy (measured in months) and dummy codes for the Jacobson-Truax framework classification. Interaction effects between the time terms and the framework classifications were also included as fixed effects. Random effects were used to model within-client variability in change over time.

The next step in model fitting was to eliminate non-significant terms. This was accomplished using the *step* function in the *lmerTest* R package (Kuznetsova, Brockhoff, and Christensen, 2015). This package performs backward elimination of non-significant effects from the model. For fixed effects, significance was assessed using p-values calculated from an *F* test based on Sattethwaite's approximation for degrees of freedom (Satterthwaite,

1946). *P*-values from the likelihood ratio test were used for random effects. This process returned the final fitted model.

**Early Detection of Treatment Outcome.** After fitting trajectories for the entirety of treatment, we used a similar method to create models for the first three time points. This was done in order investigate early patterns of change during psychotherapy in relation to final categorization of progress.. Participants were included in the model if they had at least three scores for the given measure and the first measure was completed no later than one month after the beginning of therapy.

After the data was subsetted according to the above specifications, model fitting was completed according to the method previously outlined. However, unlike before the time variable *session number* was used instead of *month* and no quadratic term was tested in the model because this is not possible with only three timepoints.

Because of the small sample sizes for the GAD-7 and the PHA-9, this step was only performed on the BASE-6. Furthermore, this was a preliminary, investigative step in the analysis and so concrete conclusions cannot be drawn concerning predictions.

## Results

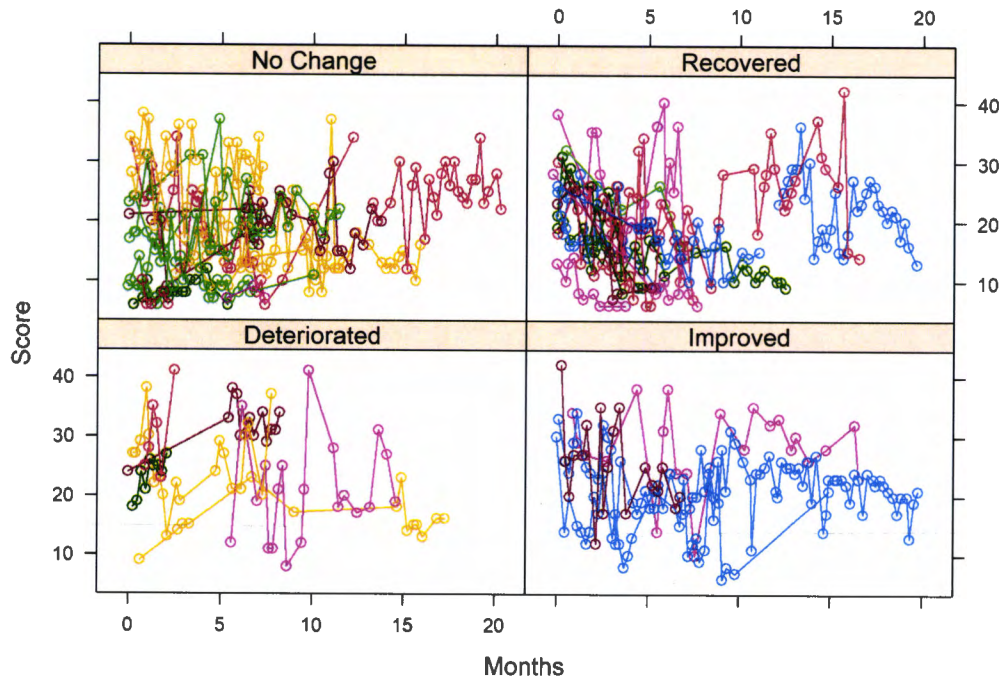
### BASE-6 Analysis

**Jacobson-Truax Classifications.** The first step in the analysis was to classify each participant according to the Jacobson-Truax framework (i.e., using just two data points). Descriptive statistics resulting from this process can be seen in Table 5. According to this framework, the majority of clients either recovered or did not significantly change. Those in the *improved* class began on average with the highest score (representing the most severe symptoms) but also improved the most on average, even slightly more than those in the *recovered* class. Those in the *no change* class started with the lowest score. The average score of 18 for the *no change* class is less than the clinical cutoff for the BASE-6 of 19. It is important to note that clients varied regarding when they began taking the measure and on the frequency with which the measure was taken, so these statistics are only repre-

sentative of the available data. This is particularly true with the statistics concerning the “initial” scores because some clients did not begin taking the measure until after months in therapy. Figure 1 shows that clients did not always began taking the BASE-6 at the beginning of treatment and that the length of the trajectory and the number of timepoints varies from client to client (Sarkar, 2008). This figure also demonstrates how complex this data is.

**Table 5:** Descriptive statistics for Jacobson-Truax classifications

Classification	# of clients	Average Starting Score (SD)	Average Ending Score (SD)	Average Difference (SD)
Recovered	15	24.27 (6.34)	10.87 (3.68)	13.40 (4.93)
Improved	4	34.75 (5.12)	21.25 (1.26)	13.50 (5.80)
No Change	15	18.00 (9.26)	17.40 (9.04)	0.60 (2.59)
Deteriorated	6	19.17 (7.41)	29.00 (10.06)	-9.83 (3.31)



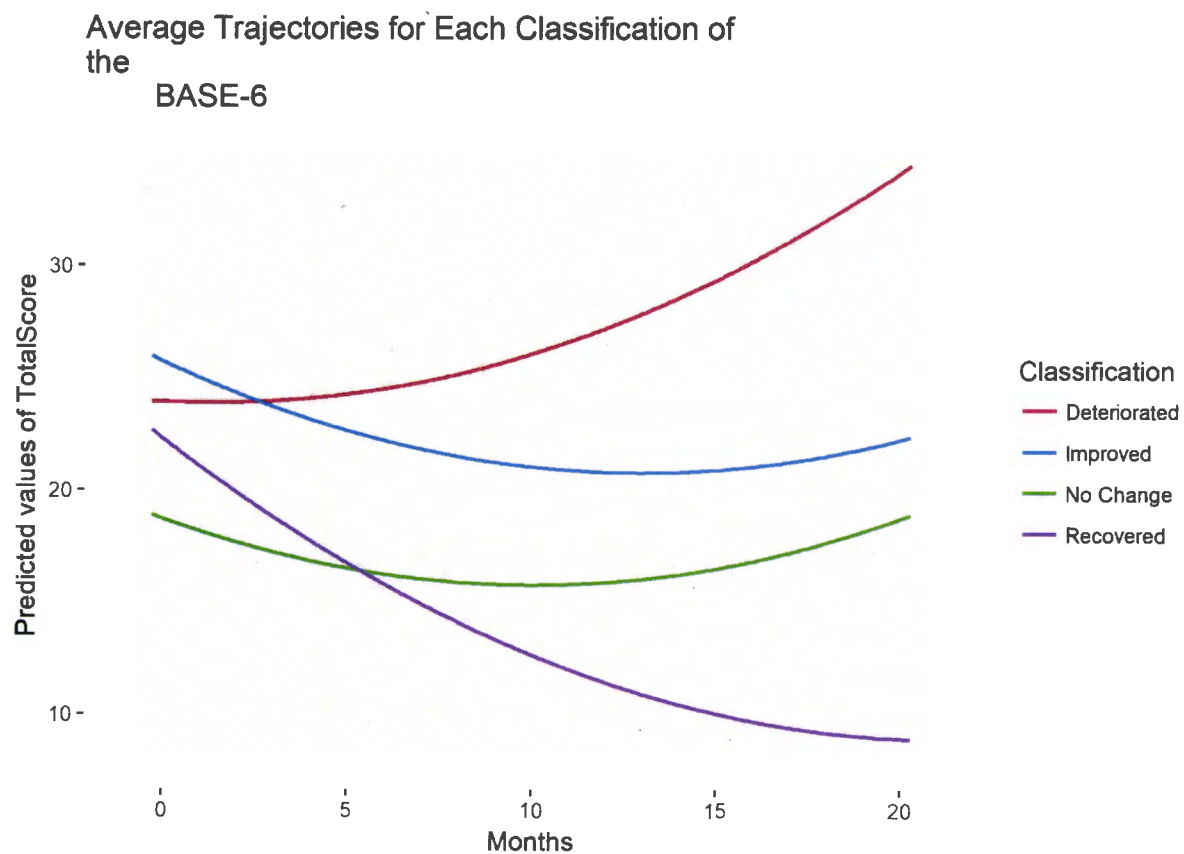
**Figure 1:** Individual trajectories for each client for the BASE-6. Different colors represent each client.

**Multilevel Modeling Across Full Treatment.** The next step in the analysis was to create average change trajectories across the complete treatment time. After eliminating insignificant terms, the remaining model included fixed effects for linear and quadratic time, dummy codes for Jacobson-Truax classifications, and a time\*group interaction. Table 11 shows a summary of the fixed effects of this model. The estimate represents the coefficient for each fixed term and gives us information concerning how each term affects the response variable (total score). For example, the estimate for months tells us that, on average, for every month in treatment, clients' score change by -0.08 points. The model is fit in relation to the *deteriorated* class and so other Jacobson-Truax class terms and interactions represent deviations in the intercept and slope for each class with respect to the *deteriorated* class. For example, clients in the *recovered* class started 1.6 points lower than clients in the *deteriorated* class and their scores decreased at a rate 1.18 per month more than clients in that class as well. The table also indicates that the only significant terms are the intercept, quadratic and linear time variables, and interaction between time and the *recovered* class. Thus, there may not be a significant difference between client scores in the beginning of treatment across classes. Furthermore, the slope of change over time may not be significantly different for any groups other than the *recovered* and *deteriorated* class.

**Table 6:** Model fit statistics for multilevel model estimating total score. All classification variables are figured in relation to the Deteriorated class variable.

	Estimate	Std. Error	df	t value	Pr(> t )
Intercept	23.93	2.69	39.11	8.91	0.00
Months <sup>2</sup>	0.03	0.01	206.30	2.78	0.01
Months	-0.08	0.32	41.96	-0.26	0.79
Improved	1.82	4.01	31.90	0.45	0.65
No Change	-5.21	3.15	37.42	-1.65	0.11
Recovered	-1.60	3.15	37.57	-0.51	0.61
Months:Improved	-0.69	0.36	18.05	-1.90	0.07
Months:No Change	-0.51	0.33	26.50	-1.56	0.13
Months:Recovered	-1.18	0.34	26.05	-3.49	0.00

Figure 2 shows the average trajectories for each of the classifications across time (Lüdtke, 2016). This curvature of the trajectories shows the effect of the quadratic time term and gives information not available when using the Jacobson-Truax method alone. For instance, the trajectories for both the *improved* and *no change* classes show that on average, clients in these classes experience early positive change but then demonstrate a resurgence in their symptoms.



**Figure 2:** Average trajectories for each Jacobson-Truax classification.

**Early Detection of Treatment Outcome.** After computing the complete trajectories across all of therapy, we computed trajectories for the first three sessions. In performing backward elimination of the model terms all of the fixed effects were retained in the model (linear time variable, the Jacobson-Truax classification term and the interaction

between these two terms). Additionally, a random intercept term for participant was included in the model, but not a term for the slope. This suggests that there was variability around the intercept for each individual client, but there was less variability in the slope (or rate of change).

Table 7 summarizes the fixed effect terms in the prediction model. As in the previous model, the classification terms were done with respect to the *deteriorated* class. We can see from the table that the p-value for each of terms in the model besides the *no change* term. This indicates that at the first timepoint, all of the classes except the *no change* have significantly different scores in comparison to the *deteriorated*, and that the slope of the trajectories for each are all significantly different from the *deteriorated* class.

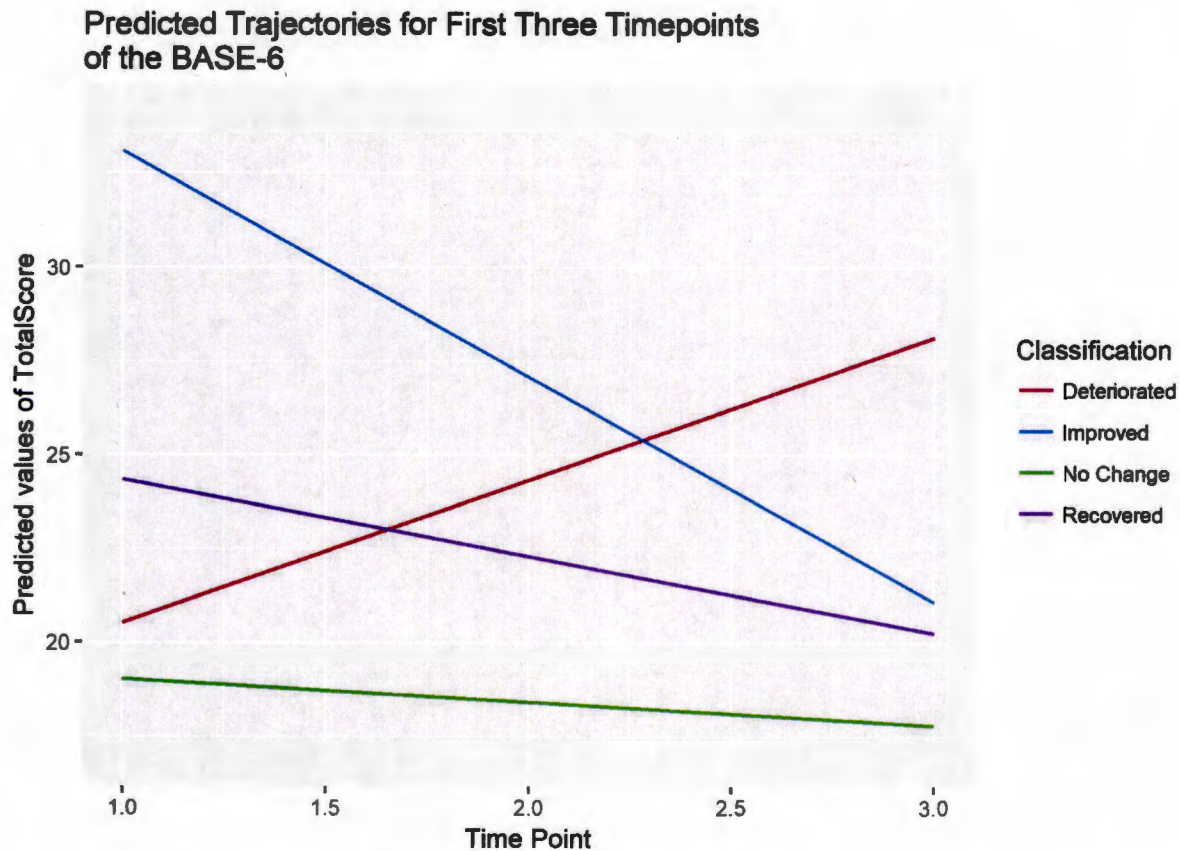
**Table 7:** Model fit statistics for multilevel model estimating total score. All classification variables are figured in relation to the Deteriorated class variable.

	Estimate	Std. Error	df	t value	Pr(> t )
Intercept	16.73	3.89	76.07	4.30	0.00
Session	3.80	1.27	64.00	2.99	0.00
Improved	22.35	5.84	76.07	3.83	0.00
No Change	2.87	4.77	76.07	0.60	0.55
Recovered	9.65	4.58	76.07	2.11	0.04
Session:Improved	-9.80	1.90	64.00	-5.15	0.00
Session:No Change	-4.40	1.56	64.00	-2.83	0.01
Session:Recovered	-5.84	1.49	64.00	-3.91	0.00

Figure 3 provides a visual representation of the model. The model shows that on average, clients in the *improved* class start at a higher score than the other classes, but they also are predicted to improve in the first three session much more rapidly than any other class. If this level of change were to persist, clients would most likely pass the clinical cutoff representing recovery. Thus, in the beginning, those in the *improved* class may see substantial change, but that this level of change may not continue to full recovery. The figure also shows us that unlike the three other classifications, the trajectory for the



*deteriorated* increases sharply. On average, members of this class have a sharp increase in symptoms within the first three sessions of therapy.



**Figure 3:** Predicted trajectories for the first three timepoints across Jacobson-Truax classifications.

### GAD Analysis

**Jacobson-Truax Classifications.** Only eight participants met the conditions necessary to be included in the analysis for the GAD-7. Two participants were classified in the *recovery* category and six in the *no change* category. Descriptive statistics for each group can be seen in Table 8. Since the clinical cutoff for the GAD-7 is a score of 10, so we can see that the average beginning score of 11.00 for the *recovered* class is just above this cutoff. The average for the beginning score for *no change* begins at lower than that of the *re-*

*covered* class (7.67), however, the average ending score of 6.50 for the *no change* class was higher than the average ending score of 2.00 for the *recovered* class.

**Table 8:** Descriptive statistics for Jacobson-Truax classifications

Classification	# of Patients	Average Starting Score (SD)	Average Ending Score (SD)	Average Difference (SD)
Recovered	2	11.00 (2.83)	2.00 (2.83)	9.00 (5.66)
No Change	6	7.67 (5.09)	6.50 (5.92)	1.17 (1.94)

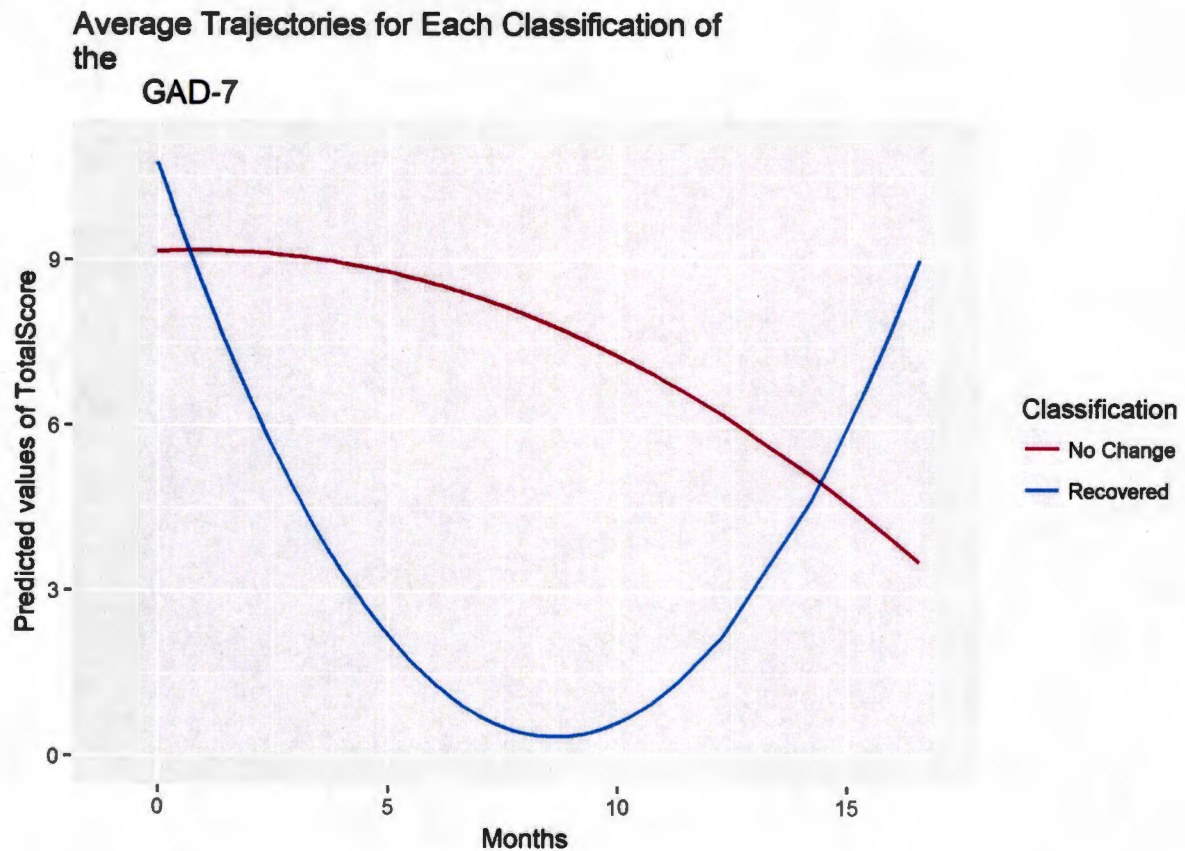
**Multilevel Modeling Across Full Treatment.** After classifying each participant according to the Jacobson-Truax classification framework, a multilevel model was fitted. The best fit model included fixed effects for linear and quadratic time, dummy codes for Jacobson-Truax classifications, a linear time:group interaction, and, unlike in the BASE-6 model, a quadratic time:group interaction. A random effect for participant intercept was also included in the model, but not a random slope term. Table 11 shows a summary of the fixed effects of this model. This model was fitted with the *no change* group as the reference group. Thus, the intercept, months, and months<sup>2</sup> terms are used to construct the trajectory of the *no change* class, the the recovered term and interaction terms provide information on how the trajectory of the *recovered* class trajectory varies from the *no change* trajectory. For example, we can see that the starting score/intercept for the *recovered* class is 1.54 points above that of the *no change* class and the score also decreases 2.43 points more per month for the *recovered* class. However, it can also be seen in the table that the only significant terms are the two interaction terms.

Figure 4 demonstrates the effect of the quadratic time:class interaction, which we did not see in the BASE-6 model. We can see that not only the slopes and the intercepts vary between the two Jacobson-Truax classes, but the shape of the change. The trajectory for the *no change* group shows that the rate of change increases as time goes on, while the trajectory for the *recovered* group shows a rapid decrease at the beginning of treatment, but then this change levels out and eventually leads to an increase in symptoms. Of

**Table 9:** Model fit statistics for multilevel model estimating total score. All classification variables are figured in relation to the Deteriorated class variable.

	Estimate	Std. Error	df	t value	Pr(> t )
Intercept	9.16	2.17	13.66	4.22	0.00
Months	0.04	0.42	93.27	0.09	0.93
Recovered	1.54	4.29	12.97	0.36	0.73
Months <sup>2</sup>	-0.02	0.03	92.62	-0.68	0.50
Months:Recovered	-2.43	0.97	93.85	-2.50	0.01
Months <sup>2</sup> : Recovered	0.16	0.06	92.90	2.70	0.01

course, it is important to remember that because these trajectories were formed from such a small sample, we should not draw any conclusions about the typical nature of change as demonstrated by this measure. For instance, when a closer look at the data shows that out of the two clients classified as recovered, one was only in therapy for 12 months. Thus, though the whole trajectory is based on very little information, there is even more error after 12 months because it is only based on one case. However, this model does give us an example of the information that can be portrayed by the combination of the Jacobson-Truax method and multilevel modeling.



**Figure 4:** Average trajectories for each Jacobson-Truax classification.

### PHQ-9 Analysis

**Jacobson-Truax Classifications.** Ten participants met the conditions necessary to be included in the analysis for the PHQ-9. Two participants were classified in the *recovery* category and eight in the *no change* category. Descriptive statistics for each group can be seen in Table 10. One of the most interesting findings in this table is that the beginning average scores for both the *recovered* and *no change* classes were less than a point apart, but the ending average ending scores for the two classes were about 10 points apart. Thus, though the two groups started at around the same symptom level, the *recovered* group's scores dropped dramatically while those in the *no change* group did not change very much.

**Table 10:** Descriptive statistics for Jacobson-Truax classifications

Classification	# of Patients	Average Starting Score (SD)	Average Ending Score (SD)	Average Difference (SD)
Recovered	2	13.00 (5.66)	1.00 (0.00)	12.00 (5.66)
No Change	8	12.38 (5.90)	10.75 (5.99)	1.62 (3.46)

**Multilevel Modeling Across Full Treatment** Unlike the models for the other two measures, the best fit multilevel model for the PHQ-9 did not contain a quadratic time variable. The fixed effects for the model included the linear time variable, dummy codes for Jacobson-Truax classifications, and a linear time:group interaction. A random effect for participant intercept was also included in the model, but not a random slope term. Again, this suggests that there was not a lot of less variability in the slope (or rate of change) between clients. Table 11 shows a summary of the fixed effects of this model. Thus, this model was relatively simple in comparison to the models for the other variables. Like the GAD-7 model, this model was fitted with the *no change* group as the reference group. We can see then that the intercept for the *no change group* is 24.61 and that the scores of members of this class decrease by 0.29 points per month on average. The intercept, for the *recovered* class is not significantly different than that of the *no change* group, but the scores of the *recovered* class did decrease by an average of 2.2 more points a month than that of the *no change* class.

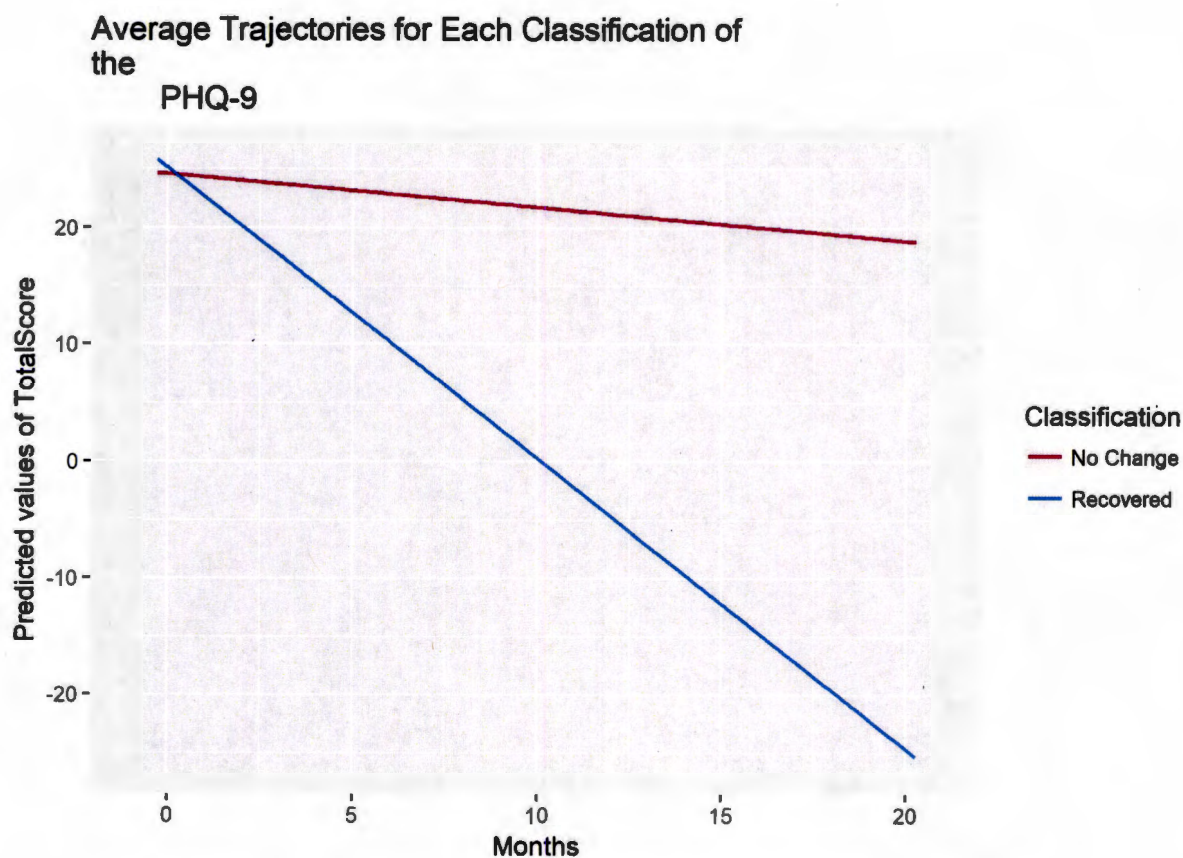
**Table 11:** Model fit statistics for multilevel model estimating total score. All classification variables are figured in relation to the Deteriorated class variable.

	Estimate	Std. Error	df	t value	Pr(> t )
Intercept	24.61	2.26	16.37	10.88	0.00
Months	-0.29	0.14	166.87	-2.11	0.04
Recovered	0.54	4.95	15.50	0.11	0.92
Months:Recovered	-2.20	0.53	186.00	-4.13	0.00

A visual description of this model can be seen in 5. We can see that because this model is linear, we obtain similar information from this model and the Jacobson-Truax de-



scriptive statistics. Both methods show that though the *no change* and *recovered* classes began with around the same average score and both classes decreased at a constant rate, the *recovered* class decreased much more rapidly than the *no change* class. As with the GAD-7, with such a small sample, we cannot not draw any definitive conclusions about the typical nature of change, but this model gives yet another example of the how multi-level modeling and the Jacobson-Truax framework can be used in conjunction with one another.



**Figure 5:** Average trajectories for each Jacobson-Truax classification.

### Discussion

The purpose of this study was to investigate how multilevel modeling and change trajectories could be used to build upon and expand the information available through the

Jacobson-Truax classification framework. Through the analysis of three different outcome measures, it was evident that while the Jacobson-Truax method can provide descriptive information concerning pre-treatment and post-treatment psychological conditions, multi-level modeling enables us to learn more about the shape of change throughout the entire treatment. Using Jacobson-Truax classifications within a multilevel model framework allows us to harness the simplicity and interpretability offered through the classifications while also learning more about the differences and similarities in symptom trajectories across time. We can examine the intercept, slope and shape of change of trajectories for each treatment outcome class as well as the differences in trajectories between the classes.

While a secondary goal of this analysis was to learn more about the shape of symptom change as measured by the BASE-6, the GAD-7, and the PHQ-9, one of the major limitations to this goal was the small participant samples for each of these measures, especially the GAD-7 and the PHQ-9. Because of this, we cannot draw any final conclusions about the nature of change across treatment when using any of these measures. However, the primary goal of this research was to understand if and how the Jacobson-Truax method and multilevel modeling could be integrated to widen our understanding of how change during therapy works. The analysis of these measures did show the information that may be available through these methods, even if the conclusions that were drawn from this analysis are not generalizable. Additionally, this study primarily provided a descriptive representation of the different trajectories of symptom change in psychotherapy.

Despite the inability to draw conclusion concerning the nature of change in psychotherapy, one of my goals in this project was to produce methods and codes to enable further investigation. This goal was met and the code used in this project was constructed in a way that future researchers will input a new dataset and the analysis will be run (Appendix A). The reproducibility of this project provides a gateway for future research in this subject matter.

One of the future directions stemming from this research is the potential to build

predictive models. We saw from the analysis of the beginning time points of the BASE-6 that clients who's therapy results in different treatment outcomes (i.e. recovery or deterioration) may exhibit very different change trajectories early in their therapy. Thus, it is realistic to believe that we could be able to predict future treatment outcome within the first few therapy sessions. However, formal methods and models need to be created in order to accomplish this.

The Jacobson-Truax method for understanding clinical change and the multilevel modeling framework are both valuable methods. While the Jacobson-Truax method creates a simple classification system that is easy to understand, multilevel modeling provides more information about the shape of change throughout treatment. This project showed that these methods do not have to be mutually exclusive. We can use multilevel modeling to gain valuable information about the similarities and differences between the Jacobson-Truax classes and to predict treatment outcome.



## Reflection

Completing a capstone project has truly been a figurative capstone to my undergraduate education. Through this experience, I have not only been able to integrate many of my intellectual interests, but I have also learned skill that will help me in my future pursuits in education and a career. I have gained insight on how to work and communicate with other researchers across multiple locations, the general research process, and how to disseminate findings through writing and formal presentations. In addition, I have also acquired skills in new statistical models and computing methods which will be valuable to me in the future.

As a student majoring in both psychology and statistics, I am always looking for ways to integrate both of my fields of study. Not only am I interested in applying statistics in the field of psychology, I also enjoy learning about the underlying mechanics of statistical methods and ways we can use these methods to gain the most information from research data. Though I have taken many courses in both psychology and statistics, in my coursework the two subjects were taught separately with very little crossover. My capstone project, along with other research projects and honors contracts, was the perfect integration of all of my interests.

I began working with Dr. Cruz a few years ago and thus I have been able to go through the entire research process with this project. I participated in every stage of this project from using past literature to formulate research questions, gaining IRB approval, acquiring and cleaning the data, creating an analysis plan, analyzing the data, and summarizing the findings. I learned the necessary skills associated with each of these stages and I know this information will be vital to me as I attend graduate school and pursue a career in research in the future.

Because we used data from a different university, doing my capstone project also taught me how to collaborate and communicate with other researchers. Because the ma-

majority of our communication with team members from different universities was through email, I learned to be specific and concise in my communications. As I tried to understand and organize the data that we received, I also learned the importance of asking questions concerning things such as how the data was collected and recorded. This helped to prevent misinterpretations of the data.

In addition to the lessons I learned from collaborating with other researchers, I also gained a great deal from my student-mentor relationship with Dr. Cruz. Having someone to go to with my questions about the research process, statistical methods, the graduate school application process, or anything else provided me with a vital resource in navigating through academia. I appreciated Dr. Cruz's patience and guidance as I tried to meet deadlines and understand confusing theories and ideas. This taught me the importance of seeking out mentors as I continue in my schooling and career. I also learned effective ways to work with other researchers from Dr. Cruz's example.

Dr. Cruz also helped me learn how to communicate the findings of my research. I presented four posters in connection to my research with Dr. Cruz, one being specifically related to my capstone project. I learned how to orally describe my research and to adjust this description based on the level of knowledge of those who I spoke with. I also learned important lessons about visually representing data through this experience. Though challenging, the written thesis portion of the capstone project also taught me how to succinctly summarize my ideas in a way that was understandable to others. I also learned how to use different technology to assist me in producing a quality finished product. For my thesis, I learned how to use the document preparation system LaTeX within the *R* software environment.

Of course, one of the most important lessons that I learned through my honors capstone project was how to work through setbacks and challenges. As in most research studies, the timeline of this project did not always go as I planned. Going through the IRB process, acquiring the needed data, and preparing the data for analysis was a process that

took multiple years from start to finish. The analysis process was rigorous as well. Because I was new to many of the statistical methods which I was using as well as relatively new to programming in *R*, I would sometimes work for hours or even days on analysis just to find out that there was a way to accomplish what I needed using one line of code. Similarly, as I tried to find out the best methods for analyzing the data we had, I would often hours researching a potential analysis method only to find out that it would not work for our set of data. Though this was frustrating, I had to remind myself that I was learning a great deal through this process, even if I did not feel as productive as I wanted to.

The challenges and setbacks, as well as the triumphs of my capstone experience, have prepared me for my future pursuits in academia and a career. I will be attending graduate school in the Fall to study biostatistics. Working with Dr. Cruz on this and other research projects was what ultimately helped me decide to pursue further education in the field of biostatistics. I will be working as a research assistant in graduate school and I know that the research skills that I gained through my capstone experience will be vital in my success in this position. I also hope to have a career in medical research in the future. My experiences in honors have laid a foundation that I can build on as I pursue further education and a career.

I feel that the honors capstone project was the perfect way to finish my undergraduate education. For this project, I integrated both my psychology and statistics fields of study into one project. I also drew on the knowledge and skills that I had learned throughout the course of my years at Utah State University. I am grateful for this experience and the many lessons that it taught me.

## References

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.
- Bickman, L., Kelley, S. D., and Athay, M. (2012). The Technology of Measurement Feedback Systems. *Couple & Family Psychology*, 1(4), 274–284. <https://doi.org/10.1037/a0031022>
- Cannon, J. a N., Warren, J. S., Nelson, P. L., and Burlingame, G. M. (2010). Change trajectories for the Youth Outcome Questionnaire self-report: identifying youth at risk for treatment failure. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, 39(3), 289–301. <https://doi.org/10.1080/15374411003691727>
- Delgadillo, J. (2012). Depression and anxiety symptoms: Measuring reliable change in alcohol and drug users. *Advances in Dual Diagnosis*, 5(3), 102–114.
- Eisen, S. V., Dickey, B., and Sederer, L. I. (2000). A self-report symptom and problem rating scale to increase inpatients' involvement in treatment. *Psychiatric Services*, 51(3), 349–353.
- Griffiths, C. A., and Griffiths, L. J. (2015). Recovery and reliable change rates for patients scoring severe on depression, anxiety or impaired functioning in a psychological therapies service: IAPT. *Mental Health Review Journal*, 20(1), 28–35.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., and Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61(2), 155–163. <https://doi.org/10.1002/jclp.20108>
- Jacobson, N. S., and Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical*

*Psychology*, 59(1), 12.

- Jacobson, N. S., Follette, W. C., and Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15(4), 336–352.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., and McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300.
- Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16(9), 606–613.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). Package “lmerTest”. *R Package Version*, 2(0).
- Lambert, Harmon, C., Slade, K., Whipple, J., and Hawkins, E. (2005). Providing feedback to psychotherapists on their patients’ progress: Clinical results and practice suggestions. *Journal of Clinical Psychology*, 61(2), 165–174.
- Laurenceau, J.-P., Hayes, A. M., and Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, 27(6), 682–695.
- Lüdtke, D. (2016). SjPlot: Data visualization for statistics in social science. *R Package Version*, 2(0).
- Lyon, A. R., Lewis, C. C., Boyd, M. R., Hendrix, E., and Liu, F. (2016). Capabilities and characteristics of digital measurement feedback systems: Results from a comprehensive review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 441–466.
- Peterson, A. P. (2015). *Psychometric evaluation of the brief adjustment scale-6 (base-6): A new measure of general psychological adjustment* (PhD thesis).
- Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. Springer Science & Busi-

ness Media.

- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), 110–114.
- Scott, K., and Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice*, 22(1), 49–59.
- Shimokawa, K., Lambert, M. J., and 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. <https://doi.org/10.1037/a0019247>
- Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The gad-7. *Archives of Internal Medicine*, 166(10), 1092–1097.
- Stodden, V. C. (2010). Reproducible research: Addressing the need for data and code sharing in computational science. *Computing in Science & Engineering*, 12(5), 8–12.
- Tang, T. Z., and DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894.
- Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. *arXiv Preprint arXiv:1308.5499*.
- Youn, S. J., Kraus, D. R., and Castonguay, L. G. (2012). The treatment outcome package: Facilitating practice and clinically relevant research. *Psychotherapy*, 49(2), 115–122. <https://doi.org/10.1037/a0027932>

### **Author Biography**

Elizabeth Wynn is graduating from Utah State University with a double major in statistics and psychology. After graduation, she will begin graduate school at the University of Colorado in the fall where she will study biostatistics. Elizabeth developed a passion for research early in her collegiate career and has worked in a number of different labs. She enjoys using statistics to solve real world problems and has worked on projects with topics ranging from eating disorders to water usage. Outside of her school work, Elizabeth likes trail running and hiking, playing and watching sports, reading, and spending time with her family.

## Appendix A

### R Code from Analysis

Required: R Packages

```
if (!require(dplyr)){
  install.packages("dplyr")
  library(dplyr)
}
if (!require(knitr)){
  install.packages("knitr")
  library(knitr)
}
if (!require(lmerTest)){
  install.packages("lmerTest")
  library(lmerTest)
}
if (!require(sjPlot)){
  install.packages("sjPlot")
  library(sjPlot)
}
if (!require(qwraps2)){
  install.packages("qwraps2")
  library(qwraps2)
}
if (!require(kableExtra)){
  install.packages("kableExtra")
  library(kableExtra)
}
if (!require(lattice)){
  install.packages("lattice")
  library(lattice)
}
```

Calculation of BASE-6 RCI. See page 9.

```
#Read in data set of BASE 6 (Aggregated scores from Mturk and Vtech)
base6<-read.csv("Data/BASE_aggregated_for_sd.csv")

#Subset Data based on counseling (counseling=1, no counseling=2, vtech=3)
base6_NoCounseling<-subset(base6, Counseling==2)
base6_Counseling<-subset(base6, Counseling==1|Counseling==3)
base6_Mturk_Counseling<-subset(base6, Counseling==1)
base6_Vtech_Pre<-subset(base6, Counseling==3)

#Calculate Standard Deviations
SDNoCounseling<-sd(base6_NoCounseling$BASE.6)
```



```

SDCounseling<-sd(base6_Counseling$BASE.6)
SDMturk<-sd(base6_Mturk_Counseling$BASE.6)
SDVtech<-sd(base6_Vtech_Pre$BASE.6)
SDAllObservations<-sd(base6$BASE.6)

#Reliability for Base
BASERel<-.93

#Data Frame of all SD's
StanDevs<-data.frame(Label=c("All Observations", "No Counseling",
                             "Counseling", "Vtech Pre-Treatment", "M-Turk Counseling"),
                      SD=c(SDAllObservations, SDNoCounseling, SDCounseling, SDVtech, SDMturk))

#Function to find RCI cutoff
RCI<-function(SD){
  SE<-SD*(sqrt(1-BASERel))
  sdiff<-sqrt(2*(SE*SE))
  RCIcutoff<-1.96*sdiff
  RCIcutoff<-round(RCIcutoff,2)
}

#Apply RCI cutoff function to all SD's
RCICutoffs<-as.data.frame(sapply(StanDevs$SD, RCI))

#Make dataframe of cutoffs
RCICutoffs$Label<-StanDevs$Label
RCICutoffs$SD<-round(StanDevs$SD, 2)
RCICutoffs<-RCICutoffs[c(2,3, 1)]
names(RCICutoffs)[3]<-paste("RCI Cutoffs")

Jacobson-Truax Classification for all measures. See page 10.

#Read in files
BASE<-read.csv("Data/ToSend_Utah_BASE-6_Full_Export_10.14.16.csv")
PHQ9<-read.csv("Data/ToSend_UtahPHQ-9_Full_Export_10.14.16.csv")
GAD<-read.csv("Data/ToSend_Utah_GAD-7_Full_Export_10.11.16.csv")

#Filter out unneeded observations
FilterFun<-function(measure){
  #Filter out observations without a score
  measure<-measure[%>%filter(TotalScore!=-99)

  #ResearchID as factor
  measure$ResearchID<-as.factor(measure$ResearchID)
  #Filter observations without two timepoints
  keep <- levels(measure$ResearchID)[table(measure$ResearchID) > 1]
  measure <- measure[measure$ResearchID %in% keep, ]
  return(measure)
}

```

```

}
BASE<-FilterFun(BASE)
PHQ9<-FilterFun(PHQ9)
GAD<-FilterFun(GAD)

#Measure clinical cutoffs
BASEcut<-19
PHQcut<-10
GADcut<-10

#Measure RCI cutoff
BASE.RCI.Cut<-6
PHQ.RCI.Cut<-8
GAD.RCI.Cut<-5

#Function to calculate classifications for RCI and Clinical Cutoff and
#add it to original dataframe. Inputs are dataframe, cutscore and RCI
RCICalc<-function(measure, clinicalCutoff, RCICut){

  #Find last and first measurements for each client. COmbine in dataframe
  last<-by(measure, measure$ResearchID, tail, n=1)
  first<-by(measure, measure$ResearchID, head, n=1)
  last<-do.call("rbind", as.list(last))
  first<-do.call("rbind", as.list(first))
  first<-first%>%select(ResearchID, TotalScore)
  colnames(first)[2]<-"FirstScore"
  last<-last%>%select(ResearchID, TotalScore)
  colnames(last)[2]<-"LastScore"
  combined <- merge(first,last,by="ResearchID")

  #Calculate RCI for each client Add clin change variable
  #ClinChange: 1=Above Clinical Cutoff to Below Clinical cutoff
  #           2 = Below to Above
  #           3 = Above to Above
  #           4 = Below to Below
  combined<-mutate(combined, Scorediff=FirstScore-LastScore)
  combined<-mutate(combined, PreAboveClin=
    ifelse(FirstScore>=clinicalCutoff, TRUE, FALSE),
    PostAboveClin=ifelse(LastScore>=clinicalCutoff,
      TRUE, FALSE),
    ClinChange= ifelse(PreAboveClin==TRUE &
      PostAboveClin==FALSE, 1,
      ifelse(PreAboveClin==FALSE &
        PostAboveClin==TRUE, 2,
        ifelse(PreAboveClin==TRUE

```

```

& PostAboveClin==
  TRUE,3,
  ifelse(
    PreAboveClin==
      FALSE &
      PostAboveClin
      ==FALSE, 4,NA
  ))))
#RCI Coding: 1= Reliable improvement
#           2 = No Reliable change
#           3 = Reliable Deteriation
combined<-mutate(combined, RCI.Cut=ifelse(Scorediff>=RCICut, 1,
                                          ifelse(Scorediff<RCICut &
                                                  Scorediff>-RCICut, 2,
                                                  ifelse(Scorediff<=-RCICut,
                                                        3, NA))))
#ClinRCI: 1.1 = Clinically Reliable Improvement
#          1.2 = Clinical improvement but not reliable
#          2.2 = Clinical Deterioation but not reliable
#          2.3 = Clinical Reliable Deteriation
#          3.1 = Reliable Improvement but no clinical change
#               (AboveCutoff)
#          3.2 = No clinical or reliable change (Above Clinical
#               Cutoff)
#          3.3 = Reliable deterioration but no clinical change (Above
#               Clinical Cutoff)
#          4.1 = Reliable Improvement but no clinical change (Below
#               Cutoff)
#          4.2 = No clinical or reliable change (Below Clinical
#               Cutoff)
#          4.3 = Reliable deterioration but no clinical change (Below
#               #Clinical Cutoff)
combined<-mutate(combined,ClinRCI=ifelse(ClinChange==1&
                                          RCI.Cut==1,1.1,
                                          ifelse(ClinChange==1 &
                                                  RCI.Cut==2,
                                                  1.2,ifelse(ClinChange==2
& RCI.Cut==2,
2.2,ifelse(ClinChange==2
& RCI.Cut==3,
2.3,ifelse(ClinChange==3
& RCI.Cut==1,
3.1,ifelse(ClinChange==3
& RCI.Cut==2,
3.2,ifelse(ClinChange==3
& RCI.Cut==3

```

```

        3.3,
        ifelse(ClinChange==4
              & RCI.Cut==1,
4.1, ifelse(ClinChange==4
              & RCI.Cut==2,
4.2, ifelse(ClinChange==4
              & RCI.Cut==3,
              4.3,
              NA)))))))))

#Jackson-Truax Classification

combined<-combined%>%mutate(JTClass=ifelse(ClinRCI==1.1, "Recovered",
                                           ifelse(ClinRCI==4.1,
                                                 "Recovered",
                                                 ifelse(ClinRCI==1.2,
                                                       "No Change",
                                                       ifelse(ClinRCI==2.2,
                                                             "No Change",
                                                             ifelse(ClinRCI==3.2,
                                                                   "No Change",
                                                                   ifelse(ClinRCI==4.2,
                                                                         "No Change",
                                                                         ifelse(ClinRCI==3.1,
                                                                               "Improved",
                                                                               ifelse(ClinRCI==2.3,
                                                                                     "Deteriorated",
                                                                                     ifelse(ClinRCI==3.3, "Deteriorated",
                                                                                     ifelse(ClinRCI==4.3,
                                                                                           "Deteriorated",
                                                                                           NA)))))))))
stepcombined<-select(combined, ClinChange, ResearchID, Scorediff,
                     RCI.Cut, FirstScore, LastScore, ClinRCI, JTClass)

#Add RCI into original data, return original data
return(combined)
}

BASERCI<-RCICalc(BASE, BASEcut, BASE.RCI.Cut)
PHQRCI<-RCICalc(PHQ9, PHQcut, PHQ.RCI.Cut)
GADRCI<-RCICalc(GAD, GADcut, GAD.RCI.Cut)

BASE<-merge(BASERCI, BASE, by="ResearchID")
PHQ9<-merge(PHQRCI, PHQ9, by="ResearchID")
GAD<-merge(GADRCI, GAD, by="ResearchID")

BASE-6 Analysis Code. See page 12.
## pdf

```

```

## 2
## pdf
## 2
## pdf
## 2
GAD-7 Analysis Code. See page 17.

#Read in and prepare data
GAD<-read.csv("Data/GAD_RCI_ClinCutoff.csv")
GAD$ResearchID<-as.factor(GAD$ResearchID)
GAD$JTClass<-as.factor(GAD$JTClass)

#Create months variable and rename JTClass classes
GAD<-GAD%>%
  mutate(months=DaysafterIntake/30)

#Get the first observation for each patient
HeadGAD<-do.call("rbind", as.list(by(GAD, GAD$ResearchID,
                                     head,n=1)))

#Get the summary for each classification
GADSum<-HeadGAD%>%
  group_by(JTClass)%>%
  summarise("# of Patients"=n(), "Average Starting
    Score (SD)"=mean_sd(FirstScore, denote_sd = "paren",
                        digits=2),
    "Average Ending Score (SD)"=mean_sd>LastScore,
                                                denote_sd="paren",
                                                digits=2),
    "Average Difference (SD)"=mean_sd(Scorediff,
                                      denote_sd = "paren",
                                      digits=2))

#Reorder and rename column
GADSum<-as.data.frame(GADSum[c(2,1),])
colnames(GADSum)[1]<-c("Classification")

#Full model with all terms
modelTest<-lmer(TotalScore~months*JTClass+I(months^2)*JTClass+
               (months|ResearchID),data=GAD,REML=FALSE)

#Backwards elimination of model
stepTest<-step(modelTest)
#Final model after backwards elimination
modelFinal<-lmer(TotalScore~months*JTClass+I(months^2)*JTClass+
               (1|ResearchID), data=GAD, REML=FALSE)

#Make coefficient table
ModelfinalSummary<-summary(modelFinal)
CoeffTable<-round(as.data.frame(ModelfinalSummary$coefficients),2)

```

```

rownames(CoeffTable)<-c("Intercept", "Months","Recovered", "Months^2",
                        "Months:Recovered", "Months^2: Recovered")

#Save needed coefficients for report
recovCoeff<-(CoeffTable$Estimate[3])
recovMonthsCoeff<-(-CoeffTable$Estimate[5])

#Plot of average trajectores for each classification
jpeg('plot3.jpg', width=7, height=5 , unit= 'in', res=300)
sjp.int(modelFinal, type="eff",
        axis.title = c("Months", "Score"),
        title = "Average Trajectories for Each Classification",
        legend.title = "Classification", int.term = "months*JTClass",
        swap.pred = TRUE)

dev.off()
PHQ-9 Analysis Code. See page 20.

#Read in and prepare data
PHQ9<-read.csv("Data/PHQ9_RCI_ClinCutoff.csv")
PHQ9$ResearchID<-as.factor(PHQ9$ResearchID)
PHQ9$JTClass<-as.factor(PHQ9$JTClass)

#Create months variable and rename JTClass classes
PHQ9<-PHQ9%>%
  mutate(months=DaysafterIntake/30)

#Get the first observation for each patient
HeadPHQ9<-do.call("rbind", as.list(by(PHQ9, PHQ9$ResearchID,
                                       head,n=1)))

#Get the summary for each classification
PHQ9Sum<-HeadPHQ9%>%
  group_by(JTClass)%>%
  summarise("# of Patients"=n(), "Average Starting
    Score (SD)"=mean_sd(FirstScore, denote_sd = "paren",
                        digits=2),
    "Average Ending Score (SD)"=mean_sd>LastScore,
                                                denote_sd="paren",
                                                digits=2),
    "Average Difference (SD)"=mean_sd(Scorediff,
                                       denote_sd = "paren",
                                       digits=2))

#Reorder and rename column
PHQ9Sum<-as.data.frame(PHQ9Sum[c(2,1),])
colnames(PHQ9Sum)[1]<-c("Classification")

#Full model with all terms

```