2014-08-01

Examining Trajectories and Predictors of Individual Change for Youths Treated in Usual Care

Ashley M. Smith
University of Miami, a.smith63@umiami.edu

UNIVERSITY OF MIAMI

EXAMINING TRAJECTORIES AND PREDICTORS OF INDIVIDUAL CHANGE
FOR YOUTHS TREATED IN USUAL CARE

By

Ashley M. Smith

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

August 2014

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

EXAMINING TRAJECTORIES AND PREDICTORS OF INDIVIDUAL CHANGE
FOR YOUTHS TREATED IN USUAL CARE

Ashley M. Smith

Approved:

____
Amanda Jensen-Doss, Ph.D.
Associate Professor of Psychology

____
Brian D. Doss, Ph.D.
Associate Professor of
Psychology

____
Cynthia L. Rowe, Ph.D.
Associate Professor of Epidemiology
and Public Health

____
M. Brian Blake, Ph.D.
Dean of the Graduate School

SMITH, ASHLEY M. (M.S., Psychology)
Examining Trajectories and Predictors of Individual Change for Youths Treated in Usual Care (August 2014)

Abstract of a thesis at the University of Miami.

Thesis supervised by Associate Professor of Psychology Amanda Jensen-Doss.
No. of pages in text. (62)

Improving mental health services for youths in usual care (UC) is one of the most critical issues in mental health services research. Identification of change trajectories in UC (e.g., improvement, no response, deterioration) can help researchers gain a richer understanding of UC and facilitate efforts to tailor UC to individuals. This study used multilevel growth mixture modeling (MGMM) to examine trajectories of change for two outcome measures (i.e., problem severity and functioning) in a sample of youths (N = 722) treated in UC served at four clinics operating under a large county-wide public mental health authority. Multinomial logistic regression was used to predict trajectory group membership from youth demographic and clinical variables, to identify the types of clients most likely to have positive or negative treatment outcomes. Results evidenced three distinct trajectories of change on a measure of problem severity: 1) Remained High (12.2%), 2) Remained Moderate (85.0%), and 3) Moderate Improvement (2.8%). Two distinct trajectories of change were identified on a measure of functioning: 1) No Change (98.7%), and 2) Moderate Improvement (1.3%). Predictors of trajectory group membership indicate that baseline problem severity and functioning, and associated clinical variables, significantly predicted trajectory group membership. Findings are discussed in terms of understanding change in youth UC and informing treatment targets.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Chapter 1: Introduction

There is an urgent need to improve mental health services for the millions of youths being treated in community outpatient psychotherapy, known as usual care (UC). Limited research examining treatment outcomes in UC is discouraging, generally reporting minimal improvement in symptoms and functioning (Bickman, Lambert, Andrade, & Penaloza, 2000; Weiss, Catron, Harris, & Phung, 1999; Weisz, 2004; Weisz, Donenberg, Han, & Weiss, 1995). At the same time, there is evidence demonstrating the efficacy of research-based youth treatments relative to controls (Casey & Berman, 1985; Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995), suggesting that perhaps moving these treatments into practice settings might improve UC outcomes. Unfortunately, when these treatments have been implemented in real-world settings, their effects are often much less positive (Weisz, Donenberg, et al., 1995). A meta-analytic review comparing outcomes between evidence-based treatments (EBTs) and UC suggested that although EBTs produced better outcomes than UC (Weisz, Jensen-Doss, & Hawley, 2006), the effect sizes were generally small. Thus, it appears that simply moving EBTs into community settings will not completely bridge the gap between the effectiveness of UC and the efficacy of research-supported psychotherapies. Consequently, in order to improve mental health services for youths served in the community, there is a need for research that examines individual response to treatment in real-world settings.

Although hundreds of efficacious treatments have been developed for youths, concerns remain about the generalizability of these studies to UC. UC differs in several important ways from research trials, including demographic and clinical differences observed in patients and families (Baker-Ericzén, Hurlburt, Brookman-Frazee, Jenkins, & Hough, 2010; Southam-Gerow, Velez, & Kendall, 2003; Weersing & Weisz, 2002), therapist characteristics, as well as training and supervision, (Weisz, Donenberg, et al., 1995), and variation in treatment practices (Garland, Bickman, & Chorpita, 2010). These differences may potentially diminish the generalizability and applicability of findings from research trials to community settings (Kazdin, 2008; Weisz, Jensen-Doss, & Hawley, 2005).

Much of what is known about UC currently comes from effectiveness trials in which UC has served as the control condition. Effectiveness trials examine the effects of a previously tested efficacious intervention in a more naturalistic environment using a more heterogeneous sample (Hoagwood, Hibbs, Brent, & Jensen, 1995). That is, effectiveness studies take place in environments (e.g., community mental health clinic, school, home) and under conditions (e.g., heterogeneous sample, real-world therapists) that resemble real-world usual care settings, which allows researchers to understand the generalizability of an intervention (Barlow, 1996; Nathan, Stuart, & Dolan, 2000). The goal of effectiveness studies is to test whether or not a treatment or intervention that has been proven efficacious in a highly controlled trial is effective when transported and implemented in the real-world (Barlow, 1996; Carroll & Rounsaville, 2003; Chambless & Hollon, 1998) and can help answer questions such as for whom and under what conditions a treatment or intervention is effective (Flay, 1986; Flay et al., 2005; Paul, 1967; Sexton & Kelley, 2010).

Although conditions of effectiveness studies are meant to mimic real-world care, important concerns remain about the generalizability of results (Hunsley & Lee, 2006;

Lonigan, Elbert, & Johnson, 1998; Norcross, Beutler, & Levant, 2006). Even though effectiveness studies lack the stringent study inclusion and exclusion participant criteria seen in efficacy studies, this does not mean that inclusion and exclusion criteria are absent from effectiveness studies, as these studies are still designed within the context of research (Norcross et al., 2006). For example, effectiveness studies will often exclude participants if immediate attention is required (e.g., in the case of a suicidal threat), if another condition takes precedent (e.g., substance abuse), or if the participant is currently receiving another treatment (Hunsley, 2007; Hunsley & Lee, 2007). Effectiveness studies also involve a high degree of treatment monitoring to ensure treatment adherence and fidelity. However, these same experimental controls are typically not found in nonresearch contexts because of the increased burden required (Clarke, 1995). Thus, although effectiveness studies are intended to be more naturalistic in design, important limitations call into question the extent to which they are truly similar UC (Kazdin, 2008). Indeed, it is becoming increasingly important to utilize naturalistic data that were not collected for research purposes to gather a more accurate representation of what is going on in UC in the absence of interference from research.

Methodological Concerns

Our understanding of UC is also limited because effectiveness trials often use traditional pre/post designs that focus on the aggregate effects of treatment (Howard, Moras, Brill, Martinovich, & Lutz, 1996). In these research designs, little or no consideration is given to variation in *individual responses* to treatment, because the focus is on the *group level responses* (Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010). Thus, information about unique individual variability is lost and little is known about differential individual responses to a treatment. This suggests that although the overall effect of a treatment can be determined for a sample population, it is unclear how the results from the study apply to a particular individual, or if the same treatment effects would generalize to other individuals. Consequently, the field is limited in how it can apply findings to real-world settings and individuals because of these study designs that focus on the group level effect. Further, because pre/post designs do not consider individual responses to treatment, researchers are not able to identify how treatments can be improved or adapted to be most effective for individuals in UC.

As such, the field is moving toward an emphasis on research that focuses on individual client outcomes (Howard et al., 1996). In 2006, the APA Presidential Task Force on Evidence-Based Practice emphasized that individual client differences can have a large impact on the overall success of a treatment, and therefore cannot be ignored. Similarly, according to Howard and colleagues (1996), important questions to ask regarding treatment include whether a treatment is working for a *particular client*. A focus on the latter question moves beyond aggregate effects from pre-test/post-test research and examines individual patient progress in response to treatment by looking at trajectories of individual change over time.

Patient-focused research is one attempt to help solve this problem (Howard et al., 1996). Patient-focused research builds upon traditional research designs, but takes an additional step further to understand *what works for an individual* throughout treatment. Central to patient-focused research is an emphasis on measuring and monitoring individual progress to make adjustments to treatment (Lutz, Martinovich, Howard, & Leon, 2002), which has implications for both research and practice. With regards to research, the patient-focused paradigm holds the potential to explain variability in response to treatment through examination of individual change trajectories. For example, differences in change trajectories and differences in the timing or magnitude of change allows researchers to explore change as related to interventions or demographic and clinical characteristics (e.g., diagnosis) and can help to inform theory (Nelson, 2011). Patient-focused research has been applied to a variety of problems in the field. For example, rather than simply examining the average group response to treatment, different outcome classes of individuals can be identified based on groups of similar change trajectories (e.g., patients who demonstrate rapid improvement, gradual improvement, no change, or deterioration) (Warren et al., 2010). In addition, to help identify patients who might be at risk for deterioration or treatment failure, empirically-driven algorithms have been developed to function as "warning systems" that will alert therapists when patients are at risk for negative treatment outcomes (Bishop et al., 2005; Bybee, Lambert, & Eggett, 2007; Warren, Nelson, Burlingame, & Mondragon, 2012).

Change Trajectories Identified in the Literature

One way to gain a richer understanding of UC is to use principles of patient-focused research to identify individual trajectories of change and predictors of trajectory group membership in UC. As mentioned previously, the limited extant literature examining treatment outcomes for youths in community mental health settings has typically found mean effect sizes near zero (Angold, Costello, Burns, Erkanli, & Farmer, 2000; Weiss et al., 1999; Weisz, 2004; Weisz, Donenberg, et al., 1995). However, little is known about what contributes to these small mean effects. As prior studies have traditionally focused on examining the group level effects of treatment, individual variability in response to treatment and change processes has been largely ignored. It is very likely that surrounding that average effect of zero there are some clients who improve and others that do not – variation that gets lost when data are reduced to differences at the group level. Consequently, we are limited in our ability to understand how to improve the quality of UC for *individuals*. One way to gain a richer understating of UC is to learn about trajectories of change present in UC youth psychotherapy and to understand classes of individuals that experience change in similar ways. Identification of change trajectories (e.g., rapid response, intermediate response, nonresponse) can improve UC to be more targeted (i.e., by identifying and changing services for individuals who are not doing well in UC) and can highlight situations in which the field might benefit from emulating UC practices (e.g., by learning what services are being provided to individuals who are responding well to UC).

To date, several studies have examined types of change in research samples. In the adult and child literatures, there is evidence for several patterns and classes of change. For instance, an *early rapid response* pattern has been identified in cognitive-behavioral therapies for adult (Cuijpers, van Lier, van Straten, & Donker, 2005; Ilardi & Craighead, 1994) and adolescent (Renaud et al., 1998) depression, which is typified by a significant decrease in symptoms early on in treatment (e.g., before the fourth treatment session), followed by a period in which change levels off (Renaud et al., 1998). This similar rapid response pattern has also been seen with panic disorder (Penava, Otto, Maki, & Pollack, 1998), bulimia (Grilo, Masheb, & Wilson, 2006), and substance abuse (Breslin, Sobell, Sobell, Buchan, & Cunningham, 1997) in the adult literature. An *intermediate response* pattern, characterized by a mild decline in symptoms and improvement in functioning, has also been identified in the adult and adolescent depression literatures (Renaud et al., 1998). Further, adult and child literatures have identified an *initial nonresponse* pattern of change, distinguishable by an initial score on measures of symptoms and functioning that fluctuates very little or even tends to increase slightly (Renaud et al., 1998).

Only a single study has examined trajectories of change for youths in UC. Specifically, this study compared symptom trajectories and outcome classes between youths in UC and managed care (Warren et al., 2010). Rather than taking an exploratory approach to identifying naturalistic change groups, the authors of this study identified four outcome groups based on whether the clients evidenced a "reliable change" of 13 points on their outcome measure: deterioration (i.e., a post-treatment score that was at least 13 points higher than pretreatment), no reliable change (i.e., a difference of less than 13 points between pre- and post-treatment), improvement (i.e., a 13 point improvement between pre- and post-treatment), and recovery (i.e., a post-treatment score that fell in the subclinical range according to the measure's norms). Results indicated that of the youths in the community and managed care samples, respectively, 21.6% and 13.3% deteriorated, 37.1% and 30.8% experienced no reliable change, 25.4% and 30.5% improved, and 15.9% and 25.4% recovered. They also examined individual change trajectories, but did not examine whether there were groups of clients who experienced similar trajectories of change over time. This study was also limited in that outcome groups were identified "a priori", and were not classified based on *naturalistic* change. As well, limited data were available to examine predictors of outcome. However, neither gender nor age had an effect on either change trajectories or rates of change. Higher symptom severity was also predictive of worse outcomes. As well, greater number of sessions and greater total number of outcome assessments were both associated with more positive outcomes. Consequently, there remains a critical need for more in-depth examination of change trajectories in UC settings, particularly in combination with an examination of predictors of those trajectories.

Predictors Associated with Youth Outcomes and Change Trajectories

Identification of predictors of change trajectories and outcomes is important to help tailor UC to individuals. For example, it is possible that demographic factors such as gender, age and ethnicity might be associated with treatment response. Although limited work has been conducted on predictors specifically within UC settings, much more work has been done in RCT samples that might inform hypotheses about predictors of UC outcomes. Further, few studies have examined predictors of trajectory group membership; therefore, the broader literature on predictors of youth outcomes in UC or RCTs will also be used to inform hypotheses in the current study. The literature generally suggests that demographic variables are not consistently related to youth outcomes (Emslie, Mayes, Laptook, & Batt, 2003; Phillips et al., 2000). As mentioned previously, Warren and colleagues (2010) did not find that gender was associated with treatment outcomes and no other studies of youth UC have examined this question. Data from RCTs on the effects of gender on outcomes is mixed. Some have found that gender is not predictive of youth outcomes (Barkley, Guevremont, Anastopoulos, & Fletcher, 1992; Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008; Rohde, Lewinsohn, & Seeley, 1994), while others have found that gender is a significant predictor of outcomes, with girls evidencing worse outcomes than boys (Hops, Lewisohn, & Roberts, 1990). Consistent with Warren and colleagues' (2010) finding that age was not related to

UC outcomes, RCTs have also typically found that age is not predictive of youth outcomes (Barkley et al., 1992; Hops et al., 1990; Kendall et al., 2008; Rohde et al., 1994). Findings in the UC and RCT literatures on ethnicity have suggested that ethnic minority status is associated with smaller treatment gains (Weersing & Weisz, 2002) and negative outcomes (e.g., attrition) (Kendall & Sugarman, 1997). Further, studies examining socioeconomic status (SES) have generally shown that youths served in UC settings are oftentimes from lower SES families compared to youths served in other settings, and generally evidence worse outcomes (Baker-Ericzén et al., 2010; Southam-Gerow et al., 2003; Warren et al., 2010).

Clinical predictors of outcomes, including type of diagnosis, symptom severity and diagnostic comorbidity, also have the potential to influence change trajectories and outcomes in UC. A review of youth psychotherapy examined whether there is a differential effect of psychotherapy for different diagnoses (e.g., primary anxiety, depression, conduct) and found some evidence that outcomes are dependent on diagnosis, specifically that primary diagnoses requiring behavior therapy (e.g., ADHD) evidenced slightly better outcomes (Casey & Berman, 1985). However, these effects were small and may be the result of several other factors (e.g., therapist, type of therapy). More recent work in this area has not found any evidence that diagnosis is predictive of outcomes (Ash & Weis, 2009). The findings on initial symptom severity are also mixed. As mentioned above, Warren and colleagues (2010) found that higher initial symptom severity was predictive of worse outcomes (i.e., deterioration). Another study of UC found that greater problem severity was predictive of greater improvement, but was also associated with higher post-treatment symptomology (Jensen-Doss & Weisz, 2006). In the RCT literature, some studies have found greater initial symptom severity to be associated with worse outcomes (Kazdin & Wassell, 2000; Renaud et al., 1998; Southam-Gerow, Kendall, & Weersing, 2001); however, others have found the opposite (Flannery-Schroeder & Kendall, 2000; Kendall & Sugarman, 1997), and still others have shown nonsignificant relationships between initial symptom severity and outcomes (Rohde et al., 1994). Regarding treatment response, further evidence has found that lower initial self-reported symptom severity was predictive of not only more positive outcomes, but also a more rapid trajectory of change (i.e., rapid responders tended to have lower initial symptom severity compared to initial nonresponders) (Renaud et al., 1998).

Comorbidity is generally thought to be one of the factors that differentiates youths in usual care settings from those found in RCTs, with an assumption that UC clients are more difficult to treat. However, results examining comorbidities are mixed, and often suggest that the predictive utility of comorbid diagnoses depends on the combination of diagnoses found in a clinical profile. For example, one study examining treatment as usual found that comorbidity was not predictive of outcomes (Jensen-Doss & Weisz, 2006). In the RCT literature, some studies have also shown that comorbid diagnoses do not significantly predict outcomes (Barkley et al., 1992; Clarke et al., 1992; Kendall et al., 1997; Renaud et al., 1998), although others have found that comorbid diagnoses are predictive of worse treatment outcomes (Brent et al., 1998). Mixed findings on demographic predictors suggest the importance of considering individual differences and constellations of characteristics that might affect treatment response.

Treatment attendance has also been examined as a predictor of outcomes. The Warren and colleagues (2010) study described above yielded mixed findings across different definitions of outcome: attending more treatment sessions was associated with better outcomes when outcome was operationalized as a youth's final score on the outcome measure, but not when operationalized as trajectory group membership. Given that RCTs often specify the number of treatment sessions to be delivered, they are limited in their ability to inform theories about the relationship between dose and outcomes in UC. Although a number of predictors have been studied in the context of youth outcomes, most research has focused on youth outcomes in effectiveness trials and few studies have examined predictors of outcomes in UC.

In sum, there is little evidence on trajectories of change in youth UC and how individual variability influences treatment outcomes. Further, little outcome data is truly naturalistic and unaffected by observer effects intrinsic in research studies. This study contributes to the literature by identifying and describing individual change trajectories and change groups in youth psychotherapy, in addition to predictors of trajectory group membership. This information will provide a richer understanding of what is taking place in usual care (e.g., patterns and shapes of change present in UC psychotherapy), and may inform decisions on how to direct efforts at improving UC and tailor treatment for individuals.

Current Study

The aims of the present study were two-fold. In Aim 1, naturally occurring trajectories of change within a sample of youths participating in UC were identified. It was hypothesized that three trajectories of change would emerge in our sample: improvement, no response, and deterioration. Assessments occurred every 90 days, which would not allow for the identification of patterns such as rapid response. Aim 2 examined predictors associated with the identified change groups. Given the mixed extant literature on age and gender as predictors of youth outcomes, and the limited literature on how these predictors related to trajectories of change in youth UC, age and gender were examined as exploratory predictors and no specific hypotheses were made regarding the direction of these relationships. Consistent with research in the UC and RCT literature, it was hypothesized that ethnic minority status would be associated with less positive outcomes. For the purposes of the current study, insurance status was used as a proxy for SES as the two have been found to be highly correlated (Halfon, Inkelas, & Wood, 1995). It was hypothesized that having insurance would be associated with more positive outcomes, such that youth from families who are currently insured will evidence more positive trajectories of change.

It was also hypothesized that clinical characteristics would affect trajectories of change. Families with higher initial clinician ratings of overall family problems were hypothesized to evidence more negative trajectories of change. As prior literature has found little evidence for the effect of diagnosis on outcome, the present study did not generate any direct hypotheses and examined type of diagnosis as an exploratory predictor. Because the clinics in this study provided different services to children with externalizing and internalizing problems, the present study also explored the effect of receiving skills training (for externalizing problems) or therapy (for internalizing problems). It was hypothesized that youths with greater initial parent-rated symptom severity would evidence less positive outcomes (i.e., more likely to fall into a trajectory of change exhibiting either no response or deterioration) compared to youths who exhibited lower symptom severity at intake (Warren et al., 2010). Further, it was hypothesized that youths rated to have lower functioning by parents at intake would evidence worse trajectories of change (e.g., no change) or less rapid change. Due to mixed results on the effects of comorbidity on youth outcomes, no direct hypotheses were generated and comorbidity was examined as an exploratory predictor. As the UC literature suggests mixed findings on the effect of number of treatment sessions on outcomes and trajectory class membership (Warren et al., 2010), these variables (i.e., number of treatment sessions, number of weeks in therapy) were examined as exploratory predictors. Specifically, due to the large degree of variability in treatment received, including variables targeting both treatment dose and duration offered a more accurate representation of the amount of therapy received by individuals.

Chapter 2: Method

Participants

The study sample included administrative data collected through routine clinical care of 722 youths, ages 4-18 (M = 11.2, SD = 3.8), served at four clinics operating under a large county-wide public mental health authority. Study youths are 59.8% male; racial-ethnic makeup includes 42.2% African American, 16.9% Caucasian, 38.0% Hispanic, and 2.9% Asian/Other. Youths presenting for a first episode of care during the period of September 1, 2004, and September 30, 2006, and who received psychosocial treatment are included in the study. The mental health authority defined psychosocial treatment as either "therapy" (treatment for internalizing disorders provided by trained mental health professionals) or "skills training" (treatment for externalizing disorders provided by paraprofessionals). Youths who did not receive either skills training or therapy or who did not present for a first episode of care during this period were excluded from the study.

Procedure

Data were obtained through an electronic medical records data extraction. Demographic and clinical data gathered through the routine clinic intake and outcome monitoring procedures were extracted and de-identified by clinic staff before being provided to the research team. All procedures were approved by the Texas A&M Institutional Review Board and the mental health authority's Human Subjects Protection Committee. Youths receiving services at the clinics were administered symptom and functioning measures at intake and approximately every 90 days throughout treatment; however, the number of assessments varied by client and ranged from 1-23 assessments throughout treatment (M = 8.25, SD = 4.78). On average, youths received an assessment every 11.29 weeks (every 79.03 days) during treatment. Please see Measures below for a description of the assessment tools used by the clinics. Parents of youths completed the Ohio Youth Problem Severity (PS) and Functioning Scales (Ogles, Dowell, Hatfield, Melendez, & Carlston, 2004) to gather information about problem severity and functioning. At intake, youth and family demographic information was collected, including gender, age, ethnicity, and insurance status.

The Usual Care context for this study.

For this study, "usual care" was defined as the psychosocial treatment being delivered in the Texas public mental health system. Like the vast majority of states (Chambers, Ringeisen, & Hickman, 2005), Texas is making efforts to incorporate evidence-based treatments into its usual care practices, in part due to a legislative mandate to improve the quality of its mental health services ("House Bill 2292," 2003). This effort, called Resiliency and Disease Management (RDM), was an unfunded mandate to use EBTs in mental health services statewide. Under RDM, seven psychosocial EBTs were selected for use with child and adolescent clients. Selected EBTs for externalizing disorders included Barkley's *Defiant Children* (Barkley, 1997) and *Defiant Teens* (Barkley, Robin, & Edwards, 1999), as well as *Skills Training for Children with Behavior Disorders* (Bloomquist, 1996). Programs selected for internalizing disorders included *Taking Action* (Stark & Kendall, 1996) program for depressed children, the *Adolescent Coping with Depression Course* (Clarke, Lewinsohn, & Hops, 1990), the *Coping Cat* (Kendall, 2000) for anxious children, and *The C.A.T. Project* (Kendall, Choudhury, Hudson, & Webb, 2002) for anxious adolescents. Additionally, clinicians treating youths presenting with diagnoses not covered by one of the seven selected EBTs were encouraged to reference a provided list of EBTs to select a program appropriate for the youth.

Despite the efforts to use EBTs in this setting, there are several reasons this should still be considered a "usual care" sample. First, no researchers were involved in this effort, so these were services delivered in routine services without the intervention of a research project. Second, the effort included regular providers and youths that would be typical of real-world mental health settings. Youths entering the clinics were not specifically selected, and there were also not any inclusion or exclusion criteria for services. Third, very few resources were provided to support the use of the EBTs. Clinician training in the EBTs was minimal; all clinicians employed in the clinics attended a 2-day workshop to receive training in either skills training or therapy EBTs. Providers hired into the clinic after the initial training effort received a 1-day training which was provided in tandem with their employee orientation. In addition, the effort did not include a large amount of supervision or feedback and monitoring to ensure treatment adherence and fidelity, as would generally be seen in research contexts.

Measures

Ohio Youth Functioning, Problem Severity, and Satisfaction Scales (Ogles et al., 2004). The Ohio Scales consist of 48 items that assess four domains: functioning, problem severity, hopefulness, and satisfaction with services. The present study utilized the Ohio Youth Functioning and Problem Severity Scales, parent-report forms. The Functioning scale consists of 20 items rated on a 5-point Likert scale ranging from 0 (Extreme troubles) to 4 (Doing very well), and evaluates how well a child is able to maintain relationships and complete daily activities. Higher scores on the Functioning scale indicate better functioning. Specifically, scores of >54 on the functioning scale were estimated to indicate normal functioning for this sample, while scores of <44 indicated a clinical range, and scores ≥45 and ≤53 indicated the borderline range for this sample, based on a pilot study of the measure in clinics in Texas (Texas Department of State Health Services, 2003). The Problems scale measures symptom severity and consists of 20 items, scored on a 6-point Likert scale of how severe and how frequent the problem (i.e., symptoms) has been within the past 30 days. Responses range from Not At All (0) to All the Time (5) with higher scores indicating worse symptoms. For this sample, it was estimated that scores ≥ 30 on the problem severity scale were considered to be clinically meaningful and scores ≤12 indicated a minimally symptomatic state (Texas Department of State Health Services, 2003). Within the Problem Severity scale are three subscales including: Externalizing problems (8 items), Internalizing problems (9 items), and Delinquency problems (3 items). The Ohio Scales demonstrate good reliability and validity, and the original measure demonstrates adequate 1-week test-retest reliability (Ogles et al., 2004). Item-level data were not available to assess reliability in the current sample; however, a pilot study of the measure in clinics in Texas found adequate reliability for both scales (α's > .9; Texas Department of State Health Services, 2003).

Predictor variables. Demographic (gender, age and ethnicity, insurance status, family problems) and clinical information (initial symptom and functioning severity as rated by both parents and clinicians, type of diagnosis, number of diagnoses, weeks in treatment, hours in treatment, skills training or therapy) were obtained from the electronic medical records database. A description of demographic and clinical characteristics of the total sample is presented in Table 1.

Chapter 3: Data Analytic Approach

All study analyses were conducted within a multilevel framework (Raudenbush, Bryk, & Congdon, 2002), as the data consist of repeated measures (level 1), nested within children (level 2), nested within clinicians (level 3). Clinicians were also nested within 4 clinics; however, with so few clinics, clinic was added as a predictor of the level 3 intercept in the models, rather than being treated as a fourth level of nesting.

Analyses for Aim 1: Multilevel growth mixture modeling (MGMM) (Muthén, 2004, 2008; Muthén & Asparouhov, 2008, 2011) was used to examine trajectories of change for symptoms and functioning as reported on the Ohio Problem Severity and Functioning Scales (Ogles et al., 2004) parent reports. First, data were examined to test for significant variability, which suggested variability in trajectories of change. Next, a subsample was plotted (i.e., in linear, quadratic, and loglinear forms) and examined, to look for present patterns and subgroups of change before beginning initial analyses.

The next steps, before conducting MGMM, included specifying several different models to guide the MGMM analyses. First a single-class (univariate) growth model was specified, which estimates a single average growth trajectory and a single estimate of variance for the growth parameters (Jung & Wickrama, 2008). An important limitation is that this approach does not allow for unobserved subpopulations, but rather only examines change at the aggregate level. To address this limitation, the next step in model building included specifying models using LCGA and GMM to explore the number of trajectory classes prior to using MGMM.

LCGA is a type of GMM that has the advantage of fixing the growth factor variances and covariances to zero (i.e., it specifies no within-class variance), thus helping with more clear identification of trajectory classes which makes it a useful starting point for conducting GMM, and later, MGMM (Kreuter & Muthén, 2008). Although LCGA can allow for individually-varying times of observations, specifying this option (i.e., TIMESCORES) renders several critical fit statistics unavailable (i.e., LMR and BLRT, comparing the *k*-class model to the *k*-1 class model). LCGA is also limited by its inability to describe variability around the average growth parameters for each class and cannot account for nested data. Therefore, growth mixture modeling (GMM) was also used in order to describe variation around the growth parameters and to gain needed fit statistics. A requirement of GMM that allows access to needed fit statistics is that participants complete measures at similar time intervals, meaning this approach cannot account for individually-varying times of observation if the LMR or BLRT statistics are needed. Therefore, a variant of growth mixture modeling (GMM) was used, such that data were restructured into 90-day time "buckets," around each assessment point. This included identifying assessments every 90 days throughout treatment and sorting any assessments into a "bucket" that fell 45 days prior to or after a designated 90-day time point. Any cases with multiple assessments per time bucket were averaged to create a single value for each 90-day time bucket. This time-structured approach allowed us to examine the LMR-LRT and BLRT statistics to compare the *k*-class model to the *k*-1 class model, which were not available with the LCGA approach. Comparison of results between the LCGA and GMM models was used to help to guide the MGMM analyses and to make decisions about the number of trajectory classes.

Multilevel growth mixture modeling (MGMM) was conducted using Mplus statistical software (version 7.11; Muthen & Muthen, 2013) with full information maximum likelihood (FIML) to account for missing data (Kline, 2010). FIML creates individual functions from each individual's set of data, and thus allows for incomplete data where each individual contributes a varying amount of information to parameter estimates (Kline, 2010). MGMM has the added benefit over other approaches by being able to account for nested data. Further, MGMM addresses the limitations of previous approaches by allowing individually-varying times of observations and estimating within-class variances. However, allowing estimation of within-class variance adds considerable computational burden and can often lead to model nonconvergence or instability. Therefore, it is not uncommon to estimate only the variance around the intercepts, while constraining the slope variance estimates to be equal within-class (Jung & Wickrama, 2008).

Several fit statistics were consulted when considering the number of trajectory classes (Muthén, 2003). First, the Bayesian information criterion (BIC) statistic and Akaike information criterion (AIC) (Akaike, 1973) were used to help determine which model fits the data best, such that the likelihood is maximized while keeping the model parsimonious. Specifically, a model with the smallest BIC value indicates a well-fitting model. The BIC is very sensitive to sample size and penalizes for increased sample size, whereas the AIC penalizes for complexity by adding twice the number of parameters. As well, the Lo, Mendell, and Rubin (2001) likelihood ratio test (LMR-LRT) statistic and bootstrap likelihood ratio test (BLRT) were utilized to assess model fit and decide among competing models with different latent classes (i.e., it uses a corrected likelihood ratio reference distribution to compare the *k*-class and *k*-1 class model) (Yang, 1999). Second, determining the number of classes was guided by theoretical justification, the present research question, and interpretability. Trajectory classes were determined using fit indices and theory. Additionally, other considerations to determine trajectory classes included the following: 1) convergence of the model, 2) a high entropy value (range is 0.00 to 1.00) (Jedidi, Ramaswamy, & DeSarbo, 1993) which assesses whether individuals were neatly classified into one and only one category, with values above 0.80 indicating good classification (Muthén, 2004), 3) no less than 1% of total count in a class (if proportion in a class is less than .01, consider combining with another class) (Jung & Wickrama, 2008), and 4) high posterior probabilities (near 1.0) (Jung & Wickrama, 2008).

Analyses for Aim 2. After trajectory classes were identified for each outcome measure using MGMM, multinomial logistic regression was used to examine how trajectory class membership related to each predictor variable (Kwak & Clayton-Matthews, 2002; Menard, 2001) using HLM 7.1 software (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). Trajectory group membership was the dependent variable, and the independent variables included all demographic and clinical predictors (age, gender, ethnic minority status, type of diagnosis, insurance status, initial symptom and functioning severity as rated by parents, initial measures of functioning as rated by clinicians [i.e., risk of self-harm, family problems, juvenile justice involvement, school problems, substance use], number of treatment sessions, weeks in treatment, skills training or therapy). All continuous predictors were centered around their grand means to reduce multicollinearity and aid interpretation. First, all predictors were examined separately. Then, significant predictors were entered simultaneously into a single model to assess their independent contributions to predicting group membership.

Chapter 4: Results

Multilevel Growth Mixture Modeling

For Aim 1, it was hypothesized that at least three trajectories of change would emerge in our sample (i.e., improvement, no response, and deterioration) for both measures of outcome. As mentioned above, first, a univariate growth curve model was specified. Results indicated that Problem Severity scores were, on average 32.76, $p<.001$, at intake, and significantly decreased by an average of -1.44, $p<.001$, every ninety days. This suggests that, on average for this sample, it would take approximately 173 days for an individual's problem severity score to fall below the clinical cutoff. Results also indicated that Functioning scores were, on average 39.67, $p<.001$, at intake, and significantly increased by an average of .90, $p<.001$, every ninety days. This indicates that it would take 433 days, on average, for individuals in this sample to rise above the clinical cutoff for functioning. Taken together, these results suggested a modest decline in problem severity and a modest improvement in functioning over the course of treatment, on average. This univariate approach to understanding aggregate change offered an important first step in model building; however, next steps focused on understanding trajectory classes within the sample.

After fitting the baseline model, LCGA models were run as a first step in model building to examine trajectory classes. LCGA is a particularly useful first step in specifying GMM and MGMM models because it does not allow individual variation around the growth parameters, which helps with identification of trajectory classes. The LMR-LRT and BLRT fit statistics were not available for these analyses due to the data not being time-structured. Therefore, these analyses offered an exploratory examination of trajectory classes. Using the BIC and entropy to guide decisions about the number of trajectory classes, results suggested that for Problem Severity, the two-class solution had a lower BIC value (BIC= 31,271.59) and higher entropy value (*E*= .70) than the three-class model (BIC= 31,852.32; *E*= .60), favoring the two-class model over the three-class solution. For Functioning, the five-class solution evidenced the best model fit (BIC= 31,432.98; *E*= .62); however, comparisons could not be made to the six-class solution due to one of the classes including an empty class (i.e., 0% of the sample). Table 2 provides the estimated mean intercepts and linear slopes for the trajectory classes for results from the LCGA models for both Problem Severity and Functioning.

Next, GMM models were run as a second approach to guide decisions about trajectory classes. As mentioned previously, a variant of the GMM approach was used such that a “bucket” of time was designated every 90 days throughout treatment, and any assessments that fell 45 days prior to or after a designated 90-day time point were grouped into the time bucket and averaged. Using this approach, GMM offered the needed fit statistics to compare the *k*-class model to the *k*-1 class model and also allowed for unique variability around the growth parameters for trajectory classes, both of which were not available with LCGA. However, this approach was not able to account for nested data and did not allow for individually-varying times of observation due to the use of the bucketing approach. Results for Problem Severity evidenced a lower BIC (BIC= 25,013.99) and higher entropy value (*E*= .62) for the two-class solution compared to the three-class solution (BIC= 25,025.47; *E*= .62). As well, the BLRT favored the two-class model over the three-class solution (p<.001). Discrepant from results of the LCGA, results of the GMM model for Functioning indicated that the three-class solution (BIC=

24,779.95; *E*= .46) fit the data better compared to the four-class solution (BIC= 24,797.30; *E*= .44). The BLRT (*p*<.001) also favored this decision. Table 3 provides the estimated mean intercepts and linear slopes for the trajectory classes for results from the GMM models for both Problem Severity and Functioning.

Using the results from the LCGA and GMM analyses as a guide to determine the number of trajectory classes, the next step included running the MGMM models. MGMM has the advantage over LCGA and GMM being able to account for nested data, allowing individually-varying times of observation, and generating a full range of fit statistics needed to compare trajectory classes. Due to difficulty with model convergence, the within-class variance estimates of the slopes for both Problem Severity and Functioning were fixed to zero (Jung & Wickrama, 2008). Results for Problem Severity indicated that *both* the two-class and three-class models evidenced good model fit. The two-class model had a lower BIC value (BIC= 31,766.97) compared to the three-class model (BIC= 31,771.39); however, the entropy value for the two-class solution (*E*= .66) was lower than the entropy value for the three-class solution (*E*= .72). Further, the BLRT (*p*= .095) for the three-class solution indicated that the three-class should be favored over the two-class solution. Although both the LCGA and GMM analyses suggested a two-class solution, results from the MGMM for Problem Severity indicated the three-class model is the best solution. Thus, the three-class model for Problem Severity was selected as the best solution, based on results from the MGMM and theory. Similar to Problem Severity, the results from the MGMM analyses for Functioning also evidenced discrepant results compared to results from the LCGA and GMM. Results from the MGMM analyses suggested the two-class model was the most favorable solution (BIC= 31,385.64; *E*= .88) compared to the three-class solution (BIC= 31,398.78; *E*= .65). The BLRT value comparing the three-class model to the two-class solution was nonsignificant (p=.33), however the BLRT was significant when comparing the two-class solution over a single-class solution (p<.001), and thus the two-model solution was selected.

Describing Trajectory Classes

Table 4 provides the estimated mean intercepts and linear slopes for the trajectory classes for results from the MGMM for Problem Severity. Figure 1 offers a graphical presentation of the mean trends over time for the Problem Severity trajectory classes. For Problem Severity, 12.2% of individuals had high initial baseline parent-rated problem severity and experienced a very small downward linear trend or a negligible decrease in problem severity over time (Class 1: Remained High). The average intercept for this trajectory class indicated that individuals had a baseline score of 46.06, p<.001, and significantly decreased by -.90, p<.001, every ninety days, on average. Further, these results indicate that it would take approximately 1,601 days for days for individuals in this group to fall below the clinical cutoff for Problem Severity. In contrast, 85.0% of individuals were initially right around the clinical cutoff for Problem Severity and experienced little to no decrease in problem severity over time, on average (Class 2: Remained Moderate). The intercept for this class indicates that these individuals had a baseline Problem Severity score of 29.08, p<.001, and decreased by -1.26, p<.001, every ninety days, on average. The final trajectory class included a small number of individuals (2.8%) who were initially rated very high on problem severity, and experienced a steep downward linear trend in problem severity over time (Class 3: Moderate Improvement). At intake, on average, this trajectory class had a Problem Severity score of 57.64, p<.001, and experienced a significant decrease in severity, -6.66, $p<.001$, every ninety days. These results suggest that, due to the steep negative slope, it would take these individuals 374 days, on average, to fall below the Problem Severity clinical cutoff.

Similarly, there were distinct differences in the growth parameters for the two trajectory classes for Functioning (see Table 4 for estimated mean intercepts and linear slopes for results from MGMM). Figure 2 provides a graphical presentation of the mean trends over time for the Functioning trajectory classes. The majority of individuals (98.7%) were initially rated as low on functioning and experienced a negligible increase in functioning over time (Class 1: No Change). Specifically, the intercept for this trajectory class was at a Functioning score of 39.71, $p<.001$, which increased by .72, $p<.001$, every ninety days, on average. Thus, results suggest that it would take approximately 537 days, on average, for these individuals to rise above the clinical cutoff for functioning. In comparison, a small proportion of individuals (1.3%) were initially rated very low on functioning; however, these individuals experienced a steep upward linear trend, indicating large increases in functioning over time (Class 2: Moderate Improvement). The intercept for this trajectory class was at a Functioning score of 23.66, $p<.001$, which increased by 7.74, $p<.001$, every ninety days, on average, indicating large improvements in functioning during treatment. Due to the steep slope of change, results indicate that it would only take these individuals an average of 237 days to rise above the clinical cutoff for functioning. In sum, these results suggest that the majority of individuals were initially rated low on functioning and experienced small or negligible increases in functioning throughout treatment. On the other hand, a very small proportion of individuals who were rated to have very low functioning experienced much steeper improvements in functioning throughout treatment.

Chi square analyses were conducted to examine the overlap between Problem Severity and Functioning trajectory classes (see Table 5). Results suggested that there was not a strong relationship between the class solutions for Problem Severity and Functioning, $\chi^2(2)= 3.40$, $p= .18$. Further, due to the small number of individuals classified into the Moderate Improvement change trajectory for Functioning, interpretation of the relationship between these measures of Problem Severity and Functioning is limited. Although the majority of individuals in the Remained High trajectory for Problem Severity (97.7%) were categorized into the No Change trajectory for Functioning, a small number (2.3%) experienced Moderate Improvement in Functioning. Comparatively, only 1.0% of youths categorized into the Remained Moderate change trajectory for Problem Severity were categorized into the Moderate Improvement trajectory for Functioning. The Moderate Improvement trajectory class for Problem Severity had the greatest proportion of youths who were categorized into the Moderate Improvement trajectory for Functioning (5.0%).

Multinomial Logistic Regression: Problem Severity

Results of the multinomial logistic regression analyses for Problem Severity are presented in Table 6. To allow comparisons between all three groups, the models were run twice: once with the Moderate Improvement group as the reference group and once with the Remained High group as the reference group. No demographic characteristics were significant predictors of group membership. Results regarding clinical characteristics for Problem Severity indicated that youths with an anxiety disorder diagnosis predicted greater likelihood of falling into the Remained Moderate trajectory class (*OR*= 4.10, $p<.001$) versus the Remained High trajectory class, while youths with a diagnosis of conduct disorder or serious mental illness predicted lower likelihood of being classified into the Remained Moderate trajectory (*OR*= .67 [reverse coded *OR*= 1.50], $p<.05$; *OR*= .48 [reverse coded *OR*= 2.11], $p<.01$) versus the Remained High trajectory class. As well, results also indicated that a diagnosis of conduct disorder predicted lower likelihood of being classified into the Remained Moderate trajectory class (*OR*= .39 [reverse coded *OR*= 2.56], $p<.05$) versus the Moderate Improvement trajectory class. Comorbidity also emerged as a significant predictor of trajectory group membership. Specifically, youths with a greater number of comorbid diagnoses were less likely to be classified into the Remained Moderate change trajectory (*OR*= .68 [reverse coded *OR*= 1.47], $p<.05$; *OR*= .74 [reverse coded *OR*= 1.35], $p<.001$) versus the Moderate Improvement and Remained High change trajectories, respectively.

As expected, results indicated that baseline parent-ratings of problem severity and functioning predicted trajectory class membership. Specifically, youths with higher initial parent-rated symptom severity had a lower likelihood of falling into the Remained High (*OR*= .94 [reverse coded *OR*= 1.06], $p<.001$) and Remained Moderate (*OR*= .87 [reverse coded *OR*= 1.15], $p<.001$) trajectory classes, versus the Moderate Improvement trajectory class. Further, youths with higher initial problem severity were also less likely to be classified into the Remained Moderate trajectory class (*OR*= .92 [reverse coded *OR*= 1.09], $p<.001$) versus the Remained High trajectory class. On the other hand, youths with higher initial parent-rated functioning were more likely to fall into the Remained

Moderate trajectory class (*OR*= 1.07, *p*<.001; *OR*= 1.05, *p*<.001) versus the Moderate Improvement and Remained High trajectory classes, respectively.

Consistent with hypotheses, results on the measure of Problem Severity also indicated that youths with higher initial clinician-rated risk of self-harm, family problems, and school problems predicted lower likelihood of being classified into the Remained Moderate trajectory class (*OR*= .70 [reverse coded *OR*= 1.42], *p*<.01; *OR*= .62 [reverse coded *OR*= 1.61], *p*<.01; *OR*= .67 [reverse coded *OR*= 1.49], *p*<.001) versus the Remained High trajectory class. Youths with higher initial clinician-rated school problems were also less likely to be classified into the Remained Moderate trajectory class (*OR*= .65 [reverse coded *OR*= 1.54], *p*< .05) versus the Moderate Improvement trajectory class. In sum, these results suggest that higher initial clinician-rated problems significantly predicted classification into trajectory class, with youths who were rated higher on initial clinician-rated problems being more likely to fall into more negative trajectories of change (e.g., Remained High). Further, results suggest that predictor variables associated with baseline severity (i.e., type of diagnosis, clinician-rated problems) best differentiated the Remained High and Remained Moderate trajectory groups.

When all significant predictors (i.e., baseline parent-rated problem severity and functioning; clinician-rated baseline risk of self-harm, family problems, and school problems; diagnosis of anxiety, conduct, or serious mental illness) of Problem Severity trajectory class membership were entered simultaneously into a model, a number of the predictors remained significant. Results are presented in Table 7. No diagnostic predictors (i.e., anxiety, conduct, serious mental illness, comorbidity) remained significant. Youths with higher initial parent-rated symptom severity continued to predict lower likelihood of falling into the Remained High (*OR*= .93 [reverse coded *OR*= 1.08], $p<.001$) and Remained Moderate (*OR*= .86 [reverse coded *OR*= 1.16], $p<.001$) trajectory classes, versus the Moderate Improvement trajectory class. As well, youths with higher initial problem severity were also less likely to be classified into the Remained Moderate trajectory class (*OR*= .93 [reverse coded *OR*= 1.08], $p<.001$) versus the Remained High trajectory class. Though initial problem severity continued to be a significant predictor, initial parent-rated functioning no longer remained a significant predictor of trajectory class membership in this model. Results also indicated that youths with higher baseline clinician-rated school problems predicted lower likelihood of being classified into the Remained Moderate trajectory class (*OR*= .77 [reverse coded *OR*= 1.30]) versus the Remained High trajectory class. Clinician-rated risk of self-harm and family problems no longer remained significant predictors of trajectory group membership.

Multinomial Logistic Regression: Functioning

Results of the multinomial logistic regression analyses for Functioning are presented in Table 8. Of note, multinomial logistic regression analyses could not be completed for ethnic minority status, diagnostic match, and clinician-rated juvenile justice involvement and substance use at baseline, due to having not having enough members in both trajectory classes for the analyses to run. No demographic characteristics or diagnostic variables were significant predictors of group membership. Consistent with hypotheses, higher parent-rated problem severity at baseline predicted lower likelihood of being classified into the No Change trajectory class (*OR*= .97 [reverse coded *OR*= 1.04], $p=.01$) versus the Moderate Improvement trajectory class. As well, youths with higher initial parent-rated functioning predicted greater likelihood of being classified into the No Change trajectory class (*OR*= 1.07, *p*<.01) versus the Moderate Improvement trajectory class. Most clinician-rated problems at baseline were not significant predictors of trajectory class membership. However, higher clinician-rated risk of self-harm at baseline predicted lower likelihood of being classified into the No Change trajectory class (*OR*= .52 [reverse coded *OR*= 1.92], *p*<.05) versus the Moderate Improvement trajectory class. These results indicate that youths with higher initial parent- or clinician-rated problem severity and/or lower functioning were more likely to fall into a more negative change trajectory.

Results also indicated that amount of treatment received, specifically, number of treatment sessions and weeks in treatment, significantly predicted Functioning trajectory class membership. Specifically, greater number of treatment sessions and more weeks in treatment predicted lower likelihood of being classified into the No Change trajectory class (*OR*= .93 [reverse coded *OR*= 1.08], *p*=.005; *OR*= .94 [reverse coded *OR*= 1.06], *p*<.05) versus the Moderate Improvement trajectory class. This suggests that youths with lower initial functioning, who were also more likely to have higher initial problem severity, were both more likely to require a greater number of treatment sessions and to spend more weeks in treatment.

In a model containing all significant predictors (i.e., baseline parent-rated problem severity and functioning, clinician-rated baseline risk of self-harm, number of weeks in treatment, number of treatment sessions) of trajectory class membership for Functioning, two predictors remained significant. Results are presented in Table 9. Parent-rated problem severity at baseline and treatment dose and duration (i.e., number of treatment sessions and number of weeks in treatment) no longer remained significant predictors of trajectory group membership. Results indicated that higher initial parent-rated functioning continued to predict greater likelihood of being classified into the No Change trajectory class (*OR*= 1.06, *p*<.05) versus the Moderate Improvement trajectory class. As well, clinician-rated risk of self-harm at baseline continued to predict lower likelihood of being classified into the No Change trajectory class (*OR*= .49 [reverse coded *OR*= 2.05], *p*<.05) versus the Moderate Improvement trajectory class. These results are consistent with the hypothesis that youths with lower initial parent-rated functioning and clinician-rated problem severity would be more likely to be classified into a more negative trajectory of change. In this case, the Moderate Improvement trajectory class has a lower baseline (intercept) score for functioning compared to the No Change trajectory class.

Chapter 5: Discussion

This study facilitates understanding of change trajectories for youths treated in UC, and predictors of trajectory group membership. This study provided an extension of previous research on change trajectories in youth UC, using two measures of outcomes (i.e., Problem Severity and Functioning) in a naturalistic sample. Previous research has explored change using theoretically-derived change trajectories and outcome categories determined a priori, however this study is the first to use a data-driven approach to identify groups of individuals who respond in similar ways. Results of this study are consistent with previous literature reporting less than encouraging outcomes of UC.

As hypothesized, several (i.e., three) distinct trajectories of change were identified on a measure of Problem Severity that included the following: Remained High, Remained Moderate, Moderate Improvement. Though three distinct trajectories were identified, inconsistent with hypotheses, no trajectory emerged that classified individuals who deteriorated during treatment, as was seen in the study by Warren and colleagues (2010). However, the trajectories identified in this study largely indicated that youths in UC are making little or no improvement. Two of the groups, representing 97.3% of the sample, experienced little to no meaningful improvement in symptoms. One group, the Remained High group, started treatment well above the clinical cutoff and basically remained so – the slope estimate for this group indicated that these individuals would have needed over 4 years of treatment to fall below the clinical cutoff on the problem severity measure. The other, the Remained Moderate group, representing the majority of the sample, started treatment just below the clinical cutoff and remained stable. Finally, a very small number of individuals were classified into the Moderate Improvement trajectory class; they were initially rated high on Problem Severity and experienced greater improvement during treatment compared to the other two classes. Though this class improved, they still would have needed nearly a year of treatment to fall below the clinical cutoff, which exceeded the average length of treatment for this sample (approximately 40 weeks).

This study also identified two trajectories of change on a measure of Functioning: No Change and Moderate Improvement. Similar to the patterns found for Problem Severity, nearly all participants (98.7%) were classified into the No Change trajectory class. These youths experienced a very small increase in functioning over the course of treatment and would have needed to greatly exceed the average length of treatment to fall within the range of Normal Functioning. A second group was identified: a Moderate Improvement trajectory class that started with very low functioning and showed moderate improvements in functioning throughout treatment. However, this group made up a very small proportion of the sample (1.3%), so, although the large sample employed was adequately powered to detect this group, it is not clear that this is a clinically meaningful subgroup and that the main conclusion to be drawn regarding functional improvement in this sample is that there was not any.

When considered together, results of trajectory class membership for Problem Severity and Functioning are consistent with previous literature that, generally, youths are not experiencing positive outcomes in UC (e.g., Weersing & Weisz, 2002). Though a small percentage of the sample improved, these improvements were generally small and required a longer than average duration of treatment in order to cross the clinical cutoff and/or to fall into the range of normal symptoms/functioning. Regarding the change trajectories identified in the present study, these differed somewhat from trajectories in previous literature, which may have been due to different approaches to examining change (i.e., theory-driven versus data-driven approach) or the nature of the data in this study. As mentioned previously, the timing of assessments in the present study would have precluded the detection of sudden gains or rapid response trajectories, as youths, on average, received an assessment every 11.29 weeks (every 79.03 days) during treatment. The nature of the data and UC sample may have also made it difficult to detect some trajectories of change (i.e., deterioration) due to dropout. That is, if youths were not experiencing positive treatment gains they may have been more likely to drop out of treatment early, making it difficult to detect a trajectory of individuals who may have been likely to experience deterioration.

Though no study has examined EBT trajectories, randomized trials examining the effects of EBTs for youths (e.g., Kendall et al., 1997; Lewinsohn, Clarke, Hops, & Andrews, 1990; Walkup et al, 2008; Weisz et al., 1987) have generally found that, on average, youths respond positively to treatment (i.e., EBTs). Further, a study by Weersing and Weisz (2002) used a benchmarking procedure to compare trajectories of depressed youth treated in the community compared to trajectories in RCTs. Results from this study indicated that youths treated in the community evidenced worse outcomes and improved more slowly compared to youths in RCTs, and appeared to follow a trajectory that was more similar to the control group in the study. This may suggest that we could expect change trajectories in UC to be less positive than change in RCTs; however, research in this area is needed to better understand differences in change trajectories in UC and EBT samples. In this study, findings for both measures of outcomes supported the hypothesis for the existence of distinct trajectories of change, indicating differences in individual response to treatment. However, the extent to which these all of the identified trajectory classes are meaningful (e.g., Moderate Improvement trajectory class for Functioning) or representative of change during UC treatment remains unclear.

Interestingly, in addition to identifying a different numbers of classes for the Problem Severity and Functioning measures, there was also a lack of concordance between how these two measures classified specific individuals. Examination of the cross-tabulations of group memberships suggested that improvements in symptoms do not necessarily lead to improvements in functioning. Most of the individuals (95.0%) who showed improvements (i.e., Moderate Improvement trajectory class) on the Problem Severity scale fell into the No Change group for Functioning. As well, only 9 individuals were classified into the Moderate Improvement trajectory class for Functioning, further suggesting that improvements in symptoms may not directly influence functioning. Thus, in addition to the need to improve UC in general, there seems to be a particular need to improve treatment in a way that leads to functional improvement in addition to symptom improvement.

Examination of predictors of trajectory class membership indicated that trajectory classes were primarily determined by baseline problem severity and functioning, as well as other indicators of severity (e.g., comorbidity, school problems). Given that most of the differentiation between the trajectory groups was in the intercept (i.e., the estimate of baseline severity), this finding is not surprising. When significant predictors were entered simultaneously, there were only two predictors other than initial severity that remained significant. For Problem Severity, clinician ratings of school problems differentiated the two trajectory groups that showed no change (i.e., Remained High, Remained Moderate); for Functioning, youths whose clinicians rated them as at-risk for self-harm were more likely to fall in the Moderate Improvement group. Thus, other than baseline severity levels, the present set of predictors provided little information about how to predict which youths are at risk for treatment failure, which is likely in part due to the very small number of individuals identified as evidencing improvement.

The present study has several notable strengths. First, this study expanded our current understanding of youth UC by examining trajectories of change and predictors of trajectory group membership. This study was also the first of its kind to use a naturalistic sample to examine study aims. Further, unlike previous investigations that have examined change using theoretically-derived change trajectories and outcome categories determined a priori, this study used a data-driven approach to identify trajectories of change. Finally, compared to previous literature, this study included an expanded set of demographic and clinical predictors.

The results of the current study should be interpreted in light of several limitations. First, this study included only parent-report measures of outcomes, which may not fully or accurately describe youth problem severity and functioning, as well as change during treatment. Due to important differences between parent-, self-, and clinician-report, future work should include self- and clinician-report on measures of outcomes to more fully capture perspectives of change throughout treatment. Second, despite using a data-driven approach to classifying change, this study did not categorize individuals based on final outcomes (i.e., classifying individuals using a measure of reliable change in addition to examination of trajectory class). Therefore, although information about change during the treatment *process* is known, final classification of outcomes for individuals remains unknown. Future research should therefore examine *both* trajectories of change as well as treatment outcomes in order to gain a richer understanding of change during the treatment process as well as final outcomes. Finally, while the data set included a large set of predictor variables, it is possible that other variables not measured here might help identify youths at risk for treatment failure. For example, perhaps client readiness to change and motivation to change would be better able to predict which youths are more likely to experience treatment success. Additionally, therapeutic alliance and consensus on goals, as well as therapist characteristics (e.g., level of education, theoretical orientation), may also be important predictors of outcomes.

In conclusion, this study contributes to the extant literature by providing some evidence for change trajectories and predictors of trajectory group membership using a diverse, naturalistic sample, and data-driven methodology. Results indicate several possibilities for future research. First, findings are consistent with previous literature that outcomes of youth UC are not encouraging, and the predictors examined here did not differentiate the small percentage of participants who improved from those who did not, other than intake levels of symptoms and functioning. However, it may be the case that the identified trajectory classes do not fully capture how youths are changing in this sample. MGMM models used to determine the final group classifications indicated significant variability in the intercepts within-classes. However, within-class variance estimates for the slopes were fixed to zero to help with model convergence; therefore, the current study does not allow examination of the slope variance estimates. Thus, within each group there may be variability in the way youths are changing that was not captured using this approach. Therefore, future research should use different methodological approaches to capture and describe change to more accurately, in order to better understand change in UC. For example, approaches employing *both* data- and theoretically-driven methodologies to describe change groups may better capture and describe change in youth UC. As well, new methodologies may allow for more accurate examination of predictors of trajectory group membership, which could aid in better understanding targets for treatment.

It may also be the case that the measures used to track change may not be well-suited to capture change for this sample. The literature suggests that although the Ohio Scales are sensitive to change in clinical samples, sensitivity to change in naturalistic samples has not yet been examined (Ogles et al., 2004). As well, there is some evidence that the Functioning scales are less sensitive to capturing change longitudinally in a clinical sample compared to the Problem Severity scale (Ogles et al., 2004). Therefore, future research using different measures of symptoms and functioning may be warranted, as it is possible that perhaps the measures used in this study were not adequately able to capture change, or might have needed to be administered more frequently. Taken together, this study suggests that youths in UC are generally not demonstrating encouraging improvements. Therefore, it will be important for future research to continue to explore change in youth UC using different methodologies and measures to answer questions about whether it is possible to identify specific targets for treatment (i.e., by using different methodologies to understand whether there are groups of youths who respond differentially to UC). As well, this study underscores the need for continued efforts aimed at developing strategies to effectively implement EBTs in UC settings.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *International Symposium on Information Theory, 2nd, Tsahkadsor, Armenian SSR*, 267-281.

Angold, A., Costello, J. E., Burns, B. J., Erkanli, A., & Farmer, E. M. Z. (2000). Effectiveness of nonresidential specialty mental health services for children and adolescents in the 'real world'. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(2), 154-160. doi: 10.1097/00004583-200002000-00013

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271-285.

Ash, S. E., & Weis, R. (2009). Recovery among youths referred to outpatient psychotherapy: Reliable change, clinical significance, and predictors of outcome. *Child & Adolescent Social Work Journal, 26*(5), 399-413. doi: 10.1007/s10560-009-0171-3

Baker-Ericzén, M. J., Hurlburt, M. S., Brookman-Frazee, L., Jenkins, M. M., & Hough, R. L. (2010). Comparing child, parent, and family characteristics in usual care and empirically supported treatment research samples for children with disruptive behavior disorders. *Journal of Emotional and Behavioral Disorders, 18*(2), 82-99. doi: 10.1177/1063426609336956

Barkley, R. A. (1997). *Defiant children: A clinician's manual for assessment and parent training* (2nd ed.). New York: Guilford Press.

Barkley, R. A., Guevremont, D. C., Anastopoulos, A. D., & Fletcher, K. E. (1992). A comparison of three family therapy programs for treating family conflicts in adolescents with attention-deficit hyperactivity disorder. *Journal of Consulting and Clinical Psychology, 60*(3), 450.

Barkley, R. A., Robin, A. L., & Edwards, G. H. (1999). *Defiant teens: A clinician's manual for assessment and family intervention*. New York: Guilford Press.

Barlow, D. H. (1996). Health care policy, psychotherapy research, and the future of psychotherapy. *American Psychologist, 51*(10), 1050-1058. doi: 10.1037/0003-066x.51.10.1050

Bickman, L., Lambert, E. W., Andrade, A. R., & Penaloza, R. V. (2000). The Fort Bragg continuum of care for children and adolescents: Mental health outcomes over 5 years. *Journal of Consulting and Clinical Psychology, 68*(4), 710-716. doi: 10.1037/0022-006x.68.4.710

Bishop, M., Bybee, T., Lambert, M., Burlingame, G., Wells, M. G., & Poppleton, L. (2005). Accuracy of a rationally derived method for identifying treatment failure in children and adolescents. *Journal of Child and Family Studies, 14*(2), 207-222. doi: 10.1007/s10826-005-5049-1

Bloomquist, M. L. (1996). *Skills training for children with behaivor disorders*. New York, NY: Guilford Press.

Brent, D. A., Kolko, D. J., Birmaher, B., Baugher, M., Bridge, J., Roth, C., & Holder, D. (1998). Predictors of treatment efficacy in a clinical trial of three psychosocial treatments for adolescent depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(9), 906-914.

Breslin, F. C., Sobell, M. B., Sobell, L. C., Buchan, G., & Cunningham, J. A. (1997). Toward a stepped care approach to treating problem drinkers: The predictive utility of within-treatment variables and therapist prognostic ratings. *Addiction, 92*(11), 1479-1489. doi: 10.1111/j.1360-0443.1997.tb02869.x

Bybee, T., Lambert, M., & Eggett, D. (2007). Curves of expected recovery and their predictive validity for identifying treatment failure. *Tijdschrift voor Psychotherapie, 33*(6), 272-281. doi: 10.1007/bf03062308

Carroll, K. M., & Rounsaville, B. J. (2003). Bridging the gap: A hybrid model to link efficacy and effectiveness research in substance abuse treatment. *Psychiatric Services, 54*, 333-339.

Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin, 98*(2), 388-400. doi: 10.1037/0033-2909.98.2.388

Chambers, D. A., Ringeisen, H., & Hickman, E. E. (2005). Federal, state, and foundation initiatives around evidence-based practices for child and adolescent mental health. *Child and Adolescent Psychiatric Clinics of North America, 14*(2), 307-327. doi: 10.1016/j.chc.2004.04.006

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7-18. doi: 10.1037/0022-006x.66.1.7

Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63*(5), 718-725. doi: 10.1037/0022-006x.63.5.718

Clarke, G. N., Hops, H., Lewinsohn, P. M., Andrews, J., Seeley, J. R., & Williams, J. (1992). Cognitive-behavioral group treatment of adolescent depression: Prediction of outcome. *Behavior Therapy, 23*(3), 341-354.

Clarke, G. N., Lewinsohn, P., & Hops, H. (1990). *Leader's manual for adolescent groups: Adolescent coping with depression course.* Portland, OR: Kaiser Permanente Center for Health Research.

Cuijpers, P., van Lier, P. A. C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders, 89*(1–3), 137-146. doi: 10.1016/j.jad.2005.09.001

Emslie, G. J., Mayes, T. L., Laptook, R. S., & Batt, M. (2003). Predictors of response to treatment in children and adolescents with mood disorders. *Psychiatric Clinics of North America, 26*(2), 435-456. doi: 10.1016/s0193-953x(02)00110-7

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*(3), 251-278.

Flay, B. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*(5), 451-474. doi: 10.1016/0091-7435(86)90024-1

Flay, B., Biglan, A., Boruch, R., Castro, F., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151-175. doi: 10.1007/s11121-005-5553-y

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research, 37*(1-2), 15-26. doi: 10.1007/s10488-010-0279-y

Grilo, C. M., Masheb, R. M., & Wilson, G. T. (2006). Rapid response to treatment for binge eating disorder. *Journal of Consulting and Clinical Psychology, 74*(3), 602-613. doi: 10.1037/0022-006x.74.3.602

Halfon, N., Inkelas, M., & Wood, D. (1995). Nonfinancial barriers to care for children and youth. *Annual Review of Public Health, 16*(1), 447-472.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the Special Section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 683-687. doi: 10.1037/0022-006x.63.5.683

Hops, H., Lewisohn, P. M., & Roberts, R. E. (1990). Psychological correlates of depressive symptomology among high school students. *Journal of Clinical Child Psychology, 19*(3), 211-220.

House Bill 2292, State of Texas (2003).

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*(10), 1059-1064. doi: 10.1037/0003-066x.51.10.1059

Hunsley, J. (2007). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice, 38*(2), 113-121. doi: 10.1037/0735-7028.38.2.113

Hunsley, J., & Lee, C. M. (2006). *Introduction to clinical psychology: An evidence-based approach*. Toronto, Ontario, Canada: Wiley.

Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*(1), 21-33. doi: 10.1037/0735-7028.38.1.21

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice, 1*(2), 138-155. doi: 10.1111/j.1468-2850.1994.tb00016.x

Jedidi, K., Ramaswamy, V., & DeSarbo, W. S. (1993). A maximum likelihood method for latent class regression involving a censored dependent variable. *Psychometrika, 58*(3), 375-394.

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*(3), 416-425. doi: 10.1037/0022-006x.74.3.416

Jung, T., & Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146-159. doi: 10.1037/0003-066x.63.3.146

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58*(6), 729-740. doi: 10.1037/0022-006x.58.6.729

Kazdin, A. E., & Wassell, G. (2000). Therapeutic changes in children, parents, and families resulting from treatment of children with conduct problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(4), 414-420. doi: 10.1097/00004583-200004000-00009

Kendall, P. C. (2000). *Cognitive behavioral therapy for anxious children therapist manual* (2nd ed.). Ardmore, PA: Workbook Publishing, Inc.

Kendall, P. C., Choudhury, M., Hudson, J., & Webb, A. (2002). *"The C.A.T. project" workbook for the cognitive behavioral treatment of anxious adolescents*. Ardmore, PA: Workbook Publishing, Inc.

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 clincal trial. *Journal of Consulting and Clinical Psychology, 65*(3), 366-380. doi: 10.1037/0022-006x.65.3.366

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*(2), 282.

Kendall, P. C., & Sugarman, A. (1997). Attrition in the treatment of childhood anxiety disorders. *Journal of Consulting and Clinical Psychology, 65*(5), 883-888. doi: 10.1037/0022-006x.65.5.883

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kreuter, F., & Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology, 24*(1), 1-31.

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing research, 51*(6), 404-410.

Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy, 21*(4), 385-401. doi: http://dx.doi.org/10.1016/S0005-7894(05)80353-3

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767-778.

Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27*(2), 138-145. doi: 10.1207/s15374424jccp2702_1

Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology, 58*(10), 1291-1304. doi: 10.1002/jclp.10070

Menard, S. (2001). *Applied logistic regression analysis* (Vol. 106). Thousand Oaks, CA: Sage Publications, Inc.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003).

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences. Thousand Oaks, CA: Sage Publications, Inc.*, 345-368.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. *Advances in latent variable mixture models, 1*, 1-24.

Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. *Longitudinal data analysis*, 143-165.

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. *Handbook of advanced multilevel analysis*, 15-40.

Muthen, L. K., & Muthen, B. O. (1998). *Mplus user's guide.* Los Angeles: Muthen & Muthen.

Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin, 126*(6), 964-981. doi: 10.1037/0033-2909.126.6.964

Nelson, P. L. (2011). Change trajectories and early warning system to identify youth at risk for negative psychotherapy outcome. *71*. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-99120-415&site=ehost-live

Norcross, J. C., Beutler, L. E., & Levant, R. F. (2006). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions.* Washington, DC: American Psychological Association.

Ogles, B. M., Dowell, K., Hatfield, D., Melendez, G., & Carlston, D. L. (2004). The Ohio Scales. *The use of psychological testing for treatment planning and outcomes assessment, 2*, 275-304.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*(2), 109-118. doi: 10.1037/h0024436

Penava, S. J., Otto, M. W., Maki, K. M., & Pollack, M. H. (1998). Rate of improvement during cognitive–behavioral group treatment for panic disorder. *Behaviour Research and Therapy, 36*(7–8), 665-673. doi: 10.1016/S0005-7967(98)00035-7

Phillips, S. D., Hargis, M. B., Kramer, T. L., Lensing, S. Y., Taylor, J. L., Burns, B. J., & Robbins, J. M. (2000). Toward a level playing field: Predictive factors for the outcomes of mental health treatment for adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(12), 1485-1495. doi: 10.1097/00004583-200012000-00008

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2011). HLM 7. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (Eds.). (2002). *Hierarchical linear modeling*. Thousands Oaks, CA: Sage Publications, Inc.

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

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1994). Response of depressed adolescents to cognitive-behavioral treatment: Do differences in initial severity clarify the comparison of treatments? *Journal of Consulting and Clinical Psychology, 62*(4), 851.

Texas Department of State Health Services. (2003). Validation and norms for the Ohio Scales among children served by the Texas Department of Mental Health and Mental Retardation. Austin, TX.

Sexton, T. L., & Kelley, S. (2010). Finding the common core: Evidence-based practices, clinically relevant evidence, and core mechanisms of change. *Administration and Policy in Mental Health and Mental Health Services Research, 37*(1), 81-88. doi: 10.1007/s10488-010-0277-0

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.

Southam-Gerow, M. A., Velez, J. R., & Kendall, P. C. (2003). Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. *Journal of Clinical Child & Adolescent Psychology, 32*(3), 375.

Stark, K., & Kendall, P. C. (1996). *Treating depressed children: Therapist manual for "taking action"*. Ardmore, PA: Workbook Publishing, Inc.

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. *New England Journal of Medicine, 359*(26), 2753-2766. doi: doi:10.1056/NEJMoa0804633

Warren, J. S., Nelson, P. L., Burlingame, G. M., & Mondragon, S. A. (2012). Predicting patient deterioration in youth mental health services: Community mental health vs. managed care settings. *Journal of Clinical Psychology, 68*(1), 24-40. doi: 10.1002/jclp.20831

Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology, 78*(2), 144-155. doi: 10.1037/a0018544

Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology, 70*(2), 299-310. doi: 10.1037/0022-006x.70.2.299

Weiss, B., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67*(1), 82-94. doi: 10.1037/0022-006x.67.1.82

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*: Cambridge University Press.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688-701. doi: 10.1037/0022-006x.63.5.688

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56*, 337-363. doi: 10.1146/annurev.psych.55.090902.141449

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*(7), 671-689. doi: 10.1037/0003-066x.61.7.671

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55*(4), 542-549. doi: 10.1037/0022-006x.55.4.542

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*(3), 450-468. doi: 10.1037/0033-2909.117.3.450

Yang, C. C. (1999). *Finite mixture model selection with psychometric applications.* ProQuest Information & Learning.

Figure 1

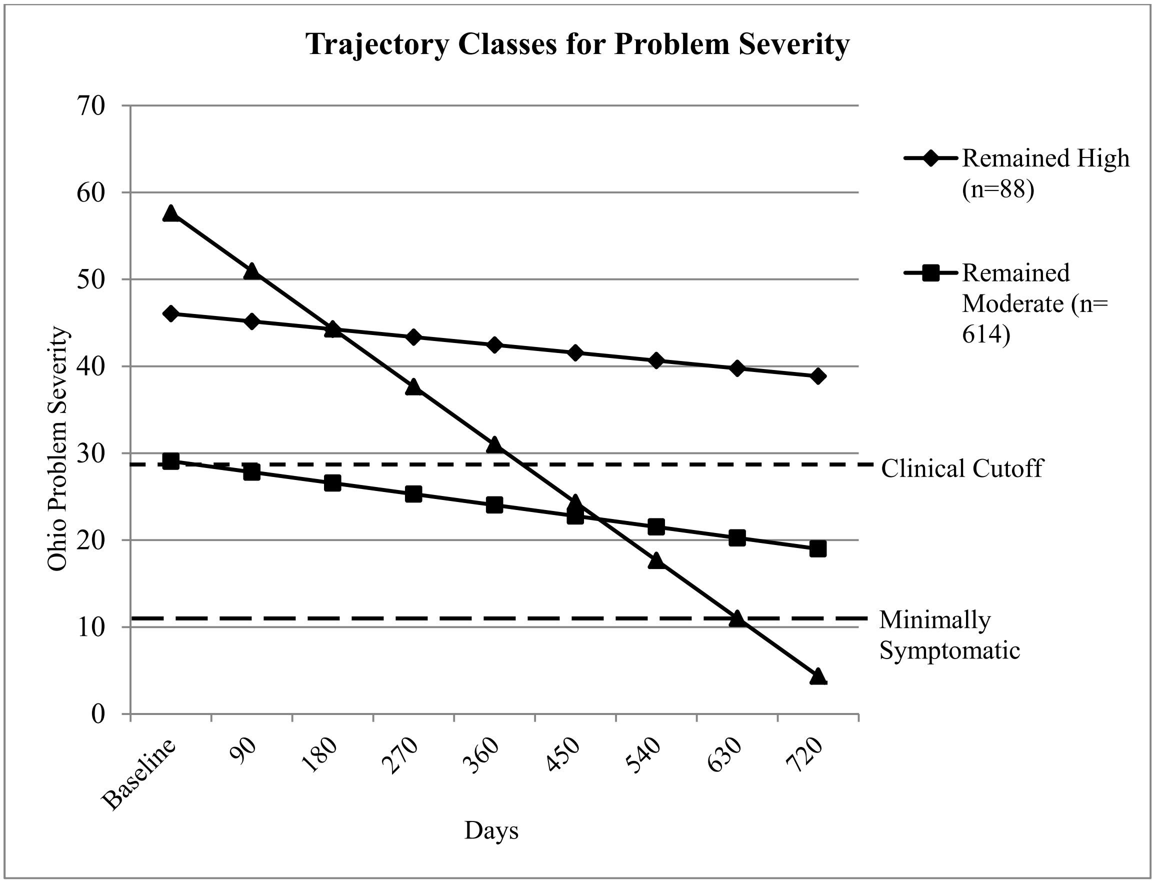
Trajectory Classes for Problem Severity
70
60
50
40
30
20
10
0
Ohio Problem Severity
Baseline
90
180
270
360
450
540
630
720
Days
Remained High (n=88)
Remained Moderate (n= 614)
Clinical Cutoff
Minimally Symptomatic

Figure 2

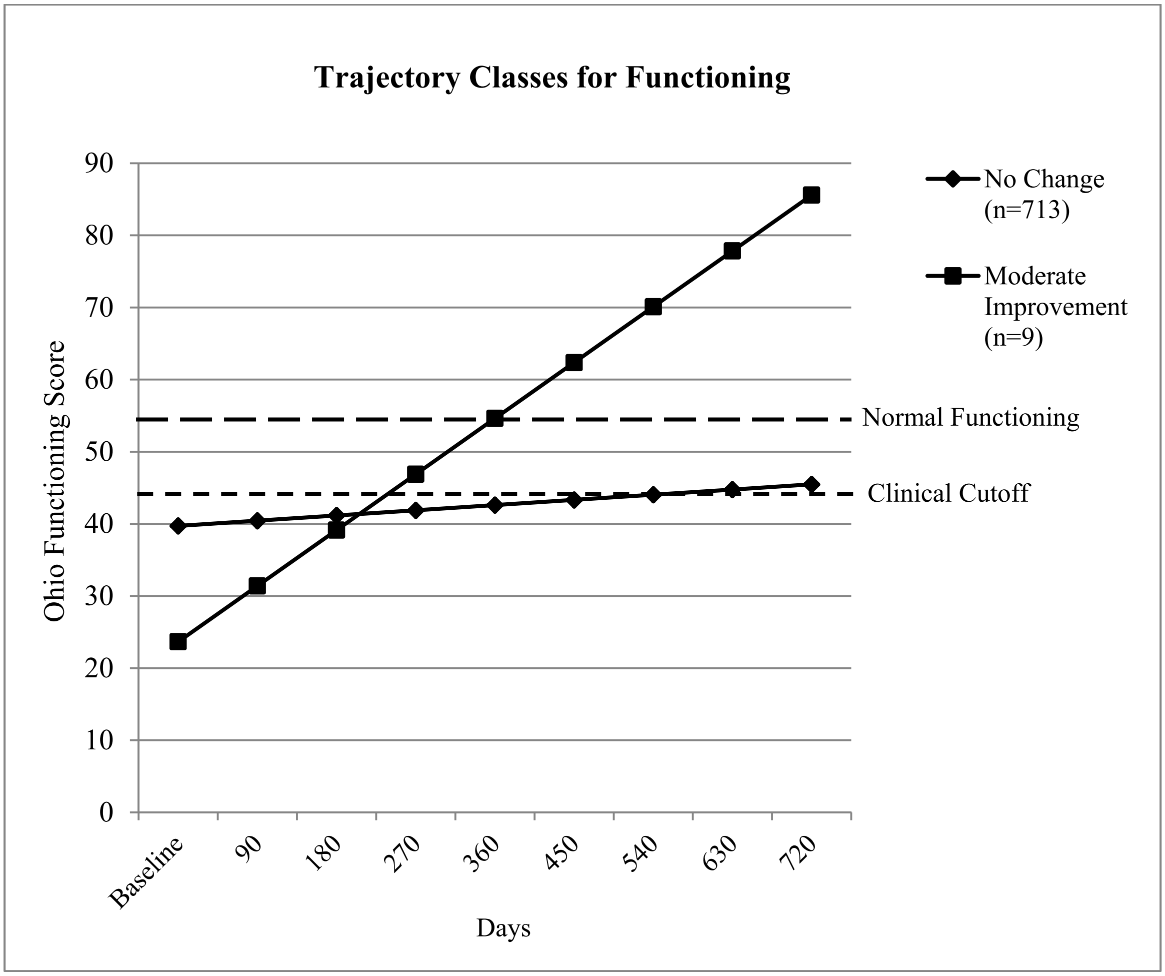
Trajectory Classes for Functioning
Ohio Functioning Score
90
80
70
60
50
40
30
20
10
0
Baseline
90
180
270
360
450
540
630
720
Days
No Change (n=713)
Moderate Improvement (n=9)
Normal Functioning
Clinical Cutoff

Table 1
Demographic and Clinical Characteristics of Total Sample

	Total Sample (N= 722)
Demographic Variables	
Mean (SD) Age	11.2 (3.8)
% Male	59.8%
% Caucasian	16.9%
% African American	42.2%
% Hispanic	38.0%
% Asian/Other	2.9%
% Ethnic Minority Status	83.1%
% Insured	66.5%
Clinical Variables	
Diagnosis	
% ADHD [a]	50.0%
% Anxiety [a]	7.6%
% Conduct [a]	27.1%
% Depression [a]	37.3%
% Serious Mental Illness [a]	26.9%
% Other [a]	24.2%
Mean (SD) Comorbidity	1.7 (1.0)
% Diagnostic Match	82.5%
% Therapy [a]	48.8%
% Skills Training [a]	58.3%
Parent Ratings at Baseline	
Mean (SD) Problem Severity	38.1 (17.6)
Mean (SD) Functioning	36.6 (14.7)
Clinician Ratings at Baseline	
Mean (SD) Risk of Self-Harm	1.3 (.7)
Mean (SD) Family Problems	2.6 (.9)
Mean (SD) Juvenile Justice	1.1 (.5)
Mean (SD) School Problems	3.1 (1.2)
Mean (SD) Substance Use	1.2 (.7)
Mean (SD) Number of Treatment Sessions	8.2 (7.8)
Mean (SD) Weeks in Treatment	39.6 (14.7)

[a] Percentage value denotes % within category; numbers do not add to 100

Table 2
LCGA Parameter Estimates of Trajectory Classes for Problem Severity and Functioning

Parameter Estimates	*Problem Severity Class* Small Improvement (33.5%)	No Change (66.5%)
Mean intercept	45.53***	25.56***
Mean linear slope [a]	-1.71**	-.90***

Parameter Estimates	*Functioning Class* Normal Functioning (2.4%)	Borderline Functioning (40.2%)	No Change (41.2%)	Small Improvement (13.8%)	Moderate Improvement (2.3%)
Mean intercept	60.44***	47.60***	36.95***	24.94***	18.63***
Mean linear slope [a]	.18*	.27*	.99***	1.44***	7.74***

[a] Slope indicates average change every 90 days
**p*<.05; ***p*<.01; ****p*<.001

Table 3
GMM Parameter Estimates of Trajectory Classes for Problem Severity and Functioning

	Problem Severity Class	
Parameter Estimates	Small Improvement (25.2%)	No Change (74.8%)
Mean intercept	47.81***	26.65***
Mean linear slope [a]	-2.26***	-.90***

	Functioning Class		
Parameter Estimates	Remained High (40.4%)	Remained Moderate (58.6%)	Moderate Improvement (1.0%)
Mean intercept	43.65***	38.04***	15.10*
Mean linear slope [a]	-.53*	1.68***	9.49***

[a] Slope indicates average change every 90 days
*p<.05; **p<.01; ***p<.001

Table 4
MGMM Parameter Estimates of Trajectory Classes for Problem Severity and Functioning

	Problem Severity Class		
Parameter Estimates	Remained High (12.2%)	Remained Moderate (85.0%)	Moderate Improvement (2.8%)
Mean intercept	46.06***	29.08***	57.64***
Mean linear slope [a]	-.90***	-1.26***	-6.66***

	Functioning Class	
Parameter Estimates	No Change (98.7%)	Moderate Improvement (1.3%)
Mean intercept	39.71***	23.66***
Mean linear slope [a]	.72***	7.74***

[a] Slope indicates average change every 90 days
****p*<.001

Table 5
Cross-Tabulation of the Three Problem Severity Trajectory Classes and the Two Functioning Trajectory Classes

	Functioning Trajectory Classes	
Problem Severity Trajectory Classes (n)	No Change	Moderate Improvement
Remained High (88)	97.7%	2.3%
Remained Moderate (614)	99.0%	1.0%
Moderate Improvement (20)	95.0%	5.0%

Note. Cells contain percentages. Values indicate the percentages of Problem Severity trajectory classes within Functioning trajectory classes (e.g., within columns).

Table 6

Multinomial Logistic Regression of Predictors of Trajectory Group Membership for Ohio Problem Severity

	Remained High vs. Moderate Improvement [a]		Remained Moderate vs. Moderate Improvement [a]		Remained Moderate vs. Remained High [b]	
Predictor Variable	β	Odds Ratio (95% CI)	β	Odds Ratio (95% CI)	β	Odds Ratio (95% CI)
Demographic Predictors						
Age	-.02	.98 (.86, 1.12)	.01	1.01 (.89, 1.16)	.04	1.04 (.99, 1.08)
Gender	-.37	.69 (.22, 2.17)	.06	1.06 (.40, 2.86)	.43	1.53 (.90, 2.59)
Ethnic Minority Status	.45	1.57 (.70, 3.52)	.16	1.17 (.48, 2.87)	-.29	.75 (.43, 1.32)
Insurance Status	.63	1.88 (.72, 4.95)	.23	1.26 (.56, 2.85)	-.40	.67 (.42, 1.08)
Clinical Predictors						
Diagnosis [c]						
ADHD	.71	2.03 (.66, 6.22)	.51	1.67 (.56, 4.93)	-.20	.82 (.57, 1.18)
Anxiety	-.90	.41 (.04, 3.94)	.51	1.66 (.23, 12.05)	1.41**	4.10 (1.55, 10.83)
Conduct	-.54	.58 (.24, 1.40)	-.94*	.39 (.16, .97)	.40*	.67 (.47, .96)
Depression	-.44	.64 (.22, 1.84)	-.60	.55 (.22, 1.36)	-.16	.85 (.59, 1.24)
Serious Mental Illness	.38	1.47 (.45, 4.78)	-.36	.70 (.22, 2.19)	-.74**	.48 (.30, .74)
Other	-.87	.42 (.16, 1.10)	-.79	.45 (.20, 1.05)	.08	1.08 (.66, 1.76)
Comorbidity	-.08	.92 (.70, 1.21)	-.38*	.68 (.54, .87)	-.30***	.74 (.63, .87)
Diagnostic Match	-.33	.72 (.20, 2.60)	-.11	.89 (.28, 2.88)	.22	1.25 (.73, 2.12)
Therapy	-.22	.80 (.29, 2.19)	-.41	.66 (.25, 1.78)	-.19	.83 (.55, 1.23)
Skills Training	-.07	.93 (.34, 2.60)	.01	1.01 (.36, 2.89)	.08	1.09 (.78, 1.52)

Parent Ratings at Baseline						
Problem Severity	-.06***	.94 (.92, .97)	-.14***	.87 (.84, .90)	-.08***	.92 (.90, .94)
Functioning	.01	1.01 (.99, 1.05)	.07***	1.07 (1.04, 1.10)	.05***	1.05 (1.04, 1.07)
Clinician Ratings at Baseline						
Risk of Self-Harm	-.05	.95 (.57, 1.58)	-.40	.67 (.41, 1.10)	-.35**	.70 (.54, .91)
Family Problems	.15	1.16 (.63, 2.13)	-.33	.72 (.44, 1.19)	-.48**	.62 (.47, .83)
Juvenile Justice	-.48	.62 (.26, 1.49)	-.61	.54 (.29, 1.02)	-.14	.87 (.50, 1.51)
School Problems	-.03	.97 (.65, 1.44)	-.43*	.65 (.43, .98)	-.40***	.67 (.54, .83)
Substance Use	-.14	.87 (.39, 1.93)	-.19	.83 (.44, 1.58)	-.05	.95 (.67, 1.36)
Number of Treatment Sessions	.02	1.02 (.96, 1.08)	.01	1.01 (.95, 1.06)	-.01	.99 (.97, 1.02)
Weeks in Treatment	.01	1.01 (.97, 1.05)	.00	1.00 (.96, 1.04)	-.01	.99 (.98, 1.01)

[a] The reference category is: Moderate Improvement trajectory class

[b] The reference category is: Remained High trajectory class

[c] Primary diagnosis assigned at intake

**p*<.05; ***p*<.01; ****p*<.001

Table 7

Multinomial Logistic Regression of Significant Predictors of Trajectory Group Membership for Ohio Problem Severity Entered Simultaneously

	Remained High vs. Moderate Improvement [a]		Remained Moderate vs. Moderate Improvement [a]		Remained Moderate vs. Remained High [b]	
Predictor Variable	β	Odds Ratio (95% CI)	β	Odds Ratio (95% CI)	β	Odds Ratio (95% CI)
Clinical Predictors						
Diagnosis [c]						
Anxiety	-1.60	.20 (.02, 1.87)	-.72	.49 (.06, 3.71)	.88	2.41 (1.00, 5.84)
Conduct	-.33	.72 (.24, 2.11)	-.27	.76 (.25, 2.35)	.06	1.06 (.63, 1.80)
Serious Mental Illness	.43	1.53 (.41,	-.09	.92 (.26, 3.22)	-.51	.60 (.34, 1.05)
Comorbidity	-.13	.88 (.49, 1.55)	-.30	.74 (.47, 1.17)	-.17	.84 (.63, 1.13)
Parent Ratings at						
Problem Severity	-.08***	.93 (.90, .96)	-.15***	.86 (.83, .90)	-.08***	.93 (.91, .94)
Functioning	-.01	.99 (.95, 1.02)	-.01	.99 (.96, 1.04)	.001	1.01 (.99, 1.03)
Clinician Ratings at						
Risk of Self-Harm	-.05	.95 (.446,	-.27	.77 (.37, 1.60)	-.22	.81 (.62, 1.05)
Family Problems	.47	1.61 (.84,	-.47	1.59 (.89, 2.84)	-.01	.99 (.73, 1.36)
School Problems	.06	1.06 (.72,	-.21	.81 (.53, 1.25)	-.26*	.77 (.60, .99)

[a] The reference category is: Moderate Improvement trajectory class

[b] The reference category is: Remained High trajectory class

[c] Primary diagnosis assigned at intake

**p*<.05; ***p*<.01; ****p*<.001

Table 8
Multinomial Logistic Regression of Predictors of Trajectory Group Membership for Ohio Functioning [a]

Predictor Variable	β	S.E (β)	χ^2	Odds Ratio (95% CI)
Demographic Predictors				
Age	.13	.09	66.77	1.14 (.96, 1.35)
Gender	-.61	.68	85.28	.54 (.14, 2.05)
Insurance Status	-1.36	1.03	67.71	.26 (.03, 1.96)
Clinical Predictors				
Diagnosis [b]				
ADHD	-.73	.73	67.35	.48 (.11, 2.05)
Anxiety	-.42	1.12	83.92	.66 (.07, 5.93)
Conduct	.33	.80	82.49	1.39 (.29, 6.70)
Depression	.79	.81	70.45	2.20 (.44, 10.90)
Serious Mental Illness	-1.18	.63	99.95	.31 (.09, 1.06)
Other	.98	1.06	75.96	2.66 (.33, 21.12)
Comorbidity	.29	.36	74.14	1.33 (.65, 2.72)
Therapy	.26	.69	74.69	1.30 (.33, 5.03)
Skills Training	-1.85	1.08	57.35	.16 (.02, 1.31)
Parent Ratings at Baseline				
Problem Severity	-.04**	.01	126.50	.97 (.94, .99)
Functioning	.07**	.02	177.98	1.07 (1.03, 1.12)
Clinician Ratings at Baseline				
Risk of Self-Harm	-.65*	.27	52.13	.52 (.31, .88)
Family Problems	-.17	.29	113.80	.85 (.48, 1.49)
School Problems	-.35	.34	118.53	.70 (.36, 1.37)
Number of Treatment Sessions	-.07**	.03	83.37	.93 (.89, .98)
Weeks in Treatment	-.06*	.03	86.67	.94 (.89, .99)

[a] The reference category is: Moderate Improvement trajectory (No Change trajectory class vs. Moderate Improvement trajectory class)
[b] Primary diagnosis assigned at intake
**p*<.05; ***p*<.01; ****p*<.001

Table 9
Multinomial Logistic Regression of Significant Predictors of Trajectory Group Membership for Ohio Functioning Entered Simultaneously [a]

Predictor Variable	β	S.E (β)	Odds Ratio (95% CI)
Clinical Predictors			
Parent Ratings at Baseline			
Problem Severity	-.01	.02	.99 (.94, 1.04)
Functioning	.06*	.03	1.06 (1.00, 1.12)
Clinician Ratings at Baseline			
Risk of Self-Harm	-.72*	.35	.49 (.25, .97)
Number of Treatment Sessions	-.07	.04	.93 (.87, 1.01)
Weeks in Treatment	-.03	.03	.97 (.92, 1.02)

[a] The reference category is: Moderate Improvement trajectory (No Change trajectory class vs. Moderate Improvement trajectory class)
**p*<.05; ***p*<.01; ****p*<.001