

©2015

Laura Christine Skriner

ALL RIGHTS RESERVED

TRAJECTORIES AND PREDICTORS OF TREATMENT RESPONSE IN CBT FOR YOUTH
ANXIETY: AN INTEGRATIVE DATA ANALYSIS APPROACH

By

LAURA CHRISTINE SKRINER, M.S.

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Dr. Brian Chu

And approved by

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE DISSERTATION

Trajectories and Predictors of Treatment Response in CBT for Youth Anxiety:

An Integrative Data Analysis Approach

By LAURA C. SKRINER

Dissertation Director:

Dr. Brian Chu

Objective: To enhance knowledge of symptom change and treatment response in cognitive behavioral therapy (CBT) for youth anxiety by adopting an Integrative Data Analysis (IDA) approach. The study used an IDA framework to combine existing data from nine clinical trials of CBT for anxious youth ($N = 832$) and identify distinct trajectories of anxiety symptoms during and following treatment, as well as predictors of these trajectories. *Method:* Youth and parent-reported anxiety symptom data from the nine contributing clinical trials were combined using item response theory (IRT) models. Growth mixture modeling was used to identify distinct trajectories of treatment response using IRT-scored anxiety symptom data across four time points including: pre-, mid-, and post-treatment, and 1-year follow-up. Once identified, several pre-treatment client demographic and clinical profile traits were tested as predictors of trajectory classes. *Results:* Growth mixture modeling identified three trajectory classes based on parent-reported symptoms: steady responders (71.0%), rapid responders (7.2%), and delayed improvement (21.4%). Four classes were identified based on youth-reported symptoms: steady responders (55.1%), rapid responders (12.6%), delayed improvement (7.1%), and

low symptom responders (25.3%). Number of diagnoses, youth age, treatment type, and youth gender predicted trajectory class in both child- and parent-reported anxiety models. Delayed improvement classes were predicted by number of pre-treatment diagnoses (based on parent and youth report); receiving family versus individual CBT predicted membership in the delayed improvement compared to all other response classes and also in the steady responder compared to rapid responder class (based on youth report); rapid responders were predicted by older age (parent report) and number of pre-treatment diagnoses (parent report); and low symptom responders were more likely to be male (youth report) compared with those in the steady responder class. *Conclusions:* The use of an IDA framework allowed for the identification of distinct patterns of symptom change during and following CBT for youth anxiety that have not been previously identified in individual trials. Diagnostic complexity, age, gender, and treatment modality differentiated response classes.

Dedication

For my mom, who always has been the biggest supporter of my education, and whose unwavering love and support helped me to get through the difficult moments.

Acknowledgments

This dissertation would have never come to fruition without the support and encouragement from a great number of individuals, and it is with pleasure that I acknowledge their efforts. First, I am indebted to the investigators who graciously shared data with me from their clinical trials. This project could not have been conducted without the many years of work with anxious youth completed by Denise Bodden, Brian Chu, Philip Kendall, Maaïke Nauta, Wendy Silverman, and Jeffrey Wood. I am deeply grateful to my committee chair and doctoral advisor, Brian Chu. His guidance has made this a thoughtful and rewarding journey. I would also like to thank my dissertation committee members, Jami Young, Min-ge Xie, and Jimmy de la Torre, for their enthusiastic support of an ambitious study. These individuals provided a wealth of insightful comments and analytic guidance. In addition, I am grateful to Mehmet Kaplan and David Barker for their statistical consulting and support. I would also like to extend thanks to all of the members of the Youth Anxiety and Depression Clinic. Being surrounded by such caring friends and colleagues provided needed encouragement as well as comic relief to help get through graduate school. Finally, I want to thank my family for their ongoing love and support as I pursued my educational goals year after year.

Table of Contents

	PAGE
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
 CHAPTER	
1. INTRODUCTION	1
The Current State of Treatment Outcome Research	3
Defining Treatment Response	4
Examining Rates of Response and Non-response	5
Distinct Classes of Response and Non-response	8
Evidence for Response Classes in CBT for Youth Anxiety	11
Current State of Predictor Research	13
Methodological Issues and Examining Trajectory Classes	16
The Current Study	18
Study Aim 1	19
Study Aim 2	20
Study Aim 3	20
2. METHODS	21
Sample	21
Measures	24
Analytic Overview	28
Missing Data	29
Power	30
3. RESULTS	31
Phase I: Data Cleaning and IRT	31
Phase 2: Growth Modeling of Anxiety Symptom Trajectories	34
Parent Reported Anxiety Models	37

Youth Reported Anxiety Models	39
Phase 3: Predictors of Trajectory Classes	40
4. DISCUSSION	44
Identified Response Trajectories	45
Predictors of Response Trajectories	48
Study Strengths and Limitations	55
Conclusions	60
REFERENCES	62
ENDNOTES	70

List of Tables

	PAGE
Table 1. Characteristics of Included Clinical Trials	71
Table 2. Descriptive Statistics for Demographics and Pre-treatment Clinical Variables ..	72
Table 3. Overall Structure of Anxiety Symptom Data from each Trial	74
Table 4. Descriptive Statistics for Item Parameters from IRT Calibration.....	75
Table 5. Means and Standard Deviations for IRT Anxiety Scale Scores Across Trials ...	76
Table 6. Fit Indices for Growth Models.....	77
Table 7. Class Counts and Parameter Estimates for Trajectory Classes.....	78
Table 8. Predictors of Trajectory Classes – Multinomial Logit Models.....	79

List of Figures

	PAGE
Figure 1. Diagnostic response rates from post-treatment through 1-year follow-up from 18 clinical trials of CBT for youth anxiety.....	80
Figure 2. Latent growth mixture model trajectories for parent reported anxiety from pre-treatment through 1-year follow up.....	81
Figure 3. Estimated mean latent class growth trajectories and observed values for each of the 3 parent-trajectory classes	82
Figure 4. Latent growth mixture model trajectories for youth-reported anxiety from pre-treatment through 1-year follow-up	83
Figure 5. Estimated mean latent class growth trajectories and observed values for each of the 4 youth trajectory classes.....	84

CHAPTER 1

INTRODUCTION

Trajectories and Predictors of Treatment Response in CBT for Youth Anxiety:

An Integrative Data Analysis Approach

Anxiety disorders are among the most prevalent psychiatric conditions affecting children and adolescents, eclipsing both depression and behavior disorders in most studies (for reviews see Cartwright-Hatton, McNicol, & Doubleday, 2006; Curry, March, & Harvey, 2004). Recent epidemiological studies estimate a cumulative prevalence of anxiety disorders of 10% by 16 years of age (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). These disorders represent serious mental health problems for children and are associated with substantial psychosocial difficulties, including impairment in social relationships, academic difficulties, low self-esteem, and depression (Pine, 1997). If untreated, childhood anxiety disorders tend to be chronic in their course and predict the development of other psychopathology later in life (Bittner, Egger, Erkanli, Costello, Foley, & Angold, 2007). The successful treatment of youth anxiety has the potential to reduce the incidence of a range of psychiatric disorders and is thus a top health research priority.

Clinical research on psychotherapy for anxiety disorders for youth has advanced considerably over the past decade. Cognitive-behavioral therapies (CBT) have been the most rigorously studied psychotherapies for anxiety disorders and individual randomized controlled trials (RCTs) and meta-analyses consistently support the efficacy of CBT for anxiety-disordered youth (Silverman, Pina, & Viswvaran, 2008). Despite advances in treatment over the past decade the average remission rate of anxious youth receiving

CBT has been reported to be about 63% (Silverman, et al., 2008). Thus, a substantial number of children do not improve, even after receiving a course of an empirically supported treatment (EST).

Research that identifies unique classes of responders to treatment and the factors that predict those classes are essential to improving ESTs. To date, most predictor research has relied on analysis of mean group trends within single treatment samples. Such an approach has yielded highly inconsistent and inconclusive results. The limited heterogeneity present in individual treatment samples may be one factor limiting the field's ability to model predictor-outcome relations in a consistent way. For example, most treatment samples have limited ethnic diversity and restricted pre-treatment diagnostic complexity, and most treatment interventions produce a relatively narrow range of treatment outcomes (most participants typically improve) (Halliday-Boykins, Henggeler, Rowland, & DeLucia, 2004). The current study will take an integrative data analysis approach, (IDA; Curran & Hussong, 2009) making use of existing data from multiple clinical trials, to overcome the shortcomings of individual studies. Using item-response theory (IRT) to develop a common metric across measures from more than one study, IDA allows for increased statistical power, greater sample heterogeneity, a broader psychometric assessment of constructs, and the ability to estimate a variety of models that would not be possible within any single data set (Curran & Hussong, 2009). Within this framework, the primary goal of the current study was to identify distinct classes of trajectories of treatment responders and non-responders. The identification of classes of response can account for heterogeneity and provide a foundation to better identify

predictors of response. To this end, the secondary goal was to identify predictors of those trajectory classes.

The Current State of Treatment Outcome Research

CBTs have been the most rigorously studied psychotherapies for anxiety disorders in youth. Commonly used child CBT programs teach youth to recognize emotional and physiological signs of anxiety and to employ somatic and cognitive strategies for managing these symptoms, in addition to encouraging children to gradually expose themselves to increasingly feared stimuli. Individual RCTs and meta-analyses consistently support the efficacy of CBT for anxiety-disordered youth (Silverman et al., 2008). In the earliest efficacy studies, Kendall (1994) and Kendall et al. (1997) reported that 64% of youth receiving individual CBT (ICBT) no longer met criteria for their principal anxiety disorder by post-treatment compared to just 5% in the wait-list condition. In a recent meta-analysis of evidence-based treatments for youth anxiety, ICBT demonstrated an average success rate of 59% across trials (Silverman et al., 2008).

Adaptations to standard child-focused CBT have been made to address developmental considerations, such as the inclusion of parents in treatment (CBT/P) and different forms of family-based CBT (FCBT). While these formats have garnered evidence of success (e.g., Barrett, Dadds & Rapee, 1996; Bogels & Siqueland, 2006; Wood, Piacentini, Southam-Gerow, Chu, & Sigman, 2006) the bulk of evidence fails to find significant differences between ICBT, CBT/P, and FCBT (Ishikawa, Okajima, Matsuoka, & Sakano, 2007; Liber et al., 2008; henceforth, CBT/P and FCBT will be referred to as FCBT). Results from meta-analyses (Silverman et al., 2008) report overall mean response rates of 68% for treatments with parent involvement, suggesting that

FCBT is comparable to individual treatment. Despite advances in CBT for youth anxiety over the past two decades, not all youth respond to CBT, with numerous factors contributing to relapse or non-response to treatment. Therefore, an important step for improving the treatment of youth anxiety is to focus research on those youth who do not respond.

Defining Treatment Response

Summarizing the literature can be difficult because of the lack of clear consensus on what criteria constitute treatment response. This lack of consensus makes it difficult to accurately compare response rates across studies. Measuring treatment response can be difficult because response occurs on a continuum and is relative to the specific outcome measure used (Peris & Piacentini, 2008). For example, outcome can be measured by symptom severity, diagnostic status, or functional impairment. Further, these measures can be conceptualized as either categorical or continuous variables. Most treatment studies use multiple methods to measure treatment response, including change in diagnostic status (determined by a structured diagnostic interview), dimensional youth- and parent-report symptoms scales, and clinician-rated scales such as the Clinical Global Impression-Improvement Scale (CGI-S/I; Guy, 1976). These methods are not uniform across studies and different investigators use a variety of criteria to define success. When using diagnostic status as the main outcome measure, some studies define treatment response as the percentage of youth who no longer meet criteria for their primary diagnosis at post-treatment (e.g., Kendall, 1994; Kendall et al., 1997; Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008; Silverman et al., 1999a; Silverman et al., 1999b; Silverman, Kurtines, Jaccard, & Pina, 2009) while others define response as the

percentage of youth who no longer meet criteria for any anxiety diagnosis at post-treatment (Barrett, 1998; Barrett et al., 1996; Nauta, Scholing, Emmelkamp, & Minderaa, 2003; Spence, Donovan, & Brechman-Toussaint, 2000; Wood et al., 2006). Some studies report response-rates based on both definitions (e.g., Bodden, 2008; Liber et al., 2008). These differences in criteria of treatment response have made summarizing the literature challenging.

Youth- and parent-reported symptom measures are also used as an indicator of response. Sometimes, statistically significant change scores between pre- and post-treatment are interpreted as evidence of treatment response. Other times, investigators use normative cutoff scores from standardized measures to indicate when a youth returns to a non-deviant range of functioning (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999). Different measurement approaches and definitions have implications for which individuals are classified as responders and how the overall success rates of a treatment are summarized. These differences can also explain how comparing rates across studies can be misleading.

Examining Rates of Response and Non-response

Taking these varying definitions into account, the clinical trials literature reflects a complex story. Approximately 40 clinical trials of CBT for children and adolescents with phobic and anxiety disorders [excluding studies focused on obsessive-compulsive disorder (OCD) and post-traumatic stress disorder (PTSD)] have been reported in the literature. To inform hypotheses for the current study, a review of these trials was conducted to examine the rates of response across studies and treatment groups. The review included the 32 trials identified in the comprehensive review by Silverman and

colleagues (2008) and eight additional RCTs published since this meta-analysis was published. These studies include trials with either wait-list or active control groups and CBT across different modalities including ICBT, and FCBT, and group-based CBT (GCBT). In addition, the majority of studies tended to pool youth with social anxiety disorder (SAD), generalized anxiety disorder (GAD), and separation anxiety disorder (SEP), although a number of studies focused on a specific disorder, typically SAD. Response rate was most often defined in terms of diagnostic status, but studies differed in which criteria were used. Using the absence of primary diagnosis as the criterion, the review revealed post-treatment response rates ranging from 27% to 78.4% ($M = 55.0\%$; $N=21$). Using absence of all anxiety diagnoses as the criterion, post-treatment response rates ranged from 22.5% to 78.9% ($M = 54.2\%$; $N=14$), indicating a slightly larger range of response. Seven studies did not report diagnostic response rates. These success rates indicate that an average of 45% - 46% of youth could be classified as non-responders with a range of 21% - 77% non-responders across trials. These rates further highlight the substantial variation in response rates across treatment conditions and studies, suggesting heterogeneity in youth response to CBT. Figure 1 illustrates this heterogeneity in post-treatment response rates.

In addition to looking at post-treatment rates of response, it is important to look at response rates at follow-up assessment points to evaluate if treatment gains are maintained, deteriorate, or if youth continue to improve after treatment ends. First, diversity in the timing of follow-up assessment points was identified. The most frequent timing for follow-up was 12 months post-treatment, followed by three- and six-month follow-ups. Few studies reported multiple follow-up time points (e.g., Barrett et al., 1996;

Bogels & Siqueland, 2006; Silverman et al., 1999a; Silverman et al., 1999b; Spence et al., 2011). Second, reporting of diagnostic success rates at the follow-up assessments varied. Some studies did not provide the exact diagnostic response rates at follow-up; instead reporting whether or not treatment gains were maintained and whether there were significant differences in response between treatment conditions. Of the studies that did report diagnostic response rates during follow-up, average rates of youth with absence of principal diagnosis were 66.9%, 69.4%, and 73.8% at 3-month ($n = 5$), 6-month ($n = 5$), and 12-month ($n = 7$) follow-up, respectively. Range of response rates was 53-79% at 3-month, 50-85% at 6-month, and 60- 91% at 12-month follow-up. Figure 1 depicts response rates for 18 trials across post-treatment and follow-up assessments. Data suggest that substantial variation in response rates still exists at various follow-up assessments, but the range visually appears narrower than at post-treatment.

The picture may get more complicated when using symptom change as the metric for response. Nearly all studies included in the above review reported significant pre- to post-treatment reductions on anxiety symptom measures. Further, some studies reported continued significant improvement in anxiety symptoms from post- to follow-up (e.g., Barrett et al., 1996; Barrett, 1998; Hudson et al., 2009; Kendall et al., 2008; Silverman et al., 2009), while others reported non-significant changes (often reported as maintenance of treatment gains) (e.g., Bodden et al., 2008; Kendall, 1994; Kendall et al., 1997; Nauta et al., 2003). Symptom change may be critical to evaluate as some investigators reported dimensional symptom improvement, even amongst treatment recipients who did not demonstrate diagnostic change. For example, Kendall et al. (1997) reported that youth who still met criteria for their primary diagnosis showed significant improvement across

most parent- and child- symptom measures. In contrast, Bodden et al. (2008) reported that some portion of her sample showed no improvement at all, noting that 11% of their sample ($n = 14$) showed no improvement in severity of anxiety diagnosis at post-treatment.

In sum, this review highlights the lack of uniformity in criterion used for defining treatment response and in reporting styles. Second, this review provides evidence that diagnostic response rates and mean symptom levels from pre- to post-treatment and follow-ups indicate significant heterogeneity. From pre- to post-treatment, there are at least some youth who show symptom improvement but continue to meet criteria for at least one anxiety disorder while other youth show no improvement. From post- through follow-ups, there are youth who show additional symptomatic improvement, youth who appear to maintain level of response, and some youth who may worsen over the course of follow-up.

Distinct Classes of Response and Non-response

Despite the apparent heterogeneity in response to CBT, the majority of studies have focused on studying average tendencies by only reporting mean response (across individuals within a treatment condition). This approach places the focal point on the treatment condition and less on individual trajectories. However, such an approach might conceal subgroups of youth whose response trajectories differ from the overall mean pattern. It may be useful, therefore, to identify the range and nature of response patterns. No studies in youth anxiety have evaluated the possibility that there might be several qualitatively distinct trajectories of treatment response. There might be groups of individuals who recover quickly and stay symptom free, recover and relapse, partially

recover, or do not respond at all. Collapsing these disparate groups might lead researchers to overlook important patterns or draw inaccurate conclusions (Stein, Dickstein, Schuster, Litz, & Resick, 2012).

Moreover, examining distinct patterns of treatment response may yield the best returns in terms of identifying factors that contribute to non-response. As it stands, predictor analysis in the treatment literature has been highly inconsistent and inconclusive. One issue that limits the predictor literature is the possibility of multiple response classes. The above review suggests that there is heterogeneity at the study level, showing that whole treatment conditions vary in their response across time. The logical extension of this observation is that individual variation also likely exists within study and within treatment conditions. If distinct classes of individuals can be identified, then each class may have distinct predictors of treatment outcome. For example, those who exhibit less comorbidity or lower symptom severity at pre-treatment may recover quickly and maintain recovery whereas those who exhibit greater comorbidity or higher symptom severity may only partially improve. Thus, it may be important to identify distinct response classes first before examining which variable(s) predict treatment outcome as it may depend on what kind of response trajectory one follows.

Research with adult clinical samples provides examples. Some studies have estimated patterns of response with frequent symptom measures during treatment while others have estimated patterns using measurements at pre- and post-treatment and follow-ups. These studies have also focused on identifying predictors of response classes once classes are identified. Elliott, Biddle, Hawthorne, Forbes, & Creamer (2005) employed growth mixture modeling (GMM) to identify patterns of response for patients in CBT for

post-traumatic stress disorder (PTSD) across four time-points (intake, 6-, 12-, and 24-month follow-up). They found three nonlinear response patterns: two with moderate improvement (at different rates) and one with no improvement. In another study of patients receiving treatment for PTSD, Stein et al. (2012) used GMM to identify two linear latent class trajectories from symptom measures across nine time-points during and following treatment: responders (decreasing symptoms over time) and non-responders (persistent, high symptoms over time). Several psychopharmacology studies have also employed these methods. Gildengers et al. (2005) identified two response trajectories (cubic and quadratic) in elderly patients with major depressive disorder (MDD) who were treated with antidepressants and interpersonal therapy. The trajectories followed a similar course of decreasing symptoms, but differed in rates of decline. Gueorguieva, Mallinckrodt, and Krystal (2011) identified two trajectories (responders and nonresponders) of patients treated with antidepressants. These authors tested linear, quadratic, cubic, and piecewise models using bi-weekly data over eight weeks before settling on the piecewise models with linear change during the first two weeks of treatment and quadratic change after week two.

Two treatment studies with youth were identified. Halliday-Boykins et al. (2004) examined symptom trajectories among youth over 16 months following psychiatric crisis and identified five trajectories: high severity-improvement, high severity-unimproved, borderline clinical-improved, borderline clinical-unimproved, and subclinical. Four of the trajectory classes were linear and one was quadratic. In addition, these authors found that trajectory group membership (i.e., whether in an improved versus unimproved group) predicted whether a youth was placed out of the home (e.g., psychiatric hospitalization)

in the six-months following the final assessment point. Henderson, Dakof, Greenbaum, and Liddle (2010) identified two linear classes among adolescents receiving multidimensional family therapy for drug abuse and delinquency: low severity and high severity. Together, these studies demonstrate how assessing for multiple classes of symptom change can account for the heterogeneity of response among treatment samples and provide a foundation to enhance predictors research. In addition, they provide evidence that predictor variables can indeed distinguish response classes. For example, Stein et al. (2012) found that PTSD patients who met criteria for MDD or had high levels of hyperarousal at pre-treatment were more likely to belong to the non-responder group. Response class membership was also found to have utility in predicting distal outcomes (e.g., out-of-home placement; Halliday-Boykins et al., 2004). To better understand the heterogeneity in treatment response, especially among non-responders, and to ultimately increase confidence in predictor research, studies using these methods are needed in youth anxiety.

Evidence for Response Classes in CBT for Youth Anxiety

No youth anxiety CBT trials have assessed for multiple trajectory classes, but some have assessed the overall mean pattern of change during treatment and follow-up. Results from these studies provide suggestions for what trajectory classes may exist. In an RCT evaluating the efficacy of two exposure-based treatments for youth with phobic disorders, Silverman et al. (1999b) used Hierarchical Linear Modeling (HLM) to assess the mean patterns of pre- to post-treatment gains and maintenance at 3-, 6-, and 12-month follow-ups based on youth-reported measures. Results indicated a quadratic trend whereby youth showed a marked decrease in symptoms from pre- to post-treatment

followed by slight continued improvement before a “leveling off” by the last follow-up period. Between-group analyses revealed that children in one of the treatment conditions (exposure with contingency management) showed a slight relapse at 6-month follow-up before showing continued improvement at the 1-year follow-up. These results suggest that at least some youth might show symptom decline both during and following treatment but that a subgroup of individuals might show an increase in symptoms during the follow-up period. Silverman et al. only evaluated mean trajectories, but by conducting separate analysis for each treatment condition, they identified de facto response classes, each with distinct response trajectories. Had this approach not been used, these differences may have gone unrecognized.

Herbert et al. (2009) used HLM to examine trends in change over time within an RCT testing the relative effectiveness of ICBT and GCBT for adolescents with SAD. Results based on youth-rated symptom measures showed a significant decrease in symptoms from pre-treatment through follow-up with significant variability. The significant variability suggested individual differences existed in symptom course across participants and indicated a potential for distinct groups of youth within the treatment group. However, the authors did not conduct analyses to identify classes of trajectories to characterize different symptom courses.

Researchers have also started to examine overall mean trajectories of symptom change during treatment using multiple assessment points. The Child-Adolescent Multimodal Study (Walkup et al., 2008) compared CBT, medication, and their combination. Using data from four time points (Weeks 0, 4, 8, 12), HLM indicated a linear trajectory for CBT compared with placebo. Chu, Skrinker, and Zandberg (2013)

modeled the mean trajectory of change for youth who received 16-20 weeks of CBT using session-by-session symptom data. Results of HLM indicated a curvilinear cubic model suggesting a typical course of symptom change whereby anxiety declined relatively rapidly over the first half of treatment, flattened during early exposures sessions, and dropped again in later sessions. However, Chu et al. (2013) indicated that multiple individual trajectories and classes of trajectories were plausible. For example, significant individual variation existed around the mean cubic trajectory and qualitative independent rater coding identified four classes of symptom trajectory: linear (23%), quadratic (35%), cubic (33%), and no visible pattern (9%).

Together, these studies suggest heterogeneity in symptom change both during and after treatment among anxious youth receiving CBT. Silverman et al.'s (1999b) study identified de facto response classes by analyzing mean trajectories separately for two treatment conditions within the same study. Herbert et al. (2009) and Chu et al. (2013) identified significant individual variation around a single mean trajectory and Chu et al. identified qualitatively distinct response patterns that occurred during treatment. These studies suggest that a number of response classes are plausible (both within and after treatment) in individual CBT for anxious youth. Still, each study focused on analyzing mean trajectories, and the small sample sizes limit the confidence in any response classes that have been identified (e.g., Chu et al., 2013; Silverman et al., 1999b).

Current State of Predictor Research

Once distinct response categories have been identified, then predictor analyses can commence. Identifying predictors of outcomes can answer questions about which youth are likely to benefit from treatment and provide information about prognosis

(Southam-Gerow, Kendall, & Weersing, 2001). Predictor studies are common, but results are inconsistent. For every study that provides support for a specific factor predicting poor treatment outcome there is at least one other study showing opposite results. Even when predictors of outcome at post-treatment are identified, these same effects are often not replicated at follow-up assessments (see Hudson, 2005 for review). Predictor-outcome associations are replicated inconsistently mainly because clinical trials are typically powered to detect large effects (Compton et al., 2004). Interactions among predictors are also possible. For example, Barrett and colleagues (Barrett et al., 1996) found that younger children (aged 7 to 10) and female children were more likely to benefit from the additional family component compared to older, male counterparts. Evidence for predictors remains tentative however, due to the exploratory nature of these analyses and inadequate sample sizes. Therefore, further knowledge regarding the relationship between youth and clinical characteristics and differential patterns of treatment response is needed (Berman, Weems, Silverman, & Kurtines, 2000). The current study investigated several pre-treatment client demographic (age, gender, ethnicity) and clinical profile traits (number of diagnoses, comorbidity with either externalizing or depressive disorders) that have been associated with treatment outcomes. We also assessed whether treatment modality (FCBT, ICBT) was associated with different patterns of response.

Older age has been associated with poorer treatment response in CBT for youth anxiety (e.g., Southam-Gerow et al., 2001), but results are inconsistent. A recent study by Bennett et al. (2013) attempted to bring some clarity to the question of whether there are age differences in response to CBT by conducting an individual-level data meta-analyses

based on data from 16 CBT efficacy trials. Results showed that age did not moderate CBT treatment outcomes at post-treatment. These results are encouraging, especially given the rigorous methodology employed, and indicate that CBT can be equally effective for adolescents and younger youth. However, age effects at follow-up time points were not investigated.

Most studies investigating gender as a predictor of outcome have yielded non-significant results yet some have reported gender effects (see Nilson, Eisemann, & Kvernmo, 2013 for review). Several predictor and moderator studies found female gender to be associated with poorer treatment response (Cobham, Dadds, & Spence, 1998) while others found the opposite effect with males showing less response (Ost, Svensson, Hellstrom, & Lindwall, 2001). Studies examining racial and ethnic background as predictors are rare. The studies that have investigated these variables tend to show non-significant results (see Nilson et al., 2013 for review) with one study reporting a greater decrease in anxiety symptoms for European-American youths than for Hispanic youths (Pina, Silverman, Fuentes, Kurtines, & Weems, 2003).

Prior research exploring the effect of diagnostic comorbidity and clinical complexity on CBT outcomes reports mixed and somewhat contradictory findings. Some studies have found that increased diagnostic comorbidity negatively impacts treatment outcomes (Berman et al., 2000; Crawley et al., 2008; Liber et al., O'Neil & Kendall, 2012; Rapee et al., 2013), a few found that greater diagnostic comorbidity may actually be associated with more favorable outcomes, (MTA Cooperative Group, 1999; Rhode, Clarke, Lewinsohn, Seeley, & Kaufman, 2001), while many find no effect of comorbidity on treatment outcome (see Nilsen et al., 2013 for review). In terms of specific comorbid

conditions, depression has been most often associated with negative effects on anxiety treatment outcome (e.g., Berman et al., 2000; O’Neil & Kendall, 2012) while externalizing disorders, including attention deficit-hyperactivity disorder (ADHD) and oppositional defiance disorder (ODD), have typically not been shown to effect outcome (see Nilson et al., 2013 for review). With regard to treatment modality, both FCBT and ICBT have been shown effective at post-treatment and to produce lasting reductions in youth anxiety. However, studies comparing the relative efficacy of FCBT versus ICBT have yielded inconsistent results with several studies reported better outcomes for FCBT (e.g., Barrett et al., 1996; Wood et al., 2006) while others find no differences (e.g., Kendall et al., 2008; Nauta et al., 2003).

Methodological Issues and Examining Trajectory Classes

What remains unknown in the treatment literature is challenging for any single treatment study to address, in part because most clinical trials employ few measures, are often underpowered, produce un-interpretable statistically non-significant results, and produce results that are inconsistent from one clinical trial to another (Kraemer, Frank, & Kupfer, 2006). Multi-level trajectory research has rarely been completed due to limited sample size and heterogeneity of traditional RCTs with the majority of RCTs examining CBT for anxious youth having N’s under 100 (range: 35-488; only one study had more than 160). The latent modeling required to complete analyses testing for multiple trajectory classes generally requires a minimum sample size of 300 (Muthén, 2004). Beyond sample size, variance (heterogeneity) in both outcomes (response/non-response) and predictors (e.g., demographic and trait variables) are required for accurate classification and prediction (Kraemer et al., 2006). Most existing clinical trials include

relatively homogenous samples, relatively small samples, limited statistical power, and low base rates of subgroups (e.g., ethnic minorities, age groups, primary diagnosis; Hudson, 2005). In sum, individual studies rarely have large enough samples to be heterogeneous in terms of participants, settings, treatments, outcome measures, and methodology that are necessary for confident generalization (Kraemer et al., 2006).

The most noteworthy attempts to build a more robust, generalizable CBT knowledge base are studies that use meta-analytic techniques. Meta-analytic techniques are most commonly used to generate summary statistics based on group-level data and statistical tests (e.g., pooling effect sizes across studies.) This is useful when raw data from the original studies are not available (e.g., Ishikawa et al., 2007), but there are many advantages to fitting models directly to the original raw data (Cooper & Patall, 2009; Curran & Hussong, 2009). For instance, by increasing sample heterogeneity and sample size, a pooled sample allows for comparison of subgroups of individuals that are not possible within individual studies (Curran & Hussong, 2009). Indeed, several studies comparing the substantive conclusions reached by group- versus person-level meta-analytic methods have consistently favored the use of person-level data (e.g., Lambert, Sutton, Abrams, & Jones, 2002). Meta-analyses using person-level data have been shown to produce less biased and more precise effect sizes of treatment efficacy than group-level meta-analyses. Person-level meta-analyses have also identified subgroups in which treatment was either more or less effective that group-level analyses failed to detect. This enhanced detection is due to increased power associated with using person-level data and to avoiding ecological bias introduced by group-level analyses (Cooper & Patall, 2009). Despite the availability of these methods, very few studies in clinical psychology in

general, and only one in youth anxiety have utilized such methods. In youth anxiety, Bennett et al. (2013) recently conducted an individual patient data meta-analysis focused on identifying age effects in CBT for youth anxiety. Pooling of individual diagnosis severity data from multiple trials allowed Bennett and colleagues (2013) to more closely examine whether age moderates CBT effect size by testing an interaction between age and the effect of CBT. Their results provide evidence that individual-level meta-analyses can help clarify the mixed or differing results produced by individual studies. Moreover, group-level meta-analysis does not allow for certain research questions to be addressed. For example, identifying multiple trajectories of change during and following therapy require longitudinal multilevel modeling (MLM). Multiple raw data points are required to conduct such analysis.

The Current Study

In an attempt to overcome limitations of single studies and group-level meta-analyses, the current study aims to enhance knowledge of symptom change in anxiety CBT by adopting an IDA approach. IDA can be defined as “the statistical analysis of a single data set that consists of two or more separate samples that have been pooled into one” (Curran & Hussong, 2009, p. 82). IDA is not a single analytic technique but rather a global methodological approach to draw inferences from multiple studies. This approach is an alternative to meta-analysis using group-level summary statistics that is feasible when raw data are available and overlap exists in key design features, such as measures and samples, across independent studies (Curran & Hussong, 2009; Curran et al., 2008). In the context of the current study, by pooling across nine clinical trials of CBT for youth anxiety, our analyses benefit from a larger sample and allow multiple trajectories to be

modeled. A secondary goal of the current study was to test whether predictors of response trajectories can be identified.

The trials contributing to the current study have several methodological strengths that lend confidence to the use of an integrative approach and to the use of IRT models. Regarding sampling and measurement, all studies were designed to sample the same general population (youth presenting with a primary anxiety disorder), used the same diagnostic assessment, used multiple reporters of symptomatology (e.g., parent and youth), and demonstrate sufficient overlap in measurement (e.g., anxiety symptom measures) to allow linking of measurement across studies. In test linking, as long as each study shares a set of common items with at least one other study the scores can be “chained together” across studies (Bauer & Hussong, 2009). Regarding treatment characteristics, six of the nine studies used the Coping Cat treatment package (Kendall & Hedtke, 2006) as the ICBT intervention, while three studies used protocols similar in content to the Coping Cat. The FCBT did not follow a standard approach across studies; however, there was significant overlap in format (all used adjunctive parent involvement; one study included siblings) and content (youth-parent communication, parental behaviors, thoughts, and feelings towards youth, parental anxiety).

The current study pursued three specific aims:

Aim 1: To develop a common metric of anxiety symptom scores from measures used across multiple studies using an item response theory model: Measurement data from nine clinical trials ($N = 798$) were analyzed to develop consistent measurement of anxiety symptoms for youth and parent-rated measures. The items, which were calibrated at

baseline, allowed for person-specific scale scores to be created that incorporate information about study group membership and individual youth characteristics.

Aim 2: To identify classes of treatment response by modeling trajectories of symptom change during and following CBT for anxiety disordered youth using the IRT-derived scores. Growth mixture models were used to assess for multiple, distinct trajectories of symptom change. Trajectories were modeled across four time points: pre-, mid-, and post-treatment, and 1-year follow-up.

Hypothesis 1a. Distinct trajectory classes of anxiety symptom change over time will emerge for treatment responders and non-responders based on parent report.

Hypothesis 1b. Distinct trajectory classes of anxiety symptom change over time will emerge for treatment responders and non-responders based on youth report.

Aim 3: To examine the relationships between pre-treatment and treatment predictor variables and trajectories of treatment response and non-response.

Predictor variables assessed include demographics (age, gender, ethnicity), diagnostic characteristics (total number of diagnoses at pre-treatment, comorbidity with either externalizing or depressive disorders), and CBT modality (individual, family). Total number of diagnoses included all comorbidities, both additional anxiety and non-anxiety (depressive and externalizing) disorders. Externalizing disorders included ADHD and ODD.

Hypothesis 2. Older age, number of pre-treatment diagnoses, and comorbidity of depression will predict trajectories of non-response (i.e. youth with these characteristic will show little improvement in anxiety symptoms over time).

CHAPTER 2

METHOD

Sample

Data for this study were drawn from nine trials, eight completed RCTs of ICBT and/or FCBT, and one open clinical trial of ICBT for youth anxiety. Several recent comprehensive reviews and meta-analyses were reviewed to identify relevant trials (e.g., Cartwright-Hatton, Roberts, Chitsabesan, Fothergill, & Harrington, 2004; Compton et al., 2004; Silverman et al., 2008). Studies were also identified through online databases and by reference trails that resulted from identified studies.

To directly analyze pooled raw data and draw inferences across studies, overlap is needed in key study aspects, including chronological time, measurement, and design. Thus, trials were selected for inclusion based on the following criteria: a) study participants were selected for a primary Diagnostic and Statistical Manual of Mental Disorders-III-R (DSM-III-R) or DSM-IV/TR diagnosis of anxiety including, SEP, GAD, SAD, panic disorder (PD), or specific phobia (SP); b) the treatment intervention was ICBT or FCBT; c) sample age range was between 6 and 18; d) studies had person-level symptom data for at least two time points including pre- and post-treatment; e) studies had data for hypothesized predictors (e.g., age, gender, comorbidity status including number of diagnoses, and whether there was a comorbid depressive and/or externalizing disorder); and e) study investigators agreed to share their raw data for inclusion in the project. Exclusion criteria included: a) group-CBT interventions, b) inclusion of participants with subclinical anxiety, c) prevention trials, and d) primary diagnosis of OCD or PTSD. Six investigators agreed to share data from nine clinical trials, totaling an

overall sample of 832 youth who received short-term (between 12 and 20 sessions) ICBT ($n = 557$) or FCBT ($n = 275$). Table 1 describes individual study characteristics, including sample, design, procedural characteristics, and each sample's percent diagnostic-response rates. Full descriptions of study designs, sample characteristics, treatment protocols, and primary outcome findings can be obtained from the original clinical trial publications (Bodden et al., 2008; Chu, 2013; Kendall, 1994; Kendall et al., 1997; Kendall et al., 2008; Nauta et al., 2003; Silverman et al., 1999b; Silverman et al., 2009; Wood et al., 2006).

Regarding study design, all studies included an ICBT condition, five studies (Bodden et al., 2008; Kendall et al., 2008; Nauta et al., 2003; Silverman et al., 2009; Wood et al., 2006) compared ICBT to FCBT, three studies (Kendall, 1994; Kendall et al., 1997; Nauta et al., 2003) included a wait-list group, two studies (Kendall et al., 2008; Silverman et al., 1999b) included an active or attention control condition, and one was an uncontrolled open-trial (Chu, 2013). Initial IRT analyses used to pool anxiety symptom data across studies included all available data, including individuals in the control conditions. These individuals did not differ from their treatment group counterparts on pre-treatment variables (see Kendall et al., 2008; Silverman et al., 1999b) and allowed for a larger sample size. These individuals were later excluded from growth modeling analyses because they never received CBT. As indicated by inclusion criteria, all nine studies assessed anxiety symptoms at pre- and post-treatment. Four studies (Chu, 2013; Kendall et al., 1997; Kendall et al., 2008; Wood et al., 2006) assessed anxiety at mid-treatment while seven studies (Kendall, 1994; Kendall et al., 1997; Kendall et al., 2008; Nauta et al., 2003; Silverman et al., 1999b; Silverman et al., 2009; Wood et al., 2006)

assessed anxiety at 1-year follow-up. Diagnostic response rates from the five studies that defined response as absence of primary anxiety diagnosis at post-treatment ranged from 53% to 78% across treatment conditions. Diagnostic response rates from the three studies that defined response as absence of all anxiety disorders at post-treatment ranged from 28% to 54%.

Table 2 presents descriptive statistics for basic demographic and pre-treatment clinical characteristics for each individual study. Across all nine studies, 816 parents reported on youth anxiety symptoms and 824 youth self-reported anxiety symptoms for at least two time points. For parent-reported symptoms, mothers provided the majority of data. The full sample (either parent or youth-reported anxiety available) of 832 youth participants were 49.9% female ($n = 415$); 76.2% White, 13.1% Hispanic, 4.8% African American, 1.3% Asian, and 3.8% Other/Mixed. The majority (82.5%) were diagnosed with at least one additional disorder and the mean number of diagnoses was 2.78 ($SD = 1.29$). Of those with more than one diagnosis, 13.6% ($n = 113$) had a comorbid depressive disorder and 20.3% ($n = 169$) had a comorbid externalizing disorder (i.e. ADHD or ODD). Although study group membership was accounted for in all main analyses, descriptive and inferential analyses were conducted to assess for potential differences among studies on pre-treatment variables. In general, while there was some degree of heterogeneity among studies, no one study appeared markedly different from the others on the variables assessed. These results suggest that all nine studies could remain in analyses.¹

Measures

Anxiety Disorders Interview Schedule for Children (ADIS-IV) Child/Parent

Interviews. The ADIS-IV (Silverman & Albano, 1996) is a semi-structured interview that assesses presence and severity of DSM–IV–TR diagnoses. Diagnostic profiles include parent, child, and consensus diagnosis. Impairment (Clinician’s Severity Rating; CSR) is rated per disorder on a 0 (*not at all*) to 8 (*debilitating*) scale, where 4 represents clinical threshold. All studies contributing to the proposed study used the ADIS to determine presence and severity of diagnoses for inclusion of youth and as the primary measure of treatment response at post-treatment. The ADIS-IV has good to excellent reliability for specific diagnoses and symptom patterns as well as strong correspondence with youths’ anxiety self-ratings (Silverman, Saavedra, & Pina, 2001). Data from the ADIS were used for all comorbidity/severity predictor variables: number of diagnoses, presence of a depressive disorder, and presence of an externalizing disorder. Number of diagnoses was defined as the total number of all diagnoses, including comorbid anxiety, depressive, and externalizing disorders. Number of diagnoses was limited to five. This decision was based on the available data. Five studies allowed for five or more diagnoses while one study allowed for only four diagnoses.

Anxiety Symptom Measures for Primary Analyses

The primary outcome measure was anxiety symptom scores. All studies utilized at least one measure of anxiety symptoms and several studies utilized more than one measure. Across studies there were four distinct anxiety measures used (two measures were considered the same due to identical items). Table 3 provides a visual representation of the structure of the anxiety symptoms data available across study groups. All studies

shared at least one anxiety measure with at least one other study. Even when different measures were used item constructs are similar. With a pool of items that tap similar content it is feasible to develop a new set of standard scores that cut across studies and time (Curran et al., 2008). By utilizing IRT-based models, this study calculated a set of latent trait scores that are anchored on a standard metric across studies, as well as time, using the following measures:

Revised Children's Manifest Anxiety Scale – Child/Parent Versions

(RCMAS-C/P). The RCMAS-C/P (Reynolds & Richmond, 1978) are 28-item youth and parent report questionnaires. Each item is rated *Yes* (1) or *No* (0) and the items are summed to yield a Total Anxiety score. The RCMAS is the most widely used child self-rating scale in the youth anxiety treatment research literature and has been tested in large, nationwide samples of youth. Pela and Reynolds (1982) reported a 3-week test-retest reliability of .98 for the Total Anxiety scale and the RCMAS has been shown to have strong convergent validity with other measures of anxiety (Ollendick, 1983). Results of exploratory (Reynolds & Paget, 1981) and confirmatory (White & Farrell, 2001) factor analyses have provided support for a general anxiety factor, with confirmatory factor analysis (CFA) results providing evidence of a three-factor solution reflecting anxious arousal, social evaluation, and worry, with a higher order anxiety factor accounting for the correlations among factors (White & Farrell, 2001). White and Farrell (2001) also reported adequate internal consistency and evidence of measurement invariance across genders.

Multidimensional Anxiety Scale for Children – Child/Parent Versions

(MASC-C/P). The MASC-C/P (March, Parker, Sullivan, Stallings, & Conners, 1997) are

39-item youth and parent report questionnaires. The MASC-C/P asks how the child has been thinking, feeling, or acting over the last 2 weeks on a 1 (*never*) to 4 (*often*) four-point scale, with higher scores reflecting higher levels of anxiety. For the current study, we reduced the number of response categories to three by collapsing two of the categories because of sparse endorsement, which would introduce estimation problems and model instability during initial IRT analyses. The scale assesses four factors: physical symptoms, social anxiety, harm avoidance, and separation anxiety and a total score can be calculated. Multiple studies have assessed the psychometric properties of the MASC. Both youth and parent versions of the MASC have been shown to have excellent internal consistency of the total score and strong convergent validity with other measures of anxiety (Baldwin & Dadds, 2007; March et al., 1997; Muris, Merckelback, Ollendick, King, & Bogie, 2002). CFA results indicate that a second-order factor (i.e. an overall anxiety factor) accounts for the intercorrelations between the four factors (Baldwin & Dadds, 2007).

State-Trait Anxiety Inventory for Children–Trait-Child/Parent Versions (STAIC, STAIC-P). The STAIC-T (Spielberger, 1973) is a 20-item child self-report scale that measures enduring trait anxiety. The STAIC-T-P (Strauss, 1987) is a 26-item parent report version. Both are rated on a 1 (*hardly ever*) to 3 (*often*) three-point scale, with higher scores reflecting higher levels of anxiety. Both child and parent versions of the STAIC-T have been showed to have good to excellent internal consistency and evidence of the STAIC's discriminant and convergent validity has been reported (Muris et al., 2002; Southam-Gerow & Chorpita, 2007; Southam-Gerow, Flannery-Schroeder,

Kendall, 2003). Factor analyses on the child version supported the unidimensionality of the anxiety trait scale (Dorr, 1981).

Spence Child Anxiety Scale – Child/Parent Versions (SCAS-C/P). The SCAS (Spence, 1998) is 44-item self-report measure with 38 items designed to assess symptoms consistent with the DSM-IV anxiety disorders and seven items measuring social desirability. Items are rated on a 0 (*never*) to 3 (*always*) four-point scale, with higher scores reflecting higher levels of anxiety. For the current study, we reduced the number of response categories to three by collapsing two of the categories because of sparse endorsement, which would introduce estimation problems and model instability during initial IRT analyses. The scale assesses six factors that generally map onto the DSM-IV anxiety disorders: separation, social, and general anxiety, obsessive-compulsive, panic/agoraphobia, and fear of physical injury. The child version has shown high internal consistency, good convergent and divergent validity, and satisfactory test-retest reliability (Arendt, Hougaard, & Thastum, 2014; Nauta et al., 2004; Spence, 1998; Spence, Barrett, & Turner, 2003). The SCAS-parent version (SCAS-p) contains the 38 items assessing anxiety on the SCAS-c and has also demonstrated good psychometric properties (Arendt et al., 2014; Nauta et al., 2004). CFA provides support for both a single-factor solution and for a second-order factor of anxiety in general accounting for the intercorrelations among the six factors (Spence, 1998). Twenty-nine of the items on the SCAS are also included in the RCADS (see below). These overlapping items were used in the current study.

Revised Child Anxiety and Depression Scale – Child/Parent Versions (RCADS-C/P). The RCADS-C/P (Chorpita, Yim, Moffitt, Unemoto, & Francis, 2000)

are 47-item youth and parent report questionnaires of symptoms consistent with DSM-IV anxiety and depression symptoms. A Total Anxiety Scale score can be calculated from 37 of the items. Items are rated on a 0 (*never*) to 3 (*always*) four-point scale, with higher scores reflecting higher levels of anxiety. For the current study, we reduced the number of response categories to three by collapsing two of the categories because of sparse endorsement, which would introduce estimation problems and model instability during initial IRT analyses. The RCADS assesses five anxiety factors that map onto DSM-IV anxiety disorders: separation, social, and general anxiety, obsessive-compulsive, and panic/agoraphobic. The parent and child versions have demonstrated good internal consistency and strong convergent and discriminant validity (Chorpita, Moffitt, & Gray, 2005; Ebesutani, Bernstein, Nakamura, Chorpita, & Weisz, 2010; Muris et al., 2002), and longitudinal factorial invariance across a range of ages has been established (Mathyssek et al., 2013). CFA indicate that although a multi-factor model fits best, a model with a single, general anxiety factor also fit the data well (Chorpita et al., 2005). The RCADS was developed as a revised version of the SCAS. Thirty of the 37 anxiety items are identical to items on the SCAS. The seven different items assess excessive worry across several domains. The items from the SCAS that the RCADS does not include are those items ($n = 7$) assessing specific phobias and items ($n = 6$) comprising a social desirability scale. The 29 RCADS items that overlap with the SCAS were used in the current study.

Analytic Overview

Following recommendations from a recent series of papers utilizing and describing IDA (Curran & Hussong, 2009; Hussong, Flora, Curran, Chassin, & Zucker, 2008), data analyses proceeded in three primary phases. In the first phase, IRT analyses

were used to calculate anxiety scale scores across all individuals and time points. The process of creating a common scale is usually referred to as scale linking in IRT analyses. In the current study, linking was used to place the four separate measures onto a common metric. We linked the samples using “simultaneous (concurrent) IRT calibration” such that the single-population item parameters were estimated from the combined dataset containing data from all contributing studies (Wingersky & Lord, 1984). All IRT analyses were conducted with PARSCALE, a specialty IRT program. In the second phase of analyses, the IRT-derived anxiety scores were used to model trajectories of anxiety symptoms over four time points from pre-treatment through 1-year follow-up. These growth models were used to evaluate whether distinct trajectories of treatment response could be identified. Finally, in the third phase of analyses multinomial logistic regressions were performed to examine whether individual and treatment factors at pre-treatment differentially predicted to symptom trajectory classes. Analyses for phases 2 and 3 were conducted using Mplus version 7.0 (Muthén & Muthén, 1998-2012), a comprehensive latent variable modeling program. Multinomial logistic regressions were performed using the 3-step method available in Mplus to determine significant predictors.

Missing data. There were two reasons for missing data in the current study. First, there was missing data due to individual study design (i.e., not all contributing studies assessed anxiety at all four time points). Second, given the longitudinal nature of the data, participant attrition increased across time. Across the entire sample of 832 youth, 99.3% of the sample had anxiety data at pre-treatment, 33.2% at mid-treatment, 85% at post-treatment, and 45.4% at 1-year follow-up. Previous studies that successfully used IDA methods have also reported high rates of missing data at some time points due to the data

available from contributing studies (e.g., Hussong et al., 2008; Mun et al., 2015). For example, Hussong et al. (2008) pooled data from two studies to investigate the development of internalizing trajectories using data from ages 2 through 17, with only one study providing data for youth at ages 2 through 9 (i.e. one of the studies had “missing data” for half of the years included as time points). Indeed, being able to combine studies with differing data collection schedules is one of the advantages of IDA (Curran & Hussong, 2009). Missing data were handled differently at each phase of analyses. During the first phase, maximum likelihood estimation was used to handle missing data. The second phase of analyses (growth modeling) was conducted in Mplus, which also uses a maximum likelihood approach (full information maximum likelihood; FIML) to handle missing data. FIML allows the inclusion of participants with missing data on the dependent variable (anxiety scores across time) to be included in analyses. The third phase of analyses (predictor analyses) was also conducted in Mplus; however, because FIML estimation is not possible for missing data in predictors (Muthén and Muthén, 1998-2012), a few participants ($n = 2$) were excluded in the regression analyses.

Power. The current study features a large sample with repeated measures across time. With regards to estimation of trajectory classes, Muñoz and Acuña (1999) showed that sample sizes of around 500 were adequate for mixture analysis methods. Regarding predictor estimation, the sample size of the current study is over five times larger than the sample size of any included clinical trial. In addition, the use of person-level data allows for an increased sample size making possible investigation of finer-grain effects than those reported in each independent study.

CHAPTER 3

RESULTS

Phase 1: Data Cleaning and IRT

To accomplish Specific Aim 1, IRT test linking and equating procedures were used to develop a common metric of anxiety symptom scores using item level data from anxiety symptom measures (e.g., STAIC, MASC). This analysis proceeded in multiple steps including data cleaning and checking, item calibration at pre-treatment, and scoring of items across reporters and time points using results from item calibration.

The first step included checking of assumptions and preparing the data for IRT analyses. A core assumption of IRT models is unidimensionality, meaning that a single dimension underlies a set of observed items (Embretson & Reise, 2000). Due to the differences in response scales, it was difficult to assess the unidimensionality of the combined sample through a factor analysis of the combined set of items. Nevertheless, unidimensionality has been demonstrated separately for each of the five anxiety measures being combined. Factor analyses with the STAIC-T indicated that all items loaded on a single factor (Dorr, 1981). Although each of the other four measures has evidence of being multifactorial, CFAs have provided support for a second-order general (or overall) anxiety factor in each of these measures (for the RCMAS see Reynolds & Paget, 1981; White & Farrell, 2001; for the MASC see Baldwin & Dadds, 2007; for the SCAS see Spence, 1998; and for the RCADS, see Chorpita et al., 2005). Based on the results of these studies, unidimensionality of the entire, combined anxiety scale was assumed. Items from each study and each measure were then compiled, creating an 116-item scale of anxiety symptomatology comprised of 28 items from the RCMAS, 20 items from the

STAIC-T, 39 items from the MASC, and 29 items from the overlapping items of the RCADS and SCAS.

The next step was to calibrate the items to calculate item parameters for creating anxiety scale scores. We calibrated items using the pre-treatment parent-reported raw anxiety data (i.e., data from the distinct anxiety measures). Calibration included data from all available participants: those who received treatment as either part of a main treatment group or a wait-list control group as well as those in active or attention control groups (Kendall et al., 2008 and Silverman et al., 1999b included control groups). The results from the calibration were then used to create scale scores for the remaining time points for parent-reported data and for all time points for youth-reported data. We chose to use a common set of item parameters (from the parent calibration) so that parent and youth scale scores would be on the same metric, and thus scores would be comparable. Of note, we also calibrated data based on youth report and found similar overall results both in anxiety scores and when testing for trajectory classes in the second phase of analyses. We ultimately chose to use parent data for calibration based on the presence of greater variance in IRT-derived anxiety scores when using parent data compared to youth data. Alternatively, we could have calibrated scores separately for youth and parents. Justification for this approach could be made on the basis that parent and youth report of psychopathology are commonly discrepant (De Los Reyes & Kazdin, 2005). However, scores would not then be comparable between reporters. For this study, we chose to calibrate data on parent report so that parent and youth scores could be compared (and differences in trajectories interpreted) and because analyses on youth data resulted in comparable findings.

A generalized partial credit IRT model (Muraki, 1992), which allows for polytomous item responses, was used to calibrate the items. Analyses were conducted in PARSCALE-4.1 (Muraki & Bock, 2002). This program also allows for items to have different numbers of response categories. In the current study, initially 28 items were dichotomous (from the RCMAS) and 88 items had three response categories (from the STAIC, MASC, and SCAS/RCADS). With this IRT model, two types of item parameters are estimated: item discrimination/slope (α) and item difficulty/location (β). For items with more than two response categories, a separate difficulty parameter was estimated for each score category. Item characteristic curves (ICCs) are generated for each item based on item discrimination and difficulty. During initial calibration, ICCs and the distribution of responses across categories for items with more than two response categories were examined. Results indicated that some items with three response categories needed to be dichotomized because either response categories did not differentiate from one another or certain categories had too few responses. In total, 21 of the 88 items with three response categories were dichotomized: one item from the STAIC, 12 items from the MASC, and 8 items from the SCAS/RCADS. The partial credit model was then fit again, and new items parameters were estimated. During calibration, we assumed a mean of 0 and a standard deviation of 1 for the estimated latent trait scores. Descriptive summary statistics of item parameter estimates and their standard errors are presented in Table 4.

In the scoring phase, the item parameter estimates were used to calculate IRT-scaled anxiety scores for each participant's set of repeated observations based on his or her item responses. These parameter estimates were used to score the remaining time

points (mid- and post-treatment and 1-year follow-up) for parent-reported anxiety and all time points for youth-reported anxiety (pre-, mid-, and post-treatment and 1-year follow-up). We used the expected a posteriori (EAP) scoring method (Bock & Mislevy, 1982) for each participant's set of repeated measures to obtain the person parameter (theta). It is important to note that these scores are anchored to a shared underlying standard normal metric (Curran et al., 2008), despite the fact that we are combining separate samples in which individuals did not respond to all items. During item scoring, we relaxed the assumption of a mean of 0 and standard deviation of 1 to allow for variation associated with study group membership. These scores served as the observed dependent variables in the growth models described below.

Phase 2: Growth Modeling of Anxiety Symptom Trajectories

To accomplish Specific Aim 2, growth mixture modeling (GMM) was used with the anxiety IRT scores to evaluate whether distinct trajectories (i.e., subgroups) of anxiety symptom change (treatment response) over four time points from pre-treatment through 1-year follow-up could be identified. This analysis proceeded in multiple steps. After examining the distributions of IRT anxiety scale scores a series of unconditional growth mixture models were fit to the data. Analyses were conducted separately for parent- and youth-reports. We included both parent and youth report because multi-reporter assessment is typical and recommended in youth treatment studies (Achenbach, 2006; Barbosa, Tannock, & Manassis, 2002; Klein, 1991) given that disagreement between parent and youth reports of symptoms is common (De Los Reyes & Kazdin, 2005; Grills & Ollendick, 2002; Klein, 1991). Parent and youth reports of anxiety symptoms tend to be only modestly correlated (Achenbach, McConaughy, & Howell,

1987; Grills & Ollendick, 2002; Safford, Kendall, Flannery-Schroeder, Webb, & Sommer, 2005) but each reporter may contribute useful information about the youth's symptoms and functioning.

Mean values by study and for the full sample of IRT-scaled anxiety scores are presented in Table 5 for both parent- and youth-reports. Scores were generally centered around zero (there was slight variation due to the relaxed assumption of a mean of 0 and SD of 1 during the scoring phase); lower scores indicate lower anxiety and higher scores indicate higher anxiety. For parent-reported anxiety, scores ranged from -3.42 to 2.47 across time points. For youth-reported anxiety, scores ranged from -3.76 to 3.34 across time points. As expected, bivariate correlations indicated moderate associations between youth and parent anxiety scores. Correlations across pre, mid, post-treatment and 1-year follow-up time points were 0.32, 0.43, 0.39, and 0.33, respectively ($p < .01$ for all correlations). Prior to fitting growth models, the distributions of anxiety scale scores were examined across studies using histograms. In general, because of the IRT analyses, the distribution of scores in each study was approximately normal, with scores centered around zero. Skewness and kurtosis values were non-significant across all studies and all time points.

Based on the results reported by each of the contributing clinical trials and on initial examination (distributions) of anxiety scores, we decided to specify a model with a linear slope for the first three time points (from pre-treatment, through mid-treatment, and then to post-treatment) while allowing the fourth time point (1-year follow-up) to be freely estimated (i.e., there was no specification on the way in which the fourth time point related to the previous linear trajectory). In this way, 1-year follow-up anxiety could

continue with a downward trajectory or follow a more quadratic type of pattern.² To account for the nesting of participants within study, a dummy variable for study group membership was created and entered into all growth models. Three-level models were initially considered to account for study group membership but were not possible due to the relatively small number ($n = 9$) of studies (Hox, 1998; Hussong, Curran, & Bauer, 2013). As previously stated, FIML estimation was used to account for missing data.

Growth mixture modeling is a method built on the conceptual underpinnings of latent growth curve modeling (LGCM) and latent class analysis (LCA) that identifies classes/profiles of individuals that share a common growth trajectory over time for a target variable (e.g., anxiety score; Muthén, 2004; Nagin, 2005). This approach is similar to the former because individual values for a target variable are modeled longitudinally (i.e., each individual has his/her own growth trajectory); this approach is similar to the latter in that a categorical latent variable is created that represents individuals who share a common growth trajectory (and thus would be assigned to the same class) but differ from individuals who have been assigned to other classes. Thus, GMM refers to modeling with categorical latent variables (latent trajectory classes), which allow for different groups of individual growth trajectories to vary around different means (Jung & Wickrama, 2008).

To identify the number of classes a series of models with progressively greater number of trajectory classes were estimated separately for parent- and youth-reported anxiety. First, a one class solution was specified which served as a baseline model and estimated the average growth rate and shape across all participants. Then, models with 2 through 5 classes were estimated. It is recommended that a number of criteria be considered when deciding the number of latent trajectory classes (Muthén & Muthén,

2000). Thus, multiple indices of model fit were used. For model selection we relied on the sample-size adjusted Bayesian Information Criterion (ssaBIC; Tofighi & Enders, 2006), the Lo-Mendell-Rubin likelihood ratio test (LMR LRT) and the bootstrapped likelihood ratio test (BLRT; Nyland, Asparouhov, & Muthén, 2007). For the ssaBIC, smaller values indicate a better model fit. The ssaBIC can be used with nested and non-nested models and penalizes for model complexity. For the LMR LRT and BLRT a small p-value suggests that the model with k classes is preferred over k-1 classes. These fit indices were selected based on results and recommendations from simulation studies of class enumeration in mixture models (see Henson, Reise, & Kim, 2007; Nyland et al., 2007; Tofighi & Enders, 2006; Yang, 1998). The entropy of the classification was also taken into consideration. Entropy is used to quantify the uncertainty of classification of subjects into latent classes (Jung & Wickrama, 2008; Muthén & Muthén, 2000). Entropy values range from 0 to 1, with 0 corresponding to randomness and 1 to a perfect classification (Celeux & Soromenho, 1996). However, entropy tends to perform poorly in mixture models; it has been shown to be unreliable as a fit index with unbalanced class size (Henson et al., 2007). Thus, while we report entropy, we more heavily relied on the ssaBIC and BLRT when deciding on the number of classes. In addition to evaluating fit indices, the class size, interpretability, and theoretical/clinical justification of classes were considered (Muthén, 2004).

Parent-reported anxiety models. Results from the parent-reported one-class baseline model indicated that on average, there was a significant decline in anxiety symptoms in the initial slope from pre- through mid- and to post-treatment (slope = -0.42, SE = 0.02, $p < .001$) and from post to 1-year follow-up (slope = -0.20).³ There was

significant variance around the initial slope (slope variance = 0.05, SE = 0.10, $p < .001$) indicating variation around the mean growth trajectory. Such variation provides support for examining whether there are distinct groups of individual change trajectories that differ from the mean trajectory. As shown in Table 6, a 3-class solution had the most support based on our statistical criteria. The ssaBIC and BLRT supported the 3-class solution, although the LMR supported the 2-class solution. The ssaBIC and BLRT were weighted more heavily in decisions about number of classes based on results from recent simulation studies demonstrating that they outperformed (extracted the correct number of classes) a range of fit indices (Nyland et al., 2007; Tofighi & Enders, 2006). Entropy was better for the 2-class model; however, entropy has been shown to perform poorly with unbalanced class sizes (Henson et al., 2007), which were present in the current results. The 3-class solution was also supported based on our theoretical criteria including adequate class size and interpretability of classes. The 3-class solution class counts based on posterior probabilities were adequate (smallest was 7.2% of the entire sample) and the 3-class model was more interpretable than the 2-class model, in that it yielded a relatively small but interesting class of individuals who did not experience much decline in anxiety symptoms over the course of treatment; this class was obscured in the 2-class solution, which subsumed these youth into a larger class that contained youth who did experience decline in symptoms. Thus, inspection of the 3-class solution indicated that these trajectories appeared to make clinical sense.

The parameter estimates and class counts for the three latent classes identified for parent-reported anxiety are presented in Table 7. The classes were named based on the overall shape of the trajectory (see Figure 2): (1) *steady responders* (71.0%), whose

members tended to show gradual, steady decline in anxiety symptoms from pre- through post-treatment and then slowed in symptom change through 1-year follow-up; (2) *delayed improvement* (21.8%), whose members consistently scored high on anxiety symptoms from pre- through post-treatment but then showed some decrease in symptoms by follow-up; and (3) *rapid responders* (7.2%) whose members experienced a sharp and rapid decline in symptoms from pre to post-treatment followed by a slowing in symptom change through follow-up. Figure 3 displays the three trajectory classes with the estimated mean trajectory and observed individual trajectories around the mean curve.

Youth-reported anxiety models. Results from the youth-reported one-class baseline model indicated that on average, there was a significant decline in anxiety symptoms in the initial slope from pre- through mid- and to post-treatment (slope = -0.45, SE = 0.02, $p < .001$) and from post-treatment to 1-year follow-up (slope = -0.22).³ There was significant variance around the initial slope (slope variance = 0.05, SE = .10, $p < .001$) indicating variation around the mean growth trajectory. For youth-reported anxiety, the 4-class solution had the most support. Although the LMR was non-significant for the 4-class solution (thus, supporting the 3-class solution), the ssaBIC and BLRT indicated that a 4-class model fit the data best (Table 6). Entropy was again low, yet the difference between entropy values between the 3-, 4-, and 5-class models was relatively small. The 4-class model had adequate class sizes (smallest class was 7.1% of the total sample) and added in a class that appeared to make clinical sense; youth who reported relatively low symptoms throughout but who did show a decline in symptoms.

The parameter estimates and class counts for the four latent classes identified for youth-reported anxiety are presented in Table 7. The four classes identified for youth-

reported anxiety (Figure 4) were (1) *steady responders* (55.1%) whose members showed consistent, steady decline in symptoms across all four time points; (2) *low symptom responders* (25.3%) whose members appeared to start at lower levels of anxiety symptoms, show a decline in symptoms through post-treatment, and then maintain treatment gains through follow-up; (3) *rapid responders* (12.6%) who experienced a sharp decline in symptoms through post-treatment followed by a slight increase in symptoms at follow-up; and (4) *delayed improvement* (7.1%) whose members did not show much change in anxiety symptoms until a sharp decrease from post-treatment to 1-year follow-up. Figure 5 displays the four trajectory classes with the estimated mean trajectory and observed individual trajectories around the mean curve.

Phase 3: Predictors of Trajectory Classes

To accomplish Specific Aim 3, multinomial logistic regression analyses to assess potential predictors of trajectory class membership were conducted. Pre-treatment diagnostic variables and demographic factors that were available across all studies and that might differentiate between classes were investigated. These analyses were again conducted separately for parent- and youth-reported anxiety. Analyses were conducted within the growth mixture models that were determined to be the “best fitting” during the previous phase of analyses. For parent-reported anxiety this meant testing predictors within the 3-class model and for youth-reported anxiety within the 4-class model. Prior to predictor analyses we assessed for multicollinearity among pre-treatment diagnostic variables (total number of diagnoses, comorbid depressive disorder, comorbid externalizing disorder) due the inclusion of depressive and externalizing comorbidities within the total number of diagnoses variable. Correlations and tests of the collinearity

assumption indicated that multicollinearity was not a concern⁴ (Across parent and youth samples: Bivariate correlations range = 0.07 to 0.44, Tolerance range = 0.70 to 0.86, VIF range = 1.15 to 1.66, for the three diagnostic variables).

We used the “3-step method” (Asparouhov & Muthén, 2013) to estimate multinomial logistic regressions within the growth models. This 3-step approach was used instead of a standard 1-step approach (Asparouhov & Muthén, 2013). In the 1-step approach, after the optimal number of classes is chosen, a conditional model with predictors is then conducted. However, this conditional model may affect the latent class formation and the trajectory classes may change their meaning with the inclusion of covariates. In the 3-step approach, the logistic regressions are performed without allowing the covariates to affect the trajectories (Asparouhov & Muthén, 2013). In addition, classification uncertainty probabilities (i.e., measurement error) are incorporated into the multinomial logistic regressions. In this way, the latent trajectory classes remain unchanged from the original model, yet the classification probabilities (measurement error) are accounted for while the parameters for the multinomial regression are estimated. We used the R3STEP command in Mplus to automatically perform this 3-step method. Hypothesized predictors were entered simultaneously into the models. Predictors included pre-treatment clinical variables (total number of diagnoses, whether a comorbid depressive disorder, whether a comorbid externalizing disorder), a treatment modality (ICBT or FCBT) variable, and basic demographic variables including age, gender, and race. Due to limited racial diversity among the studies included only broad distinctions in race were possible (for adequate power); we dichotomized race as either White/Non-

Hispanic or Non-White (which included Hispanic, African American, Asian, and Other or mixed-race individuals).

For parent-reported anxiety, the total number of diagnoses at pre-treatment was a significant predictor of class membership (Table 8). First, number of diagnoses predicted membership in the delayed improvement class such that each additional diagnoses was associated with a 0.83 increase in the relative log odds of being the delayed improvement versus the steady responder class. Number of diagnoses also predicted membership in the rapid responder class, with each additional diagnosis associated with a 0.81 increase in the log odds of being in the rapid compared with the steady responder class. Membership in the rapid responder class was predicted by age, with a 1-unit increase in age associated with a 0.27 increase in log odds of being in the rapid versus the steady responder class. Thus, older youth were more likely to be in the rapid responder class compared to the steady responder class.

For youth-reported anxiety classes, treatment modality, gender, and the total number of pre-treatment diagnoses emerged as predictors (Table 8). Participant youth who received FBCT had an increase in the relative log odds of being in the delayed improvement class compared with the steady, rapid, and low symptom responder classes. Specifically, receiving FCBT was associated with a 1.21 increase in the log odds of being in the delayed improvement versus steady responder class, a 2.60 increase in the log odds of being the delayed improvement versus the rapid responder class, and a 1.40 increase in the log odds of being in the delayed improvement versus the low symptom responder class. In other words, those in FCBT were more likely to be in the delayed improvement class compared to all other classes. Those in FCBT also had a 1.39 decrease in log odds

of being in the rapid responders class compared to the steady responders class. Gender and number of pre-treatment diagnoses predicted membership in the low symptom responder class. Specifically, each additional diagnosis was associated with a 0.63 decrease in the log odds of membership in the low symptom versus the delayed improvement class. Being female was associated with a 0.72 decreased log odds of being in the low symptom versus the steady responders class. In other words, males were more likely and females were less likely to be in the low symptom class compared to the steady responders.

CHAPTER 4

DISCUSSION

The present study took an integrative data analytic approach to address questions about treatment response and predictors of response that are difficult for any one study to answer primarily due to inadequate power. Raw data from nine clinical trials of CBT for youth anxiety were pooled and examined for meaningful response trajectory classes. Once distinct classes were identified, several pre-treatment client demographic and clinical profile traits successfully predicted class belonging. The study is methodologically novel in several ways. It is the first to combine individual item-level symptom data from multiple trials of CBT for youth anxiety. It is also the first we know of to identify multiple trajectories of response in CBT for youth anxiety. Substantively, number of diagnoses, youth age, treatment type, and youth gender predicted trajectory class for either youth- or parent-reported anxiety. This knowledge can help clinicians identify at-risk clients and help prepare for clinical challenges. It can also serve as the foundation for research that identifies meaningful moderators of response across treatment types as well as mediators of change.

The study proceeded in three main phases. First, we used IRT models to pool the item-level data from four distinct measures of anxiety symptomatology and created anxiety scale scores that were in a common metric. Second, we used the IRT-modeled anxiety scale scores to model trajectories of symptom change (response) over four time points from pre-treatment to 1-year follow-up based on both parent and youth report. In the third phase, we assessed whether pre-treatment clinical and demographic variables differentially predicted to the response trajectories.

Identified Response Trajectories

As hypothesized, considerable heterogeneity was found for both parent- and youth-reported anxiety symptoms, which allowed for multiple trajectories of change to be identified. For both parent and youth models, we identified a class whose members tended to follow a pattern of steady response (steady decline in symptoms). In both parent and youth models, this class was the largest and included about half of the sample for youth-report and nearly three-fourths of the parent-reported sample. Hence, this class represents an “average” or typical class of individuals. A much smaller class in both models was a rapid responder class (7.2%, 12.6% in parent and youth samples, respectively), whose members evidenced an overall steeper rate of improvement from pre- to post-treatment compared to the other classes. These youth tended to start with relatively high anxiety levels but then show a great deal of change by post-treatment. We were also able to identify a class, delayed improvement, (21.8%, 7.1% in parent and youth samples, respectively), whose members showed less response to treatment overall, based on both parent and youth report. In the parent model, these youth tended to show no significant change in anxiety symptoms from pre- to post-treatment (the mean slope was non-significant) and then show some reduction in symptoms by 1-year follow-up. In the youth-reported model, this group tended to show a slight *increase* in symptoms by post-treatment (significant, positive slope) before showing a marked decline in symptoms by 1-year follow-up. The youth-reported model identified an additional class, representing a quarter of the sample, whose symptoms appear relatively low throughout the course of treatment. The youth in this class did show a reduction in symptoms over time, although the initial starting point of anxiety was approximately 1-standard deviation

below the steady responder class. Across both parent and youth models, with the exception of the delayed improvement classes, trajectories indicated that, on average, little change in symptoms occurred from post-treatment to 1-year follow-up.

Although this is the first study to model multiple response trajectory classes of symptom change, previous research supports the identified classes. Previous studies with adult PTSD samples have similarly identified between 2 and 4 classes of responders, with multiple “responder” classes characterized by different rates in response (e.g., Gidycz et al., 2005; Stein et al., 2012). Thus, it seems reasonable that we found multiple response classes (steady responders, rapid responders, low symptom responders) all showing improvement but at different rates. Specific to youth anxiety studies, the steady responders classes and the rapid responder classes could be expected given the reported response rates over time in CBT trials. It makes sense that the largest group was the steady responder class. Review of diagnostic response rates (see Figure 1) and mean trajectories (e.g., Chu et al., 2013; Silverman et al., 2009, Walkup et al., 2008) during and following treatment indicate that the majority of youth show some degree of symptom improvement over the course treatment. Similarly, previous results have shown significant improvements in symptoms on parent- and youth-reported measures despite youth still meeting criteria for their primary anxiety diagnosis (Kendall et al., 1997). Thus, it is possible that youth in the steady response class still meet diagnostic criteria for at least one disorder, yet still demonstrated symptom improvement.

That the responder (steady, rapid, low symptoms) classes indicated little change in symptoms from post-treatment to 1-year follow-up is in line with previous research. In reviewing response rates from trials of CBT for anxiety many suggested quadratic change

during follow-up periods with response rates either remaining stable or slightly decreasing (e.g., Barrett et al., 1996; Bodden et al., 2008; Kendall, 1994; Kendall et al., 1997; Nauta et al., 2003, Silverman et al., 1999b). The identification of the delayed improvement classes are somewhat new, but could also be expected. Bodden et al. (2008; trial included in the current study) reported that about 11% of the sample showed no improvement by post-treatment but significant change at follow-up, and multiple trials of CBT for youth anxiety have reported significant improvement in anxiety symptoms from post-treatment to follow-up (e.g., Barrett, 1998; Hudson et al., 2009; Kendall et al., 2008; Silverman et al., 2009). Thus, it seems reasonable that a group of youth showed little to no response at post treatment but then had a reduction in symptoms by 1-year follow-up. The low symptoms responder class also has not been identified elsewhere. That the rapid response, delayed improvement, and low symptom response classes were relatively small may have made them difficult to identify previously. Combining trials and using GMM enabled us detect different patterns of response. Knowledge of these different patterns can be used to inform clinicians that it is typical for symptom change to occur at differing rates and can help set expectations about change.

Two classes may be partially explained by floor effects (i.e., initial low symptoms). The youth-reported low symptoms class did not have much room for further reduction from the start of treatment and both rapid responder classes did not have much room for further reduction after post-treatment. In the “average”, steady responder class, there was less evidence of a floor effect indicating there may still be “room for improvement” among these youth.

We did not find a class where the youth showed no improvement by the 1-year follow-up. On one hand, this is encouraging in that even those youth who showed little to no response at post-treatment made gains by the 1-year follow-up assessment. On the other hand, it is possible that some of the non-responders dropped out of the studies by 1-year follow-up. Unfortunately, we were not able to investigate study attrition as a predictor of response trajectories. Yet another explanation is that anxious youth may show partial recovery over time without any intervention (i.e. regression to the mean) (Beesdo, Knappe, & Pine, 2009). Validation of these results in future studies is necessary to assess how stable the trajectory class definitions are.

Predictors of Response Trajectories

Pre-treatment number of diagnoses (including additional anxiety disorders, depressive disorders, and externalizing disorders) and treatment modality proved to be the most robust predictors of response class. Number of pre-treatment diagnoses can be seen as an indicator of clinical severity or complexity. Total number of diagnoses continued to act as a predictor even when all models were adjusted for comorbid depression and externalizing disorders. In both parent and youth trajectories, a greater number of pre-treatment diagnoses were associated with membership in the delayed improvement class. Intuitively, this suggests that more complex cases may respond to treatment more slowly or not at all because they may have more problem areas to address or the treatment protocols were not robust enough to affect this magnitude of pathology. Contrary to expectations, a greater number of pre-treatment comorbidities were also associated with placement in the rapid responder class compared to the steady responder class for parent-reported anxiety. This result was surprising. Of note, when diagnostic

variables (number of diagnoses, comorbid depression, comorbid externalizing disorder) were entered into the model separately during preliminary analyses, comorbid depression was a significant predictor of some response classes. When total number of diagnoses was included in the model, comorbid depression no longer remained significant. We chose to combine all comorbidities into one variable (number of diagnoses) as an indicator of severity. Future research should consider separating comorbid anxiety and number of anxiety disorders from non-anxiety comorbid conditions.

Prior research exploring the effect of diagnostic comorbidity and clinical complexity has found equally contradictory findings, where some studies have found that increased diagnostic comorbidity negatively impacts treatment outcomes (Berman et al., 2000; Crawley, Beidas, Benjamin, Martin, & Kendall, 2008; Liber et al., 2010; O'Neil & Kendall, 2012; Rapee et al., 2013) and other studies find no effect of comorbidity on treatment outcome (see Nilsen et al., 2013 for review). For instance, Rapee and colleagues (2013) recently reported on the effects of comorbidity on the response to CBT in anxious youth among a large sample (with an N over 700). They explored potential differences in treatment response at post-treatment and at a follow-up time point based on dividing youth into comorbidity groups (no comorbid, comorbid anxiety only, comorbid depression, and comorbid externalizing). The “good” news was that their group by time interactions did not suggest any negative effects of comorbidity; even youth with comorbid disorders showed improvement. However, they also reported a more complicated picture where youth with comorbidities reported higher levels of symptom severity and were less likely to be free of their primary anxiety disorder at post-treatment and follow-up. Rapee et al. (2013) interpreted results to indicate that both comorbid and

non-comorbid youth improved at a similar rate but that youth with complex diagnostic profiles may need additional treatment since these youth initiate treatment as more severe. In line with Rapee et al.'s (2013) results, our findings indicate a complex picture, where some comorbid youth show improvement, but to a lesser degree than do those with fewer comorbid disorders. In the current study, diagnostic complexity was associated with placement in the delayed improvement group. Like Rapee et al., highly comorbid youth showed resistance to complete improvement even as youth improved to a modest degree from post-treatment to 1-year-follow-up.

In contrast, other research has suggested that greater diagnostic complexity may actually be associated with better outcomes compared to less impaired youth. For example, Rhode et al.'s (2001) study of adolescent depression and the Multimodal Treatment Study of Children with ADHD (MTA; MTA Cooperative Group, 1999) found better outcomes (greater decrease in targeted symptoms) associated with comorbid anxiety disorders. In the current study, diagnostic complexity predicted inclusion in the rapid responder class, indicating that these youth were improving rapidly and to a greater degree than the average class. Analytic reasons may explain these results in part. Youth with greater pathology may start at a higher initial severity (Jenson Doss & Weisz, 2006; O'Neil & Kendall, 2012; Rhode et al., 2001), providing more room for improvement over the course of therapy. Thus, placement in the rapid responder class may, in part, owe its steepness to regression to the mean. However, other researchers have suggested that youth with greater clinical severity may come to treatment with greater motivation for change, greater awareness for the necessity of treatment, and greater retention in treatment (Jenson Doss & Weisz, 2006).

Together, the literature and this study's results present a complex picture for the impact of comorbidity on treatment outcomes. Findings seem to indicate that youth with greater clinical complexity (i.e. multiple diagnoses) are unlikely to respond to CBT in the typical, average way. In this way, a more complex diagnostic profile can potentially be a useful indicator of response trajectory. It is critical for clinicians to be aware of the potential for each path. Clinicians can use this knowledge to make decisions about treatment planning. First, those who are likely to show less or delayed improvement could be identified early by an extremely slow or nonexistent decrease in symptoms and changes in either the treatment approach or treatment dosing may be useful. That the comorbid youth in the delayed improvement classes did eventually show a decline in symptoms may suggest that a higher dose of treatment (longer or more frequent treatment sessions) could help lead to improvement more quickly. A longer treatment course or "booster" sessions might also enhance response for more diagnostically complex youth as it appears that many who initially are slow to improve do show a decrease in symptoms by 1-year follow-up.

Comorbidity also serves as an important marker for future research to disentangle the ways in which diagnostic complexity can lead very different treatment responses and what factors account for some comorbid youth showing rapid response while others show little or slower response. There may be client-level factors such as motivation or specific coping styles as well as clinician behaviors associated with better or worse response patterns. Though unable to investigate coping styles here, Chu et al. (2013, one of the included trials) found an effect of pre-treatment coping style on the mean trajectory of change during treatment. Such factors could be specifically targeted early if they were

known markers of a non-response or delayed improvement trajectory. It is important for researchers to continue investigating the effects of treatment and specific clinical practices on comorbidity given that such youth are a truly heterogeneous group. Future research would also benefit from evaluating the co-occurring trajectories of comorbid depressive and externalizing symptoms along with anxiety symptoms. This could shed light on the relationship between how changes in anxiety symptoms covary with symptom change in comorbid conditions and whether this relationship has an effect on trajectory class membership.

Treatment modality (individual or family treatment) also had a strong effect, though only based on youth-reported anxiety. In the youth model, receiving FBCT significantly increased the probability of membership in the delayed improvement class compared to all other classes and increased the probability of being in the steady responder class compared with the rapid responder class. This finding is somewhat different than recent meta-analytic results showing no differences in the effectiveness of treatment based on level of parent involvement (Reynolds, Wilson, Austin, & Hooper, 2012). At the same time, while the current finding may seem to suggest that FCBT is less effective than ICBT, there are two important caveats. First, the members of the delayed improvement class in the youth model show substantial decline in symptoms by 1-year follow-up, indicating that on average these youth do seem to show some level of response. Second, this class was the smallest and was only 7.1% of the sample. Results can also be viewed in terms of the fact that treatment type was only a predictor based on youth report. Not surprisingly, these differences among reporters are similar to findings from Wood et al. (2006, 2009; a trial included in the current study sample) as well as

other studies comparing family and individual CBT (Barratt et al., 1996). Specifically, Wood et al. (2006, 2009) reported that FCBT outperformed ICBT at post-treatment and 1-year follow up on some diagnostician and parent-report scales, but not on child-report. Wood et al. (2009) provided several possible interpretations for this difference in reporting: the difference could indicate that the interventions have comparable effects on children's experiences of anxiety and may also reflect the limitations of self-report assessments in school-age samples, but it could also indicate that parents were biased based on participation in the treatment. The same interpretations can be applied to the current results. There is some distinction between the current findings and those reported by Wood et al. (2006, 2009) in that in the current study FCBT predicted less response compared to greater response trajectories (rather than equivalent response to FCBT and ICBT). However, we did not take a moderator approach and specific differences in efficacy between ICBT and FCBT cannot be made.

At the same time, the sample size and inclusion of multiple trials in the current study may have helped identify response patterns affected by parent involvement in treatment that individual studies could not. The current findings may have identified new trends worth note. For example, most of the ICBT protocols included in the current analyses did permit (and even encouraged) some degree of parent involvement over the course of therapy. ICBT protocols simply did not mandate consistent parental involvement. Most ICBT protocols mandated 1-2 parent sessions, but then typically left further parent involvement up to the clinician and supervisor. In this way, a main distinguishing feature between treatment modalities was whether to allow parental involvement to be mandated or individually considered. Though speculative, the decision

to mandate parent involvement in treatment may be more complex than originally conceived. The present results suggest that the decision to include substantial parent involvement may benefit from taking a more individualized, case conceptualization approach. Using ICBT as a base, there may be important times when a clinician should limit a parent's involvement in treatment (e.g., to enhance client privacy and autonomy). Such a decision may lead to better (faster, greater magnitude) response for a small subset of youth, or prevent the occurrence of a delay in improvement, than mandated family involvement.

Gender was also a significant predictor in the youth-reported model. Results show that males tended to be in the low symptom responder compared with the steady responder class. In line with the majority of studies (see Nilsen for review, 2013), this finding does not indicate gender differences in response overall to CBT for anxiety; as individual in both the steady decreasing and low decreasing classes showed improvement in symptoms by post-treatment. This finding does suggest that males reported somewhat less anxiety than their female counterparts (i.e., reported a lower level of anxiety at the start and throughout treatment). This finding seems consistent with past research showing that males are less likely to manifest anxiety (Albano & Krain, 2005) and at lower levels than females (Carter, Silverman, & Jaccard, 2011).

We found that for the parent-reported anxiety trajectories, older age was associated with membership in the rapid responder class compared with the steady responder class. Although some individual studies investigating age as a predictor or moderator have found the reverse, with older age associated with poorer treatment response (Southam-Gerow et al., 2001), recent meta-analyses have reported mixed

findings, albeit with somewhat different samples of studies. Bennett et al. (2013) made use of individual level data to test age as a moderator in CBT for youth anxiety and found no effect of age. Another recent meta-analysis (Reynolds et al., 2012) of psychotherapy for youth with anxiety found that the degree of symptom change reported by adolescents was actually greater than that reported by younger children, a finding similar to the current one. However, this second meta-analysis (Reynolds et al., 2012) did include somewhat of a different sample of studies than the current study or the study by Bennett and colleagues (2013) in that they included trials that were focused on youth with OCD and trauma-related disorder in addition to other anxiety disorder (e.g., GAD, SAD). These differences notwithstanding, current results together with previous research seem to make it clear that anxious adolescents are not at a particular risk of poor response to CBT. In addition, although the mechanism is not known, at least some older youth appear to have the potential to show a quicker, more rapid response. While only speculative, this may be because older youth are better able to engage in psychotherapy in general, or might have greater cognitive ability to understand and try out some of the coping skills learned (e.g., cognitive restructuring) within the CBT framework.

Study Strengths and Limitations

The present study has a number of strengths that enhance its contribution to the literature. First, there were several methodological strengths owed to taking an IDA approach including the pooling of data across multiple trials, which allowed for a large sample size; combining data not only at the individual but also at the item level; and integrating a person-centered approach, which models heterogeneity in trajectories, with a variable-centered approach, which permits the prediction of class membership from

salient demographic and clinical factors. In addition, we included symptom report from both youth and parents. These strengths allowed for the identification of multiple trajectories of treatment response and for more reliable conclusions about predictors of treatment response to be made. We are also the first study to implement IRT models to combine different youth anxiety symptom measures, which strengthens the measurement of the anxiety construct and allows for direct comparison of anxiety symptom levels across studies that employed different measures. The application of IRT models within an IDA framework provides support for the use of these methods in future studies. Finally, the differential response patterns highlight the benefits of going beyond overall group means when investigating treatment outcomes.

At the same time, these strengths are tempered by several limitations. First, we examined a limited range of predictor variables. Only a small number of potential predictor variables were assessed in similar ways across all of the included trials. Thus, we were only able to look at basic demographics including gender, age, and race as well as basic clinical and treatment characteristics including comorbidity and treatment modality. Other factors that could have been examined include socioeconomic status, household composition, and parental psychopathology. Parental psychopathology, in particular, has been associated with both reduced treatment response (Cobham et al., 1998; Legestee et al., 2010) and more favorable response (Legestee et al., 2010) among anxious youth samples. Additional clinical and early treatment predictors of interest include pre-treatment coping styles, psychoactive medication, therapeutic alliance, and treatment attrition. Further, although we did include race as a predictor, we were only

able to assess White/Non-Hispanic versus Non-white groupings. Despite only including a limited number of predictors, meaningful associations were detected.

Second, we investigated response trajectories based on self-reported symptom measures. We could also have investigated trajectories based on diagnostic status as the outcome variable. Indeed many studies primarily define treatment response in terms of diagnostic status. Although basing trajectory analyses on symptom measures made it possible to include more time points than if we had relied on diagnostic status (there was not available diagnostic data for mid-treatment), conducting trajectory analysis on diagnostic status in future research could allow for comparisons between types of outcome measures as well as for response rates based on diagnostic status as reported in clinical trials.

Third, although the use of symptom measures allowed for additional time points to be included, we were still limited in the number of time points available for inclusion in growth analyses. Recent investigations of mean trajectories with more frequent time points, especially more frequent time points during treatment (e.g., weekly assessments suggest non-linear patterns of change (Chu et al., 2013) during CBT for anxiety. In the current study we were limited to specifying a linear trajectory from pre- to post-treatment, when in fact, symptom change during treatment may be more complex. Indeed, when examining the trajectory classes identified (see Figures 2 and 4), there remains variation around each class, most notably for the largest class (steady responders) in both models. This variation hints that with more time points, additional classes might have been identified. Nevertheless, we were able to identify several clinically meaningful patterns of change with only four time points. Fourth, there was a

substantial portion of missing data at the mid-treatment and 1-year follow up time points due to the data available from contributing studies and to treatment attrition. Though missing data were handled using FIML, conclusions about trajectories at 1-year follow up would be strengthened if there were data available at all time points from all contributing studies.

Fifth, the included trials were somewhat heterogeneous in their study design (e.g., open-trial versus RCT, type of treatment comparison group), sample size, and available time points. A certain degree of heterogeneity is needed to detect latent classes and predictors and can lead to increased confidence in the generalization of findings. To increase the likelihood that studies were adequately comparable the included studies were similar to one another in many key ways (clinical severity and level of comorbidity at treatment entry, selection criteria and assessment methods [all included structured diagnostic interviews], treatment structure and adherence monitoring, training and supervision models of study therapists) and we used statistical analyses (GMM) that are designed to deal with heterogeneous samples. Also, the level of heterogeneity allowed us to specify well-fitting models that produced interpretable classes as well as identify reliable predictors.

However, if there is too much variation between studies, it may not seem warranted to combine them and consider them as one group. The heterogeneity in the current study is similar to that found in meta-analyses (e.g., Bennett, 2013; Silverman, et al., 2008), but conclusions should remain tentative. For instance, there were fewer youth who received FCBT compared to ICBT across studies and there was some diversity in the family treatment protocols. Though study group was accounted for, it is possible that one

or more studies accounted for the results. In particular, one study, Bodden et al. (2008) accounted for 23% of all family cases and reported a lower than expected response rate (28%). Analysis was conducted to determine if analyses comparing FCBT and ICBT were impacted by this disproportionate representation. Follow-up analysis indicated that cases from Bodden et al.'s (2008) sample did have the highest proportion (29%) of representation in the delayed improvement class (in the youth model), yet cases from other contributing studies were also represented, with the other four FCBT trials representing between 3.2 and 22.6% of the delayed improvement class. Thus, although the Bodden et al. sample likely significantly contributed to the results of the FCBT and ICBT comparison, cases from other studies also showed a delayed improvement pattern of symptom change. Future studies could conduct sensitivity analyses to more thoroughly assess whether any one individual study biased predictor analyses. Similarly, we were not able to account for potential differences in specific number of treatment sessions between individuals. Treatment duration may be an important predictor of response trajectory. While included studies were similar in planned number of sessions, there was some variation. The majority of studies specified treatment protocols with between 12 and 16 sessions, though one study's (Silverman, 1999b) protocol specified 10 sessions and another (Chu, 2015) specified between 16 and 20 sessions. Further, even within a study, we do not know whether all youth received the planned number of sessions.

Sixth, this study is still one step short of providing information that could guide treatment-matching decisions. For example, knowing which interventions would best help youth who fell into different response classes would provide clinicians with a powerful tool. The current study had a limited number of treatment interventions to study

(only ICBT and FCBT) and did not conduct moderator analyses (e.g., intervention * response class) as a central aim of the current study was to assess for multiple trajectories of symptom change and combining across treatment groups allowed for large sample size required for such analysis. Still, the current results take an important first step by outlining different trajectories that can inform clients and therapists about prognosis. Clinicians can use this knowledge to set proper client expectations and to recognize positive and negative response trajectories early on. It would be yet another step to research what interventions work best to intervene with poor response classes. It is also important to remember that there is not 100% accuracy in class membership (i.e. there is still variation around each trajectory class), and results need to be interpreted with caution. Future research will benefit from identifying moderators of response trajectories across treatment types to strengthen the current findings.

Finally, ideally more studies would have been included in order to include study group as its own nested level in the growth models and for study level to be estimated as a random effect. Traditionally 20-30 studies would be required to test nesting effects of study; however, we did include dummy codes for study in all growth models, which accounted for the individual variation across studies.

Conclusions

The current study provides a rigorous examination of trajectories of response to CBT for youth anxiety based on data drawn from nine studies and from both youth and parent report. The results suggest that heterogeneity exists with regard to patterns of symptom change during and after CBT for youth anxiety. We identified clinically meaningful subgroups of youth including a group of youth who show delayed

improvement following CBT. We also found subgroup differences on several pre-treatment clinical and demographic factors. Predictor results provide some clarity to predictors of response and non-response and also raise some questions for future research. Specifically, greater diagnostic complexity was associated with classes of delayed improvement (based on both parent and youth report), as was inclusion in an FCBT treatment group (by youth report only). Results also indicated that according to parent report, older youth were more likely to be in a rapid response class and that male youth were more likely to rate low levels of symptoms during and following treatment. These results represent an initial step toward identifying individuals at the start of treatment who would require additional sessions, more intense services, or alternative interventions. Conducting moderator analysis is an important next step to help identify which interventions would best serve youth who fall into different response classes. However, the current findings provide clinicians and clients with valuable information about prognosis.

In addition, future research including a greater number of studies would strengthen current results and generalization of findings. While results provide evidence of heterogeneity in response patterns, they also raise questions regarding additional factors that may account for the observed differences. Future research would benefit from the inclusion of more time points, though few existing trials have frequent assessments, as well as the inclusion of time-varying covariates within a growth modeling approach to examine the covariation and causal relationships among response trajectories and proposed mechanisms of change.

References

- Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15*, 94-98.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.
- Albano, A. M., & Krain, A. (2005). Anxiety and anxiety disorders in girls. In: D. J. Bell, S. L. Foster, & E. J. Mash (Eds.), *Handbook of behavioral and emotional problems in girls (79-116)*. New York: Kluwer Academics / Plenum Publishers.
- Arendt, K., Hougaard, E., & Thastum, M. (2014). Psychometric properties of the child and parent versions of Spence Children's Anxiety Scale in a Danish community and clinical sample. *Journal of Anxiety Disorders, 28*, 947-956.
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. Mplus web notes: No. 15, Version 8, August 5, 2014. Retrieved from <https://www.statmodel.com/download/webnotes/webnote15.pdf>.
- Baldwin, J. S., & Dadds, M. R. (2007). Reliability and validity of parent and child versions of the Multidimensional Anxiety Scale for Children in community samples. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 252-260.
- Barbosa, J., Tannock, R., & Manassis, K. (2002). Measuring anxiety: Parent-child reporting differences in clinical samples. *Depression and Anxiety, 15*, 61-65.
- Bennett, K., Manassis, K., Walter, S. D., Cheung, A., Wilansky-Traynor, P., Diaz-Granados, N., . . . Rice, M. (2013). Cognitive behavioral therapy age effects in child and adolescent anxiety: An individual patient data metaanalysis. *Depression and Anxiety, 30*, 829-841.
- Barrett, P. M. (1998). Evaluation of cognitive-behavioral group treatments for childhood anxiety disorders. *Journal of Clinical Child Psychology, 27*, 459-468.
- Barrett, P. M., Dadds, M. R., & Rapee, R. M. (1996). Family treatment of childhood anxiety: A controlled trial. *Journal of Consulting and Clinical Psychology, 64*, 333-342.
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric Approaches for Developing Commensurate Measures Across Independent Studies: Traditional and New Models. *Psychological Methods, 14*, 101-125.
- Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: Developmental issues and implications for DSM-V. *Psychiatry Clinics of North America, 32*, 483-524.
- Berman, S. L., Weems, C. F., Silverman, W. K., & Kurtines W. M. (2000). Predictors of outcome in exposure based cognitive and behavioral treatments for phobic and anxiety disorders in children. *Behavior Therapy, 31*, 713-731.
- Bittner, A., Egger, H. L., Erkanli, A., Costello, J. E., Foley, D. L., & Angold, A. (2007). What do childhood anxiety disorders predict? *Journal of Child Psychology and Psychiatry, 48*, 1174-1183.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 34*, 197-211.

- Bodden, D. H., Bogels, S. M., Nauta, M. H., De Haan, E., Ringrose, J., Appelboom, C.,... Appelboom-Geerts, K. (2008). Child versus family cognitive-behavioral therapy in clinically anxious youth: An efficacy and partial effectiveness study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 1384-1394.
- Bogels, S. M., & Siqueland, L. (2006). Family cognitive behavioral therapy for children and adolescents with clinical anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 134-141.
- Carter, R., Silverman, W. K., & Jaccard, J. (2011). Sex variations in youth anxiety symptoms: Effects of pubertal development and gender role orientation. *Journal of Clinical Child and Adolescent Psychology*, 40, 730-741.
- Cartwright-Hatton, S., McNicol, K., & Doubleday, E. (2006) Anxiety in a neglected population: Prevalence of anxiety disorders in pre-adolescent children. *Clinical Psychology Review*, 26, 817-833.
- Cartwright-Hatton S., Roberts C., Chitsabesan P., Fothergill C., & Harrington R. (2004). Systematic review of the efficacy of cognitive behaviour therapies for childhood and adolescent anxiety disorders. *British Journal of Clinical Psychology*, 43, 421-436.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212.
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38, 835-855.
- Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*, 43, 309-322.
- Chu, B. C., Skriner, L. C., & Zandberg, L. J. (2013). Shape of change in cognitive behavioral therapy for youth anxiety: Symptom trajectory and predictors of change. *Journal of Consulting and Clinical Psychology*, 81, 573-587.
- Cobham, V. E., Dadds, M. R., & Spence, S. H. (1998). The role of parental anxiety in the treatment of childhood anxiety. *Journal of Consulting and Clinical Psychology*, 66, 893-905.
- Compton, S. N., March, J. S., Brent, D., Albano, A. M., Weersing, V. R., & Curry, J. (2004). Cognitive-behavioral psychotherapy for anxiety and depressive disorders in children and adolescents: An evidence-based medicine review. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 930-959.
- Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, 14, 165-176.
- Costello, E., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and Development of Psychiatric Disorders in Childhood and Adolescence, *Archives of General Psychiatry*, 60, 837-844.
- Crawley, S. A., Beidas, R. S., Benjamin, C. L., Martin, E., & Kendall, P. C. (2008). Treating socially phobic youth with CBT: Differential outcomes and treatment considerations. *Behavioural and Cognitive Psychotherapy*, 36, 379-389.

- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*, 81-100.
- Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K., & Zucker, A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology, 44*, 365-380.
- Curry, J., March, J. S., & Harvey, A. S. (2004). Comorbidity of childhood and adolescent anxiety disorders: Prevalence and implications. In: J.S. March (Ed.), *Phobic and anxiety disorders in children and adolescents: A clinician's guide to effective psychosocial and pharmacological interventions (116-140)*. New York: Oxford University Press.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509.
- Dorr, D. (1981). Factor structure of the State-Trait Anxiety Inventory for Children. *Personality and Individual Differences, 2*, 113-117.
- Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., & Weisz, J. R. (2010). A psychometric analysis of the Revised Child Anxiety and Depression Scale-Parent Version in a clinical sample. *Journal of Abnormal Child Psychology, 38*, 249-260.
- Elliott, P., Biddle, D., Hawthorne, G., Forbes, D., & Creamer, M. (2005). Patterns of treatment response in chronic posttraumatic stress disorder: An application of latent growth mixture modeling. *Journal of Traumatic Stress, 18*, 303-311.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Flannery-Schroeder, E. C., & Kendall, P. C. (2000). Group and individual cognitive-behavioral treatments for youth with anxiety disorders: A randomized clinical trial. *Cognitive Therapy and Research, 24*, 251-278.
- Gildengers, A. G., Houck, P. R., Mulsant, B. H., Dew, M. A., Aizenstein, H. J., Jones, B. L., ... Reynolds III, C. A., (2005). Trajectories of treatment response in late-life depression: Psychosocial and clinical correlates. *Journal of Clinical Psychopharmacology, 25*, 8-13.
- Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement. The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review, 5*, 57-83.
- Gueorguieva, R., Mallinckrodt, C., & Krystal, J. H. (2011). Trajectories of depression severity in clinical trials of duloxetine: Insights into antidepressant and placebo responses. *Archives of General Psychiatry, 68*, 1227-1237.
- Guy, W. (1976). *Clinical global improvement scale: Assessment manual of psychopharmacology*. Rockville, MD: National Institute of Health.
- Halliday-Boykins, C. A., Henggeler, S. W., Rowland, M. D., & DeLucia, C. (2004). Heterogeneity in youth symptom trajectories following psychiatric crisis: Predictors and placement outcomes. *Journal of Consulting and Clinical Psychology, 72*, 993-1003.
- Henderson, C. E., Dakof, G. A., Greenbaum, P. E., & Liddle, H. A. (2010). Effectiveness of multidimensional family therapy with higher severity substance-abusing adolescents: Report from two randomized controlled trials. *Journal of Consulting and Clinical Psychology, 78*, 885-897.

- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, 14, 202-226.
- Herbert, J. D., Gaudino, B. A., Rheingold, A. A., Moitra, E., Myers, V. H., Dalrymple, K. L., & Brandsma, L. L. (2009). Cognitive behavior therapy for generalized social anxiety disorder in adolescents: A randomized controlled trial. *Journal of Anxiety Disorders*, 23, 167-177.
- Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schaker (Eds.), *Classification, Data Analyses, and Data Highways* (pp. 147-154). New York: Springer.
- Hudson, J. L. (2005). Efficacy of cognitive-behavioural therapy for children and adolescents with anxiety disorders. *Behaviour Change*, 22, 55-70.
- Hudson, J. L., Rapee, R. M., Deveney, C., Schniering, C. A., Lyneham, H. J., & Bovopoulos, N. (2009). Cognitive-behavioral treatment versus active control for children and adolescents with anxiety disorders: A randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 533-544.
- Hussong, A. M., Curran, P. J., Bauer, D. J. (2013) Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, 9, 61-89.
- Hussong, A. M., Flora, D. B., Curran, P. J., Chassin, L. A., & Zucker, R. A. (2008). Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents. *Development and Psychopathology*, 20, 165-193.
- Ishikawa, S., Okajima, I., Matsuoka, H., & Sakano, Y. (2007). Cognitive behavioural therapy for anxiety disorders in children and adolescents: A meta-analysis. *Child and Adolescent Mental Health*, 12, 164-172.
- Jensen Doss, A., & Weisz, J. R. (2006). Syndrome co-occurrence and treatment outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology*, 74, 416-425.
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2, 302-317.
- Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 62, 100-110.
- Kendall P. C., Flannery-Schroeder E., Panichelli-Mindel S. M., Southam-Gerow M., Henin A., & Warman M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 65, 366-380.
- Kendall, P. C., & Hedtke, K.A. (2006). *Cognitive-behavioral therapy for anxious children: Therapist manual (3rd edition)*. Ardmore, PA: Workbook.
- Kendall, P. C., Hudson, J. L., Gosch, E., Flannery-Schroeder, E., & Suveg, C. (2008). Cognitive-behavioral therapy for anxiety disordered youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology*, 76, 282-297.
- Kendall, P. C., Marrs-Garcia, A., Nath, S., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 285-299.

- Klein, R. G. (1991). Parent-child agreement in clinical assessment of anxiety and other psychopathology: A review. *Journal of Anxiety Disorders*, 5, 187-198.
- Kovacs, M. (1992). *Children's Depression Inventory Manual*. North Tonawanda, NY: Multi-Health Systems.
- Kraemer, H. C., Frank, E., & Kupfer D. J. (2006). Moderators of treatment outcomes. *JAMA: The Journal of the American Medical Association*, 296, 1286-1289.
- Lambert, P. C., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, 55, 86-94.
- Legerstee, J. S., Tulen, J. H. M., Dierckx, B., Treffers, P. D. A., Verhulst, F. C., & Utens, E. (2010). CBT for childhood anxiety disorders: differential changes in selective attention between treatment responders and non-responders. *Journal of Child Psychology and Psychiatry*, 51, 162-172.
- Liber, J. M., Van Eidenfelt, B. M., Utens, E., Ferdinand, R. F., Van der Leeden, A., Van Gastel, W., & Treffers, P. (2008). No difference between groups versus individual treatment of childhood anxiety disorders in a randomized clinical trial. *Journal of Child Psychology and Psychiatry*, 49, 886-893.
- Liber, J. M., van Widenfelt, B. M., van der Leeden, A. J. M., Goedhart, A. W., Utens, E. M. W. J., & Treffers, P. D. A. (2010). The relation of severity and comorbidity to treatment outcome with cognitive behavioral therapy for childhood anxiety disorders. *Journal of Abnormal Child Psychology*, 38, 683-694.
- Manassis, K., Mendlowitz, S. L., Scapillato, D., Avery, D., Fiksenbaum, L., Freire, M., ... Owens, M. (2002). Group and individual cognitive behavioral therapy for childhood anxiety disorders: A randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 1423-1430.
- March, J. S., Parker J. D. A., Sullivan K., Stallings P., & Conners C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 554-565.
- Mathyssek, C. M., Olino, T. M., Hartman, C. A., Ormel, J., Verhulst, C., & Van Oort, F. V. A. (2013). Does the Revised Child Anxiety and Depression Scale (RCADS) measure anxiety symptoms consistently across adolescence? The TRAILS study. *International Journal of Methods in Psychiatric Research*, 22, 27-35.
- Mun, E-Y., de la Torre, J., Atkins, D.C., White, H. R., Ray, A. E., Kim, S-Y., Jiao, Y., Clarke, N., Huo, Y., Larimer, M. E., & Huh, D. (2015). Project INTEGRATE: An integrative study of brief alcohol interventions for college students. *Psychology of Addictive Behaviors*, 29, 34-48.
- Muñoz, M. A., & Acuña, J. D. (1999). Sample size requirements of a mixture analysis method with applications in systematic biology. *Journal of Theoretical Biology*, 196, 263-265.
- Muris, P., Merckelback, H., Ollendick, T., King, N., & Bogie, N. (2002). Three traditional and three new childhood anxiety questionnaires: Their reliability and validity in a normal adolescent sample. *Behaviour Research and Therapy*, 40, 753-772.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative*

- methodology for the social sciences* (pp.345-368). Newbury Park, CA: Sage Publications.
- Muthén, L., & Muthén, B. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychology Measurement*, 16, 159-176.
- Muraki, E. & Bock, D. (2002). PARSCLE 4.1 Computer program. Chicago: Scientific Software International, Inc.
- Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Nauta, M. H., Scholing, A., Emmelkamp, P. M., & Minderaa, R. B. (2003). Cognitive-behavioral therapy for children with anxiety disorders in a clinical setting: No additional effect of a cognitive parent training. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1270-1278.
- Nauta, M. H., Scholing, A., Rapee, R. M., Abbott, M., Spence, S. H., & Waters, A. (2004). A parent-report measure of children's anxiety: psychometric properties and comparison with child-report in a clinic and normal sample. *Behaviour Research and Therapy*, 42, 813-839.
- Nilson, T. S., Eisemann, M., & Kvernmo, S. (2013). Predictors and moderator of outcome in child and adolescent anxiety and depression: A systematic review of psychological treatment studies. *European Child and Adolescent Psychiatry*, 22, 69-87.
- Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535-569.
- O'Neil, K. A., & Kendall, P. C. (2012). Role of comorbid depression and co-occurring depressive symptom in outcomes for anxiety-disordered youth treated with cognitive-behavior therapy. *Child and Family Behavior Therapy*, 34, 197-209.
- Ost, L. G., Svensson, L., Hellstrom, K., Lindwall, R. (2001) One-session treatment of specific phobia in youths: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 69, 814-824.
- Pela, O. A., & Reynolds, C. R. (1982). Cross-cultural application of the Revised-Children's Manifest Anxiety Scale: Normative and reliability data for Nigerian primary school children. *Psychological Reports*, 51, 1135-1138.
- Peris, T., & Piacentini, J. (2009). Adjustments in treatment for limited or nonresponding adolescents in contemporary CBT. In D. McKay, & E. A. Storch (Eds.), *Cognitive-behavior therapy for children: Treating complex and refractory cases* (pp. 47-77).
- Pine, D.S. (1997). Childhood anxiety disorders. *Current Opinion in Pediatrics*, 9, 329-338.
- Pina, A. A., Silverman, W. K., Fuentes, R. M., Kurtines, W. M., & Weems, C. F. (2003). Exposure-based cognitive-behavioral treatment for phobic and anxiety disorders: Treatment effects and maintenance for Hispanic/Latino relative to European-American youths. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1179-1187.

- Rapee, R. M., Lyneham, H. J., Hudson, J. L., Kangas, M., Wuthrich, V. M., & Schiering, C. A. (2013). Effect of comorbidity on treatment of anxious children and adolescents: Results from a large, combined, sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 52, 47-56.
- Reynolds, C. R., & Paget, K. D. (1981). Factor analysis of the Revised Children's Manifest Anxiety Scale for blacks, whites, males, and females with a national normative sample. *Journal of Consulting and Clinical Psychology*, 49, 352-359.
- Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6, 271-280.
- Reynolds, S., Wilson, C., Austin, J., & Hooper, L. (2012). Effects of psychotherapy for anxiety in children and adolescents: A meta-analytic review. *Clinical Psychology Review*, 32, 251-262.
- Rohde, P., Clarke, G. N., Lewinsohn, P. M., Seeley, J. R., & Kaufman, N. K. (2001). Impact of comorbidity on a cognitive-behavioral group treatment for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 795-802.
- Silverman, W. K., & Albano, A. M. (1996). *Anxiety Disorders Interview Schedule for DSM-IV: Child and parent versions*. San Antonio, TX: Graywind.
- Silverman, W. K., Kurtines, W. M., Ginsburg, G. S., Weems, C. F., Lumpkin, P. W., & Carmichael, D. H. (1999a). Treating anxiety disorders in children with group cognitive-behavioral therapy: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 67, 995-1003.
- Silverman, W. K., Kurtines, W. M., Ginsburg, G. S., Weems, C. F., Rabian, B., & Serafini, L. T. (1999b). Contingency management, self-control, and education support in the treatment of childhood phobic disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 67, 675-687.
- Silverman, W. K., Kurtines, W. M., Jaccard, J., & Pina, A. A. (2009). Directionality of Change in Youth Anxiety Treatment Involving Parents: An Initial Examination. *Journal of Consulting and Clinical Psychology*, 77, 474-485.
- Silverman, W. K., Pina, A. A., & Viswesvaran, C. (2008). Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 37, 105-130.
- Silverman, W. K., Saavedra, L. M., Pina, A. A. (2001). Test-retest reliability of anxiety symptoms and Diagnoses with the Anxiety Disorders Interview Schedule for DSM-IV: Child and parent versions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 937-944.
- Southam-Gerow, M. A. & Chorpita, B. F. (2007). Anxiety in children and adolescents. In E. J. Mash & R. A. Barkley (Eds.), *Assessment of childhood disorders* (4th ed., pp. 347-397). New York: Guilford Press.
- Southam-Gerow, M. A., Flannery-Schroeder, E. C., & Kendall, P. C. (2003). A psychometric evaluation of the parent report form of the State-Trait Anxiety Inventory for Children – Trait Version. *Anxiety Disorders*, 17, 427-446.
- Southam-Gerow, M. A., Kendall, P. C., & Weersing, V. R. (2001). Examining Outcome Variability: Correlates of Treatment Response in a Child and Adolescent Anxiety Clinic. *Journal of Clinical Child Psychology*, 30(3), 422-436.

- Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, 36, 545-566.
- Spence, S. H., Barrett, P. M., & Turner, C. M. (2003). Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. *Anxiety Disorders*, 17, 605-625.
- Spence, S. H., Donovan, C., & Brechman-Toussaint, M. (2000). The treatment of childhood social phobia: The effectiveness of a social skills training-based, cognitive-behavioral intervention, with and without parental involvement. *Journal of Child Psychology and Psychiatry*, 41, 713-726.
- Spence, S. H., Donovan, C. L., March, S., Gamble, A., Anderson, R. E., Prosser, S., & Kenardy, J. (2011). A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. *Journal of Consulting and Clinical Psychology*, 79, 629-642.
- Spielberger, C. D. (1973). *Manual for the state-trait anxiety inventory for children*. Palo Alto, CA: Consulting Psychologists Press.
- Stein, N. R., Dickstein, B. D., Schuster, J., Litz, B. T., & Resick, P. A. (2012). Trajectories of response to treatment for posttraumatic stress disorder. *Behavior Therapy*, 43, 790-800.
- Strauss, C. (1987). *Modification of trait portion of State-Trait Anxiety Inventory for Children-Parent Form*. Unpublished manuscript, University of Florida, Gainesville.
- Tabachnick, B. G., & Fidell, F. S. (2007). *Using multivariate statistics*, 5th edition. Boston: Pearson.
- Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture modeling. In G. R. Handcock & K. M. Samuelsen (Eds.) *Advances in latent variable mixture models* (pp. 317-341). Information Age.
- Walkup, J. T., Albano, A. M., Piacentini, J., Birmaher, B., Compton, S. N., Sherrill, J. T., . . . Kendall, P. C. (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *The New England Journal of Medicine*, 359, 2753-2766.
- White, K. S., & Farrell, A. D. (2001). Structure of anxiety symptoms in urban children: Competing factor models of the Revised Children's Manifest Anxiety Scale. *Journal of Consulting and Clinical Psychology*, 69, 333-337.
- Wingersky, M. S., & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, 8, 347-364.
- Wood, J. J., McLeod, B. D., Piacentini, J. C., & Sigman, M. (2009). One-year follow-up of family versus child CBT for anxiety disorders: Exploring the roles of child age and parental intrusiveness. *Child Psychiatry and Human Development*, 40, 301-316.
- Wood, J. J., Piacentini, J. C., Southam-Gerow, M., Chu, B. C., Sigman, M. (2006). Family cognitive behavioral therapy for child anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 314-321.

Endnotes

1. A series of chi-squares and ANOVAs were calculated between each pair of studies to assess for differences in pre-treatment diagnostic variables including total number of diagnoses, severity of primary anxiety diagnosis (based on the ADIS-IV), and number of participants with a comorbid depressive or externalizing disorder. Regarding total number of diagnoses, the number of diagnoses was limited to five, and one study only allowed for four total diagnoses. Analyses indicated that pairs of studies differed about half of the time (i.e. half of the pair comparisons) across diagnostic variables tested. In general, there was no discernable pattern within analyses of a particular variable or across all tested variables. No one study appeared to be more different from the other studies than any other study. It is noteworthy that one study (Study 6; Silverman et al., 1999b) differed in sample inclusion criteria, such that the majority of their sample had a primary diagnosis of a specific phobia. In light of this difference, final growth models were tested with and without the inclusion of this study. There were no substantive differences when this study was excluded from analyses. Therefore, only results including this study are presented.

2. A piecewise model with one slope specified from pre- to post-treatment and a second slope specified from post-treatment to 1-year follow-up was also considered. However, specifying a piecewise led to problems with model convergence, which then required more constraints to be placed on the growth model. Specifically, when a piecewise model was tested, the variances of both slope factors as well as the covariance among the slopes and intercepts had to be fixed to zero. Because there was only 1 time point after post-treatment, allowing that time point to be freely estimated offered the flexibility for a non-linear pattern of change in anxiety while avoiding the additional computational burden and potential convergence problems of adding in a second slope. By specifying a model with only 1 slope and a freely estimated 1-year follow-up time point we were able to retain the slope parameter as a random effect rather than as a fixed effect in the models.

3. Slope from post-treatment to 1-year follow-up was calculated by multiplying the difference between the post-treatment (T3) and 1-year follow-up (T4) latent slope factor loadings by the mean of the slope for that class. For example, for class 1 in the parent model we used the following formula: (T4 loading – T3 loading) * -0.43. Because of calculating the slope in this way, the SE for the slope was not able to be calculated.

4. For “tolerance,” very small values indicate that a predictor is redundant, and values that are less than 0.10 may merit further investigation. The VIF, which stands for variance inflation factor, is the inverse of tolerance and as a rule of thumb, a variable whose VIF value is greater than 5 may merit further investigation (Tabachnick & Fidell, 2007).

Table 1

Characteristics of Included Clinical Trials

Study	Sample & design	Age range (% female)	Diagnoses included	Treatment duration	Assessment points available	% diagnostic-response	
						ICBT	FCBT
1. Bodden et al., (2008)	ICBT vs. FCBT	8-18 (59)	SOP, SAD, GAD, SP, PD	13 ss, 60-90 min.	Pre, post	53**	28
2. Chu (2013)	ICBT; open trial	7-16 (51)	GAD, SAD, SOP, SP	16-20 ss, 60 min.	Pre, mid, post	61*	n/a
3. Kendall (1994)	ICBT vs. WL	9-13 (40)	GAD, SAD, SOP	16 ss, 60 min.	Pre, post, 1-yr FU	64*	n/a
4. Kendall et al., (1997)	ICBT vs. WL	9-13 (48)	GAD, SAD, SOP	16 ss, 60 min.	Pre, mid, post, 1-yr FU	53.19*	n/a
5. Kendall et al., (2008)	ICBT vs. CBT/P vs. Control ^a	7-14 (44)	GAD, SAD, SOP	16 ss, 60 min	Pre, mid, post, 1-yr FU	64* (across both conditions)	
6. Nauta et al., (2003)	ICBT vs. CBT/P vs. WL	7-18 (51)	SAD, SOP, GAD, PD	12 ss., 60 min.	Pre, post, 1-year FU	54**	59
7. Silverman et al. (2009)	ICBT vs. CBT/P	7-16 (57)	GAD, SAD, SOP, SP	12-14 ss, 60 min.	Pre, post, 1-yr FU	78* (across both conditions)	
8. Silverman et al., (1999b)	ICBT vs. Control ^a	6-16 (51)	SP	10 ss, 80 min.	Pre, post, 1-yr FU	69*	n/a
9. Wood et al., (2006)	ICBT vs. FCBT	6-13 (40)	SAD, SOP, GAD	12-16 ss, 60-80 min.	Pre, mid, post, 1-yr FU	52.6*	78.9

^a. Education/attention control conditions were included in IRT analyses but not in growth modeling. * Diagnostic-response rate reflects percent of youth with primary anxiety disorder absent at post-treatment. ** Diagnostic-response rate reflects percent of youth with all anxiety disorders absent at post-treatment.

Table 2

Descriptive Statistics for Demographics and Pre-treatment Clinical Variables

	(1) Bodden (2008)	(2) Chu (2013)	(3) Kendall (1994)	(4) Kendall (1997)	(5) Kendall (2008)
Sample Size	126	113	59	111	111
Age range	8-17	7-17	8-15	8-14	7-14
Mean age (SD)	12.43 (2.59)	11.38 (2.60)	11.70 (1.78)	11.07 (1.37)	10.39 (1.80)
Percent female	59.7	55.8	37.3	43.2	43.2
Ethnicity: (%)					
White	100.0	75.2	81.4	80.2	82.0
African American	0	4.4	15.3	9.9	11.7
Hispanic	0	1.8	3.4	2.7	0.9
Asian	0	5.3	0	1.8	0
Other	0	12.4	0	5.4	3.6
Total # diagnoses: M (SD)	2.75 (1.31)	3.16 (1.28)	2.69 (1.09)	2.92 (1.2)	3.53 (1.28)
Primary diagnosis					
ADIS severity: M (SD)	7.07 (1.02)	5.96 (.86)	6.42 (1.50)	6.73 (1.12)	5.88 (.90)
ADIS severity range	4-8	4-8	4-8	4-8	4-8
Primary Diagnosis (%)					
Generalized Anxiety Disorder	18.0	46	64.4	55.9	46.8
Separation Anxiety Disorder	25.8	12.4	16.9	25.2	23.4
Social Anxiety Disorder	31.3	26.5	18.6	18.9	29.7
Specific Phobia	16.9	7.1	0	0	0
Panic Disorder / Agoraphobia	7.1	8.0	0	0	0
Comorbid Depression (%)	18.8	23	20.3	5.4	9.9
Comorbid Externalizing Disorder (%)	9.7	35.3	16.9	19.8	38.7

Table 2

Descriptive Statistics for Demographics and Pre-treatment Clinical Variables (continued)

	(6) Nauta (2003)	(7) Silverman (1999)	(8) Silverman (2009)	(9) Wood (2006)
Sample Size	76	78	118	40
Age range	7-17	6-16	6-18	6-13
Mean age (SD)	11.0 (2.44)	9.90 (2.46)	9.96 (2.74)	9.90 (2.12)
Percent female	50.0	48.7	56.8	40.0
Ethnicity: (%)				
White	100.0	65.4	33.9	70.0
African American	0	0	0.8	2.5
Hispanic	0	32.1	61.0	10.0
Asian	0	0	0	7.5
Other	0	1.3	4.2	5.0
Total # diagnoses: M (SD)	2.08 (.80)	2.46 (1.33)	2.57 (1.36)	2.08 (.80)
Primary diagnosis				
ADIS severity: M (SD)	6.12 (1.01)	6.65 (1.24)	--	4.85 (.70)
ADIS severity range	4-8	4-8	--	4-6
Primary Diagnosis (%)				
Generalized Anxiety Disorder	22.4	0	21.2	10.0
Separation Anxiety Disorder	30.3	0	22.0	47.5
Social Anxiety Disorder	38.2	9.0	18.6	37.5
Specific Phobia	0	83.3	36.4	0
Panic Disorder / Agoraphobia	9.2	7.7	1.7	5.0
Comorbid Depression (%)	18.4	6.5	9.3	10.0
Comorbid Externalizing Disorder (%)	5.3	23.1	12.7	12.5

Table 3

Overall Structure of Anxiety Symptom Data from each Trial

Clinical Trial	RCMAS	STAIC	MASC	SCAS/ RCADS-A ^a
1. Bodden (2008)		X		
2. Chu		X	X	X
3. Kendall (1994)	X	X		
4. Kendall (1997)	X	X		
5. Kendall (2008)	X		X	
6. Nauta (2003)				X
7. Silverman (1999b)	X			
8. Silverman (2009)	X			
9. Wood (2006)			X	

Note: RCMAS = Revised Children's Manifest Anxiety Scale; STAIC = State-Trait Anxiety Inventory for Children; MASC = Multidimensional Anxiety Scale for Children; SCAS = Spence Child Anxiety Scale; RCADS-A = Revised Children's Anxiety and Depression Scale – Anxiety Subscale.

^a The SCAS and RCADS-A are comprised of the same items and same response scale.

Table 4

Descriptive Statistics for Item Parameters from IRT Calibration

Measure (# items)	Slope (α)				Difficulty (β)			
	Min.	Max.	Mean (SD)	Mean of S.E.s (SD)	Min.	Max.	Mean (SD)	Mean of S.E.s (SD)
RCMAS (28)	0.52	1.11	0.83 (.17)	0.23 (.05)	-1.17	3.15	0.35 (.92)	.33 (.20)
STAIC (20)	0.18	1.45	0.65 (.28)	0.07 (.03)	-1.93	1.67	-0.30 (.77)	0.12 (.07)
MASC (39)	0.41	1.44	0.72 (.23)	0.24 (.08)	-3.02	1.52	-0.21 (1.04)	0.57 (.36)
SCAS/RCADS (29)	0.35	1.13	0.71 (.21)	0.17 (.05)	-1.38	1.89	0.31 (.88)	0.24 (.12)

Note: RCMAS = Revised Children's Manifest Anxiety Scale; STAIC = State-Trait Anxiety Inventory for Children; MASC = Multidimensional Anxiety Scale for Children; SCAS = Spence Child Anxiety Scale; RCADS = Revised Children's Anxiety and Depression Scales – Anxiety Subscale.

Table 5

Means and Standard Deviations for IRT Anxiety Scale Scores Across Studies

	<u>Pre-treatment</u>		<u>Mid-treatment</u>		<u>Post-treatment</u>		<u>1-year follow-up</u>	
	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)
<u>Parent-reported anxiety</u>								
Pooled sample	743	-.02 (.82)	272	-.35 (.91)	675	-.85 (.93)	358	-.93 (.85)
Study								
1 Bodden	128	-.25 (.98)	--	--	115	-1.15 (.97)	--	--
2 Chu	110	-0.07 (.76)	97	-0.82 (1.06)	92	-1.19 (.95)	--	--
3 Kendall 94	59	.26 (.72)	40	-0.18 (.85)	45	-1.07 (.92)	40	-1.21 (.87)
4 Kendall 97	108	.15 (.78)	98	.03 (.62)	94	-.78 (.88)	83	-.94 (.88)
5 Kendall 08	109	-.01 (.89)	--	--	100	-.78 (.95)	66	-1.00 (.85)
6 Nauta	74	.02 (.65)	--	--	71	-.51 (.69)	63	-.71 (.78)
8 Silverman 99	--	--	--	--	60	-.52 (.90)	32	-.98 (.84)
7 Silverman 09	116	-.01 (.83)	--	--	60	-.75 (.95)	42	-1.02 (.98)
9 Wood	39	-.19 (.58)	37	-.35 (.59)	38	-.48 (.74)	32	-.68 (.71)
<u>Child reported anxiety</u>								
Pooled sample	806	-.27 (.98)	274	-.96 (1.11)	683	-1.17 (1.10)	363	-1.3 (.99)
Study								
1 Bodden	124	-.63 (1.00)	--	--	115	-1.38 (1.05)	--	--
2 Chu	109	-.31 (1.11)	97	-1.4 (1.14)	92	-1.82 (1.2)	--	--
3 Kendall 94	59	-.12 (.87)	41	-.57 (1.01)	48	-1.43 (.98)	41	-1.41 (.99)
4 Kendall 97	111	-.09 (.98)	100	-.72 (1.11)	95	-1.31 (1.22)	88	-1.38 (1.13)
5 Kendall 08	103	-.56 (.91)	--	--	97	-1.02 (.98)	69	-1.34 (.95)
6 Nauta	76	.05 (.78)	--	--	72	-.59 (.85)	62	-.93 (1.03)
7 Silverman 99	77	-.16 (.91)	--	--	57	-.76 (.85)	29	-1.4 (.80)
8 Silverman 09	111	.01 (.97)	--	--	80	-.82 (1.12)	42	-1.34 (.87)
9 Wood	36	-.67 (.79)	36	-.85 (.71)	37	-1.06 (.80)	32	-1.34 (.78)

Table 6

Fit Indices for Growth Models

Classes	SSAdj BIC ^a	Entropy ^b	LMR-LRT ^c		BLRT ^c	
Parent-report			Value	<i>p</i> -value	Value	<i>p</i> -value
1	4811.79	NA	NA	NA	NA	NA
2	4793.92	0.81	31.97	<.001	31.97	<.001
3	4791.80	0.56	16.19	0.057	16.18	0.02
4	4794.58	0.57	11.35	0.16	11.35	0.12
5	4801.04	0.62	11.45	0.20	11.45	0.17
Youth-report						
1	5558.78	NA	NA	NA	NA	NA
2	5543.22	0.66	29.7	0.06	29.7	<.001
3	5529.85	0.63	27.52	0.03	27.52	<.001
4	5523.09	0.55	20.91	0.17	20.91	<.001
5	5523.11	0.59	14.14	0.26	14.14	0.07

a = Lower SSAdj. BIC values indicate better fit. SSAdjBIC = sample-size adjusted Bayesian information criterion.

b = Values should be greater than 0.7. Values closer to 1 are better. Entropy refers to the average classification accuracy in assigning individuals to classes.

c = Significant values indicate k-classes fit better than k-1 classes. LMR = Low-Mendell-Rubin adjusted likelihood ratio test. BLRT = Bootstrap Likelihood Ratio Test.

Table 7

Class Counts and Parameter Estimates for Trajectory Classes

Class (class size: N, %)	Intercept (SE)	Slope (SE) (slope from pre- through post- treatment)	Slope ^a (slope from post- treatment to 1-yr follow-up)
<u>Parent 3-class model</u>			
1 Steady responders (577, 71.0%)	-0.38 (.09)***	-0.43 (.03)***	-0.07***
2 Delayed improvement (178, 21.8%)	0.04 (.15)	0.009 (.01)	-0.56*
3 Rapid responders (58, 7.2%)	0.49 (.13)***	-1.31 (.10)***	-0.25***
<u>Child 4-parent model</u>			
1 Steady responders (453, 55.1%)	-0.47 (.11)***	-0.31 (.04)***	-0.41***
2 Low symptom responders (208, 25.3%)	-1.42 (.25)***	-0.48 (.10)***	0.07***
3 Rapid responders (103, 12.6%)	-0.04 (.21)	-1.27 (.15)***	0.31***
4 Delayed improvement (58, 7.1%)	0.05 (.18)	0.097 (.04)**	-1.13**

^a Slope from post-treatment to 1-year follow-up was calculated by multiplying the difference between the post-treatment (T3) and 1-year follow-up (T4) latent slope factor loadings by the mean of the slope for that class. For example, for class 1 in the parent model we used the following formula:
 (T4 loading – T3 loading) * -0.43.

* $p < .05$

** $p < .01$

*** $p < .001$

Table 8

Predictors of Trajectory Classes – Multinomial Logit Models

Parent 3-class model	Rapid vs. Steady Responders	Delayed vs. Steady responders	Rapid vs. Delayed improvement
	Log odds (SE)	Log odds (SE)	Log odds (SE)
# Diagnoses	0.81 (.23)***	0.83 (.29)**	0.02 (.23)
Comorbid mood dx	-0.17 (.73)	0.55 (.59)	0.73 (.66)
Comorbid externalizing dx	-0.84 (.78)	-0.61 (.60)	0.23 (.66)
Treatment type	0.17 (.52)	0.58 (.48)	0.42 (.52)
Age	0.27 (.09)**	0.10 (.11)	-0.17 (.10)
Gender	-0.43 (.53)	-0.19 (.49)	0.24 (.53)
Ethnicity	-0.92 (1.26)	0.31 (.79)	1.23 (1.38)

Child 4-class model	Delayed vs. Steady Responders	Low symptoms vs. Steady Responders	Rapid vs. Steady Responders	Low symptoms vs. Delayed improvement	Rapid vs. Delayed improvement	Low symptoms vs. Delayed improvement
	Log odds (SE)	Log odds (SE)	Log odds (SE)	Log odds (SE)	Log odds (SE)	Log odds (SE)
# Diagnoses	0.47 (.31)	-0.16 (.18)	0.27 (.17)	-0.63 (.30)*	-0.20 (.32)	-0.44 (.24)
Comorbid mood dx	0.15 (.70)	-1.09 (.84)	0.20 (.47)	-1.24 (.96)	0.05 (.70)	-1.29 (.92)
Comorbid externalizing dx	0.12 (.80)	-0.17 (.50)	0.35 (.54)	-0.29 (.79)	0.22 (.81)	-0.51 (.72)
Treatment type	1.21 (.61)*	-0.20 (.33)	-1.39 (.64)*	-1.40 (.60)*	-2.60 (.76)***	1.19 (.70)
Age	0.22 (.19)	-0.002 (.07)	0.05 (.07)	-0.22 (.18)	-0.17 (.18)	-0.05 (.09)
Gender	-0.23 (.82)	-0.72 (.33)*	-0.09 (.40)	-0.49 (.77)	0.13 (.79)	-0.63 (.48)
Ethnicity	0.06 (.65)	0.51 (.38)	-0.51 (.47)	0.45 (.64)	-.58 (.72)	-1.02 (.59)

Note: Dx = diagnoses.

* $p < .05$, ** $p < .01$, *** $p < .001$

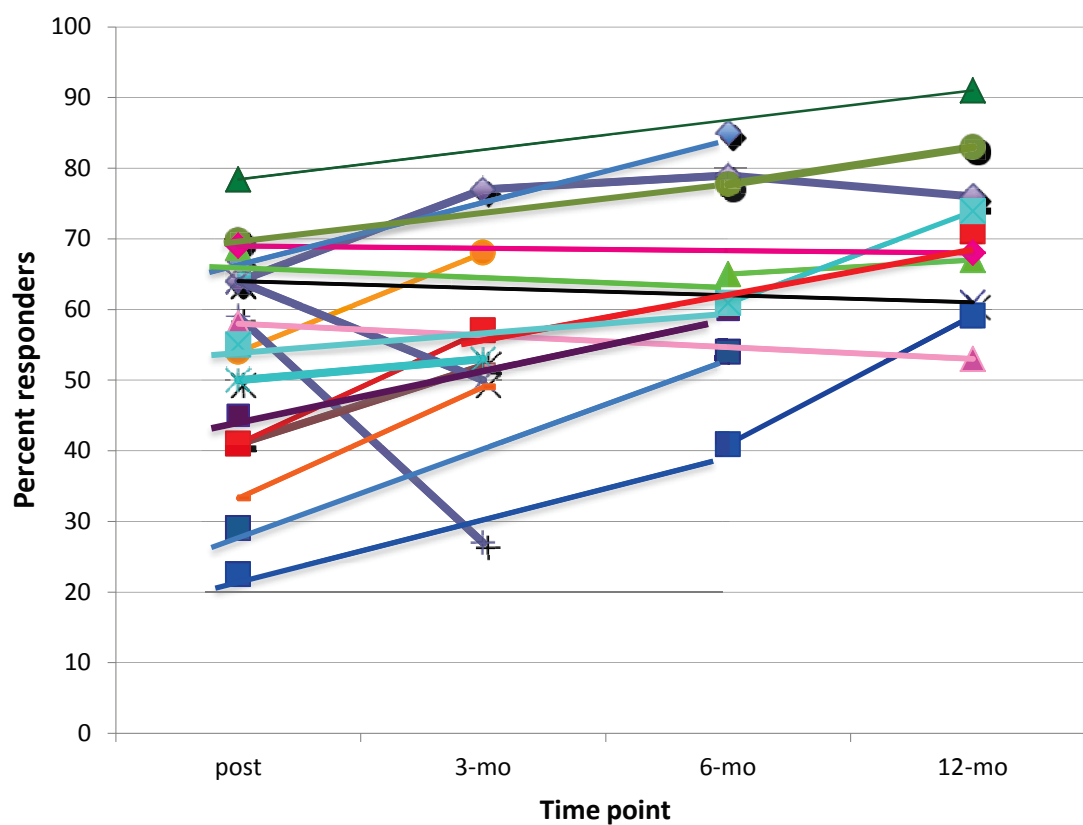


Figure 1. Diagnostic response-rates (defined as either absence of primary disorder or absence of all anxiety disorders) from post-treatment through 1-year follow-up from 18 trials of CBT for youth anxiety.

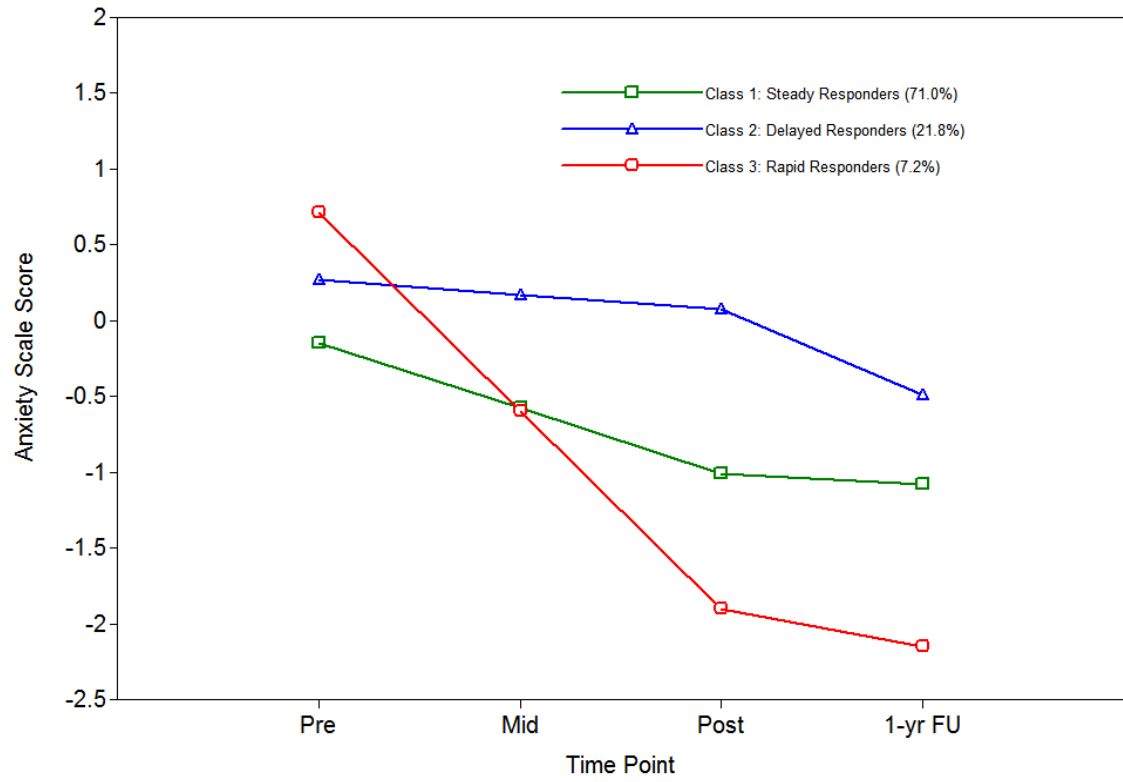


Figure 2. Latent growth mixture model trajectories for parent reported anxiety from pre-treatment through 1-year follow-up.

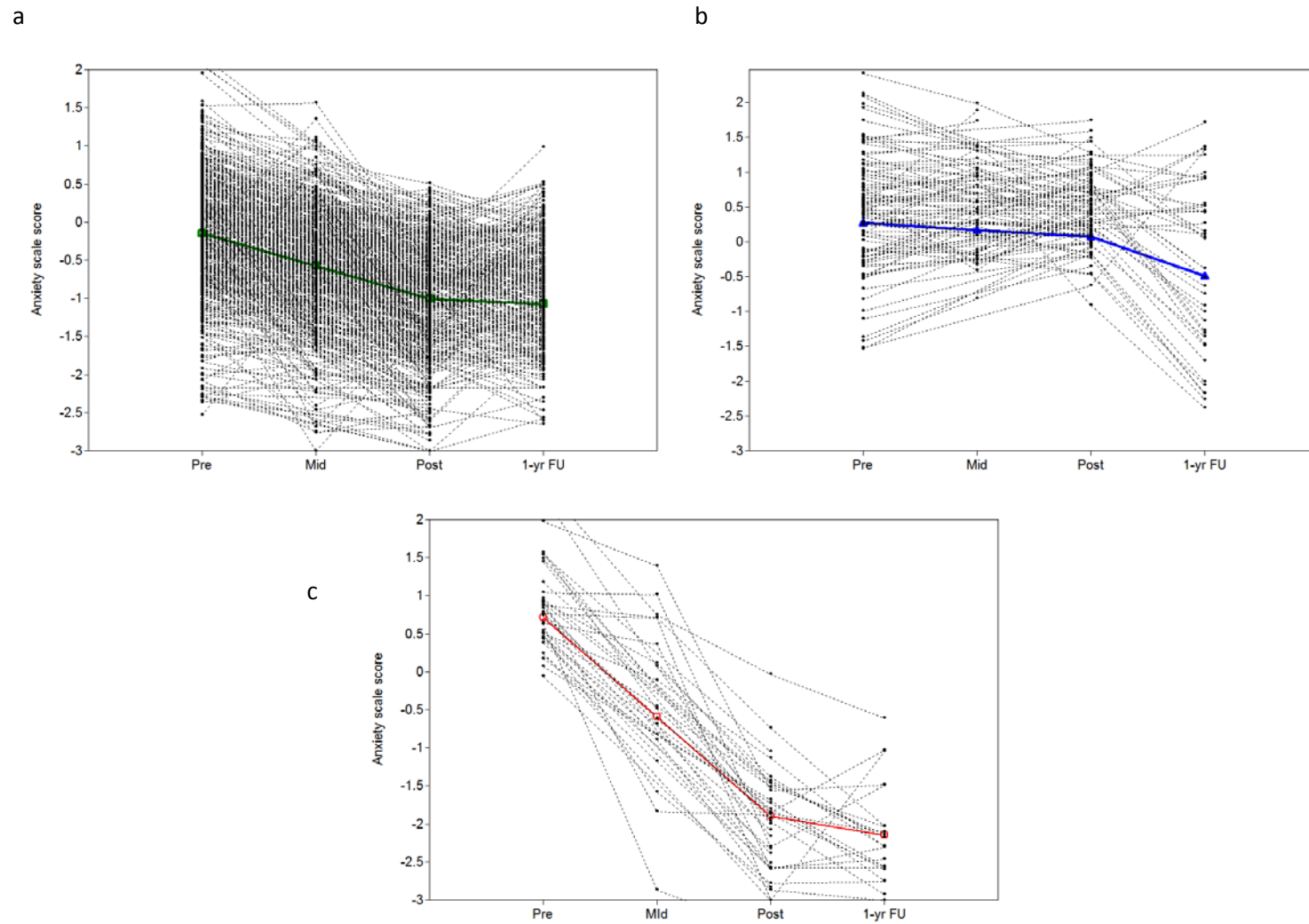


Figure 3. Estimated mean latent class growth trajectories and observed values for each of the 3 parent trajectory classes. Panels represent (a) steady responders, (b) delayed improvement, and (c) rapid responders.

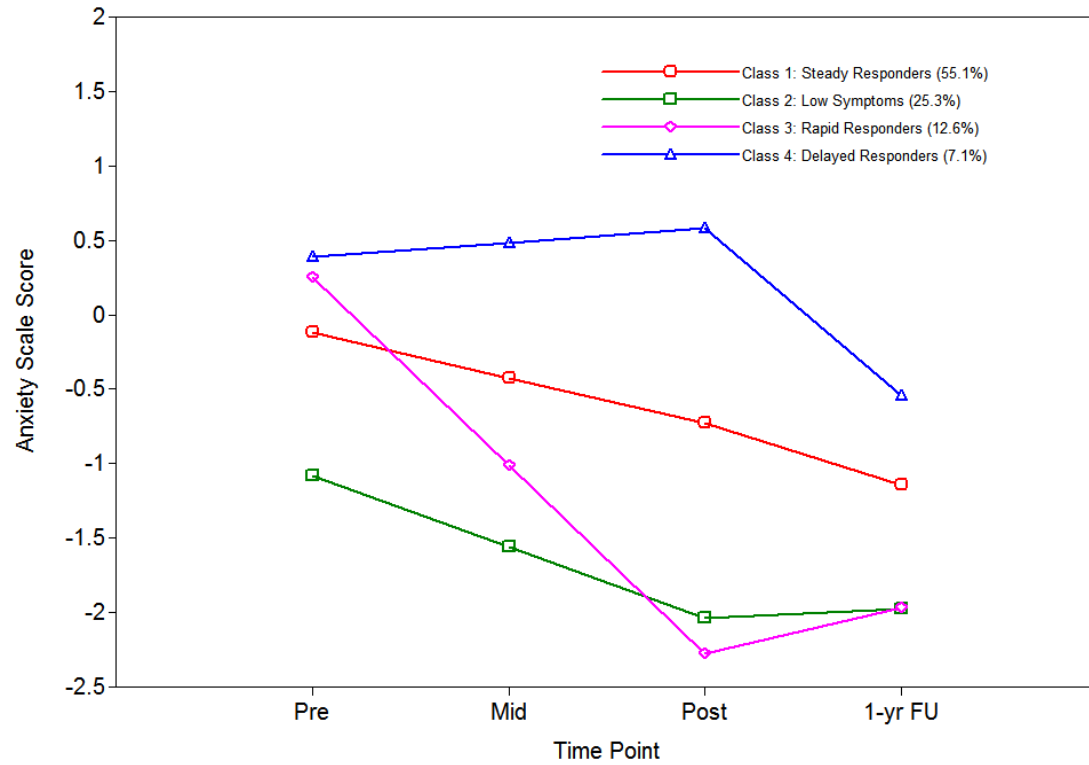


Figure 4. Latent growth mixture model trajectories for youth-reported anxiety from pre-treatment through 1-year follow up.

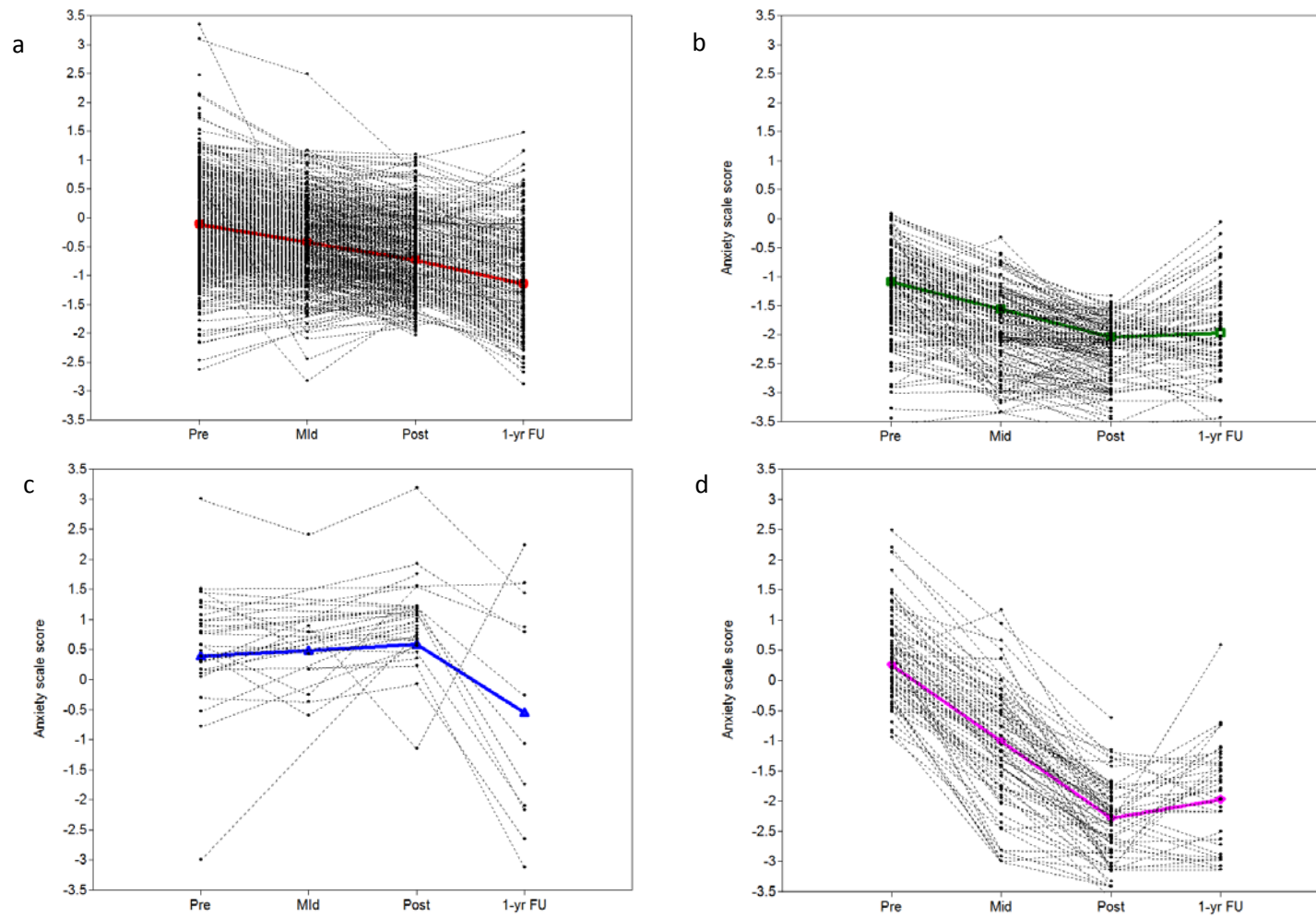


Figure 5. Estimated mean latent class growth trajectories and observed values for each of the 4 youth trajectory classes. Panels represent (a) steady responders, (b) low symptoms responders, (c) delayed improvement, and (d) rapid responders.