6-28-2021

Detecting risk for treatment nonresponse among families of young children with behavior problems: Candidate tailoring variables and early decision points for adaptive interventions

Natalie Hong
Florida International University, nhong004@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DETECTING RISK FOR TREATMENT NONRESPONSE AMONG FAMILIES OF YOUNG CHILDREN WITH BEHAVIOR PROBLEMS: CANDIDATE TAILORING VARIABLES AND EARLY DECISION POINTS FOR ADAPTIVE INTERVENTIONS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Natalie Hong

2021

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Natalie Hong, and entitled Detecting Risk for Treatment Nonresponse Among Families of Young Children with Behavior Problems: Candidate Tailoring Variables and Early Decision Points for Adaptive Interventions, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Jeremy W. Pettit

Daniel M. Bagner

Adela C. Timmons

Elizabeth D. Cramer

Jonathan S. Comer, Major Professor

Date of Defense: June 28, 2021

The dissertation of Natalie Hong is approved.

Dean Michael R. Heithaus
College of Arts, Sciences and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2021

ABSTRACT OF THE DISSERTATION

DETECTING RISK FOR TREATMENT NONRESPONSE AMONG FAMILIES OF YOUNG CHILDREN WITH BEHAVIOR PROBLEMS: CANDIDATE TAILORING VARIABLES AND EARLY DECISION POINTS FOR ADAPTIVE INTERVENTIONS

by

Natalie Hong

Florida International University, 2021

Miami, Florida

Professor Jonathan S. Comer, Major Professor

Heterogeneity in mental health treatment outcomes and high rates of treatment nonresponse highlight the need for adaptive interventions that align with precision mental health care approaches to tailor treatments according to individual differences in progress over time. Modern clinical trial methodologies and analytic strategies can inform dynamic mental health treatment decisions, but the potential to improve patient outcomes is only as strong as the extent to which selected tailoring variables (i.e., interim response factors that dictate whether treatment should shift course) accurately detect risk for treatment nonresponse. Identifying empirically informed tailoring variables and the most appropriate timepoint(s) to assess them (i.e., critical decision points) is essential in order to design adaptive interventions.

This dissertation is comprised of three manuscripts focused on the use of early interim progress data to detect risk for mental health treatment nonresponse. First, I detail a strategy that leverages secondary data analysis to examine candidate tailoring variables at candidate critical decision points, and their relationships with treatment nonresponse. Then, I directly apply this strategy to a pooled sample of families who presented for treatment of early childhood behavior problems (*N*=153). This study shows that using dichotomous classifications of early interim treatment progress yield limited

utility in differentially predicting post-treatment response when predictors are examined in isolation from one another. Thus, I subsequently adopt a continuous approach to measuring early treatment progress and examine whether interactions between early indicators of treatment response predict symptom trajectories in a sample of families who participated in a behavioral parenting intervention (BPI) for early childhood developmental delay and behavior problems (*N*=70).

Findings from the third paper suggest that examining the interaction between caregiver skills and child behavior problems within the first six sessions of a BPI can predict symptom response trajectories across the entire course of treatment. This collection of work encourages the use of routine outcome monitoring to assess multiple domains of early interim treatment progress. To improve the efficiency and effectiveness of mental health care, future work should continue to use analytic approaches that capture the dynamic interplay among multiple early interim response factors that can optimally inform clinical decision-making practices throughout treatment.

TABLE OF CONTENTS

LIST OF TABLES

TABLES PAGE

LIST OF FIGURES

FIGURES PAGE

CHAPTER I

Introduction

Heterogeneity in mental health treatment outcomes and high rates of treatment nonresponse highlight the need for adaptive interventions that align with precision mental health care approaches to tailor treatments according to individual differences in progress over time (Bickman et al., 2016; Hong et al., 2019; Lei et al., 2012; Weisz et al., 2017). Modern clinical trial methodologies, such as the sequential, multiple assignment, randomized trial (SMART; Almirall & Chronis-Tuscano, 2016), can inform dynamic mental health treatment decisions and the design of adaptive interventions, but their potential to improve patient outcomes is only as strong as the extent to which selected tailoring variables (i.e., interim response factors that dictate whether treatment should shift course) accurately detect risk for treatment nonresponse (Hong et al., 2019). Identifying empirically informed tailoring variables that can differentially predict treatment response and the critical decision point(s) at which tailoring variables should be assessed is essential in order to design adaptive interventions. Nonetheless, this work has received relatively little attention in the literature.

My dissertation is comprised of three manuscripts focused on the use of early interim progress data to detect risk for mental health treatment nonresponse. In the first paper (Chapter II), I present a conceptual paper that details a methodological and analytic strategy that leverages randomized controlled trial (RCT) data to simultaneously examine and compare candidate tailoring variables at candidate decision points, and their relationships with treatment nonresponse. I suggest that using an experimental therapeutics framework to evaluate putative mechanisms of treatment response should inform the identification of interim treatment progress factors, given changes on these variables are expected to precede desired changes in the ultimate outcome of interest (i.e., symptom remission) and are therefore more likely to occur early in treatment.

Identifying differentially predictive early treatment response indicators can lead to subsequent examinations of how adapting treatment in response to these factors leads to enhanced outcomes. In my second paper (Chapter III), I directly apply this strategy to a pooled sample of typically and atypically developing young children whose families presented for treatment of early childhood behavior problems and participated in one of four RCTs (*N*=153 families). Specifically, I use ROC analyses and corresponding classification indices to (a) examine whether early interim progress data (i.e., candidate tailoring variables) collected within the first six sessions of a behavioral parenting intervention (BPI) could accurately detect families at risk for post-treatment nonresponse, and (b) compare the clinical utility of candidate tailoring variables and their optimal cut scores at candidate critical decision points. In this study, early interim treatment progress data include measures of out-of-session engagement, caregiver skills, and child symptoms collected at sessions 3 and 6. Pivoting from the findings from this second paper, I adopt a continuous approach to measuring both early interim treatment progress and later treatment response in my third paper (Chapter IV). In this study, I utilize multilevel growth modeling to examine the roles of caregiver skills and child behavior problems measured at sessions 3 and 6 (as well as their interactions) in predicting child externalizing symptom response trajectories across treatment in a sample of (*N*=70) families participating in a BPI for their young child with developmental delay and behavior problems.

The present collection of work highlights how routine outcome monitoring can be used to assess multiple domains of early interim progress throughout mental health treatment. The identification of relations among early interim progress data and later treatment response can inform the selection of optimal tailoring variables and critical decision points, and in turn, aid in the design of adaptive interventions and precision mental health care approaches. Directions for future research that may yield more

accurate and nuanced predictions of mental health treatment outcomes are discussed, noting that tradeoffs between predictive accuracy and ease of interpretability relate to the use of actuarial methods to inform clinical decision-making in routine practice.

CHAPTER II

Coal-mine canaries in clinical psychology: Getting better at identifying early signals of treatment nonresponse

This manuscript is published in Clinical Psychological Science.

Hong, N., Cornacchio, D., Pettit, J.W., & Comer, J.S. (2019). Coal-mine canaries in clinical psychology: Getting better at identifying early signals of treatment nonresponse. *Clinical Psychological Science, 7*(6), 1207-1221. doi: 10.1177/2167702619858111

Abstract

Sequential, multiple-assignment, randomized trials (SMARTs) have emerged as a preferred design strategy with which to inform dynamic mental-health treatment decisions and adaptive interventions, yet their potential to improve patient outcomes is only as strong as the extent to which selected tailoring variables (i.e., interim response factors that dictate whether treatment shifts course) do indeed predict ultimate response. To date, tailoring variable selection has rarely drawn on adequately powered findings or conceptual links to interim target mechanisms underlying treatment response. Building on early work in this area, we detail a strategy that leverages randomized controlled trial data to simultaneously compare candidate tailoring variables at candidate decision points and their relationships with treatment response. Findings from such efforts can improve the conceptual clarity and efficiency of SMARTs, laying a foundation for modern clinical trials to ask, “Are treatment-related change mechanisms being affected and, if not, what is the most appropriate next treatment strategy?”

Keywords: adaptive interventions, clinical decision-making, SMART designs, experimental therapeutics, clinical trials.

Despite the proliferation of evidence-based treatments (EBTs) for reducing psychopathology (Barlow, Bullis, Comer, & Armetaj, 2013), heterogeneity across treatment outcomes remains a critical issue (e.g., Vittengl et al., 2016). On average, roughly 30% to 50% of individuals treated with EBTs show inadequate response and remain symptomatic at the end of treatment (e.g., Cuijpers et al., 2014; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012), and a small percentage of clients actually fare worse after treatment (Barlow, 2010). Although early intervention is critical, a recent meta-analysis found that across roughly 450 published trials of child treatments, there has been an almost 40% probability that, after treatment, randomly selected children in active-treatment conditions were no better off than randomly selected children in control conditions (Weisz et al., 2017).

These striking statistics call to mind Gordon Paul's (1967) seminal challenge to the field to identify what treatments, delivered by whom, are most effective for which patients, and under what set of circumstances. Given that heterogeneity in outcomes is the norm, and given that our best treatments seem to hit an overall response rate asymptote of around 50% to 65% (Cuijpers et al., 2014; Hofmann et al., 2012), research has begun to move beyond main effects to identify moderators of treatment response. Moderators are factors that delineate the conditions under which various treatments are differentially related to diverse outcomes (Comer & Kendall, 2013). By influencing either the direction or the strength of a relationship between treatment participation and patient response, moderators identify on whom and under what circumstances specific treatments have their varying effects. Unlike broad predictors that uniformly anticipate response regardless of which specific treatment was received, moderators have differential predictive relationships with outcomes across alternative treatments (e.g., individuals high on Attribute A do not respond to Treatment X but do respond to

Treatment Y). Therefore, moderators hold promise to inform differential treatment assignment and personalized care.

It is noteworthy that the study of treatment moderation has yet to yield large-scale improvements in treatment response. First, much of the research evaluating treatment moderators has largely resulted in null findings, mixed or conflicting findings, results that have limited face validity, or findings that speak only to differential outcomes across an active versus an inactive (e.g., waitlist) condition (e.g., Compton et al., 2014; Cooper et al., 2016; Donker et al., 2013; Frank et al., 2011; MacPherson, Algorta, Mendenhall, Fields, & Fristad, 2014; Owens et al., 2003). Second, prescriptive strategies that differentially assign patients to various treatments on the basis of previously identified moderators have rarely been evaluated relative to blind randomization approaches to treatment assignment (see Kendall, Comer, & Chow, 2013), hindering our understanding of whether applying prior moderation findings to alter future treatment selection indeed has a beneficial impact.

Limited progress in identifying relevant factors that differentially predict heterogeneous clinical responses across treatments may be due, in part, to the fact that differential predictors of response have almost exclusively been measured as baseline variables in the context of traditional fixed, randomized controlled trials (RCTs). As argued elsewhere (Barlow & Comer, 2013; Lei, Nahum-Shani, Lynch, Oslin, & Murphy, 2012), the traditional fixed RCT—despite optimizing internal validity and methodological rigor—is relatively limited in the extent to which associated results can speak to dynamic clinical decision-making that unfolds throughout treatment in routine care settings. In the face of inadequate midtreatment patient response, the clinician in a traditional fixed RCT "stays the course" and continues to deliver the same treatment at a predetermined dose/intensity or manner. In contrast, in applied settings where patient care is prioritized over knowledge generation, clinicians typically adjust treatment in response to shifting

clinical needs and interim patient nonresponse (Chorpita, Daleiden, & Weisz, 2005; Lei et al., 2012). Only in the past decade or so—in light of relatively recent innovations in modern clinical trial methodology and treatment redesign (e.g., Almirall, Nahum-Shani, Sherwood, & Murphy, 2014; Chorpita et al., 2017; Collins, Nahum-Shani, & Almirall, 2014; Lei et al., 2012; Nahum-Shani et al., 2012)—have controlled research trials in clinical psychological science begun to rigorously evaluate treatment-outcome heterogeneity in the context of shifting patient needs and responsive treatment courses that more closely reflect applied practice.

In this article, we first consider the key components of modern clinical-trial methodologies that leverage controlled designs and randomization to evaluate shifting treatment courses (e.g., the increasingly popular sequential, multiple-assignment, randomized trial, or SMART; Almirall & Chronis-Tuscano, 2016). We then move to the main thesis of the article: Advances in the development of controlled methods for evaluating adaptive interventions have largely outpaced the development of corresponding methods for systematically identifying useful tailoring variables that dictate whether treatment in an adaptive intervention needs to shift course. We argue that the success of adaptive treatment regimens and SMARTs for yielding improved patient outcomes will necessarily be limited in the absence of more systematic efforts to establish empirically informed tailoring variables that evaluate interim response and identify early signs that a patient's current course of treatment is ultimately likely to be inadequate.

Just as canaries were historically used in coal mines to provide early warnings to miners that forward progression was misguided (because of carbon monoxide and other toxic gases) and thus alternative mining plans were in order, intratreatment variables must be identified that reliably provide early treatment warnings that staying the course in treatment for a particular patient is likely misguided and that alternative treatment

strategies may be indicated. We elaborate on early work in this area (e.g., Pettit, Silverman, Rey, Marin, & Jaccard, 2016; Steidtmann et al., 2013) to detail a methodological and data-analytic strategy that illustrates one possible approach that can be applied to empirically evaluate various forms of interim patient responses and their relationships with ultimate treatment response in the context of clinical trials. In particular, when considering early treatment signals that may portend ultimate treatment nonresponse, we argue that it is critical to go beyond a simplistic focus on early partial diagnostic/symptom remission and to incorporate an experimental therapeutics framework focused on candidate treatment mechanisms. For adaptive interventions and associated SMARTs to fully realize their potential, we conclude with guiding thoughts for establishing a clear research agenda that prioritizes the establishment of empirically informed tailoring variables that evaluate interim treatment responses and serve as the foundation on which successful dynamic treatment regimens can be built.

Adaptive Interventions and Modern Clinical-Trial Methodologies for Evaluating Them

Adaptive interventions are designed to strategically modify treatment courses that respond to patients' individualized and evolving needs throughout care (Murphy, Collins, & Rush, 2007). Commonly, an adaptive intervention begins with an established EBT and/or low-intensity treatment approach and incorporates interim outcome monitoring to assess unfolding patient response at regular intervals. Such monitoring, in turn, directly informs dynamic treatment decisions, such as if, when, and how treatment should shift across time (see Lei et al., 2012). For example, an adaptive intervention for the treatment of anxiety might begin with cognitive–behavioral therapy (CBT) and include weekly assessment of anxiety symptoms. The *critical decision point* in an adaptive intervention is a predetermined point in treatment (e.g., Week 8, or Session 8, or perhaps after the initiation of exposures) at which the clinician determines whether

sufficient patient progress has been made or whether a modified treatment strategy should be adopted. The *tailoring variable* is the selected interim response factor assessed at the critical decision point that dictates whether treatment needs to shift course (e.g., standardized anxiety symptom score).

Finally, the *decision rule* provides clear guidance at the critical decision point for determining (a) the range of scores on the tailoring variable that indicates acceptable interim response, (b) the range of scores on the tailoring variable that indicates insufficient interim response, and (c) the specific modification (if any) to be made to the treatment course in response to such interim response. In the anxiety treatment example, a decision rule might delineate that after 8 weeks of CBT (critical decision point), treatment should shift from CBT to a CBT + medication combination strategy when a patient's Week 8 t score on the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990) still falls above 59 (tailoring variable); in contrast, the decision rule might also delineate that monotherapy CBT should be continued if a patient's Week 8 t score falls at or below 59 but anxiety continues to interfere (for a similar example, see Pettit et al., 2016). By providing clear guidelines for what to assess in patients and when and how to proceed on the basis of the results of such interim assessment, adaptive interventions allow clinicians to make informed, responsive, and individualized treatment decisions (Murphy et al., 2007).

Intervention science has only recently advanced sophisticated experimental designs that incorporate randomization and controlled comparisons to lay an empirical foundation on which to construct and refine evidence-based adaptive interventions (Lavori & Dawson, 2004). SMARTs have risen to the forefront of these modern experimental methodologies (Almirall & Chronis-Tuscano, 2016; Lei et al., 2012). Whereas the traditional fixed RCT assigns participants to one of multiple treatments and then evaluates responses at the conclusion of full treatment courses, the signature

design element of the SMART is the rerandomization of participants to one of multiple Phase II treatment strategies. As in the adaptive intervention strategy described above, a tailoring variable is assessed at a critical decision point, but in a SMART, the decision rule based on this tailoring variable delineates how participants are randomly assigned to one of multiple, alternative, Phase II treatments, which allows for controlled Phase II treatment comparisons that eliminate selection biases and other interpretation confounders.

Figure 2.1 presents an illustration of a hypothetical SMART evaluating sequences of care for children with conduct problems. The design begins by randomizing participants to one of two Phase I treatment arms: weekly behavioral parent training (BPT) or weekly individual child CBT (ICBT), both of which have received support for treating conduct problems (Comer, Chow, Chan, Cooper-Vince, & Wilson, 2013; Kaminski & Claussen, 2017) and may be differentially appealing across various families and clinicians. In this hypothetical SMART design, Week 4 functions as the critical decision point at which participants are evaluated for interim “response” on the Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999), which functions as the tailoring variable. The decision rule for this hypothetical SMART delineates that (a) participants with Week 4 ECBI scores that fall at least 10 points below their baseline ECBI score continue to receive the Phase I treatment to which they were initially assigned for Phase II of their treatment, and (b) participants with Week 4 ECBI scores that are not at least 10 points below their baseline ECBI score are rerandomized to one of three adjusted treatment strategies in Phase II. Specifically, Week 4 nonresponders (i.e., Week 4 ECBI < 10 points below baseline ECBI) are rerandomized to either a switching strategy (i.e., Week 4 nonresponders initially receiving BPT switch to ICBT; Week 4 nonresponders initially receiving ICBT switch to BPT), an augmentation strategy (i.e., Week 4 nonresponders initially receiving BPT continue to receive BPT and also

receive adjunctive ICBT; Week 4 nonresponders initially receiving ICBT continue to receive ICBT and also receive adjunctive BPT), or an intensification strategy (i.e., Week 4 nonresponders initially receiving weekly BPT now receive twice-weekly BPT; Week 4 nonresponders initially receiving weekly ICBT now receive twice-weekly ICBT).

This hypothetical SMART affords the efficient and strategic evaluation of the relative acute effectiveness of brief BPT versus brief ICBT at Week 4, as well as eight different sequences of care that differentially incorporate BPT and/or ICBT—including switching, augmentation, and intensification strategies (see Table 2.1). In addition, this efficient design simultaneously embeds a far greater number of controlled comparisons than the traditional fixed RCT that are directly relevant to clinical decision-making (see Table 2.2). For example: “For cases of initial nonresponse, what is the relative effectiveness of a switching strategy versus an augmentation strategy versus an intensification strategy?” or “At the end of Phase II, do initial (i.e., Phase I) responders continue to show better outcomes than initial nonresponders whose treatment was adjusted in hopes of improving response?”

Although this hypothetical SMART focuses on BPT and ICBT for child conduct problems, and although it is designed to compare switching, augmentation, and intensification strategies following initial nonresponse, SMARTs can of course be designed for any clinical population and can be constructed to compare a wide range of Phase I treatments and Phase II modifications. Other Phase II modifications might include adjustments to the treatment format or the level of therapist involvement (e.g., shifting from a computer-based format to a face-to-face format; shifting from a self-administered format to a therapist-guided or therapist-led format). In addition, although the hypothetical SMART outlined above incorporates a single critical decision point and a single associated rerandomization, more complex SMARTs can incorporate multiple critical decision points and rerandomizations. Moreover, as with the traditional RCT, in

SMARTs there is a wide range of variables that can serve as the primary outcome (e.g., diagnoses, symptoms, impairment, consumer satisfaction, cost effectiveness).

Insufficient Empirical Attention to Early Treatment "Canaries" That Signal the Need to Shift Treatment Course

Although a number of recent SMARTs have yielded important findings, the potential of SMARTs to successfully inform improved patient outcomes is only as strong as the extent to which selected tailoring variables do indeed predict ultimate treatment response. Early or midtreatment tailoring variables that are strongly tied to posttreatment outcomes can correctly flag those in need of adjusted treatment strategies. In contrast, early or midtreatment tailoring variables that are, in fact, weakly associated with posttreatment outcomes may introduce noise, prompt unnecessary treatment shifts for adequately progressing patients, and fail to appropriately flag and adapt treatment for patients on nonresponding treatment trajectories.

In the SMART illustration evaluating BPT/ICBT sequences for child conduct problems (see Figure 2.1), Week 4 ECBI change (capturing improvements in child conduct problems) was selected as the tailoring variable on which constructed decision rules determined whether to maintain or adjust treatment. In this scenario, Week 4 ECBI change is a strong tailoring variable if change in conduct problems after only 4 weeks of treatment is a reliable predictor of ultimate response. This may be the case, but it may also be the case that decreases in child conduct problems after only 4 weeks is weakly associated with ultimate response. It is possible that after 1 month of treatment, homework adherence, treatment satisfaction, therapeutic alliance, or other early treatment variables are stronger predictors of ultimate treatment success than immediate symptom decreases. Alternatively, early/midtreatment changes in conduct problems may be a useful tailoring variable, but it may be that decreases after 6 weeks of treatment yield a more reliable predictor of ultimate response/nonresponse than decreases after

just 4 weeks. Indeed, empirical identification of strong early/midtreatment tailoring variables that predict differential treatment responses, and the time points at which they are most predictive, is critical for the success of SMARTs and for developing effective adaptive treatment strategies.

Despite the uptick in enthusiasm for adaptive treatment strategies, and despite the promise of the SMART and related designs for incorporating controlled comparisons to identify sequenced treatment strategies that respond to individualized and evolving patient needs, the empirical identification of optimal tailoring variables has been largely ignored in the literature (see Pettit et al., 2016). In the construction of SMARTs to date, tailoring variables have largely been selected on the basis of simple face validity, clinical judgment, or underpowered analyses of pilot data, and critical decision points for assessing these tailoring variables have largely been selected on the basis of standardized lengths of manualized treatments, clinical judgment, convenience, or seemingly arbitrary factors. Most commonly, SMARTs have selected partial or full remission on the primary outcome domain as the tailoring variable (e.g., Almirall et al., 2016; Kasari et al., 2014; Naar-King et al., 2016; see also Pettit et al., 2016) or a hybrid of impairment and/or symptom remission (e.g., Pelham et al., 2016) at critical decision points. This may be due, in part, to the fact that most completed SMARTs to date have been designed to evaluate effective sequences of full courses of EBTs. Despite the potential of SMARTs to improve the efficiency of care, SMARTs to date have less commonly been designed to respond to early treatment responses and to adjust treatment within the course of an intervention. Given practical realities of treatment engagement and dropout in routine care settings, there may be limited utility for adaptive trials that require patients to persist in a full course of treatment before determining whether alternative treatment strategies are necessary.

In the absence of data-based identification of tailoring variables, pretrial comparisons of alternative tailoring variables in adequately powered analyses, or pretrial comparisons of alternative critical decision points in treatment, it will remain unclear whether SMARTs designed to rapidly respond to early nonresponse are selecting the optimal “canaries” for signaling whether treatment courses require adjustment for individual patients. A very small body of research has moved beyond baseline variables to examine patterns of intratreatment patient response to identify empirically informed tailoring variables (see Pettit et al., 2016; Schueller, Kwasny, Dear, Titov, & Mohr, 2015; Shih, Patterson, & Kasari, 2016). For example, Pettit and colleagues (2016) examined patterns of midtreatment patient response to identify empirically informed tailoring variables in the context of youth anxiety treatment. In this study, latent profile analysis identified classes of treatment response at midtreatment (Session 8) and latent transition analyses examined continuity in anxiety response classes from midtreatment to posttreatment. Subsequently, classification properties were evaluated to identify cut points at midtreatment that accurately distinguish response classes. In another example, Steidtmann and colleagues (2013) sought to empirically inform the identification of critical decision points assessing symptom change within treatment for depressed adults (see also Schueller et al., 2015; Shih et al., 2016).

To date, no studies or theoretical articles have considered systematic methods for simultaneously evaluating alternative tailoring variables at different critical decision points. Tailoring variables focused on partial or full remission on the ultimate primary outcome are out of step with increasingly prominent experimental therapeutics frameworks in intervention science, which emphasize explicit treatment engagement of malleable targets and mechanisms associated with psychopathology onset and maintenance—changes that may, in turn, mediate downstream treatment-related changes in psychopathology outcomes. In the context of rapidly responding SMARTs—

in which early/midtreatment evaluations are expected to inform whether treatment is sufficiently "working" thus far—an experimental therapeutics framework would suggest the researcher selecting a tailoring variable should ask not only "Have the targeted symptoms satisfactorily declined yet?" but also "Are the appropriate mechanisms of treatment-related change being impacted?"

In theory, mediator variables may serve as "canaries" that predict response/nonresponse to treatment even earlier than change in the outcome variable or variables. However, mechanistic models of treatment response remain relatively untested, so collection of both outcome and mediator data at various points in treatment can provide an efficient and sensitive approach to testing/refining treatment-mediation models and to clarifying the directionality of treatment effects. If symptoms meaningfully reduce in the absence of changes in putative mechanisms, simultaneous examination of multiple tailoring variables can inform subsequent refinement of theory about mechanisms of treatment response and help identify other possible treatment mechanisms. For example, negative and ineffective parenting practices are believed to underlie significant variance in child conduct problems, and improved parenting (e.g., greater warmth, monitoring, consistency, predictability, follow-through) is theorized to be the putative mechanism through which BPT reduces child conduct problems (Forehand, Lafko, Parent, & Burt, 2014). Thus, when designing an adaptive treatment incorporating BPT for child conduct problems, an experimental therapeutics framework would suggest that interim parent skill acquisition (rather than interim child symptom response) may be a more appropriate tailoring variable around which to determine subsequent treatment directions. However, if child conduct problems satisfactorily declined in the absence of parent skill acquisition, the extent to which parent skill acquisition indeed underlies child BPT response would need to be reconsidered.

Researchers have increasingly examined intratreatment changes and their links with treatment endpoint response. Results of this work indicate that using measures of skill acquisition, therapeutic alliance, patient engagement, symptom difference scores between sessions, as well as trajectories/shape of symptom change across sessions may serve as useful tailoring variables within adaptive interventions for different clinical populations (Chu et al., 2015; Chu & Kendall, 2004; Correll et al., 2013; Kendall & Treadwell, 2007; Lewis, Simons, & Kim, 2012; Marker, Comer, Abramova, & Kendall, 2013). To date, however, findings regarding treatment mediation, mechanisms of change, and trajectories of treatment-related change have not been used to directly inform adaptive treatment strategies or the construction of SMARTs. In fact, the literature on SMARTs provides little explicit direction for the empirical identification of effective tailoring variables. In recent years, pilot SMARTs have begun to incorporate exploration of several different possible tailoring variables within a single pilot trial (e.g., Chronis-Tuscano, Wang, Strickland, Almirall, & Stein, 2016). However, pilot SMARTs primarily serve to evaluate feasibility and acceptability in advance of full-scale SMARTs (see Almirall, Compton, Gunlicks-Stoessel, Duan, & Murphy, 2012); they are not designed or powered to statistically compare differential associations of alternative tailoring variables and critical decision points with ultimate response. For the development of rapidly responding SMARTs, other design and analytic methods are needed to select optimal "canaries" for signaling whether treatment courses require adjustment for individual patients.

Leveraging Data From Traditional, Fixed RCTs to Empirically Identify Evidence-Based Tailoring Variables

Although pilot SMARTs focused on feasibility and acceptability (Almirall et al., 2012) are not adequately powered, secondary analysis of traditional, fixed RCTs with adequate power may afford empirical identification of effective tailoring variables and

critical decision points (e.g., Pettit et al., 2016)—precisely because in the face of inadequate response, the clinician in a traditional, fixed RCT "stays the course" and continues to deliver the same treatment at the same dose/intensity (e.g., 1 hr per week) and in the same format (e.g., face to face). Intratreatment data collected in traditional, fixed RCTs can test the extent to which candidate tailoring variables at candidate critical decision points differentially predict ultimate response. For each candidate tailoring variable, post hoc analyses of traditional, fixed RCT data can allow the investigator to empirically answer the question, "What would happen if early treatment information about this candidate tailoring variable was ignored?" (for an empirical example, see Pettit et al., 2016). Despite the centrality of tailoring variables in adaptive treatment regimens and related SMARTs, such strategies have rarely been used to empirically identify effective tailoring variables and/or critical decision points. Even when this work has been conducted (see Pettit et al., 2016; Schueller et al., 2015; Shih et al., 2016; Steidtmann et al., 2013), the focus has been on evaluating the extent to which midtreatment change in the outcome variable predicts ultimate response, without consideration that midtreatment change in targeted mechanisms may also yield important predictive information.

Let us return to the example of BPT for child conduct problems. In the previous section, we considered several candidate early treatment "canaries" that may signal that a course of BPT is not taking and that treatment adjustments are required, including (a) inadequate child symptom reduction, (b) poor parental skill acquisition, and/or (c) poor treatment engagement/homework adherence. It may be that only some (or one) of these putative early treatment predictors indeed significantly anticipate(s) ultimate nonresponse, and that early information about the others is simply noise best ignored. It may also be that all of these early treatment variables are simply noise best ignored (insofar as making decisions about shifting the course of treatment are concerned).

Alternatively, relationships between these early treatment factors and ultimate response may be more nuanced. For example, it may be that (a) in the early BPT weeks, treatment engagement is the strongest predictor of ultimate response, whereas (b) just before the midpoint of a BPT course, parental skill acquisition is the strongest predictor of ultimate response and, (c) after the midpoint of a course of BPT, child symptom improvement is the best predictor of ultimate response. Such relationships may be further moderated by key variables (e.g., baseline symptom severity, baseline parental skills).

To empirically answer these questions using data from a fixed RCT, high-quality data must have been collected on each candidate tailoring variable at each candidate decision point. For this discussion, we will consider a hypothetical fixed clinical trial of BPT in which the investigators collected weekly data on three candidate tailoring variables: (a) child symptoms measured via a weekly administered ECBI (Eyberg & Pincus, 1999), (b) parental skill acquisition measured weekly via the Dyadic Parent-Child Interaction Coding System (DPICS; Eyberg, Nelson, Ginn, Bhuiyan, & Boggs, 2013), and (c) homework adherence measured via a weekly ratio comparing the number of homework activities completed that week against the number of activities assigned. Furthermore, in this hypothetical fixed trial, let us assume that after treatment, the evaluator conducted a structured interview and generated a rating on the Clinical Global Impressions-Improvement scale (CGI-I; Guy & Bonato, 1970). Cases assigned a posttreatment CGI-I score of 1 (*very much improved*) or 2 (*much improved*) were considered “responders,” and children assigned a CGI-I score > 2 were considered “nonresponders.”

With these weekly data and CGI-I categorizations of ultimate treatment response, a series of receiver-operator characteristic (ROC) analyses can yield instructive and easily interpretable data with which to empirically compare candidate tailoring variables

and candidate critical decision points. ROC analyses (see Hong & Comer, 2019; Youngstrom, 2014) provide a comprehensive depiction of a measure's predictive accuracy and clinical utility by demonstrating the limits of a measure's ability to discriminate across the full range of possible cut scores on that measure. In prior work seeking to identify decision rules or decision points (e.g., Pettit et al., 2016; Scheuller et al., 2015; Shih et al., 2016; Steidtmann et al., 2013), ROC analyses and related classification properties have served to inform clinical decisions, but these analyses have focused on patterns of midtreatment response on outcome variables in the absence of evaluation of midtreatment response on putative mechanisms that may underlie response.

Let us consider the clinical utility of predicting ultimate treatment response in the hypothetical fixed BPT trial using early symptom improvement data versus early parental skill acquisition data versus early homework adherence data. The Supplemental Material available online defines and details the calculation of five key classification indices—sensitivity, specificity, positive predictive power, negative predictive power, and overall correct classification. The top section of Table 2.3 presents hypothetical values for these five classification indices as they relate to alternative early symptom improvement cutoffs (i.e., ECBI change) at three candidate critical decision points (i.e., Sessions 3, 4, and 5). The middle section of Table 2.3 presents hypothetical values for these five classification indices as they relate to alternative early parental skill acquisition cutoffs (i.e., percentage of introduced skills mastered in a 5-min DPICS coding period) at Sessions 3, 4, and 5. The bottom section of Table 2.3 presents hypothetical values for these five classification indices as they relate to alternative early homework adherence cutoffs (i.e., % of assigned homework tasks completed) at Sessions 3, 4, and 5. In Table 2.3, we present hypothetical values for classification indices for Sessions 3, 4, and 5, although in practice, any candidate critical decision points could, of course, be selected.

Plotting all of the cutoff sensitivities by 1 – specificities within a candidate session (i.e., critical decision point) yields a curve depicting the extent to which data on a given candidate tailoring variable (across all potential cutoffs) in a given early session correctly classify children as ultimate treatment responders or nonresponders. Using hypothetical values, Figure 2.2 presents the ROC plots of sensitivities by 1 – specificities for each potential cutoff on the measures of all three candidate tailoring variables (i.e., ECBI change from baseline, percentage of parental skill acquisition, and percentage of homework adherence) at Sessions 3, 4, and 5. These figures are organized by candidate session (Figure 2.2a) and by candidate tailoring variable (Figure 2.2b). ROC plots can be examined quantitatively by evaluating the area under the curve (AUC), which ranges from .5 (i.e., early treatment response factor at that given session does not distinguish between ultimate response groups any better than chance) to 1.0 (i.e., early treatment response factor at that given session exhibits 100% accuracy in distinguishing between ultimate response groups). This AUC consequently provides a quantitative index of the overall utility of a candidate tailoring variable at a given session in distinguishing ultimate treatment responders from nonresponders, and a significance test can evaluate whether this distinguishing property differs from chance agreement.

Presenting classification properties for alternative candidate tailoring variables and alternative candidate critical decision points within the same figures and tables affords the opportunity to directly compare the utility of alternative early treatment "canaries" in predicting ultimate treatment response. This offers a novel elaboration on the small literature in this area by incorporating an experimental therapeutics framework that considers the predictive contributions of interim putative mechanisms of treatment response to enable simultaneous comparison of candidate tailoring variables. Examination of the ROC plots in Figure 2.2, along with inspection of the classification-properties data in Table 2.3, elucidates the overall utility of each early treatment

measure at each session in distinguishing ultimate responders from ultimate nonresponders. In this hypothetical illustration, one can see that, by comparing ROC curves for the three candidate tailoring variables at Session 3 (Figure 2.2a and Table 2.4), the overall predictive utility of homework adherence (AUC = .876) is superior to both parental skill acquisition (AUC = .580) and symptom improvement (AUC = .622). However, at Session 4, the overall predictive utility of parental skill acquisition (AUC = .905) supersedes homework adherence (AUC = .708) and symptom improvement (AUC = .715). Finally, at Session 5, the overall predictive utility of symptom improvement (AUC = .878) outperforms homework adherence (AUC = .614) and parental skill acquisition (AUC = .701). Drawing conclusions from these data, one can conclude that at Session 3, homework adherence is the optimal tailoring variable for predicting ultimate treatment responders, whereas at Session 4, parental skill acquisition is the optimal tailoring variable for predicting ultimate treatment responders, and at Session 5, symptom improvement is the optimal tailoring variable for predicting ultimate treatment responders.

This information alone does not provide clinicians with clear decision rules (e.g., guidelines) related to specific cutoffs within each measure at critical decision points. Additional examination of classification properties (e.g., sensitivity, specificity, positive predictive power, negative predictive power, and overall correct classification) at each potential cutoff for each measure is informative (see Table 2.3). In this hypothetical example, let us suppose that we aim to prioritize overall correct classification (OCC) properties across the three candidate tailoring variables. Identifying the cutoff with the highest OCC results in a decision rule that prioritizes the highest number of individuals being accurately classified as ultimate responders or nonresponders at the candidate session. From the perspective of OCC, one can see that at Session 3, early homework adherence ≥ 50% is the strongest predictor of ultimate treatment response; at Session 4,

parental skill acquisition ≥ 50% is the strongest predictor of ultimate treatment response; and at Session 5, child symptom improvement ≥ 12 points on the ECBI is the strongest predictor of ultimate treatment response. Among these three contenders, Session-4 parental skill acquisition ≥ 50% yields the strongest overall correct classification index across all candidate measures and sessions (i.e., OCC = .850). Specifically, in this hypothetical data set, defining early treatment response as parents having acquired at least 50% of the taught BPT skills by Session 4, one correctly flags 94% of ultimate treatment responders (i.e., sensitivity = .940) and 76% of ultimate nonresponders (i.e., specificity = .760). Moreover, defining early response in this manner and at this time point, 80% of families meeting this early response criteria were indeed responders (i.e., positive predictive power = .787) and 93% of families not meeting this early response criteria were indeed nonresponders (i.e., negative predictive power = .927).

This examination of hypothetical data suggests that very early in BPT, one should focus on homework adherence as the earliest indicator that treatment is on track toward ultimate BPT success, but that as treatment progresses, attention should shift to parental skill acquisition and then to child symptom improvement as better predictors of ultimate BPT success. Moreover, this examination of hypothetical data suggests that among the early signals of ultimate BPT success, parental skill acquisition at Session 4 is the strongest. By leveraging fixed RCT data on various intratreatment changes by Sessions 3, 4, and 5, if OCC is to be prioritized, one could now design a SMART to evaluate adaptive interventions for youth with conduct problems in which (a) Phase I treatment is BPT, (b) the tailoring variable is parental skill acquisition, (c) the critical decision point is Session 4, and (d) the decision rule for determining Phase II treatment courses is based on whether participants are classified as early responders (i.e., parents have mastered at least 50% of taught BPT skills by Session 4) or early nonresponders (i.e., parents have mastered < 50% of taught BPT skills by Session 4).

Concluding Comments

Over the past 2 decades, scholarly and empirical enthusiasm for personalized mental health care in the context of evidence-based practice has blossomed, along with major shifts in federal funding priorities that increasingly emphasize sequenced treatment, precision medicine, and individually tailored interventions to optimize mental-health outcomes (National Institute of Mental Health, 2015). Against this backdrop, SMARTs have arisen as the design strategy of choice to inform the development of evidence-based adaptive interventions in an experimentally controlled manner (Lei et al., 2012). In this article, we outlined a critical gap in the literature and a key obstacle to the ultimate success of SMARTs: the relative absence of empirical attention focused on identifying evidence-based tailoring variables and critical decision points, which in turn form the foundation on which successful dynamic treatment regimens are designed. We elaborated on a small but growing literature in this area (e.g., Pettit et al., 2016; Steidtmann et al., 2013) to detail a strategy that leverages data from traditional, fixed RCTs to simultaneously compare alternative tailoring variables (including target mechanisms) at various critical decision points and evaluates their differential relationships with ultimate response.

Since the early precursors to the SMART (e.g., Rush et al., 2004; Sachs et al., 2003; Stroup et al., 2009), evaluations of adaptive treatments have sought to yield more pragmatic findings than traditional, fixed RCTs in order to better inform typical care characterized by shifting clinical needs and heterogeneous patient responses. Across roughly the same time that SMARTs have gained prominence as a design strategy to expand the ecological validity and clinical relevance of intervention science, an experimental therapeutics framework focused on target-mechanism engagement in treatment has gained prominence as a framework to expand the internal validity and conceptual clarity of intervention science. To date, these two trends in intervention

science have barely crossed paths: SMARTs have largely focused on interim and endpoint symptom responses, whereas clinical trials designed to evaluate target engagement and mechanisms of action have largely used fixed RCT designs.

The strategy outlined in this article offers a unique opportunity for adaptive treatment designs and the experimental therapeutics framework to intersect. By moving beyond a relatively exclusive focus on interim symptom responses in the design of adaptive treatment regimens, and instead using fixed RCT data to systematically compare possible mechanisms of action as candidate tailoring variables, a new generation of SMARTs will be poised to optimize both internal and external validity by evaluating the extent to which focusing on early engagement of target mechanisms of action can improve real-world clinical decision-making. An increased focus on empirically identifying theoretically grounded and evidence-based “canaries” among a range of candidates for signaling whether individual treatment courses require adjustment will improve the conceptual clarity of SMARTs and will lay the foundation for modern clinical trials to ask, “Are the appropriate mechanisms of treatment-related change being impacted, and if not, what then is the most appropriate next treatment strategy?”

The strategy we outlined for empirically identifying tailoring variables and critical decision points can be flexibly applied to address different patient populations, circumstances, settings, or treatments. For example, in our illustration, we prioritized OCC over other classification properties so as to select a tailoring variable that maximized both sensitivity and specificity. However, for other populations of varying clinical risk (e.g., suicidal patients, maltreating families), for different outcome priorities (e.g., cost effectiveness, resource conservation), or for alternative treatments varying in risk (e.g., medications with significant side effect profiles), the investigator might differentially prioritize sensitivity or specificity over the other. Furthermore, the present

strategy is, of course, just one of many that could help empirically identify tailoring variables, critical decision points, and improved adaptive treatment designs. Researchers will do well to additionally consider how alternative analytic strategies (e.g., growth curve modeling, latent profile analysis; Pettit et al., 2016) can also yield important information to aid dynamic clinical decision-making. Moreover, although the data presented throughout this article were hypothetical, we are aware of multiple forthcoming integrated-data analyses using this methodology with actual clinical-trial data.

In the years ahead, the selection of variables to be considered in tailoring treatments may become increasingly complex, as big data move us toward tailoring algorithms that combine actuarial (e.g., data-driven machine-learning approaches) and theoretically driven mechanisms of treatment response. Collaborations between clinical scientists and computer scientists and/or engineers that together provide the necessary integration of theoretical expertise and analytic skill will be needed to facilitate this process. Importantly, although advanced modeling techniques that continuously adapt to new and evolving information may improve on clinical prediction, it will be critical to retain a focus on generating tailoring variables and decision rules that can feasibly be applied in clinical practice. As big-data opportunities enhance the statistical power, rigor, and complexity of analyses evaluating candidate tailoring variables, caution should be taken to ensure that the products of such analysis result in decision rules that are accessible to clinicians in routine settings.

Another practical reality of this work includes consideration of how researchers can responsibly design a SMART in the absence of existing data reflecting a range of interim patient-response variables. It is encouraging that shifts in federal funding are increasingly prioritizing trials that follow an experimental-therapeutics approach (Insel & Gogtay, 2014). Accordingly, researchers seeking funding for SMARTs will increasingly be required to identify putative mechanisms of treatment response (beyond solely

considering interim symptom response), assess target engagement, and examine whether target engagement is associated with treatment response. For populations and treatments for which such existing clinical-trial data on fixed (nonadaptive) interventions are not available, one approach that is useful for identifying the utility of various tailoring variables is the *unrestricted SMART*, wherein all patients are rerandomized at subsequent stages of treatment irrespective of responder status. Whereas only nonresponders are rerandomized in most SMART designs (allowing the investigator to consider the utility of candidate tailoring variables among those nonresponders), the unrestricted SMART allows the investigator to evaluate the utility of various tailoring variables among all patients, regardless of where they fell on an a priori-decided definition of "response" (D. Almirall, personal communication, April 22, 2019).

In conclusion, despite the promise of SMARTs for improving the efficiency and personalization of care, relying on tailoring variables and decision points based exclusively on face validity, clinical judgment, convenience, and/or other arbitrary factors may actually result in less-efficient and misguided care. Relying on rationally derived (instead of empirically derived) tailoring variables and critical decision points can introduce noise, prompt unnecessary treatment shifts for adequately progressing patients, and fail to appropriately flag and adapt treatment for patients on nonresponding trajectories. To better address issues of treatment-response heterogeneity across patients, the field needs to focus efforts on the systematic, empirical identification of early treatment "canaries" that signal when treatments need to shift course and provide feasible guidelines for clinicians to follow in routine settings. The strategy proposed in this article offers a promising approach that can yield key data to better poise the field to inform dynamic clinical decision-making and meaningfully improve overall treatment outcomes.

Supplemental Material

Key Classification Indices: Definitions and Calculation Details

(in the context of the hypothetical example of BPT for child conduct problems)

Sensitivity at each possible cutoff of a candidate tailoring variable is defined as the percentage of children rated as ultimate treatment responders (i.e., those with posttreatment CGI-I scores of 1 or 2) who scored above that cutoff at the early treatment timepoint.

Specificity at each possible cutoff of a candidate tailoring variable is defined as the percentage of ultimate treatment nonresponders (i.e., those with post-treatment CGI-I scores of 3 or higher) who scored below that difference score cutoff at the early treatment timepoint.

Positive Predictive Power (PPP) at each possible cutoff is defined as the percentage of children who at a given session scored above that cutoff who actually went on to be classified as ultimate treatment responders on the CGI-I.

Negative Predictive Power (NPP) at each possible cutoff is defined as the percentage of children who at a given session scored below that cutoff who actually went on to be classified as ultimate treatment nonresponders on the CGI-I.

Overall Correct Classification (OCC) at each possible cutoff is defined as the percentage of all children who were either (a) classified as an early responder and indeed went on to be classified as an ultimate treatment responder, or (b) not classified

as an early responder and indeed went on to be classified as an ultimate treatment nonresponders.

Examining the hypothetical values for these classification indices presented in Table 2.3, one can see that when defining “early response” as a child having improved at least 18 ECBI points by Session 5 (Table 2.3, top section, cutoff = -18), 81% of ultimate treatment responders are correctly flagged as responding by Session 5 (i.e., sensitivity = .81), 82% of ultimate treatment nonresponders are correctly flagged as not responding at session 5 (i.e., specificity = .82), 82% of those classified at Session 5 as early responders indeed went on to be classified as ultimate responders on the CGI-I (i.e., PPP = .818), 81% of those classified as early nonresponders by Session 5 indeed went on to be classified as ultimate nonresponders on the CGI-I (NPP = .812), and 82% of all participants are correctly classified at Session 5 as either responding or not responding (i.e., OCC = .815).

Table 2.1. Sequences of care embedded in a hypothetical SMART incorporating behavioral parent training (BPT) and individual child cognitive-behavioral therapy (ICBT).

Sequence	Phase I	Phase II	Description
1	Weekly BPT	Weekly BPT	Continuation strategy for Phase I BPT responders
2	Weekly BPT	Weekly ICBT	Switching strategy for Phase I BPT nonresponders
3	Weekly BPT	Weekly BPT + weekly ICBT	Augmentation strategy for Phase I BPT nonresponders
4	Weekly BPT	Twice-weekly BPT	Intensification strategy for Phase I BPT nonresponders
5	Weekly ICBT	Weekly ICBT	Continuation strategy for Phase I ICBT responders
6	Weekly ICBT	Weekly BPT	Switching strategy for Phase I ICBT nonresponders
7	Weekly ICBT	Weekly ICBT + weekly BPT	Augmentation strategy for Phase I ICBT nonresponders
8	Weekly ICBT	Twice-weekly ICBT	Intensification strategy for Phase I ICBT nonresponders

Note: Continuation strategy = in Phase II, patient's treatment is not modified; switching strategy = in Phase II, patient discontinues Phase I treatment approach and initiates alternative treatment approach; augmentation strategy = in Phase II, patient continues Phase I treatment approach and also adds alternative treatment approach; intensification strategy = in Phase II, patient continues Phase I treatment approach but at a higher "dosage."

Table 2.2. Controlled comparisons embedded in a hypothetical SMART incorporating behavioral parent training (BPT) and individual child cognitive-behavioral therapy (ICBT).

Comparison	Question	Contrasts
A	What is the overall effect of initial BPT versus initial ICBT (regardless of Phase II treatment)?	Sequences 1 + 2 + 3 + 4 vs. Sequences 5 + 6 + 7 + 8
B	For cases of initial nonresponse (i.e., Phase I nonresponders), what is the relative effectiveness of a switching strategy vs. an augmentation strategy vs. an intensification strategy?	Sequences 2 + 6 vs. Sequences 3 + 7 vs. Sequences 4 + 8
C	For individuals who do not adequately respond to an initial brief course of weekly BPT, is it better to discontinue BPT and switch to ICBT, to continue IBT and also add ICBT, or to continue BPT but at a higher intensity?	Sequence 2 vs. Sequence 3 vs. Sequence 4
D	For individuals who do not adequately respond to an initial brief course of weekly ICBT, is it better to discontinue ICBT and switch to BPT, to continue ICBT and also add BPT, or to continue ICBT but at a higher intensity?	Sequence 6 vs. Sequence 7 vs. Sequence 8
E	At the end of Phase II, do initial responders (i.e., Phase I responders) continue to show better outcomes than initial nonresponders whose treatment was adjusted in hopes of improving response?	Sequences 1 + 5 vs. Sequences 2 + 3 + 4 + 6 + 7 + 8

Table 2.3. A hypothetical examination of the clinical utility properties across three alternative early treatment-response factors in the prediction of ultimate treatment response.

	Session 3					Session 4					Session 5				
Cutoff	Sen-sitivity	Speci-ficity	PPP	NPP	OCC	Sen-sitivity	Speci-ficity	PPP	NPP	OCC	Sen-sitivity	Speci-ficity	PPP	NPP	OCC
Early treatment child symptom improvement (ECBI score change from baseline) predicting ultimate treatment response															
–3	.650	.470	.551	.573	.560	.730	.480	.584	.640	.605	.950	.610	.709	.924	.780
–6	.550	.620	.591	.579	.585	.710	.550	.612	.655	.630	.920	.660	.730	.892	.790
–9	.500	.660	.595	.569	.580	.660	.590	.617	.634	.625	.910	.690	.746	.885	.800
–12	.430	.700	.589	.551	.565	.610	.740	.701	.634	.675	.900	.760	.789	.884	.830
–15	.360	.730	.571	.533	.545	.550	.780	.714	.634	.665	.860	.790	.804	.849	.825
–18	.320	.780	.593	.534	.550	.490	.860	.778	.628	.675	.810	.820	.818	.812	.815
–21	.290	.820	.617	.536	.555	.420	.880	.778	.603	.650	.700	.850	.824	.739	.775
–24	.210	.860	.600	.521	.535	.370	.890	.771	.586	.630	.620	.870	.827	.696	.745
–27	.150	.890	.577	.511	.520	.310	.900	.756	.566	.605	.570	.890	.838	.674	.730
–30	.110	.920	.579	.508	.515	.270	.920	.771	.558	.595	.520	.660	.881	.660	.725
Early parent skill acquisition (% of introduced skills mastered in a 5-min coding period) predicting ultimate treatment response															
20%	.610	.470	.535	.547	.540	1	.610	.719	1	.805	.720	.500	.590	.641	.610
30%	.450	.620	.542	.530	.535	.960	.660	.738	.943	.810	.700	.570	.619	.655	.635
40%	.400	.340	.541	.524	.530	.960	.690	.756	.943	.825	.650	.610	.625	.635	.650
50%	.340	.700	.531	.515	.520	.940	.760	.797	.927	.850	.590	.710	.670	.634	.645
60%	.270	.730	.500	.500	.500	.890	.790	.809	.878	.840	.530	.760	.688	.618	.655
70%	.240	.780	.522	.506	.510	.850	.820	.825	.845	.835	.470	.840	.746	.613	.640
80%	.220	.820	.550	.512	.520	.740	.850	.831	.766	.795	.420	.860	.750	.597	.610
Early treatment homework adherence (% of assigned homework tasks completed) predicting ultimate treatment response															
20%	.970	.590	.703	.952	.780	.720	.480	.581	.632	.600	.650	.470	.551	.573	.560
30%	.920	.670	.736	.893	.795	.710	.550	.612	.655	.630	.530	.620	.582	.569	.575
40%	.900	.700	.750	.875	.800	.670	.590	.620	.641	.630	.490	.660	.590	.564	.575
50%	.860	.760	.782	.844	.810	.620	.740	.705	.661	.680	.420	.700	.583	.547	.560
60%	.810	.790	.794	.806	.800	.560	.780	.718	.639	.670	.340	.730	.557	.525	.535
70%	.750	.820	.806	.766	.785	.490	.860	.778	.628	.675	.300	.780	.577	.527	.540
80%	.650	.850	.812	.708	.750	.420	.880	.778	.603	.650	.280	.820	.609	.532	.550
100%	.590	.870	.819	.680	.730	.370	.890	.771	.586	.630	.190	.860	.576	.515	.525

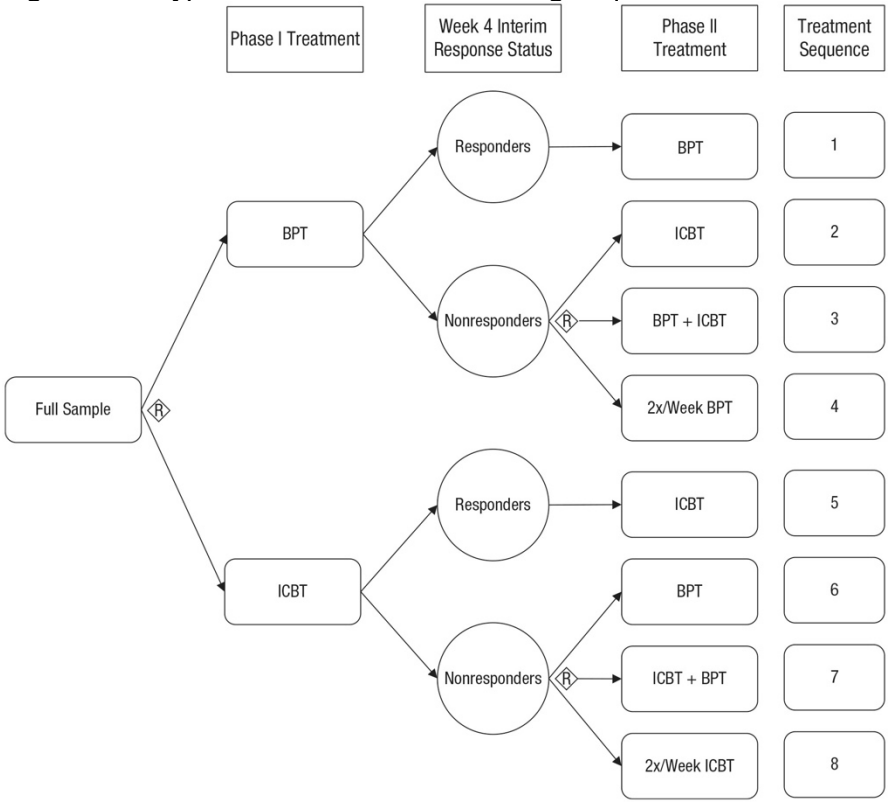
Note: Area-under-the-curve (AUC) indices correspond to plots depicted in Figure 2. PPP = positive predictive power; NPP = negative predictive power; OCC = overall correct classification.

Table 2.4. Comparing the hypothetical classification accuracy of three candidate tailoring variables at three candidate sessions.

	Session 3		Session 4		Session 5	
Candidate tailoring variable	AUC	Asymptotic significance	AUC	Asymptotic significance	AUC	Asymptotic significance
Homework adherence	.876	< .001	.708	< .001	.614	.006
Parental skill acquisition	.580	.050	.905	< .001	.701	< .001
Symptom improvement	.622	.003	.715	< .001	.878	< .001

Note: AUC = area under the curve.

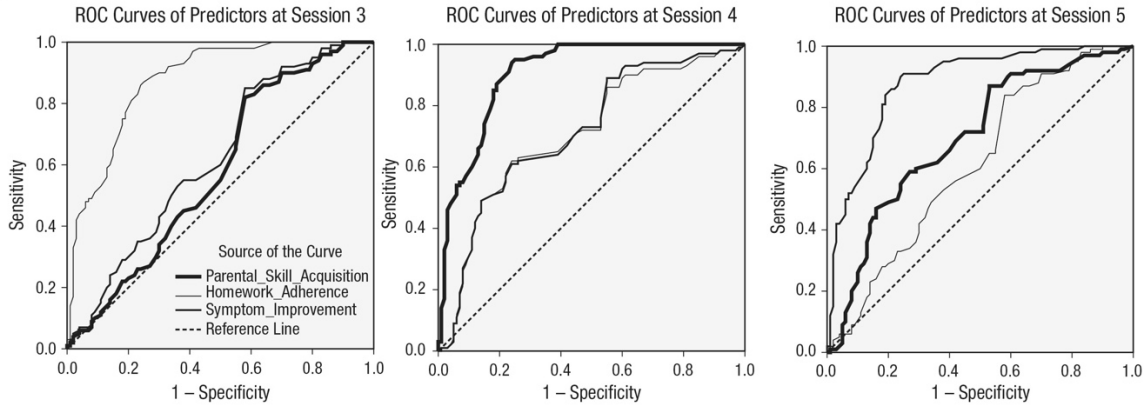
Figure 2.1. Hypothetical SMART evaluating sequences of care for children with conduct problems.



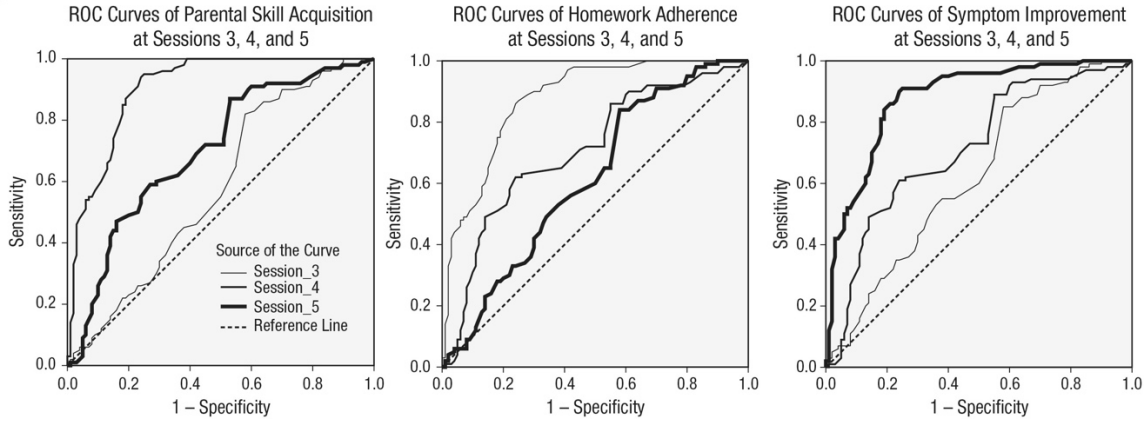
Note: Week 4 interim response status was based on assessment of the tailoring variable at the critical decision point. Responders = Week 4 ECBI score ≥ 10 points below baseline ECBI score; Nonresponders = Week 4 ECBI score < 10 points below baseline ECBI score; R = randomization.

Figure 2.2. Hypothetical receiver-operating characteristic (ROC) curve plots of sensitivity as a function of 1 – specificity for each potential cutoff on (a) candidate tailoring variables, organized by candidate session, and (b) candidate sessions, organized by candidate tailoring variable. Diagonal segments are produced by ties.

a



b



CHAPTER III

Can indicators of early treatment progress predict post-treatment nonresponse? An initial ROC analysis examining candidate tailoring variables and early decision points within a behavioral parenting intervention

Abstract

Limited research has examined the extent to which early, putative treatment mechanisms are predictive of ultimate treatment outcome in the context of behavioral parenting interventions (BPIs) for young children with behavior problems. Across the literature, there is some indication that poor engagement, poor caregiver skill acquisition, and/or limited symptom improvements demonstrated early in treatment may confer risk for BPI treatment nonresponse, but research is needed to confirm (a) which within-treatment response indicators reliably distinguish whether an individual family is at risk for treatment nonresponse (i.e., candidate tailoring variables) and (b) the optimal timepoint(s) within a treatment course at which such indicators are best measured in order to determine the subsequent direction of treatment (i.e., critical decision points; see Hong et al., 2019; Chapter II). The present study analyzed data from a pooled sample of (*N*=153) families who participated in one of four randomized controlled trials examining Parent-Child Interaction Therapy-based approaches for treating early childhood behavior problems. Using the strategy detailed by Hong and colleagues (2019; see Chapter II), receiver operating characteristic (ROC) analyses and corresponding classification indices examined several candidate tailoring variables at two candidate critical decision points and their relations with treatment outcome. Candidate tailoring variables included measures of early treatment engagement, caregiver skills, and child symptoms at sessions 3 and 6. Treatment outcome was defined dichotomously based on both clinically significant change in symptoms from pre-treatment and symptoms falling within the “normal” range at post-treatment. Across both early treatment session timepoints, none of the examined candidate tailoring variables demonstrated acceptable discriminative value in predicting post-treatment outcomes (i.e., $\geq$ 70% sensitivity and $\geq$ 70% specificity). Findings may suggest that early treatment

engagement, caregiver skill acquisition, and child symptom improvements are not significant predictors of ultimate BPI response. At the same time, the present findings also prompt concerns about how dichotomous approaches used to define satisfactory levels of early and ultimate treatment response may fail to reflect the continuous and dynamic relations among treatment response variables as they naturally present across time. Additional research examining multivariable prediction models to inform the precision of BPIs is needed.

Keywords: behavioral parenting interventions; child behavior problems; caregiver skill acquisition; clinical decision-making; experimental therapeutics.

Introduction

Early childhood behavior problems are a serious public health concern due to their high prevalence and association with considerable life impairments and negative sequelae (Costello et al., 2005; Egger & Angold, 2006; Lavigne et al., 2009; Patterson et al., 1998). These problems are well-studied in youth, although many affected children go unrecognized and/or untreated (Carach et al., 2020; Merikangas et al., 2011; Whitney & Peterson, 2019). Research supports the validity of diagnosing disruptive behavior disorders (DBDs) in preschool-age children (Bufferd et al., 2011; Egger & Angold, 2006; Egger et al., 2006; Keenan et al., 2007; Wakschlag et al., 2008), as well as the acute and long-term treatment effectiveness of behavioral parenting interventions (BPIs) for reducing externalizing problems in youth (American Academy of Pediatrics, 2016; Comer et al., 2013; Gleason et al., 2016; Keenan & Wakschlag, 2002).

BPIs—including Parent-Child Interaction Therapy (PCIT; Eyberg & Funderburk, 2011), the Incredible Years (Webster-Stratton & Reid, 2017), Triple P (Positive Parenting Program; Sanders, 1999), Helping the Noncompliant Child (McMahon & Forehand, 2005), and others—are all grounded in attachment and social learning theories that describe how child behavior problems and coercive, inconsistent, or otherwise ineffective parenting practices can maintain or exacerbate one another over time through positive and negative reinforcement (Patterson et al., 2002). As such, BPIs target child behavior problems by directly intervening on the primary social contexts and environments in their lives (e.g., caregiver behaviors, parent-child interactions). Specifically, BPIs focus on reshaping parent-child interactions by guiding caregivers to use positive attending skills that provide consistent reinforcement for appropriate child behaviors, and to use effective discipline strategies (e.g., timeout sequences) for child misbehavior. Increasing the warmth, responsiveness, predictability, and effectiveness of

caregivers in parent-child interactions can, in turn, dismantle coercive family processes that have entrenched negative child behaviors and family dysfunction (Kaminski et al., 2008; Patterson, 2002; Zisser & Eyberg, 2010).

A preponderance of research has shown that BPIs can substantially reduce behavior problems across a wide range of typically and atypically developing youth populations (Bagner et al., 2007; Bagner et al., 2010; Comer et al., 2017; Gardner & Leijten, 2017; Menting et al., 2013; Scudder et al., 2019; Thomas & Zimmer-Gembeck, 2007; Thomas et al., 2017; Ward et al., 2016). However, not all families experience positive outcomes from BPIs. A considerable proportion of treated families—often neglected in the literature—terminate treatment prematurely, demonstrate inconsistent engagement throughout treatment, and/or still report clinically significant child symptoms and impairments following a full course of treatment (Assenany & McIntosh, 2002; Bagner & Graziano, 2013; Chacko et al., 2016; Chacko et al., 2017; Danko et al., 2016). Heterogeneity in treatment effects has inspired numerous lines of work aiming to enhance the precision of mental health care. By examining a range of time-invariant factors (e.g., baseline patient characteristics) and time-varying factors (e.g., symptom trajectories) clinical scientists have sought to make sense of observed heterogeneity in treatment outcomes through prognostic research that has the potential to, in turn, inform the identification of optimal treatment strategies and the development of clinical decision-making tools that facilitate more patient-centered and dynamically responsive sequences of care (Almirall et al., 2012; Collins et al., 2014; Delgadillo & Lutz, 2020; Hong et al., 2019; Pettit et al., 2016; Rekkas et al., 2020; Varadhan et al., 2013).

Consistent with guidelines for evidence-based practice in psychology (APA Presidential Task Force on Evidence-Based Practice, 2006), routine outcome monitoring (ROM) measures are often used in clinical practice to assess progress, provide

feedback, and tailor approaches to individual patients/families throughout the course of treatment. ROM measures can be particularly useful tools for aiding clinical decision-making practices given clinicians are often biased and inaccurate in their predictions of treatment outcomes (Hannan et al., 2005; Magnavita & Lilienfeld, 2016; Perlis, 2016). Indeed, integrating ROM measures *and* leveraging automated "risk signals" for individuals "not on track" for treatment response throughout the course of treatment can assist in providing feedback informed care that reduces the likelihood of ultimate treatment nonresponse (Lambert et al., 2018). Although meta-analyses of such feedback informed treatments (Kendrick et al., 2016; Shimokawa et al., 2010) have presented inconsistent findings, research generally seems to support the use of these practices, noting that they can be particularly beneficial for individuals classified as "not on track" (Delgadillo et al., 2018; Lambert et al., 2018). Similarly, modern clinical trial methodologies—such as the sequential, multiple assignment, randomized trial (SMART; Almirall & Chronis-Tuscano, 2016)—have shown that adaptive interventions tailored according to within-treatment response indicators (i.e., "tailoring variables") assessed at mid-treatment timepoints (i.e., "critical decision points") can inform optimal sequences of care and lead to improved mental health treatment outcomes for youth. For example, Pelham and colleagues' (2016) SMART for childhood attention-deficit hyperactivity disorder (ADHD) showed that youth displayed the best outcomes when they (a) began with behavioral, as opposed to pharmacological, treatment and (b) augmented behavioral therapy with medication when children displayed insufficient response throughout the course of intervention.

Regrettably, in the treatment of early child DBDs, research examining predictors of treatment nonresponse has largely focused on baseline sociodemographic (e.g., child age, child sex, family income, caregiver education, racial and ethnic minority status)

and/or baseline clinical characteristics (e.g., child severity and comorbidity, caregiver psychopathology), and these studies have largely produced mixed findings (Baydar et al., 2003; Beuchaine et al., 2005; Comer et al., 2013; Gardner & Leijten, 2017; Gardner et al., 2010; Lundahl et al., 2006; Menting et al., 2013; Nix et al., 2009; Reyno & McGrath, 2006; Shelleby & Shaw, 2014; Ward et al., 2016; Werba et al., 2006). Less research has considered whether early process factors, such as early engagement with treatment and/or early changes in the desired outcomes of treatment, predict ultimate BPI nonresponse (Hong et al., 2019). Even when intra-treatment variables are examined as early signals of treatment nonresponse, they typically reflect measures of symptoms and/or impairment (e.g., Delgadillo et al., 2014; Durland et al., 2018; Lutz et al., 2016; Pelham et al., 2016), as opposed to considering alternative, proximal targets that characterize early treatment progress using an experimental therapeutics framework that examines early mechanisms of ultimate symptom remission (Hong et al., 2019; Southward & Sauer-Zavala, 2020).

This absence is in striking contrast to the ROM and feedback informed protocol embedded in BPIs such as PCIT, in which caregiver engagement in out-of-session homework assignments, skill acquisition, and child symptoms are assessed and discussed at each treatment session (Eyberg & Funderburk, 2011). Within BPIs, out-of-session homework completion can serve as a proxy for caregiver perspectives on treatment (e.g., motivation, buy in; Nock & Ferriter, 2005), as well as an early metric of treatment progress, insofar as the caregiver's engagement in homework leads to acquiring skills that subsequently produce improvements in child behavior (Pfeifer & Strunk, 2015). Although homework completion rates are infrequently reported in BPI research (Chacko et al., 2016), a small body of work examining patterns of BPI homework completion do show positive associations with caregiver and child outcomes

(Ros et al., 2016; Ros et al., 2017; Stokes et al., 2016). Moreover, meta-analyses have demonstrated the positive effects of homework on CBT outcomes (Kazantzis et al., 2000; Kazantzis et al., 2016; Mausbach et al., 2010) and the importance of completing homework tasks is consistently highlighted across the youth treatment literature (e.g., Cummings et al., 2014).

Despite enthusiasm for examining potential mediators of youth mental health treatment outcomes, formal tests of theorized mechanisms of treatment-related change are relatively absent from the literature (Weersing & Weisz, 2002). For example, a review of caregiver skill acquisition and utilization in evidence-based treatments for childhood behavior problems by Lindhiem and colleagues (2014) showed that although skills were frequently assessed across studies, a small minority evaluated any sort of associations between caregiver skills and other variables. Similarly, although Forehand and colleagues' (2014) review found some support for parenting practices as a mediator between BPIs and youth outcomes, they also highlighted the lack of rigorous mediation tests in the literature and the need for further research evaluating the role of parenting in BPIs. In one example, Stokes and colleagues (2016) found that although homework rates did not predict post-treatment child outcomes when controlling for pre-treatment severity, families who reported engaging in higher rates of homework early in treatment (i.e., engaging in 5-minutes of special time practicing Child-Directed Interaction skills during the first phase of PCIT) acquired skills more quickly (i.e., meeting a specific skill criterion within fewer sessions). This study also found that families who reported high rates of continued engagement with homework focused on early treatment targets during the second phase of treatment completed treatment within fewer sessions, indicating that homework was associated with not only caregiver skill acquisition, but also positive child outcomes. Indeed, other studies of child mental health treatments that incorporate

skills in treatment (e.g., CBT) show that child and caregiver skill acquisition is associated with positive youth treatment outcomes (e.g., Dishion et al., 2008; Gardner et al., 2006; Weiss, 1999).

Taken together, although very little research on BPIs has examined the extent to which early, putative treatment mechanisms are predictive of ultimate treatment outcome, there is some indication that poor engagement, poor caregiver skill acquisition, and/or limited symptom improvements early in treatment may confer risk for BPI treatment *nonresponse*. To guide clinical decision-making and improve mental health treatment outcomes for this population, research on BPIs is needed to examine the dynamic processes through which treatment unfolds, whether candidate tailoring variables such as early engagement, skill acquisition, or symptom response can accurately predict when a family is at risk for ultimate treatment nonresponse, and the critical decision points (e.g., session 3?, session 6?) at which such variables are most predictive of outcome (see Hong et al., 2019; Chapter II). Notably, building upon the clear theoretical rationale for feedback informed methods embedded within BPIs like PCIT, data-driven approaches to empirically identifying tailoring variables and critical decision points should confirm that (a) within-treatment response indicators reliably distinguish whether an individual family is at risk for treatment nonresponse and (b) the optimal timepoint(s) within a treatment course at which such tailoring variables are best measured to determine the subsequent direction of treatment (see Hong et al., 2019).

Given high BPI attrition rates (Abrahamse et al., 2016; Chacko et al., 2016) and research showing that the majority of families who drop out of treatment for youth behavior problems do so *early* in the course of treatment (i.e., within 5-7 sessions; Chacko et al., 2017; Danko et al., 2016; Kazdin & Mazurick, 1994), it is critical to identify empirically derived tailoring variables and critical decision points *early on* in the course of

treatment for youth DBDs (i.e., before the 7th session). Continuing to deliver or participate in a treatment that is not leading toward desired outcomes is inefficient, discouraging, and resource intensive for families and providers, further highlighting the need to identify probable treatment nonresponders and make necessary treatment adaptations as early as possible (Chacko et al., 2017; Delgadillo et al., 2014; Pettit et al., 2016). That said, until relatively recently there have not been clear methodologies proposed for empirically identifying reliable early treatment tailoring variables and optimal critical decision points. The methodological and analytic strategy described by Hong and colleagues (2019; see Chapter II) calls for the simultaneous comparison of classification indices (e.g., area under the receiver operating characteristic [ROC] curve, sensitivity, specificity, positive predictive power, negative predictive power, overall correct classification) for a range of candidate tailoring variables at various candidate critical decision points. Hong and colleagues (2019; see Chapter II) have suggested that researchers analyze archival data from randomized controlled trials (RCTs) on fixed (nonadaptive) treatment courses that incorporated ROM of key mechanistic response variables early in treatment. ROC analysis can then be used to examine the extent to which various candidate tailoring variables at various candidate critical decision points may differentially predict post-treatment outcomes. To date, this methodology has not been applied to the study of BPIs.

Present Study

The present study utilizes secondary data analysis to (a) examine whether early treatment progress data (i.e., candidate tailoring variables) collected within the first six sessions of a BPI can accurately detect families at risk for nonresponse at the end of treatment, and (b) compare the clinical utility of candidate tailoring variables and their optimal cut scores at candidate critical decision points. Data were analyzed from a

pooled sample of typically and atypically developing youth whose families presented for treatment of early childhood behavior problems and participated in one of four RCTs. It was hypothesized that early measures of out-of-session engagement, caregiver skills, and child symptoms would each demonstrate an ability to distinguish families who were subsequently classified as “treatment responders” versus “treatment nonresponders” at post-treatment. Specifically, it was hypothesized that ROC analyses performed on these early candidate tailoring variables would yield a significant area under the curve (AUC) for each variable at session 3 and again at session 6, as well as at least one cut score that simultaneously demonstrated at least 70% sensitivity and 70% specificity correctly classifying participants in accordance with their actual post-treatment outcome category. It was also hypothesized that candidate tailoring variables measured at session 6 would demonstrate better classification properties than those measured at session 3, given the introductory nature of the first several BPI treatment sessions, previous work suggesting caregiver skill acquisition and child behavior change is not expected prior to the fourth session (see Lieneman et al., 2019), and other examinations showing improved accuracy predicting final treatment outcomes after four or more sessions (see Delgadillo et al., 2014). Exploratory analyses compared the classification indices across candidate tailoring variables to examine their relative performance and whether any outperformed others.

Method

Sample and Procedures

Data were drawn from four RCTs examining PCIT approaches for treating a range of early childhood populations (Bagner et al., 2007; Bagner et al., 2010; Comer et al., 2017; Bagner et al., in preparation). All four trials used PCIT-based programs to treat families of children $\leq$ 6 years with externalizing behavior problems and excluded families

of children with severe sensory and/or autism spectrum disorder impairments. Table 3.1 includes a detailed overview of the shared and unique features of each study. Data from all four trials were collected at a pre-treatment assessment, at each treatment session, and at a post-treatment assessment. Families were included in the present analysis if they began treatment (i.e., attended $\geq$ 1 treatment session in one of the four studies), resulting in a pooled sample of 153 families. Table 3.2 includes sociodemographic information for the present study sample. Study procedures for all four studies were approved by university-affiliated Institutional Review Boards and all primary caregivers provided informed consent prior to family study participation.

Measures

Out-of-Session Engagement

For each family, out-of-session engagement rates during early sessions were measured via the CDI Homework Sheet included in the PCIT Protocol (Eyberg & Funderburk, 2011). Consistent with the PCIT Protocol, families were instructed to engage in daily 5-minute homework assignments between sessions (i.e., practicing parenting strategies taught in treatment). A percentage reflecting each family's cumulative out-of-session engagement was calculated by dividing the total number of days the caregiver reported engaging in out-of-session homework assignments (numerator) by the total number of days since the first treatment session (denominator). Percentages reflecting out-of-session engagement rates at sessions 3 and 6 were included as predictors in the present analyses.

Caregiver Skills

Early treatment levels of caregiver skills were assessed using the Dyadic Parent-Child Interaction Coding System – Fourth Edition (DPICS-IV; Eyberg et al., 2013). The DPICS-IV is a behavioral observation coding system designed to evaluate the quality of

parent-child interactions. The standard DPICS-IV assessment includes a 5-minute "Child-Led Play" segment in which caregivers are instructed to allow the child to choose the activity and to follow the child's lead in play. Caregiver verbalizations are coded as "neutral talk," "behavior descriptions," "reflections," "labeled praises," "unlabeled praises," "commands," "questions," and "negative talk" (see Table 3.3 for descriptions and examples of each of these types of coded verbalizations). In the Child-Led Play segment, the tally of behavior descriptions, reflections, and labeled praises (three forms of verbal positive reinforcement PCIT teaches caregivers to use during parent-child interactions) comprise a total "do" skills score, and the tally of questions, commands, and negative talk (three types of verbal attention PCIT teaches caregivers to avoid during parent-child interactions) comprise a total "don't" skills score. Higher do skill scores and lower don't skill scores during the 5-minute Child-Led Play interaction represent greater caregiver skill acquisition. A complete course of standardized PCIT requires caregivers to meet specific skill criteria in do skills (i.e., $\geq$ 10 behavior descriptions, $\geq$ reflections, and $\geq$ labeled praises) and don't skills (i.e., $\leq$ 3 total don't skills) during the 5-minute Child-Led Play coding segment at the start of session prior to advancing to the second phase of treatment. Consistent with the PCIT Protocol (Eyberg & Funderburk, 2011), the DPICS-IV was administered at the pre-treatment assessment and was used to code caregiver skills during a 5-minute child-led play segment at the beginning of most treatment sessions. Across all studies, coders were trained to 80% agreement with a criterion tape prior to coding the DPICS-IV for study purposes.

Change in caregiver do and don't skills reflect the difference between pre-treatment and session specific scores (e.g., session scores minus pre-treatment scores for do skills; pre-treatment scores minus session scores for don't skills). Higher change scores indicate greater skill acquisition. Raw skill acquisition scores at sessions 3 and 6,

and skill acquisition change scores at sessions 3 and 6, were included as predictors in the present analyses.

Child Symptoms

Early treatment levels of child symptoms were measured using the Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999). The ECBI is a 36-item caregiver-report questionnaire assessing the frequency of behavior problems in youth. Each item is rated on a 7-point Likert-type scale ranging from 1 (never) to 7 (always) and summed to yield an Intensity scale score ranging from 36 to 252. Higher scores indicate greater symptomatology. The ECBI has demonstrated good psychometric properties (Funderburk et al., 2003; Gross et al., 2007; Weis et al., 2005; α=.90-.92 across the present study samples). Consistent with the PCIT Protocol (Eyberg & Funderburk, 2011), the ECBI was collected at the pre-treatment assessment and at each treatment session.

Change in child symptoms reflect the difference between pre-treatment and session specific ECBI Intensity scale scores (i.e., pre-treatment ECBI score minus the session ECBI score). Higher change scores indicate greater improvement. Raw child symptom scores at sessions 3 and 6, and child symptom change scores at sessions 3 and 6, were included as predictors in the present analyses.

Treatment Outcome

Child treatment outcomes were characterized at the post-treatment timepoint based on the Child Behavior Checklist (CBCL) for ages 1.5-5 (Achenbach & Rescorla, 2000) or CBCL for ages 6-18 (Achenbach & Rescorla, 2001). The CBCL is a widely used caregiver-report questionnaire assessing a broad range of emotional and behavioral problems in youth. Items are rated on a 3-point Likert-type scale from 0 (not true) to 2 (very true or often true). The CBCL yields *T* Scores normed for age and sex

that correspond to a variety of domains, each with a mean of 50 and a standard deviation of 10. The present analysis used the Externalizing Problems scale, on which higher scores indicate worse behavior problems. The CBCL Externalizing Problems scale has demonstrated excellent psychometric properties (Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001; α=.82-.95 across the present study samples).

Participants were considered to have demonstrated satisfactory treatment outcomes if, at the post-treatment timepoint, they both (a) demonstrated clinically significant change from pre-treatment and (b) had scores that fell within the "normal" range on the CBCL Externalizing Problems scale. Clinically significant change was calculated using the Reliable Change Index (RCI; Jacobson & Truax, 1991), whereby the difference in pre-treatment and post-treatment Externalizing Problems *T* Scores was divided by the standard error of the difference score. RCIs > 1.96 are considered clinically significant and *T* Scores < 60 are considered within the "normal" range on the CBCL Externalizing Problems scale. Participants demonstrating these criteria were classified as "treatment responders." Participants who did not meet both of these criteria at the post-treatment timepoint were classified as "treatment nonresponders."

Analytic Plan

First, data from each eligible family across the four RCTs were pooled into one dataset to facilitate the present integrated data analysis. Next, missing data were multiply imputed using IBM SPSS Statistics (Version 26) to generate 20 imputed datasets. Specifically, multiple imputation was conducted including all variables utilized in analyses, as well as a number of auxiliary variables. Auxiliary variables included in the multiple imputation process included: study number dummy codes, treatment format (i.e., in-person versus telehealth dummy coded), child sex (dummy coded), child age at pre-treatment, CBCL Externalizing *T* Scores at pre-treatment, ECBI scores at pre-

treatment, session 1, and session 2, do skills scores at pre-treatment and session 2, don't skills scores at pre-treatment and session 2, out-of-session engagement rates at session 2, treatment dropout status (dummy coded), and the number of weeks between each family's pre-treatment and post-treatment assessments. Data imputed for session 3 and session 6 ECBI scores were constrained to values between 36 and 252. Data imputed for session 3 and session 6 out-of-session engagement rates were constrained to values between 0% and 100%. Data imputed for session 3 and session 6 do skills scores were constrained to values between 0 and 75. Data imputed for session 3 and session 6 don'ts skills scores were constrained between values of 0 and 125. Finally, data imputed for post-treatment CBCL Externalizing *T* Scores were constrained to values between 25 and 99.

Then, ROC analyses and corresponding classification indices were computed using these imputed datasets to examine the level of association between treatment outcome (dichotomously coded) and seven early treatment predictors—i.e., candidate tailoring variables; (1) child symptoms (ECBI), (2) change in child symptoms (ECBI difference scores), (3) caregiver do skills (DPICS-IV tally), (4) change in caregiver do skills (DPICS-IV tally difference scores), (5) caregiver don't skills (DPICS tally), (6) change in caregiver don't skills (DPICS-IV tally difference scores), and (7) out-of-session engagement—at two candidate critical decision points—i.e., at session 3 and session 6. Thus, a total of 14 early treatment predictors were examined. ROC analyses evaluating the AUC of each predictor variable within each of the 20 imputed datasets were conducted in IBM SPSS Statistics (Version 26) and pooled in `R`. AUC values provide a quantitative metric of a measure's overall classification accuracy across all potential cut scores, ranging from 0.5 to 1.0. Values are interpreted as follows: 0.5-0.6 = no discrimination/predictive value; 0.6-0.7 = poor discrimination/classification; 0.7-0.8 =

acceptable/fair discrimination/classification; 0.8-0.9 = excellent/good discrimination/classification; 0.9-1.0 = outstanding discrimination/classification (1.0 = perfect classification; Lantz, 2019).

Finally, the classification properties at a broad range of potential cut scores for each candidate tailoring variable, within each of the 20 imputed datasets, were analyzed and pooled in `R`. Specifically, sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP), and the overall correct classification (OCC) were calculated for each potential cut score across candidate tailoring variables at candidate critical session decision points. *Sensitivity* at each possible cut score of a candidate tailoring variable is defined as the percentage of ultimate treatment nonresponders (i.e., those who did not demonstrate clinically significant changes from pre-treatment and/or had elevated scores on the CBCL Eternalizing Problems scale at post-treatment) whose cut score indicated a "positive test result" at the candidate session decision point (i.e., scored within the particular cut score range on the candidate tailoring variable). *Specificity* at each possible cut score of a candidate tailoring variable is defined as the percentage of ultimate treatment responders (i.e., those who demonstrated clinically significant changes from pre-treatment and had scores within the "normal" range on the CBCL Eternalizing Problems scale at post-treatment) whose cut score indicated a "negative test result" at the candidate session decision point (i.e., scored outside of the particular cut score range on the candidate tailoring variable). *PPP* at each possible cut score of a candidate tailoring variable is defined as the percentage of cases who at the candidate session decision point had cut scores that indicated a "positive test result" and went on to be classified as ultimate treatment nonresponders based on their CBCL Externalizing Problems scale scores at post-treatment. *NPP* at each possible cut score of a candidate tailoring variable is defined as the percentage of cases who at the

candidate session decision point had cut scores that indicated a “negative test result” and went on to be classified as ultimate treatment responders based on their CBCL Externalizing Problems scale scores at post-treatment. Finally, *OCC* at each possible cut score of a candidate tailoring variable is defined as the percentage of all cases who were either (a) scored within the cut score range and indeed went on to be classified as an ultimate treatment nonresponder based on their CBCL Externalizing Problems scale scores at post-treatment, or (b) scored outside of the cut score range and indeed went on to be classified as an ultimate treatment responder based on their CBCL Externalizing Problems scale scores at post-treatment.

Results

Overall Classification Accuracy

ROC analyses evaluated the overall clinical utility of out-of-session engagement, (change in) caregiver do skills, (change in) caregiver don’t skills, and (change in) child symptom scores at session 3 and session 6 in accurately distinguishing treatment responders from treatment nonresponders across all possible scores for each measure. Table 3.4 presents the pooled AUC values for each measure at session 3 and session 6 across all 20 imputed datasets. The only candidate tailoring variable that performed significantly better than chance at accurately distinguishing treatment responders from treatment nonresponders was session 3 change in caregiver don't skills (*AUC* = .63, $p < .05$). Figure 3.1 presents the ROC curve for this tailoring variable at session 3—plotting the sensitivities by 1 – specificities for each possible value of session 3 change in caregiver don’t skills from the first imputed dataset.

Individual Classification Properties

The sensitivity, specificity, PPP, NPP, and OCC for each measure at session 3 and session 6 was calculated to examine comparisons across candidate tailoring

variables and decision points. Table 3.5 presents the pooled classification properties for each individual measure across various cut scores at each session.

Out-of-session engagement

Out-of-session engagement rates were examined at cut scores ranging from < 20% through < 90% at both session 3 and session 6.

Session 3. Sensitivity indices for session 3 out-of-session engagement ranged from 13% to 94% and specificity indices ranged from 12% to 93%. Sensitivity increased as cut scores increased (i.e., larger percentages of out-of-session engagement rates associated with higher sensitivity), while specificity decreased. At session 3, no out-of-session engagement cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties was found at a session 3 out-of-session engagement cut score of < 60%, which corresponded to only 58% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 57% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Session 6. Sensitivity indices for session 6 out-of-session engagement ranged from 8% to 96% and specificity indices ranged from 3% to 93%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with higher cut scores), while specificity decreased. At session 6, no out-of-session engagement cut score simultaneously achieved 70% sensitivity and 70% specificity. As with session 3 out-of-session engagement scores, at session 6, the most favorable balance between properties was found at an out-of-session engagement cut score of < 60%. At session 6, this cut score corresponded to only 53% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 48% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Caregiver skill acquisition

Caregiver do skills were examined at cut scores ranging from < 5 through < 30 at both session 3 and session 6. Additionally, change in caregiver do skills relative to pre-treatment were examined at cut scores ranging from < 5 through < 30 at both session 3 and session 6.

Caregiver don't skills were examined at cut scores ranging from > 5 through > 30 at both session 3 and session 6. Additionally, change in caregiver don't skills relative to pre-treatment were examined at cut scores ranging from < 0 to < 45 at both session 3 and session 6.

Session 3 caregiver do skills. Sensitivity indices for session 3 caregiver do skills ranged from 14% to 92% and specificity indices ranged from 7% to 90%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with a greater number of caregiver do skills), while specificity decreased. At session 3, no caregiver do skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties was found at a session 3 caregiver do skills cut score of < 15, which corresponded to only 52% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 53% of treatment responders being correctly classified with a negative test result (specificity).

Sensitivity indices for session 3 change in caregiver do skills ranged from 23% to 83% and specificity indices ranged from 17% to 79%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with greater increases in caregiver do skills), while specificity decreased. At session 3, no change in caregiver do skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 3 was found at a change in caregiver do skills cut

score of < 15, which corresponded to only 59% of ultimate treatment nonresponders beinig correctly classified with a positive test result (sensitivity) and only 48% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Session 3 caregiver don't skills. Sensitivity indices for session 3 caregiver don't skills ranged from 9% to 80% and specificity indices ranged from 29% to 89%. Sensitivity decreased as cut scores increased (i.e., higher sensitivity associated with fewer caregiver don't skills), while specificity increased. At session 3, no caregiver don't skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 3 was found at a caregiver don't skills cut score of > 10, which corresponded to only 52% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 53% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Sensitivity indices for session 3 change in caregiver don't skills ranged from 21% to 78% and specificity indices ranged from 28% to 86%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with greater decreases in caregiver don't skills), while specificity decreased. At session 3, no caregiver don't skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 3 was found at a change in caregiver don't skills cut score of < 25, which corresponded to 64% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) but only 54% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Session 6 caregiver do skills. Sensitivity indices for session 6 caregiver do skills ranged from 7% to 85% and specificity indices ranged from 27% to 96%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with a greater number of caregiver do skills), while specificity decreased. At session 6, no

caregiver do skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 6 was found at a caregiver do skills cut score of < 20, which only corresponded to 52% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 56% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Sensitivity indices for session 6 change in caregiver do skills ranged from 11% to 80% and specificity indices ranged from 26% to 90%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with greater increases in caregiver do skills), while specificity decreased. At session 6, no change in caregiver do skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 6 was found at a change in caregiver do skills cut score of < 20, which corresponded to only 59% of ultimate treatment responders being correctly classified with a positive test result (sensitivity) and only 49% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Session 6 caregiver don't skills. Sensitivity indices for session 6 caregiver don't skills ranged from 4% to 56% and specificity indices ranged from 43% to 98%. Sensitivity decreased as cut scores increased (i.e., higher sensitivity associated with fewer caregiver don't skills), while specificity increased. The most favorable balance between properties at session 6 was found at a caregiver don't skills cut score of > 5, which corresponded to only 56% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 43% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Sensitivity indices for session 6 change in caregiver don't skills ranged from 17% to 76% and specificity indices ranged from 33% to 91%. Sensitivity increased as cut

scores increased (i.e., higher sensitivity associated with greater decreases in caregiver don't skills), while specificity decreased. At session 6, no change in caregiver don't skills cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 6 was found at a change in caregiver don't skills cut score of < 30, which corresponded to only 63% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 48% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Child symptoms

Child symptoms were examined at cut scores ranging from > 80 through > 150 at both session 3 and session 6. Additionally, change in caregiver symptoms relative to pre-treatment were examined at cut scores ranging from < 0 through < 60 at both session 3 and session 6.

Session 3. Sensitivity indices for session 3 child symptoms ranged from 35% to 91% and specificity indices ranged from 8% to 78%. Sensitivity decreased as cut scores increased (i.e., higher sensitivity associated with fewer symptoms), while specificity decreased. At session 3, no child symptom cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties was found at a session 3 child symptom cut score of > 120, which corresponded to only 60% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 44% of ultimate treatment responders being correctly classified with a negative test result (specificity).

Sensitivity indices for session 3 change in child symptoms ranged from 37% to 90% and specificity indices ranged from 5% to 70%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with greater decreases in child symptoms),

while specificity decreased. At session 3, no change in child symptoms cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 3 was found at a change in child symptoms cut score of < 10, which corresponded to only 53% of ultimate treatment nonresponders being correctly classified with a positive test result (sensitivity) and only 52% of ultimate treatment respodners being correctly classified with a negative test result (specificity).

Session 6. Sensitivity indices for session 6 child symptoms ranged from 28% to 78% and specificity indices ranged from 25% to 85%. Sensitivity decreased as cut scores increased (i.e., higher sensitivity associated with fewer symptoms), while specificity decreased. At session 6, no child symptom cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 6 was found at a child symptom cut score of > 100, which corresponded to only 63% of ultimate treatment nonresponders being classified with a positive test result (sensitivity) and only 47% of ultimate treatment responders being classified with a negative test result (specificity).

Sensitivity indices for session 6 change in child symptoms ranged from 28% to 82% and specificity indices ranged from 17% to 79%. Sensitivity increased as cut scores increased (i.e., higher sensitivity associated with greater decreases in child symptoms), while specificity decreased. At session 6, no change in child symptoms cut score simultaneously achieved 70% sensitivity and 70% specificity. The most favorable balance between properties at session 6 was found at a change in child symptoms cut score of < 30, which only corresponded to 63% of ultimate treatment nonresponders being correctly classified with a positive test score (sensitivity) and only 50% of ultimate treatment responders being correctly classified with a negative test score (specificity).

Discussion

Grounded in an experimental therapeutics framework, the present study utilized a methodological and analytic strategy detailed by Hong and colleagues (2019; see Chapter II) to simultaneously compare various early interim treatment progress indicators to assess their clinical utility predicting post-treatment nonresponse. This study is the first to examine whether a range of proximal treatment targets, assessed at multiple early session timepoints, could adequately predict ultimate treatment outcomes. In a pooled sample of 153 families receiving PCIT-based treatment for early childhood behavior problems, analyses evaluated whether data related to theorized mechanisms of change—i.e., early engagement with treatment, early caregiver skill acquisition, early symptom change—could be used to identify empirically informed tailoring variables and decision points for consideration when designing personalized mental health service strategies for families of youth presenting for treatment of early childhood disruptive behavior problems.

Contrary to hypotheses, ROC analyses found little support that families' out-of-session engagement, caregiver skill acquisition, and child symptoms early in treatment could adequately distinguish which cases would progress to ultimately respond to treatment versus which cases would not. Specifically, with the exception of change in caregiver don't skills between pre-treatment and session 3, ROC analyses performed on early candidate tailoring variables yielded non-significant AUC values ($p > .05$) for all variables at all time points. The significant AUC found for session 3 change in caregiver don't skills fell within the poor (i.e., 0.6 to 0.7) discrimination range (Lantz, 2019). Additionally, contrary to hypotheses, no cut score for any candidate tailoring variable achieved the desired balance of $\geq$ 70% sensitivity and $\geq$ 70% specificity in correctly classifying participants. Finally, the hypothesis that candidate tailoring variables

measured at session 6 would demonstrate better classification properties than those measured at session 3 was not supported. In fact, the most favorable sensitivity-versus-specificity balance was observed at session 3 (i.e., change in caregiver don't skills with a cut score of < 25, which corresponded to 64% sensitivity and 54% specificity).

Although unexpected, there are several plausible explanations for the present findings. First, although theory and research indicate that engagement, caregiver skill acquisition, and child symptom improvements observed early in treatment are encouraging signs of treatment progress, it is possible these factors may not differentially inform for whom an intervention is likely to benefit or at least may not be able to do so when studied in isolation from one another (e.g., an issue with the selection of candidate tailoring variables). Failure to identify singular variables that can independently predict treatment effects has troubled clinical scientists investigating mental health care for a variety of disorders (see Cohen & DeRubeis, 2018), so this challenge is likely not specific to the identification of predictors of BPI outcomes. Alternatively, it could be that one or more of these variables are strong, independent predictors of ultimate treatment nonresponse, but that their accuracy improves over time and is not fully developed by the sixth session (e.g., an issue with the selection of candidate critical decision points). Although other research has shown that favorable treatment outcomes can be accurately predicted early in treatment (i.e., via early symptom changes), accurately identifying individuals *at risk for treatment nonresponse* has been much more equivocal (Delgadillo et al., 2018).

A number of methodological and analytical design considerations specific to the present study also merit attention. For example, drawing data from across four RCTs afforded the opportunity to increase sample size and potentially generalize study results to a broadened range of treatment-seeking families of young children with behavior

problems. Notably, families included in the present analyses shared numerous characteristics that made them well-suited for pooling for the present analyses (e.g., child age range and presenting problems, fixed courses of PCIT-based intervention, completion of ROM measures on a session-by-session basis). However, the use of multiply imputed datasets and constraints imposed by pooled ROC analysis in the present study limited the ability to control for potential covariates such as family demographics and/or other treatment-related experiences (e.g., study, format, length of treatment) that may have introduced "noise" and undermined otherwise meaningful relationships among study variables. Moreover, although common across the youth mental health treatment literature, defining treatment outcome dichotomously and limiting individual predictors to classifications based on specific cut scores introduced restrictions that may not accurately reflect distributions and relationships among early treatment experiences and post-treatment outcomes. Taken together, further examination of the selected candidate tailoring variables and critical decision points identified for this study, using broader definitions of early treatment response and outcome and more nuanced analytic strategies may be warranted in order to more accurately detect families at risk for suboptimal treatment experiences.

Continued efforts to identify empirically informed definitions of insufficient treatment progress and risk for treatment nonresponse are critical given that most research on BPIs focuses on the predicting *positive* treatment outcomes (e.g., Assemany & McIntosh, 2002) and there are serious consequences and negative sequalae demonstrated by youth who present with early onset child externalizing problems and do not receive adequate care (Patterson et al., 1998). Although it is possible that examining early levels of family engagement, caregiver skills, and child symptoms individually cannot clearly delineate for which families BPIs are leading to

satisfactory outcomes, alternative analytic approaches that model the complexities of the dynamic treatment process may show a different pattern of results (e.g., interactions between early changes in child and caregiver behaviors). Future research would do well to explore alternative strategies for examining interactions among predictors and outcomes in analyses (e.g., multilevel modeling, machine learning algorithms; Delgadillo & Lutz, 2020; Schwartz et al., 2021; Varadhan et al., 2013). Combinations of predictors, or interactions among them, may yield more meaningful predictors of ultimate treatment response. This work can aid in understanding the combined, dynamic, and interactive nature of various interim progress data, can better inform the prediction of risk for treatment nonresponse, and ultimately help to design more responsive clinical decision-making strategies that can enhance treatment outcomes for those who are otherwise unlikely to benefit from continued treatment.

Table 3.1. Shared and unique features across the four RCTs from which the presently analyzed data were drawn.

	Study 1 Bagner et al., 2007	**Study 2 Bagner et al., 2010**	**Study 3 Comer et al., 2017**	**Study 4 Bagner et al., in preparation**
Total *N*	30	28	40	150
n included in present study	20	25	38	70
Proportion of present study sample	13%	16%	25%	46%
Targeted youth externalizing behavior problems	✓	✓	✓	✓
Delivered PCIT	✓	✓	✓	✓
Included youth ≤ 6 years old	✓	✓	✓	✓
Excluded youth with severe sensory and/or ASD impairments	✓	✓	✓	✓
Collected pre- and post-treatment assessment data	✓	✓	✓	✓
Collected session-level data	✓	✓	✓	✓
Clinical features	Diagnosed with intellectual disability and ODD	Born < 37 weeks gestation	Assigned a principal diagnosis of a DSM-IV ODD, CD, and/or DBD-NOS	Developmental delay defined by prior enrollment in Part C Early Intervention services
Random assignment	Immediate or waitlist-delayed PCIT	Immediate or waitlist-delayed PCIT	Clinic- or videoconferencing-based PCIT	Videoconferencing-based PCIT or referrals to alternative services
All families offered treatment	✓	✓	✓	
Limited CDI phase of treatment	✓	✓		✓
Limited PCIT phase of treatment				✓
Offered videoconferencing-based treatment			✓	✓
Included Spanish-speaking families				✓

Table 3.2. Sociodemographic information for the present study sample.

	M	*SD*
Child age	3.62	0.99
Pre-treatment scores		
CBCL Externalizing T score	65.28	11.42
ECBI intensity raw score	140.71	38.12
DPICS-IV CLP segment sum of dos	4.34	4.42
DPICS-IV CLP segment sum of don'ts	32.26	21.85
	n	%
Child racial/ethnic minority		
Non-Hispanic, White	48	32.7
Racial/ethnic minority	99	67.3
Primary caregiver		
Female	145	94.8
Male	8	5.2
Primary caregiver education		
Completed college	69	46
Did not complete college	81	54
Annual household income		
<$50,000	75	53.2
$50,000-100,000	43	30.5
$100,001-150,000	14	9.9
>$150,000	9	6.4
Treatment format		
Clinic-based	63	41.2
Videoconferencing-based	90	58.8
Treatment completion		
Completed treatment	123	80.4
Terminated early	30	19.6

Note: CBCL = Child Behavior Checklist; ECBI = Eyberg Child Behavior Inventory; DPICS-IV = Dyadic Parent Interaction Coding System, Fourth Edition; CLP = Child-Led Play.

Table 3.3. Caregiver verbalizations and corresponding DPICS-IV coding information.

Code	Category	Definition	Examples
Neutral talk	N/A	A statement that introduces information about people, objects, events, or activities, or indicates attention to the child, but does not clearly describe or evaluate the child's current or immediately completed behavior.	"I'm making my rainbow just like yours." "It is time to clean up." "Ok."
Behavior description	"Do Skills"	A non-evaluative, declarative sentence or phrase in which the subject is the other person and the verb describes that person's ongoing or immediately completed(< 5 sec.) observable verbal or nonverbal behavior.	"You're building a truck." "You and I are making a big fort." "You seem happy that you fixed it."
Reflection		A declarative phrase or statement that has the same meaning as the child's verbalization. The reflection may repeat, paraphrase, or elaborate upon the child's verbalization but may not change the meaning of the child's statement or interpret unstated ideas.	Child: "It's a horsey." Caregiver: "It is a horse." Child: "This game is fun." Caregiver: "You like playing this game." Child: "Give me the car." Caregiver: "You want the car."
Labeled Praise		Provides a positive evaluation of a specific attribute, product, or behavior of the child.	"You did a great job building the tower." "Thank you for handing me the box." "I like the way you drew that circle."
Unlabeled Praise	N/A	Provides a positive evaluation of the child, an attribute of the child, or a nonspecific activity, behavior, or product of the child.	"I like that." "Thank you." "Perfect."
Command	"Don't Skills"	A statement in which the parent directs the behavior of the child. Commands may be direct or indirect in form. Commands include statements directing the child to perform vocal or motor behaviors, as well as mental or internal, unobservable actions (e.g., think, decide).	"Get down." "Put the crayons in the drawer." "Let's use the green piece."
Question		A verbal inquiry from the parent to the child that is distinguishable from declarative statements by having a rising inflection at the end or by having the sentence structure of a question. A Parent Question requests an answer but does not suggest that a behavior is to be performed by the child.	"Where is the dragon?" "Is this your nose?" "Whose turn is it?"
Negative Talk		A verbal expression of disapproval of the child or the child's attributes, activities, products or choices. Negative Talk also includes sassy, sarcastic, rude, or imprudent speech.	"You're working too slowly." "You put it in the wrong place." "Stop that."

Note: definition and examples are obtained from the *Dyadic Parent-Child Interaction Coding System Comprehensive Manual for Research and Training – Fourth Edition (DPICS-IV Manual; Eyberg et al., 2013).*

Table 3.4. Pooled AUC values of measures across sessions.

Session	Measure	AUC	*p*	Expected association with Treatment Nonresponse
Session 3	Out-of-session engagement	0.59	0.09	Lower scores
	Caregiver do skills	0.55	0.35	Lower scores
	Change in caregiver do skills	0.54	0.47	Lower scores
	Caregiver don't skills	0.54	0.46	Higher scores
	Change in caregiver don't skills	0.63*	0.01*	Lower scores
	Child symptoms	0.56	0.29	Higher scores
	Change in child symptoms	0.54	0.53	Lower scores
Session 6	Out-of-session engagement	0.51	0.88	Lower scores
	Caregiver do skills	0.58	0.20	Lower scores
	Change in caregiver do skills	0.57	0.33	Lower scores
	Caregiver don't skills	0.47	0.62	Higher scores
	Change in caregiver don't skills	0.60	0.08	Lower scores
	Child symptoms	0.57	0.17	Higher scores
	Change in child symptoms	0.55	0.39	Lower scores

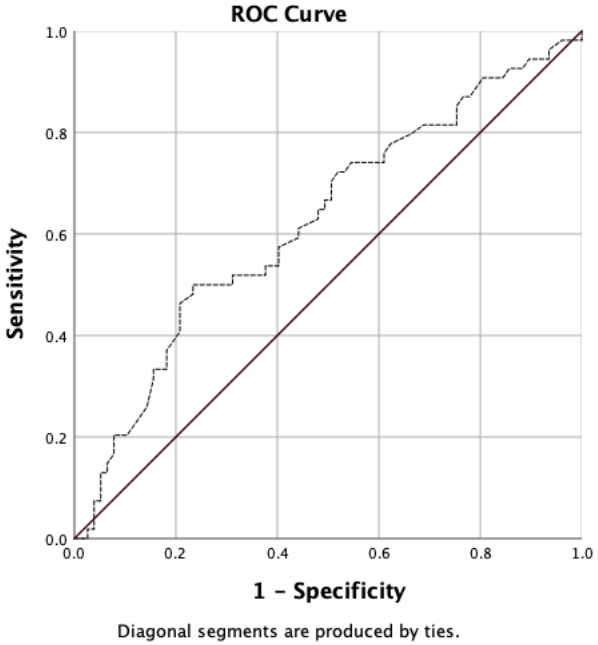
Note: An AUC value of .5 indicates the measure performs no better than chance
*$p < .05$

Table 3.5. Pooled individual classification properties of measure cut scores across sessions.

Measure	Session 3						Session 6					
	Cut score	Sens-itivity	Spec-ificity	PPP	NPP	OCC	Cut score	Sens-itivity	Spec-ificity	PPP	NPP	OCC
Out-of-session engagement	**< 20**	0.13	0.93	0.58	0.58	0.58	**< 20**	0.08	0.93	0.44	0.57	0.56
	< 30	0.23	0.86	0.56	0.59	0.59	**< 30**	0.12	0.89	0.46	0.57	0.56
	< 40	0.34	0.79	0.55	0.61	0.59	**< 40**	0.23	0.78	0.45	0.57	0.55
	< 50	0.45	0.68	0.52	0.62	0.58	**< 50**	0.34	0.69	0.45	0.58	0.54
	< 60	0.58	0.57	0.51	0.64	0.57	**< 60**	0.53	0.48	0.44	0.57	0.50
	< 70	0.68	0.44	0.48	0.64	0.54	**< 70**	0.69	0.31	0.43	0.57	0.47
	< 80	0.84	0.22	0.45	0.64	0.49	**< 80**	0.86	0.13	0.43	0.55	0.45
	< 90	0.94	0.12	0.45	0.74	0.48	**< 90**	0.96	0.03	0.43	0.50	0.43
Caregiver do skills	**< 5**	0.14	0.90	0.51	0.58	0.57	**< 5**	0.07	0.96	0.58	0.57	0.57
	< 10	0.32	0.73	0.48	0.59	0.56	**< 10**	0.15	0.89	0.51	0.58	0.57
	< 15	0.52	0.53	0.46	0.59	0.53	**< 15**	0.31	0.74	0.49	0.59	0.56
	< 20	0.73	0.37	0.47	0.64	0.52	**< 20**	0.52	0.56	0.48	0.60	0.54
	< 25	0.83	0.21	0.45	0.62	0.48	**< 25**	0.71	0.42	0.49	0.66	0.55
	< 30	0.92	0.07	0.43	0.53	0.44	**< 30**	0.85	0.27	0.47	0.70	0.52
Change in caregiver do skills	**< 5**	0.23	0.79	0.45	0.57	0.54	**< 5**	0.11	0.90	0.47	0.57	0.56
	< 10	0.42	0.59	0.44	0.57	0.52	**< 10**	0.24	0.76	0.44	0.57	0.53
	< 15	0.59	0.48	0.47	0.61	0.53	**< 15**	0.45	0.61	0.47	0.59	0.54
	< 20	0.74	0.32	0.46	0.62	0.50	**< 20**	0.59	0.49	0.47	0.61	0.53
	< 25	0.80	0.19	0.43	0.55	0.46	**< 25**	0.73	0.39	0.48	0.65	0.53
	< 30	0.83	0.17	0.43	0.57	0.46	**< 30**	0.80	0.26	0.45	0.62	0.49
Caregiver don't skills	**> 5**	0.80	0.29	0.46	0.65	0.51	**> 5**	0.56	0.43	0.43	0.56	0.49
	> 10	0.52	0.53	0.46	0.59	0.53	**> 10**	0.28	0.62	0.36	0.53	0.47
	> 15	0.37	0.70	0.48	0.59	0.56	**> 15**	0.18	0.81	0.42	0.56	0.54
	> 20	0.20	0.80	0.43	0.57	0.54	**> 20**	0.10	0.92	0.48	0.57	0.56
	> 25	0.12	0.85	0.38	0.56	0.54	**> 25**	0.05	0.97	0.59	0.57	0.57

	> 30	0.09	0.89	0.39	0.56	0.55	**> 30**	0.04	0.98	0.63	0.57	0.57
Change in caregiver don't skills	**< 0**	0.21	0.86	0.54	0.59	0.58	**< 0**	0.17	0.91	0.60	0.59	0.59
	< 5	0.32	0.83	0.60	0.62	0.61	**< 5**	0.25	0.85	0.57	0.60	0.59
	< 10	0.44	0.73	0.56	0.63	0.61	**< 10**	0.32	0.75	0.50	0.59	0.57
	< 15	0.49	0.65	0.52	0.63	0.58	**< 15**	0.38	0.70	0.49	0.59	0.56
	< 20	0.56	0.57	0.50	0.63	0.57	**< 20**	0.43	0.62	0.47	0.59	0.54
	< 25	0.64	0.54	0.51	0.66	0.58	**< 25**	0.53	0.55	0.47	0.60	0.54
	< 30	0.68	0.47	0.50	0.66	0.56	**< 30**	0.63	0.48	0.48	0.63	0.54
	< 35	0.72	0.39	0.48	0.65	0.54	**< 35**	0.71	0.43	0.49	0.65	0.55
	< 40	0.75	0.34	0.47	0.65	0.52	**< 40**	0.73	0.37	0.47	0.64	0.53
	< 45	0.78	0.28	0.45	0.62	0.50	**< 45**	0.76	0.33	0.46	0.64	0.51
Child symptoms	**> 80**	0.91	0.08	0.43	0.54	0.44	**> 80**	0.78	0.25	0.44	0.60	0.48
	> 90	0.83	0.14	0.42	0.51	0.44	**> 90**	0.71	0.32	0.44	0.59	0.49
	> 100	0.76	0.21	0.43	0.54	0.45	**> 100**	0.63	0.47	0.48	0.62	0.54
	> 110	0.70	0.34	0.45	0.60	0.50	**> 110**	0.56	0.53	0.48	0.61	0.54
	> 120	0.60	0.44	0.45	0.59	0.51	**> 120**	0.48	0.62	0.49	0.61	0.56
	> 130	0.52	0.57	0.48	0.61	0.55	**> 130**	0.40	0.66	0.48	0.59	0.55
	> 140	0.43	0.65	0.49	0.60	0.56	**> 140**	0.35	0.81	0.58	0.62	0.61
	> 150	0.35	0.78	0.54	0.61	0.59	**> 150**	0.28	0.85	0.58	0.61	0.60
Change in child symptoms	**< 0**	0.37	0.70	0.48	0.59	0.56	**< 0**	0.28	0.79	0.5	0.59	0.57
	< 10	0.53	0.52	0.46	0.59	0.52	**< 10**	0.37	0.67	0.46	0.58	0.54
	< 20	0.65	0.33	0.42	0.55	0.47	**< 20**	0.49	0.61	0.49	0.61	0.56
	< 30	0.79	0.22	0.44	0.58	0.47	**< 30**	0.63	0.50	0.49	0.64	0.55
	< 40	0.83	0.17	0.44	0.57	0.46	**< 40**	0.66	0.40	0.46	0.61	0.52
	< 50	0.89	0.14	0.44	0.63	0.47	**< 50**	0.73	0.28	0.44	0.58	0.48
	< 60	0.90	0.05	0.52	0.40	0.42	**< 60**	0.82	0.17	0.43	0.55	0.45

Figure 3.1. ROC curve depicting sensitivities by 1 – specificities for each possible value of change in caregiver don't skills between pre-treatment and session 3.



Note: ROC curve depicts data representing the first imputed dataset. Consistent with the pooled ROC analysis, *AUC*=.63, *p*<.05.

CHAPTER IV

To stay-the-course or course-correct? Using caregiver skill acquisition and child behavior problems early in treatment to predict child response trajectories in the latter part of a behavioral parenting intervention

Abstract

Parent-Child Interaction Therapy (PCIT) utilizes routine outcome monitoring (ROM) to monitor and inform the progression of treatment sequences, but research evaluating PCIT has not yet confirmed the extent to which scores on specific ROM measures (e.g., tailoring variables) assessed at specific time points in treatment (e.g., critical decision points) can predict treatment response. Building on previous work showing that narrow definitions of early interim treatment progress and/or treatment outcomes may not capture nuanced complexities of how treatment effects unfold across time, the present study adopted a continuous approach to measuring both early interim treatment progress and later treatment response. The effects of child behavior problems and caregiver skills (and their interactions) on trajectories of child externalizing symptom change were examined using multilevel growth modeling in a sample of (*N*=70) families receiving Internet-delivered PCIT for their young child with developmental delay and elevated behavior problems. Multilevel growth models showed that, on average, child externalizing symptoms improved across treatment. However, rates of child externalizing symptom improvement across treatment varied based on the interaction between levels of child behavior problems and caregiver skills at both session 3 and session 6 (examined in separate, session-specific models). Results show that despite statistical significance, differences in child externalizing symptom trajectories are subtle at session 3, and they become far more pronounced by session 6. Findings suggest families who display a combination of high child behavior problems and low caregiver skills by the sixth session of PCIT may not witness further improvements in child symptoms during the latter part of treatment in the absence of course-correction.

Keywords: behavioral parenting intervention; caregiver skill acquisition; child behavior problems; routine outcome monitoring; symptom trajectories.

Introduction

Behavioral parenting interventions (BPIs) have been considered the firstline approach to treating behavior problems in young children (Comer et al., 2013; Eyberg et al., 2008). BPIs aim to reduce coercive family interaction cycles characterized by child externalizing behavior problems and negative parenting practices by reshaping parent-child interactions to increase parental consistency and effectiveness and improve the overall warmth and positivity of family relationships (Patterson et al., 1998). By guiding caregivers to positively reinforce appropriate child behavior, remove attention from misbehavior, and provide predictable responses (i.e., consequences) to noncompliance, BPIs foster positive parent-child interactions, reduced parenting stress, and improved child behavior (Comer et al., 2013; Eyberg et al., 2008; Kaminski & Claussen, 2017).

Parent-Child Interaction Therapy (PCIT) is one BPI that has demonstrated particularly positive and large effects across a wide range of child populations (Lyon & Budd, 2010; Thomas et al., 2017; Thomas & Zimmer-Gembeck, 2007). PCIT has been broadly disseminated across many clinical, community, and service-delivery formats (e.g., Internet-delivered PCIT [Comer et al., 2017]; intensive PCIT [Graziano et al., 2020]; group-based PCIT [Niec et al., 2016]), and has been adapted to treat a range of populations (e.g., Bagner et al., 2007; Bagner et al., 2010; Comer et al., 2021; Lenze et al., 2011). Standard PCIT is comprised of two phases of treatment—Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI)—during which caregivers are taught two complementary sets of skills that help to form warm, mutually reinforcing, positive parent-child relationships and consistent follow through with effective discipline strategies (e.g., time out) that improve child compliance and decrease disruptive behavior problems over time (Eyberg & Funderburk, 2011).

Despite substantial overall empirical support for PCIT in the treatment of child behavior problems (e.g., Thomas et al., 2017), there is wide variability in treatment engagement and response among PCIT-treated families. As with all BPIs (Chacko et al., 2016; Chacko et al., 2017), a large body of work on PCIT has documented high attrition rates, variable patterns of engagement in treatment (e.g., session attendance, completion of out-of-session homework assignments), differences in caregiver skill acquisition and child symptom reductions, and inconsistencies in overall improvements in the quality of parent-child interactions and family functioning (Bagner & Graziano, 2013; Boggs et al., 2004; Danko et al., 2016; Stokes et al., 2016; Thomas & Zimmer-Gembeck, 2007; Timmer et al., 2016; Werba et al., 2006). Such heterogeneity in treatment effects is concerning and very little is known about how to accurately predict which families are least likely to benefit from PCIT nor how to appropriately adapt treatment for those at risk for treatment nonresponse (Hong et al., 2019; see Chapters II and III).

Research investigating predictors of differential patterns of PCIT treatment response has largely focused on relatively stable child, caregiver, and/or family-level characteristics measured *prior* to treatment. The broader literature on BPI engagement and response is mixed, but these studies collectively suggest that several key caregiver and/or family-level factors (e.g., family income, caregiver education, family structure, racial/ethnic minority status, caregiver psychopathology) confer risk for worse treatment outcomes (Bagner & Graziano, 2013; Lundahl et al., 2006; Reyno & McGrath, 2006). Analyses examining relations between *pre*-treatment characteristics and treatment response can aid in initial treatment selection decisions (see Cohen & DeRubreis, 2018), but are less readily applicable *within* the course of treatment, during which clinicians are faced with ongoing decisions about how to best calibrate treatment that is already

underway. In the absence of data, clinicians often proceed through treatment based on clinical instincts, rather than actuarial judgment that may more precisely position treatment for success (Perlis, 2016).

The use of routine outcome monitoring (ROM) to examine and develop data-driven, feedback-informed care that can be flexibly tailored to dynamic and evolving patient characteristics throughout the course of treatment is central to the development of precision mental health care (Almirall & Chronis-Tuscano, 2016; Delgadillo & Lutz, 2020; Hong et al., 2019; Lambert et al., 2018; Lutz et al., 2020). Such personalized approaches to mental health service delivery hold tremendous promise for improving the efficiency and effectiveness of mental health services, but have often exclusively focused on whether individuals display early patterns of change on the desired ultimate outcome (e.g., symptom remission) as the primary indicators of interim treatment response (Hong et al., 2019; Southward & Sauer-Zavala, 2020). This narrow focus on symptoms and related impairment as the sole metric of early treatment response fails to incorporate an experimental therapeutics framework, which highlights how mechanisms of treatment response must first be engaged and altered themselves, prior to subsequently leading to changes in the outcome (Hong et al., 2019).

Thus, when investigating *early* indicators of PCIT outcomes, one should consider theory regarding proposed mechanistic processes through which child improvements unfold in order to assess candidate tailoring variables that can inform whether a clinician should stay-the-course or course-correct for an individual family. In PCIT, both theoretical and empirical support suggest that although early, observable improvements in child behavior are desirable, the target mechanism through which PCIT intends to effect ultimate change in child behavior is caregiver behavioral change (Forehand et al., 2014). Nonetheless, limited work has examined the predictive utility of observed

parenting behaviors during early treatment sessions (e.g., within the first phase of treatment) on subsequent improvements in child behavior (e.g., during the latter part of treatment). Such signals of treatment nonresponse are critical for informing clinical care and ongoing clinical decision-making, but have been rarely linked to specific measures (e.g., tailoring variables) or time points (e.g., critical decision points) *during* treatment. Data-informed guidelines specifying how to determine whether or not a family is likely "on track" for future progress in PCIT are needed.

Work conducted by Hong and colleagues (2019; see also Chapters II and III) suggests that leveraging secondary data analysis can afford unique opportunities to examine the predictive clinical utility of ROM measures collected early in treatment on post-treatment outcomes. In one study (see Chapter III), a pooled ROC analysis of 153 families engaged in various PCIT-based treatments was conducted and examined classification indices corresponding to early treatment progress data (i.e., defined dichotomously at various cutscores representing "unsatisfactory" levels measures of out-of-session engagement, caregiver skill acquisition, and child symptom improvements by the third or sixth session of PCIT) could differentially predict post-treatment outcomes (defined dichotomously as "treatment response" or "treatment nonresponse"). Across both early treatment session timepoints, none of the examined candidate tailoring variables demonstrated acceptable discriminative value in predicting post-treatment outcomes (see Chapter III). Such findings may suggest that early treatment engagement, caregiver skill acquisition, and child symptom improvements are not significant predictors of ultimate BPI response, but it is also plausible that the dichotomous approaches used to define satisfactory levels of early success and ultimate treatment response in this analysis may have failed to reflect the more continuous nature and dynamic relations among these treatment response variables as they naturally

present. Moreover, the sole focus on main effects in this analysis may have failed to detect important interactions among early interim treatment response predictors that together predict whether treatment is on a favorable trajectory.

Recent work emphasizes how narrow definitions and singular measures of early treatment progress and/or treatment outcome likely do not capture the nuanced complexities of how treatment effects unfold across time and, as such, have encouraged the use of multivariable prediction models to enhance the capacity of research efforts to inform precision mental health care (Cohen & DeRubeis, 2018). Leveraging archival data analysis to examine more continuous approaches to measuring interim PCIT progress, as well as whether interactions among independent variables can predict child symptom trajectories, may offer a more dynamic and accurate model through which to understand BPI response trajectories.

Present Study

The present study examined trajectories of child externalizing symptoms from pre- to post-treatment among families receiving Internet-delivered PCIT (iPCIT) for their young with developmental delay and elevated behavior problems, and how measures of early interim treatment progress may differentially predict variations across children in the rate of child externalizing symptom change. These aims were accomplished via secondary data analysis of a sample of families who were randomly assigned to receive a course of iPCIT as part of their participation in a RCT evaluating the efficacy of iPCIT for young children with developmental delay (Bagner et al., in preparation). Examining predictors of BPI response trajectories specifically among families of children with developmental delay is important given that youth with developmental delay are at a heightened risk for displaying clinically significant behavior problems in early childhood and their caregivers experience higher levels of parenting stress that appear to be driven

by levels of child behavior problems, as opposed to developmental delay (Baker et al., 2002; Baker et al., 2003). Additionally, prior research shows that PCIT can be effective for young children with or at risk for developmental delay (e.g., Bagner et al., 2007; Bagner et al., 2010), but has not examined more nuanced and dynamic patterns of improvement among treated families.

It was hypothesized that: (a) across treated families, child externalizing symptoms would significantly decrease in a linear fashion from pre- to post-treatment; (b) child behavior problems and caregiver skills by session 3 and by session 6 (treated continuously in the models) would each moderate trajectories of child externalizing symptoms across treatment, such that fewer behavior problems and greater caregiver skills would predict steeper reductions in child externalizing symptoms across the latter part of treatment, whereas more limited early progress across these domains would predict more gradual rates of child externalizing child symptom change across the latter part of treatment; and (c) early interim progress assessed at session 6 would better predict later child symptom trajectories than would early interim progress assessed at session 3, given previous work suggesting caregiver skill acquisition and child behavior change is not reliably expected prior to the fourth session (see Lieneman et al., 2019), and other examinations showing improved accuracy predicting final treatment outcomes after four or more sessions (see Delgadillo et al., 2014). Exploratory analyses examined (d) whether the *interaction* between early child behavior problems and early caregiver skill acquisition further improves upon the prediction of child externalizing symptom trajectories across the latter part of treatment—i.e., does the utility of assessing early child behavior problems in predicting subsequent treatment response trajectories vary as a function of how well caregivers are acquiring skills in early phases of treatment?

Method

Participants

Participants were drawn from a RCT examining the efficacy of iPCIT for treating young children with developmental delay and co-occurring externalizing symptoms (Bagner et al., in preparation). Families were eligible to participate in this study if: (a) their child was aging out of Part C Early Intervention services for child developmental delay (mean age = 34.5 months), (b) their child had elevated externalizing symptoms (i.e., CBCL Externalizing Problems *T* scores $\geq$ 60), (c) the primary caregiver was willing to participate in treatment, and (d) the primary caregiver and child spoke either English or Spanish. Families were excluded from the study if: (a) the child was receiving an unstable dose of medication (i.e., changes within the past 4 weeks) to manage behavior problems; (b) the child had a history of severe physical impairment (e.g., deafness, blindness); (c) the child demonstrated severe autism spectrum disorder impairment (i.e., defined by a Social Responsiveness Scale, Second Edition [SRS-2; Constantino & Gruber, 2012] score of > 75); or (d) the primary caregiver's estimated IQ score was < 70 on a two-subtest assessment of the English or Spanish versions of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011/EIWA-III; Pons et al., 2008). Families in this RCT were randomly assigned to either iPCIT (*N*=75) or Referrals as Usual (RAU; *N*=75; in which families were provided with usual referral options offered by their Early Intervention program—e.g., child-focused therapy, family-focused therapy, parenting-focused therapy, exceptional student education classroom, psychiatrist referral; Comer et al., in preparation). Participants from the RCT were included in the present analyses if they (a) were randomized to the iPCIT condition, and (b) began treatment (i.e., attended $\geq$ 1 treatment session), resulting in the present sample of *N*=70 families. Table 4.1 presents sociodemographic information for the *N*=70 families that

made up this sample. The families included in the present analyses did not significantly differ from the other 80 families included in the RCT on pre-treatment sociodemographic or clinical variables.

Procedures

All study procedures were approved by the Florida International University Institutional Review Board, and all primary caregivers provided informed consent prior to study participation. The present analyses used data from study assessments collected at three major timepoints (i.e., pre-treatment [time 0], mid-treatment [time 1], post-treatment [time 2]; see Table 4.3). Caregiver-report questionnaires at all three timepoints were collected via REDCap, a secure online survey platform. In addition, a study assistant masked to families' study condition assignment traveled to each participating family's home at pre-treatment and post-treatment to conduct structured family observations, among other study measures. Following completion of pre-treatment assessments, families were randomly assigned to either receive treatment (i.e., iPCIT) or RAU. Families randomly assigned to receive iPCIT were offered treatment for up to 20 weeks.

Treatment

For the present study, iPCIT (Comer et al., 2017) was provided as a fully remote treatment, led by therapists in real-time using webcams and secure videoconferencing technology. Families logged into sessions from their homes while their remote therapist logged into sessions from the clinic. As in traditional PCIT (Eyberg & Funderburk, 2011), iPCIT consisted of two phases—Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI). The CDI phase began with an initial caregiver-only CDI "Teach" session, in which caregivers were oriented to the program, provided with psychoeducation about child behavior problems and patterns of positive and negative reinforcement, and were taught specific positive attending skills to use with their child.

These positive attending skills included a set of CDI “do” skills to use during parent-child interactions (described in further detail in the Measures section) and a set of parent-child interaction patterns to avoid with their child during CDI. The CDI “Teach” session was followed by up to 5 CDI “Coach” sessions attended by caregiver and children, in which the therapist observed and provided “bug-in-the-ear” coaching to caregivers on the use of the CDI skills with their child via a caregiver-worn earpiece (e.g., Bluetooth headset; see Comer et al., 2014). Next, the PDI phase began with an initial caregiver-only PDI “Teach” session, in which caregivers were taught specific structured discipline strategies and time out sequences for child noncompliance. A series of PDI “Coach” sessions attended by caregivers and children then had the therapist observe and provide “bug-in-the-ear” coaching to caregivers on the use of PDI skills with their child (see Comer et al., 2014 and Comer et al., 2017 for more details on iPCIT).

Measures

Child Behavior Checklist for Ages 1.5-5

The Child Behavior Checklist for Ages 1.5-5 (CBCL; Achenbach & Rescorla, 2000) was used to measure *child externalizing symptoms* at pre-treatment, mid-treatment, and post-treatment study timepoints. The CBCL is a 99-item caregiver-report questionnaire assessing a range of emotional and behavioral problems in young children between the ages of 1.5 and 5 years. Items are rated on a 3-point Likert-type scale from 0 (not true) to 2 (very true or often true). The CBCL yields *T* Scores normed for age and sex that correspond to variety of domains, each with a mean of 50 and standard deviation of 10; higher scores indicate greater symptomatology. It is one of the most commonly used measures of child psychopathology and has demonstrated very strong psychometric properties (Achenbach & Rescorla, 2000). The CBCL Externalizing

Problems scale was used as a repeated outcome measure within the present study (α=.95 in the present sample).

Eyberg Child Behavior Inventory

The Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999) was used to measure *child behavior problems* at pre-treatment and on a session-by-session basis throughout treatment. The ECBI is a 36-item caregiver-report questionnaire assessing the presence of child behavior problems. Items are rated on a 7-point Likert-type scale from 1 (never) to 7 (always) and summed to provide an Intensity scale score that ranges from 36 to 252; higher scores indicate greater behavior problems. The ECBI has shown strong psychometric properties in preschoolers (Funderburk et al., 2003; α=.92 in the present sample). ECBI Intensity scale scores collected session 3 and session 6 were used in the present study.

Dyadic Parent-Child Interaction Coding System

The Dyadic Parent-Child Interaction Coding System – 4th Edition (DPICS-IV; Eyberg et al., 2013) was used to measure *caregiver skills* at pre-treatment and during most treatment sessions. The DPICS-IV is a behavioral observation coding system designed to examine the quality of parent-child interactions. The standard DPICS-IV assessment includes a 5-minute “Child-Led Play” segment in which caregivers are instructed to allow the child to choose an activity and to follow the child’s lead in play. Caregiver verbalizations are coded as “neutral talk,” “behavior descriptions,” “reflections,” “labeled praises,” “unlabeled praises,” “commands,” “questions,” and “negative talk” (see Table 4.2 for descriptions and examples of each of these types of coded verbalizations). In the Child-Led Play segment, the tally of behavior descriptions, reflections, and labeled praises (three forms of verbal positive reinforcement PCIT teaches caregivers to use during parent-child interactions) comprise a total “do” skills

score. Across sessions, higher do skill scores during the 5-minute Child-Led Play interaction represent greater caregiver skill acquisition. A complete course of standardized PCIT requires caregivers to meet specific skill criteria in do skills (i.e., $\geq$ 10 behavior descriptions, $\geq$ reflections, and $\geq$ labeled praises; summing to at least 30 total do skills) during the 5-minute Child-Led Play coding segment at the start of session prior to advancing to the second phase of treatment. Caregiver do skills tallied during the Child-Led Play segment of the DPICS-IV administered at session 3 and session 6 were used in the present study. Coders were trained to 80% agreement with a criterion tape prior to coding the DPICS-IV for study purposes.

Analytic Plan

All analyses were conducted using the Statistical Package for the Social Sciences (SPSS) Statistics (Version 26). Given the nested nature of repeated measures (level-1) within individuals (level-2), multilevel growth models were used to examine change in child externalizing symptoms (i.e., CBCL Externalizing Problems *T* Scores) throughout treatment. Consistent with other studies modeling the effects of ROM measures and treatment outcomes across time (e.g., Malins et al., 2020), models included observed data only and did not impute missing data, though maximum-likelihood estimation was employed to address variability in the number of repeated outcome measures (i.e., CBCL Externalizing Problems subscales) collected across families and to facilitate comparisons of nested models differing in fixed effects through χ^2 tests of model deviance statistics (i.e., -2 log likelihood). Non-nested models were compared by examining each model's Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), for which lower values indicate better model fit statistics (Singer & Willett, 2003). As suggested by Peugh (2010), in response to non-

convergence and/or non-significance of effects, simplified models retained random intercepts, but not random slopes, for all models.

First, to calculate the intraclass correlation coefficient, *Model A* examined an unconditional model (i.e., including only a random intercept) to obtain residual within-individual and between-individual variances among child externalizing symptoms throughout treatment. Next, to examine the shape of change in child externalizing symptoms from pre- to post-treatment, *Model B* examined—at level-1 (across repeated measures; i.e., 3 timepoints)—the relationship between time (i.e., months in treatment) and child externalizing symptoms (i.e., do child externalizing symptoms decline in a linear fashion throughout the course of treatment?), as well as—at level-2 (between-individuals; i.e., *N*=70 families)—variation in initial scores (i.e., intercepts) and rates of change (i.e., slopes) across families. An unstructured covariance structure was specified. The combined linear growth model equation is: $CBCL_{ti} = \beta_{00} + \beta_{10}(months_{ti}) + u_{0i} + r_{ti}$.

Then, to evaluate the moderating effects of session 3 interim progress data on the rate of change in child externalizing symptoms throughout treatment, *Model C* built upon Model B by adding in two level-2 predictors—"session 3 child behavior problems" and "session 3 caregiver skills"—as well as the two- and three-way cross-level interactions between these 3 predictors. Finally, to examine the effects of session 6 interim progress data on the rate of change in child externalizing symptoms throughout treatment, *Model D* replaced the session 3 predictors included in Model C with the respective data from session 6. The combined equation for these two models is: $CBCL_{ti} = \beta_{00} + \beta_{01}(\text{session ECBI}_i) + \beta_{02}(\text{session DPICS}_i) + \beta_{03}(\text{session ECBI}_i \times \text{session DPICS}_i) + \beta_{10}(months_{ti}) + \beta_{11}(months_{ti} \times \text{session ECBI}_i) + \beta_{12}(months_{ti} \times \text{session DPICS}_i) + \beta_{13}(months_{ti} \times \text{session ECBI}_i \times \text{session DPICS}_i) + u_{0i} + r_{ti}$.

To ensure level-1, level-2, and cross-level effects were interpretable, meaningful zero values captured within the dataset were defined for all predictors across all models. Specifically, the time associated with each repeated outcome measure was defined as a continuous variable representing the number of months since the pre-treatment assessment, with values of 0 representing the pre-treatment assessment for each family. Child behavior problems (via the ECBI) and caregiver skills (via the DPICS-IV) were each grand-mean centered within timepoints and defined as continuous variables, with values of 0 reflecting the average score across individuals at each session. Thus, higher values on time reflected a greater number of months in treatment, higher values on child behavior problems reflected worse than the sample average child behavior problems at each specific timepoint, and higher values on caregiver skills reflected better than the sample average for caregiver skills at each specific timepoint.

Results

Preliminary Findings

Table 4.3 presents descriptive statistics on the present study data. Table 4.4 displays correlations between independent and dependent variables, showing that several correlations were significant, but none exceeded .70, easing concerns about multicollinearity (Tabachnick & Fidell, 2001). Table 4.5 presents the table of nested model coefficients for Models A through D. The initial unconditional model (*Model A*) indicated that 51% of the variance in child externalizing symptoms was accounted for by clustering (i.e., repeated measures within individuals; intraclass correlation coefficient [ICC]=.51).

Overall Child Externalizing Symptom Trajectory Across Treatment

Model B examined the shape of change over the course of treatment in child externalizing symptoms across the sample. The likelihood ratio function comparing

Model B to *Model A* indicated that *Model B* fit the data better ($\chi^2_{LR}(2)=38.70$, $p < .001$). As shown in Table 4.5, *Model B* supported a significant linear effect of time in treatment on child externalizing symptoms, with child externalizing symptoms improving, on average, over time. Figure 4.1 displays this average trajectory of child externalizing symptoms throughout treatment across the sample. Families displayed significant within-individual variation, as well as significant variation between-individuals in pre-treatment scores.

Does Interim Progress at Session 3 Predict Child Externalizing Symptom Trajectories?

Based on best practice recommendations for estimating cross-level interaction effects using multilevel modeling, *Model C* proceeded to test hypotheses regarding the roles of child behavior problems and caregiver skills measured at session 3 (and their interaction) on rates of child in child externalizing symptoms change across treatment, despite the absence of a significant random slope (Aguinis et al., 2013). The likelihood ratio function comparing *Model C* to *Model B* indicated that *Model C* fit the data better ($\chi^2_{LR}(6)=480.85$, $p < .001$). As shown in Table 4.5, *Model C* also reduced the residual and intercept variance relative to *Model B*, indicating improved prediction of child externalizing symptom trajectories. As in *Model B*, *Model C* again found that child externalizing symptoms significantly decreased in a linear fashion across treatment, with significant variation within-individual scores and between-individuals in pre-treatment scores. Additionally, session 3 child behavior problems demonstrated a significant main effect in the prediction of child externalizing symptoms and there was a significant three-way interaction between time, session 3 child behavior problems, and session 3 caregiver skills. The results of *Model C* indicated that rates of child externalizing

symptom improvement across the full course of treatment varied based on the interaction between levels of child behavior problems and caregiver skills at session 3.

To aid in the interpretation of this interaction, regression lines representing the average trajectory of child externalizing symptoms throughout treatment were plotted for families whose children displayed “high” (1 SD above the mean) versus “low” (1 SD below the mean) session 3 child behavior problems crossed with “high” (1 SD above the mean) versus “low” (1 SD below the mean) session 3 caregiver skills. As depicted in Figure 4.2, the effect of caregiver skill acquisition was not uniform across levels of child behavior problems, nor was the effect of child behavior problems uniform across levels of caregiver skills. These differing trajectories over time showed that when *both* children *and* their caregivers demonstrated unfavorable features at session 3 (i.e., the solid black line depicting those with high ECBIs and low skills), children showed less favorable patterns of externalizing symptom improvement throughout the latter part of treatment. Interestingly, a similar pattern of minimal further improvement was displayed when *both* children *and* their caregivers demonstrated *favorable* features at session 3 (i.e., the solid gray line depicting those with low ECBIs and high skills). In contrast, when families displayed mixed presentations at session 3 (i.e., the dashed/dotted lines), reflecting “room for improvement” in *either* child *or* caregiver domains (i.e., via future reduction of high ECBIs or future increases in low skills), children displayed steeper reductions in child externalizing symptoms across the latter part of treatment.

Does Interim Progress at Session 6 Predict Child Externalizing Symptom Trajectories?

Model D replaced interim progress data collected at session 3 with slightly later interim progress data, examining the roles of child behavior problems and caregiver skills measured at *session 6*, and their interaction, on rates of child externalizing

symptom change across treatment. The likelihood ratio function comparing *Model D* to *Model B* indicated that *Model D* fit the data better ($\chi^2_{LR}(6)=529.403$, $p < .001$). As shown in Table 4.5, *Model D* also reduced the residual and intercept variance relative to *Model B*, indicating improvements in predicting trajectories. Moreover, comparing the model fit statistics between Models C and D indicated that *Model D* displayed better goodness of fit (i.e., lower AIC and BIC values), indicating the model including sets of predictors from session 6 performed better than the model including sets of predictors of session 3. Overall, *Model D* displayed the same pattern of significant effects as *Model C*. Specifically, *Model D* continued to find that child externalizing symptoms significantly decreased in a linear fashion over time, with significant variation within-individual scores and between-individuals in pre-treatment scores. Additionally, similar to the effects found at session 3 (*Model C*), session 6 child behavior problems had a significant effect in the prediction of child externalizing symptom change, and there was a significant three-way interaction between time, session 6 child behavior problems, and session 6 caregiver skills. Thus, the results of *Model D* indicated that the rate of child externalizing symptom improvement across treatment varied based on the interaction between child behavior problems and caregiver skills at session 6.

To aid in the interpretation of this interaction, regression lines representing the average trajectory of child externalizing symptoms throughout treatment were again plotted for four groups at “high” and “low” values of child behavior problems and caregiver skills, this time based on the means and SDs of the session 6 data. As shown in Figure 4.3, the effect of caregiver skill acquisition was not uniform across levels of child behavior problems, nor was the effect of child behavior problems uniform across levels of caregiver skills. These differing trajectories over time showed a similar, but more pronounced, pattern of effects relative to session 3. Specifically, when *both*

children *and* their caregivers demonstrated unfavorable features at session 6 (i.e., the solid black line depicting those with high ECBIs and low skills), children were unlikely to demonstrate any subsequent improvement in child externalizing symptoms during the latter part of treatment. Again, when *both* children *and* their caregivers demonstrated *favorable* features at session 6 (i.e., the solid gray line depicting those with low ECBIs and high skills), children were also unlikely to demonstrate any subsequent improvements in externalizing symptoms during the latter part of treatment. In contrast, when families displayed mixed presentations at session 6 (i.e., the dashed/dotted lines), reflecting "room for improvement" in *either* child *or* caregiver domains (i.e., via future reduction of high ECBIs or future increases in low skills), they did demonstrate further reductions in child externalizing symptoms during the latter part of treatment. Comparing Figures 2 and 3, one can see that by session 6, the difference in trajectories is more clear, suggesting the risk for minimal improvement during the latter part of treatment becomes increasingly distinct by session 6 relative to session 3.

Discussion

Building on calls for precision medicine and the use of interim response data to inform subsequent courses of treatment for individuals (e.g., Hong et al., 2019; Pettit et al., 2016; Southward & Sauer-Zavala, 2020), the present study provides initial support that early BPI progress indicators across children and caregivers can together inform whether continued BPI is likely to yield subsequent child improvements. On average, young children with developmental delay and elevated behavior problems treated with iPCIT demonstrated significant linear improvements in child externalizing symptoms across treatment. However, rates of change in child externalizing symptoms over time varied based on an interaction between levels of child behavior problems and caregiver skills measured early in treatment.

Findings suggest that poor caregiver skill acquisition combined with persistent child behavior problems by session 6 is a significant predictor that a child's externalizing problems will not show subsequent declination if the clinician simply "stays-the-course." For these families, it appears as though the lack of early treatment progress following initial treatment sessions on both child *and* caregiver fronts signals that the child will not improve without some sort of treatment course-correction. Similarly, relatively high caregiver skill acquisition combined with relatively few child behavior problems by session 6 also predicts that a child's externalizing symptoms will not improve further without some sort of treatment course-correction. For these families, it appears as though "staying-the-course" in treatment may be misguided, as they display limited room for further improvement. In contrast, the present findings suggest that when families demonstrate *either* relatively low caregiver skills *or* relatively high child behavior problems by session 6 (but not both), staying-the-course in treatment is recommended, as it is likely the child will continue to display improvements in externalizing symptoms over time. In the context of BPIs that focus on dismantling coercive parent-child interactions and emphasize the reciprocal relationship between child behavior problems and negative parenting practices, considering the *combination* of child and caregiver features in the assessment of interim treatment progress appears to be not only logical, but necessary.

Although previous studies have suggested that detecting differential treatment outcomes may be difficult and/or unlikely within the first four sessions (e.g., Delgadillo et al., 2014; Lieneman et al., 2019), the present study showed that *combining* ROM data collected as early as session 3 could significantly predict subsequent child response trajectories. However, although patterns of significance at both session 3 and session 6 were similar, graphical depictions of the interactions show that by session 6 the

difference in subsequent symptom response trajectories becomes far more pronounced than those detectable at session 3. Therefore, whereas data collected by session 3 may provide some warning signs of minimal future improvement, the present results suggest one should wait until session 6 to make decisions about course-correcting care (e.g., augmenting, intensifying, or switching treatment).

Consistent with other work examining single variable predictors of treatment response (e.g., see Chapter III; Cohen & DeRubreis, 2018), neither child behavior problems nor caregiver skills *independently* moderated symptom response trajectories during the first phase of treatment. Thus, the present study findings provide additional support for the use of multivariable predictors in prognostic research, particularly when examining treatment nonresponse. Consistent with recent discussion regarding prognostic research and precision medicine (e.g., Cohen & DeRubreis, 2018; Gillan & Whelan, 2017), the present results suggest that it may not be possible or worthwhile to consider individual factors in isolation from one another when aiming to tailor treatment to individual needs, but that incorporating multiple measures that assess theorized mechanisms of treatment-related change throughout the first phase of treatment may provide a more clear picture of for whom treatment is unlikely to lead to further improvements during the latter part of treatment.

Specifically, in clinical practice, cursory looks early in treatment at child symptom scores or caregiver skill acquisition *independently* may lead a clinician to believe that treatment is not “working” (if child symptoms are still high or caregiver skills are still low) and that the family is therefore unlikely to benefit from further treatment. However, this interpretation, and related inclinations to course-correct, are not supported by the present findings. This study suggests that examining child or caregiver features in isolation from one another can be misleading and that it is only when considering *both*

caregiver and child interim treatment progress factors together that one can develop a clear understanding of whether staying-the-course is ill-advised.

Notably, the ROM measures examined in this study are naturally embedded within the PCIT treatment protocol, facilitating the potential for replication among other studies involving families participating in PCIT-based services. Additional work investigating how ranges of scores represented within this study correspond to other samples may also inform clinical decision-making for youth whose families participate in PCIT. However, among other BPIs, there may be tension in the relative trade-offs between adding the administration of ROM approaches that afford the opportunity to capture a more comprehensive portrait of interim treatment response and better inform clinical decisions with the time, effort, and resources it requires to do so. Indeed, research shows wide variability among clinicians in their receptiveness and use of ROM and feedback-informed care (MacKrill & Sorenson, 2020), indicating that additional work focused on how to better align ROM practices with clinician and family preferences is still needed.

Although the present findings are encouraging, study limitations warrant comment. First, despite statistically significant results, the differences in trajectories observed among families are relatively small. Although families were drawn from an RCT in which families were randomly assigned to a RAU condition, the present analyses focused on examining the effects of ROM collected during early treatment sessions to predict symptom response trajectories (as opposed to examining between condition differences). The emphasis on early treatment ROM did not allow for between group comparisons, as families randomized to the RAU condition did not complete weekly assessments. Future research should examine how evolving patient characteristics inform differences in symptom change trajectories between families who are and are not

engaged in treatment to better understand the causal relationship between symptom reductions and associated treatment-related changes.

Additionally, the present study focused on a sample of young children with developmental delay and elevated behavior problems whose caregivers agreed to participate in videoconferencing-based treatment services. It is possible that results from the present analyses may not generalize to other young children with behavior problems or families participating in alternate formats of treatment. However, previous research examining PCIT for youth with or at risk for intellectual disability and/or developmental delay (e.g., Bagner et al., 2007; Bagner et al., 2010) and the delivery of iPCIT (Comer et al., 2017) have shown positive, comparable outcomes relative to PCIT for typically developing youth and PCIT delivered within office-based settings.

Despite limitations, this study offers encouraging new findings related to the detection of risk for treatment nonresponse among families of young children presenting for treatment of early childhood disruptive behavior problems. Additional research should seek to replicate the present study findings to validate whether similar patterns are found among levels of child symptoms and caregiver skills demonstrated early in treatment and child symptom response trajectories across the latter part of treatment in other samples. Confirming relations among early interim progress data and later treatment response can inform the selection of optimal tailoring variables and critical decision points, and in turn, aid in the design of adaptive interventions and precision mental health care approaches to behavioral parenting interventions.

Table 4.1. Sociodemographic characteristics of the present study sample.

	M	*SD*
Child age, months	35.94	1.36
Caregiver age, years[a]	34.44	6.24
	N	%
Child sex		
Male	51	72.9
Female	19	27.1
Child race		
White	55	78.6
Black	13	18.6
Asian	1	1.4
Other	1	1.4
Child ethnicity		
Hispanic/Latinx	55	78.6
Non-Hispanic/Latinx	15	21.4
Caregiver sex		
Male	4	5.7
Female	66	94.3
Caregiver race		
White	55	78.6
Black	13	18.6
Asian	1	1.4
Other	1	1.4
Caregiver ethnicity		
Hispanic/Latinx	50	71.4
Non-Hispanic/Latinx	20	28.6
Caregiver preferred language		
English	39	55.7
Spanish	31	44.3
Annual household income[b]		
< $20,000	28	42.4
$20,001-$50,000	20	30.3
$50,001-$100,000	13	19.7
> $100,000	5	7.6

[a]*n*=69; [b]*n*=66

Table 4.2. Caregiver verbalizations and corresponding DPICS-IV coding information.

Code	Category	Definition	Examples
Neutral talk	N/A	A statement that introduces information about people, objects, events, or activities, or indicates attention to the child, but does not clearly describe or evaluate the child's current or immediately completed behavior.	"I'm making my rainbow just like yours." "It is time to clean up." "Ok."
Behavior description	"Do Skills"	A non-evaluative, declarative sentence or phrase in which the subject is the other person and the verb describes that person's ongoing or immediately completed(< 5 sec.) observable verbal or nonverbal behavior.	"You're building a truck." "You and I are making a big fort." "You seem happy that you fixed it."
Reflection		A declarative phrase or statement that has the same meaning as the child's verbalization. The reflection may repeat, paraphrase, or elaborate upon the child's verbalization but may not change the meaning of the child's statement or interpret unstated ideas.	Child: "It's a horsey." Caregiver: "It is a horse." Child: "This game is fun." Caregiver: "You like playing this game." Child: "Give me the car." Caregiver: "You want the car."
Labeled Praise		Provides a positive evaluation of a specific attribute, product, or behavior of the child.	"You did a great job building the tower." "Thank you for handing me the box." "I like the way you drew that circle."
Unlabeled Praise	N/A	Provides a positive evaluation of the child, an attribute of the child, or a nonspecific activity, behavior, or product of the child.	"I like that." "Thank you." "Perfect."
Command	"Don't Skills"	A statement in which the parent directs the behavior of the child. Commands may be direct or indirect in form. Commands include statements directing the child to perform vocal or motor behaviors, as well as mental or internal, unobservable actions (e.g., think, decide).	"Get down." "Put the crayons in the drawer." "Let's use the green piece."
Question		A verbal inquiry from the parent to the child that is distinguishable from declarative statements by having a rising inflection at the end or by having the sentence structure of a question. A Parent Question requests an answer but does not suggest that a behavior is to be performed by the child.	"Where is the dragon?" "Is this your nose?" "Whose turn is it?"
Negative Talk		A verbal expression of disapproval of the child or the child's attributes, activities, products or choices. Negative Talk also includes sassy, sarcastic, rude, or imprudent speech.	"You're working too slowly." "You put it in the wrong place." "Stop that."

Note: definition and examples are obtained from the *Dyadic Parent-Child Interaction Coding System Comprehensive Manual for Research and Training – Fourth Edition (DPICS-IV Manual; Eyberg et al., 2013).*

Table 4.3. Descriptive statistics of the present study variables.

Time (months since pre-treatment)	*N*	*Min*	*Max*	*M*	*SD*
Time 0	70	0	0	0	0
Time 1	66	1.5	3.75	2.13	0.49
Time 2	65	5	7.5	5.765	0.502
CBCL Externalizing *T* Score	*N*	*Min*	*Max*	*M*	*SD*
Time 0	70	43	85	60.74	10.35
Time 1	66	28	86	57.11	12.22
Time 2	64	28	79	52.56	10.90
ECBI Intensity Scale score[a]	*N*	*Min*	*Max*	*M*	*SD*
Pre-treatment	70	43	196	124.57	32.72
Session 3	49	48	184	122.31	30.53
Session 6	48	44	196	102.71	35.22
Caregiver skills[a]	*N*	*Min*	*Max*	*M*	*SD*
Pre-treatment	67	0	21	4.06	4.37
Session 3	55	0	35	12.24	8.74
Session 6	49	3	37	19.29	9.78

[a]Values reflect data prior to grand-mean centering.

Table 4.4. Correlations across predictors and pre-treatment child externalizing symptoms.

Variable	Pre-treatment child externalizing symptoms	Session 3 child behavior problems	Session 3 caregiver skills	Session 6 child behavior problems	Session 6 caregiver skills
Pre-treatment child externalizing symptoms	-				
Session 3 child behavior problems	0.492**	-			
Session 3 caregiver skills	0.095	-0.036	-		
Session 6 child behavior problems	0.330*	0.643**	0.269	-	
Session 6 caregiver skills	0.133	-0.022	0.671**	0.341*	-

*p<.05. **p<.01.

Note: session 3 and session 6 predictors are grand-mean centered.

Table 4.5. Nested model coefficients.

	Model A Unconditional (Null) Random Intercept Model	*Model B* Linear Model with Random Intercept	*Model C* Session 3 Cross-Level Interactions	*Model D* Session 6 Cross-Level Interactions
	β (SE)	β (SE)	β (SE)	β (SE)
		Fixed Effects		
Level-1				
Intercept	57.000 (1.143)**	60.491 (1.258)**	60.661 (1.302)**	59.904 (1.447)**
Months		-1.387 (0.206)**	-1.572 (0.238)**	-1.366 (0.251)**
Level-2				
Child behavior problems[a]			0.160 (0.044)**	0.098 (0.042)**
Caregiver skills[a]			0.225 (0.166)	0.094 (0.155)
Cross-level Interactions				
Months × child behavior problems[a]			-0.006 (0.008)	-0.000 (.007)
Months × caregiver skills[a]			-0.041 (0.031)	-0.026 (0.028)
Child behavior problems[a] × caregiver skills[a]			0.005 (0.005)	0.005 (0.005)
Months × child behavior problems[a] × caregiver skills[a]			-0.002 (0.001)**	-0.003 (0.001)**
		Random Effects		
Residual	64.898 (8.014)**	48.012 (5.934)**	43.675 (6.499)**	42.433 (6.482)**
Intercept	68.280 (15.563)**	74.939 (15.596)**	46.421 (12.974)**	51.671 (14.230)**
		Additional Information		
ICC	.51			
-2 Log Likelihood	1498.820	1460.118	979.266	930.715
# estimated parameters	3	4	10	10
AIC	1504.822	1468.118	999.266	950.715
BIC	1514.717	1481.311	1028.539	979.467

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

[a]reflects measures collected at session 3 for *Model C* and session 6 for *Model D*.

Figure 4.1. Trajectory of child externalizing symptoms throughout treatment across the sample.

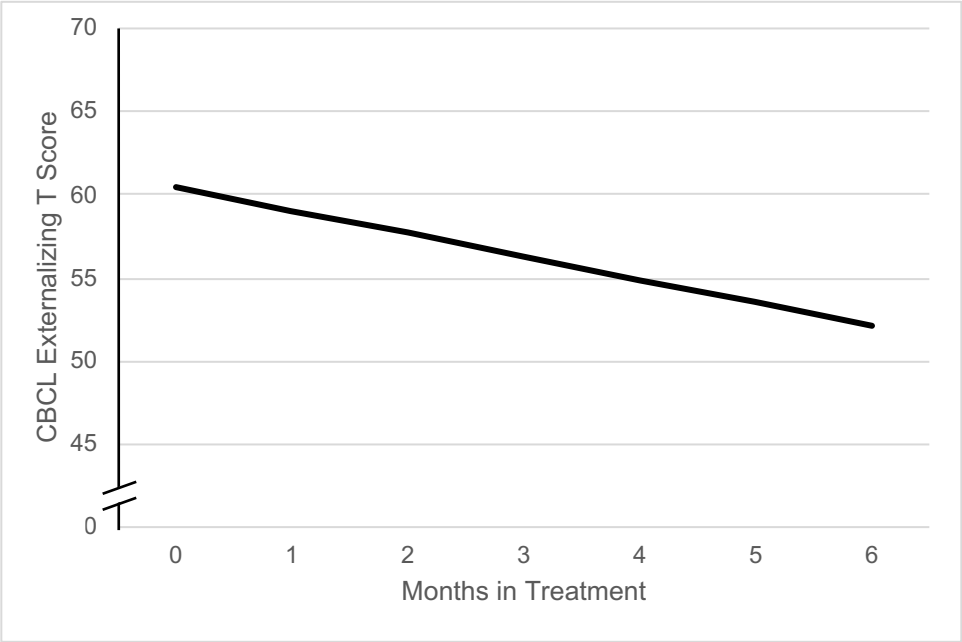


Figure 4.2. Trajectories of child externalizing symptoms throughout treatment by interim progress at session 3.

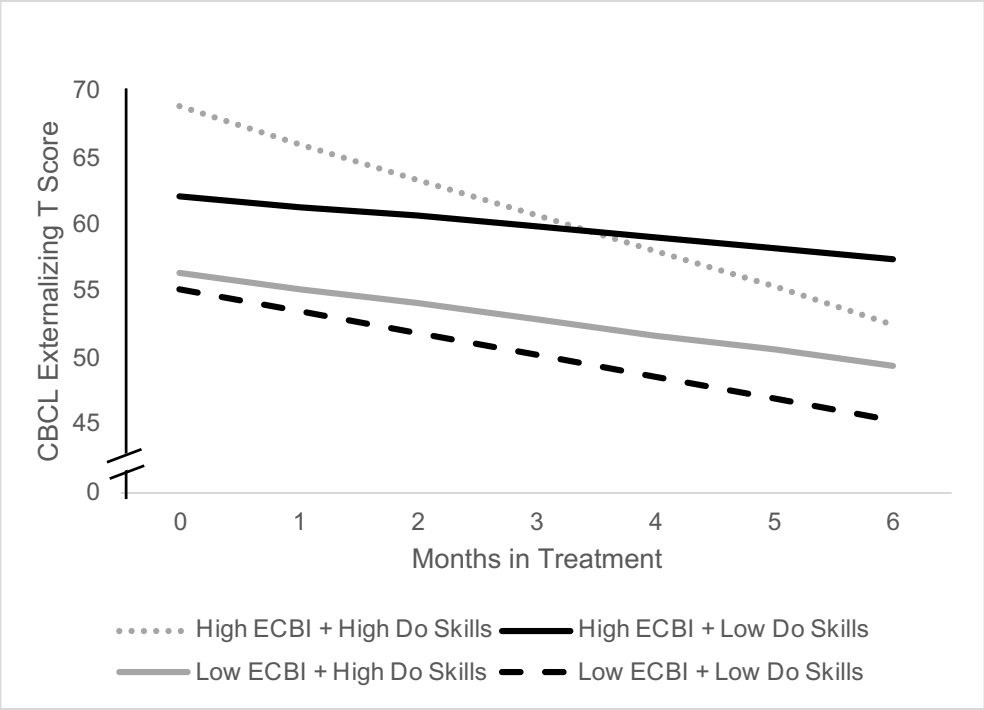
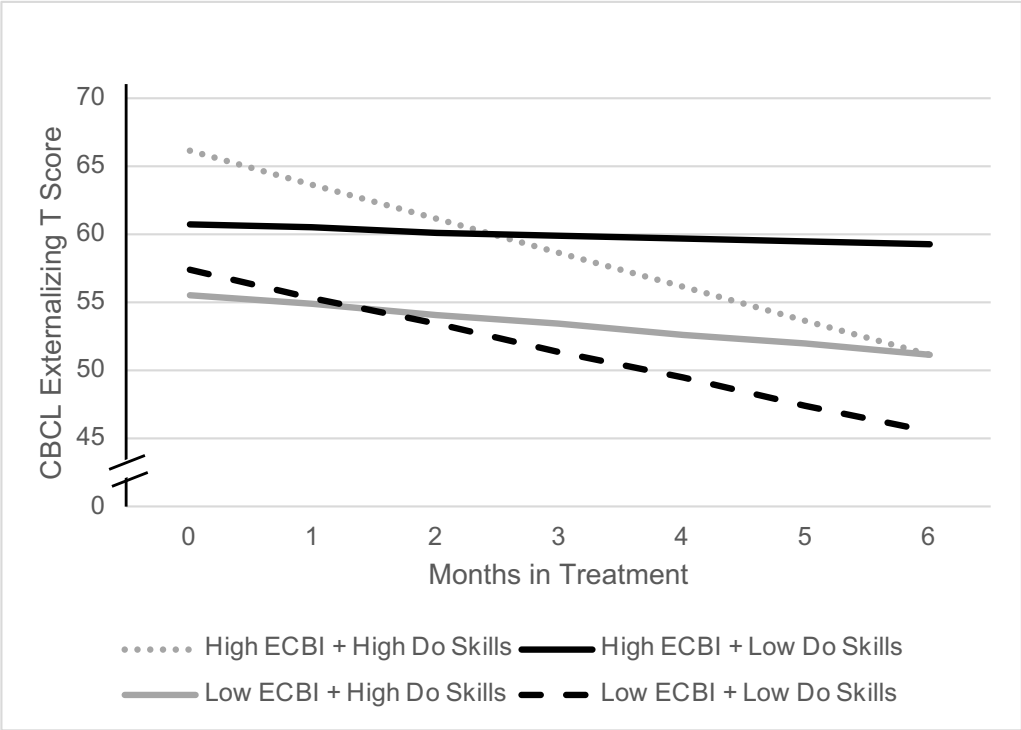


Figure 4.3. Trajectories of child externalizing symptoms throughout treatment by interim progress at session 6.



CHAPTER V

Conclusion

The present collection of work includes three manuscripts focused on the use of early interim progress data to detect risk for mental health treatment nonresponse. This work is critical given heterogeneity in treatment outcomes and high rates of treatment nonresponse reported across the mental health treatment literature (Hong et al., 2019; Lei et al., 2012; Weisz et al., 2017).

In the first manuscript (Chapter II), I discuss the critical need to identify empirically informed tailoring variables that can provide accurate detections of risk for treatment nonresponse at early critical decision points. I describe a strategy for leveraging archival data analysis of randomized controlled trials (RCTs) to examine relations among early treatment progress indicators and ultimate treatment outcome when individuals “stay-the-course.” I also highlight that most of the work examining variability in treatment effects and prognostic factor research has examined stable factors or has emphasized early changes in the ultimate outcome (i.e., symptom reductions) as the sole indicator of early treatment response. However, this approach neglects experimental therapeutics frameworks suggesting that changes within putative mechanisms of treatment response are more likely to occur early in treatment, preceding subsequent changes in the outcome of interest. The strategy detailed in this paper is not only designed to simultaneously examine and compare multiple potential predictors of treatment nonresponse, but also results in easily interpretable and clinically useful guidelines for informing clinical decision-making (i.e., if the caregiver displays x during session y, the family is at risk for treatment nonresponse and thus it may be useful to consider alternative approaches, such as z).

In my second paper (Chapter III), I directly apply the strategy outlined in Chapter II to a pooled sample of typically and atypically developing young children whose families presented for treatment of early childhood behavior problems and participated in one of four RCTs (*N*=153 families). Specifically, I used ROC analyses and corresponding classification indices to (a) examine whether early interim progress data (i.e., candidate tailoring variables) collected within the first six sessions of a behavioral parenting intervention (BPI) could accurately detect families at risk for post-treatment nonresponse, and (b) compared the clinical utility of candidate tailoring variables and their optimal cut scores at candidate critical decision points. Consistent with an experimental therapeutics approach, treatment progress data in this study included measures of out-of-session engagement, caregiver skill acquisition, and child symptoms collected at sessions 3 and 6. Findings from these analyses show that using univariate predictors and dichotomous classifications of interim treatment progress yield limited utility in differentially predicting post-treatment response when examined in isolation from one another. Thus, despite the appeal of singular metrics and clear-cut guidelines for use in clinical practice, results suggest more nuanced approaches may be necessary in order to accurately detect risk for treatment nonresponse at such early stages of intervention.

Pivoting from these results, my third paper (Chapter IV) adopted a continuous approach to measuring interim treatment progress and considered whether interactions among early indicators of treatment response could predict child symptom response trajectories in a sample of families who participated in a BPI for their young child with developmental delay and behavior problems (*N*=70). This study utilized multilevel growth modeling to examine the roles of caregiver skills and child behavior problems (and their interaction) measured at sessions 3 and 6 in predicting child externalizing symptom

change across treatment. Results from this study show that trajectories of child externalizing symptom change across treatment can be predicted by examining the interaction between levels of caregiver skills and child behavior problems displayed during the first six sessions of a BPI. These findings indicate that when ROM measures are assessed consistently (e.g., on a session-by-session basis), on continuous (as opposed to dichotomous) scales, and considered in combination with one another (rather than independently), analytic approaches are able to capture the dynamic interplay among multiple early mental health treatment response factors across time. Such analyses have the potential to inform when and under which circumstances it may be appropriate to consider course-correction (e.g., augmenting, intensifying, or switching treatment).

Taken together, the present collection of work encourages continued use of ROM to inform precision mental health care practices. Notably, much of the ongoing research examining ROM, feedback-informed care, and precision mental health care suggests that multivariable, machine learning approaches can provide particularly accurate and timely predictions (e.g., Gillan & Whelan, 2017; Schwartz et al., 2021). This work has been especially successful in guiding initial treatment selection (e.g., the Personalized Advantage Index; DeRubreis et al., 2014) and can likely be readily extended to the prediction of ongoing treatment decision-making. Thus, as the field moves toward big data approaches to precision mental health care (Bickman, 2020; Gillan & Whelan et al., 2017), identifying ways to ease tensions between clinician preferences and actuarial utility will be critical for enhancing treatment outcomes (MacKrill & Sorensen, 2020).

References

Abrahamse, M.E., Niec, L.N., Junger, M., Boer, F., & Lindauer, R.J.L. (2016). Risk factors for attrition from an evidence-based parenting program: Findings from the Netherlands. *Children and Youth Services Review, 64*, 42-50. doi: 10.1016/j.childyouth.2016.02.025

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Almirall, D., & Chronis-Tuscano, A. (2016). Adaptive interventions in child and adolescent mental health. *Journal of Clinical Child & Adolescent Psychology, 45*(4), 383-395. doi: 10.1080/15374416.2016.1152555

Almirall, D., Compton, S.N., Gunlicks-Stoessel, M., Duan, N., & Murphy, S.A. (2012). Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine, 31*(17), 1887-1902. doi: 10.1002/sim.4512

Almirall, D., DiStefano, C., Chang, Y.C., Shire, S., Kaiser, A., Lu, X., Nahum-Shani, I., Landa, R., Mathy, P., & Kasari, C. (2016). Longitudinal effects of adaptive interventions with a speech-generating device in minimally verbal children with ASD. *Journal of Clinical Child and Adolescent Psychology, 45*(4), 42-456. doi: 10.1080/15374416.2016.1138407

Almirall, D., Nahum-Shani, I., Sherwood, N.E., & Murphy, S.A. (2014). Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine, 4*(3), 260-274. doi: 10.1007/s13142-014-0265-0

American Academy of Pediatrics. (2016). Addressing early childhood emotional and behavioral problems. *Pediatrics, 138*(6), e20163023. doi: 10.1542/peds.2016-3023

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271-285. doi: 10.1037/0003-066X.61.4.271

Assenany, A.E., & McIntosh, D.E. (2002). Negative treatment outcomes of behavioral parent training programs. *Psychology in the Schools, 39*(2), 209-219. doi: 10.1002/pits.10032

Bagner, D.M., & Eyberg, S.M. (2007). Parent-child interaction therapy for disruptive behavior in children with mental retardation: A randomized controlled trial.

Journal of Clinical Child and Adolescent Psychology, 36(3), 418-429. doi: 10.1080/15374410701448448

Bagner, D.M., & Graziano, P.A. (2013). Barriers to success in parent training for young children with developmental delay: The role of cumulative risk. *Behavior Modification, 37*(3), 356-377. doi: 10.1177/0145445512465307

Bagner, D.M., Sheinkopf, S.J., Vohr, B.R., & Leser, B.M. (2010). Parenting intervention for externalizing behavior problems in children born premature: An initial examination. *Journal of Developmental & Behavioral Pediatrics, 31*(3), 209-216. doi: 10.1097/DBP.0b013e3181d5a294

Baker, B.L., Blacher, J., Crnic, K.A., & Edelbrock, C. (2002). Behavior problems and parenting stress in families of three-year-old children with and without developmental delays. *American Journal of Mental Retardation, 107*(6), 433-444. doi: 10.1352/0895-8017(2002)107<0433:BPAPSI>2.0.CO;2

Baker, B.L., McIntyre, L.L.,Blacher, J., Crnic, K., Edelbrock, C., & Low, C. (2003). Pre-school children with and without developmental delay: Behaviour problems and parenting stress over time. *Journal of Intellectual Disability Research, 47*(4-5), 217-230. doi: 10.1046/j.1365-2788.2003.00484.x

Barlow, D.H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist, 65*(1), 13-20. doi: 10.1037/a0015643

Barlow, D.H., Bullis, J.R., Comer, J.S., & Armetaj, A.A. (2013). Evidence-based psychological treatments: An update and the way forward. *Annual Review of Clinical Psychology, 9*, 1-27. doi: 10.1146/annurev-clinpsy-050212-185629

Barlow, D.H., & Comer, J.S. (2013). What are the optimal treatment courses for geriatric anxiety, and how do we find out? *American Journal of Psychiatry, 170*, 707-711. doi: 10.1176/appi.ajp.2013.13040513

Baydar, N., Reid, M.J., & Webster-Stratton, C. (2003). The role of mental health factors and program engagement in the effectiveness of a preventative parenting program for Head Start mothers. *Child Development, 74*(5), 1433-1453. doi: 10.1111/1467-8624.00616

Beauchaine, T.P., Webster-Stratton, C., & Reid, M.J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73*(3), 371-388. doi: 10.1037/0022-006X.73.3.371

Bickman, L. (2020). Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. *Administration and Policy in Mental Health and Mental Health Services Research, 47*(5), 795-843. doi: 10.1007/s10488-020-01065-8

Boggs, S.R., Eyberg, S.M., Edwards, D.L., Rayfield, A., Jacobs, J., Bagner, D.M., & Hood, K.K. (2004). Outcomes of parent-child interaction therapy: A comparison

of treatment completers and study dropouts one to three years later. *Child & Family Behavior Therapy, 26*(4), 1-22. doi: 10.1300/J019v26n04_01

Bufferd, S.J., Dougherty, L.R., Carlson, G.A., & Klein, D.N. (2011). Parent-reported mental health in preschoolers: Findings using a diagnostic interview. *Comprehensive Psychiatry, 52*(4), 359-369. doi: 10.1016/j.comppsych.2010.08.006

Chacko, A., Jensen, S.A., Lowry, L.S., Cornwell, M., & Chimklis, A. (2016). Engagement in behavioral parent training: Review of the literature and implications for practice. *Clinical Child and Family Psychology Review, 19*(3), 204-215. doi: 10.1007/s10567-016-0205-2

Chacko, A., Wymbs, B.T., Rajwan, E., Wymbs, F., & Feirsen, N. (2017). Characteristics of parents of children with ADHD who never attend, drop out, and complete behavioral parent training. *Journal of Child and Family Studies, 26*, 950-960. doi: 10.1007/s10826-016-0618-z

Charach, A., Mohammadzadeh, F., Belanger, S.A., Easson, A., Lipman, E.L., McLennan, J.D., Parkin, P., & Szatmari, P. (2020). Identification of preschool children with mental health problems in primary care: Systematic review and meta-analysis. *Journal of the Candadian Academy of Child & Adolescent Psychiatry, 29*(2), 76-105. doi: PMID: 32405310 PMCID: PMC7213917

Chorpita, B.F., Daleiden, E.L., Park, A.L., Ward, A.M., Levy, M.C., Cromley, T., Chiu, A.W., Letamendi, A.M., Tsai, K.H., & Krull, J.L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology, 85*(1), 13-25. doi: 10.1037/ccp0000133

Chorpita, B.F., Daleiden, E.L., & Weisz, J.R. (2005). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology, 11*(3), 141-156. doi: 10.1016/j.appsy.2005.05.002

Chronis-Tuscano, A., Wang, C.H., Strickland, J., Almirall, D., & Stein, M.A. (2016). Personalized treatment of mothers with ADHD and their young at-risk children: A SMART pilot. *Journal of Clinical Child and Adolescent Psychology, 45*(4), 510-521. doi: 10.1080/15374416.2015.1102069

Chu, B.C., Colognori, D.B., Yang, G., Xie, M., Bergman, R.L., & Piacentini, J. (2015). Mediators of exposure therapy for youth obsessive-compulsive disorder: specificity and temporal sequence of client and treatment factors. *Behavior Therapy, 46*(3), 395-408. doi: 10.1016/j.beth.2015.01.003

Chu, B.C., & Kendall, P.C. (2004). Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. *Journal of Consulting and Clinical Psychology, 72*(5), 821-829. doi: 10.1037/0022-006X.72.5.821

Cohen, Z.D., & DeRubreis, R.J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14*, 209-36. doi: /10.1146/annurev-clinpsy-050817-084746

Collins, L.M., Nahum-Shani, I., & Almirall, D. (2014). Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clinical Trials, 11*(4), 426-434. doi: 10.1177/1740774514536795

Comer, J.S., & Barlow, D.H. (2014). The occasional case against broad dissemination and implementation: Retaining a role for specialty care in the delivery of psychological treatments. *American Psychologist, 69*(1), 1-18. doi: 10.1037/a0033582

Comer, J.S., Chow, C., Chan, P.T., Cooper-Vince, C., & Wilson, L.A.S. (2013). Psychosocial treatment efficacy for disruptive behavior problems in very young children: A meta-analytic examination. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(1), 26–36. doi: 10.1016/j.jaac.2012.10.001

Comer, J.S., Furr, J.M., Cooper-Vince, C.E., & Madigan, R.J. (2015). Rationale and considerations for the Internet-based delivery of parent-child interaction therapy. *Cognitive and Behavioral Practice, 22*(3), 302-316. doi: 10.1016/j.cbpra.2014.07.003

Comer, J.S., Furr, J.M., Miguel, E., Cooper-Vince, C.E., Carpenter, A.L., Elkins, R.M., Kerns, C.E., Cornacchio, D., Chou, T., Coxe, S., DeSerisy, M., Sanchez, A.L., Golik, A., Martin, J., Myers, K., & Chase, R. (2017). Remotely delivering real-time parent training to the home: An initial randomized trial of Internet-delivered parent-child interaction therapy (I-PCIT). *Journal of Consulting and Clinical Psychology, 85*(9), 909-917. doi: 10.1037/ccp0000230

Comer, J.S., & Kendall, P.C. (2013). Methodology, design, and evaluation in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 6th Edition*. Hoboken, NJ: Wiley.

Compton, S.N., Peris, T.S., Almirall, D., Birmaher, B., Sherrill, J., Kendall, P.C., March, J.S., Gosch, E.A., Ginsburg, G.S., Rynn, M.A., Piacentini, J.C., McCracken, J.T., Keeton, C.P., Suveg, C.M., Aschenbrand, S.G., Sakolsky, D., Iyengar, S., Walkup, J.T., & Albano, A.M. (2014). Predictors and moderators of treatment response in childhood anxiety disorders: Results from the CAMS trial. *Journal of Consulting and Clinical Psychology, 82*(2), 212-224. doi: 10.1037/a0035458

Constantino, J. N., & Gruber, C. P. (2012). *The social responsiveness scale (2nd ed.)*. Western Psychological Services.

Cooper, Z., Allen, E., Bailey-Straebler, S., Basden, S., Murphy, R., O'Connor, M.E., & Fairburn, C.G. (2016). Predictors and moderators of response to enhanced cognitive behaviour therapy and interpersonal psychotherapy for the treatment of eating disorders. *Behaviour Research and Therapy, 84*, 9-13. doi: 10.1016/j.brat.2016.07.002

Correll, C.U., Zhao, J., Carson, W., Marcus, R., McQuade, R., Forbes, R.A., & Mankoski, R. (2013). Early antipsychotic response to aripiprazole in adolescents with schizophrenia: predictive value for clinical outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(7), 689-698.e3. doi: 10.1016/j.jaac.2013.04.018

Costello, E.J., Egger, H., & Angold, A. (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(10), 972-986. doi: 10.1097/01.chi.0000172552.41596.6f

Cuijpers, P., Karyotaki, E., Weitz, E. et al. (2014). The effects of psychotherapies for major depression in adults on remission, recovery, and improvement: A meta-analysis. *Journal of Affective Disorders, 159*, 118-126. doi: 10.1016/j.jad.2014.02.026

Cummings, C.M., Kazantzis, N., & Kendall, P. (2014). Facilitating homework and generalization of skills to the real world. In E.S. Sburlati, H.J., Lyneham, C.A., Schniering, and R.M. Rapee (Eds.), *Evidence-Based CBT for Anxiety and Depression in Children and Adolescents: A Competencies-Based Approach.* Wiley Blackwell: Malden, MA.

Danko, C.M., Garbacz, L.L., & Budd, K.S. (2016). Outcomes of Parent-Child Interaction Therapy in an urban community clinic: A comparison of treatment completers and dropouts. *Children and Youth Services Review, 60*, 42-51. doi: 10.1016/j.childyouth.2015.11.007

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M. Nevin, J., O'Hayon, H., Patel, U. Sainty, A., Spencer, P., & McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomized controlled trial. *Lancet Psychiatry, 5*(7), 564-572. doi: 10.1016/S2215-0366(18)30162-7

Delgadillo, J., & Lutz, W. (2020). A developmental pathway towards precision mental health care. *JAMA Psychiatry, 77*(9), 889-890. doi: 10.1001/jamapsychiatry.2020.1048

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

Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., Gilbody, S., Lutz, W., Rubel, R.A., & de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behavioral Research and Therapy, 99*, 89-97. doi: 10.1016/j.brat.2017.09.011

DeRubreis, R.J., Cohen, Z.D., Forand, N.R., Fournier, J.C., Gelfand, L.A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on

prediction into individualized treatment recommendations. A demonstration. *Plos One, 9*(1), e83875. doi: 10.1371/journal.pone.0083875

Dishion, T.J., Connell, A., Weaver, C., Shaw, D., Gardner, F., & Wilson, M. (2008). The family check-up with high-risk indigent families: Preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Development, 79*(5), 1395-1414. doi: 10.1111/j.1467-8624.2008.01195.x

Donker, T., Batterham, P.J., Warmerdam, L., Bennett, K., Bennett, A., Cuikpers, P., Griffiths, K.M., & Christensen, H. (2013). Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression. *Journal of Affective Disorders, 151*(1), 343-351. doi: 10.1016/j.jad.2013.06.020

Durland, P.H., Wyszynski, C.M., & Chu, B. (2018). Predictors and outcomes of sudden gains and sudden regressions in cognitive behavioral therapy for youth anxiety. *Behavior Therapy, 49*(5), 823-835. doi: 10.1016/j.beth.2018.01.006

Egger, H.L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology & Psychiatry, 47*(3), 313-317. doi: 10.1111/j.1469-7610.2006.01618.x

Egger, H.L., Erkanli, A., Keeler, G., Potts, E., Walter, B.K., & Angold, A. (2006). Test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(5), 538-549. doi: 10.1097/01.chi.0000205705.71194.b8

Eyberg, S. M., & Funderburk, B. (2011). *Parent-child interaction therapy protocol.* Gainesville, FL: PCIT International, Inc.

Eyberg, S.M., Nelson, M.M., & Boggs, S.R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child and Adolescent Psychology, 37*(1), 215-237. doi: 10.1080/15374410701820117

Eyberg, S.M., Nelson, M.M., Ginn, N.C., Bhuiyan, N., & Boggs, S.R. (2013). *Dyadic Parent-Child Interaction Coding System: Comprehensive Manual for Research and Training (4th ed.).* Gainesville, FL: PCIT International.

Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory-Revised: Professional manual.* Psychological Assessment Resources.

Forehand, R., Lafko, N., Parent, J., & Burt, K. (2014). Is parenting the mediator of change in behavioral parent training for externalizing problems of youth? *Clinical Psychology Review, 34*(8), 608-619. doi: 10.1016/j.cpr.2014.10.00

Frank, E., Cassano, G.B., Rucci, P., Thompson, W.K., Kraemer, H.C., Fagiolini, A., Maggi, L., Kupfer, D.J., Shear, M.K., Houck, P.R., Calugi, S., Grochocinski, V.J.,

Scocco, P., Buttenfield, J., & Forgione, R.N. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine, 41*(1), 151-162. doi: 10.1017/S0033291710000553

Funderburk, B.W., Eyberg, S.M., Rich, B.A., & Behar, L. (2003). Further psychometric evaluation of the Eyberg and Behar rating scales for parents and teachers of preschoolers. *Early Education and Development, 14*(1), 67-82. doi: 10.1207/s15566935eed1401_5

Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: Outcomes and mechanisms of change. *Journal of Child Psychology & Psychiatry, 47*(11), 1123-1132. doi: 10.1111/j.1469-7610.2006.01668.x

Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child & Adolescent Psychology, 39*(4), 568-580. doi: 10.1080/15374416.2010.486315

Gardner, F., & Leijten, P. (2017). Incredible Years parenting interventions: Current effectiveness research and future directions. *Current Opinion in Psychology, 15*, 99-104. doi: 10.1016/j.copsyc.2017.02.023

Gillan, C.M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences, 18,* 34-42. doi: 10.1016/j.cobeha.2017.07.003

Gleason, M.M., Goldson, E., & Yogman, M.W. (2016). Addressing early childhood emotional and behavior problems. *Pediatrics, 138*(6), e20163025. doi: 10.1542/peds.2016-3025

Graziano, P., Ros-Demarize, R., & Hare, M.M. (2020). Condensing parent training: A randomized trial comparing the efficacy of a briefer, more intensive version of Parent-Child Interaction Therapy (I-PCIT). *Journal of Consulting and Clinical Psychology, 88*(7), 669-679. doi: 10.1037/ccp0000504

Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J., Sivan, A., & Richardson. (2007). Reliability and validity of the Eyberg Child Behavior Inventory with African-American and Latino parents of young children. *Research in Nursing & Health, 30*, 213-223. doi: 10.1002/nur.20181

Guy, W. & Bonato, R. (Eds.). (1970). *CGI: The clinical global impression scale*. Chevy Chase: National Institute of Mental Health.

Hannan, C., Lambert, M.J., Harmon, C., Nielson, S.L., Smart, D.W., Shimokawa, K., & Sutton, S.W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163. doi: 10.1002/jclp.20108

Hofmann, S.G., Asnaani, A., Vonk, I.J.J., Sawyer, A.T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427-440. doi: 10.1007/s10608-012-9476-1

Hong, N. & Comer, J.S. (2018). High-end specificity of the Attention-Deficit/Hyperactivity Problems Scale of the Child Behavior Checklist for ages 1.5-5 in a sample of young children with disruptive behavior disorders. *Child Psychiatry and Human Development.* Advanced online publication. doi: 10.1007/s10578-018-0834-4

Hong, N., Cornacchio, D., Pettit, J., & Comer, J.S. (2019). Coal-mine canaries in clinical psychology: Getting better at identifying early signals of treatment nonresponse. *Clinical Psychological Science, 7*(6), 1207-1221. doi: 10.1177/2167702619858111

Hong, N., Feinberg, L., Garcia, D., Comer, J.S., & Bagner, D.M. (2018). Internet-delivered parent-child interaction therapy (I-PCIT) for children with autism spectrum disorder: Rationale, considerations, and lessons learned. In C.B. McNeil, L.B. Quetsch, and C.M. Anderson (Eds.), *Handbook of Parent-Child Interaction Therapy for Children on the Autism Spectrum* (p. 545-557). Springer: New York. doi: 10.1007/978-3-030-03213-5_30

Insel, T.R. & Gogtay, N. (2014). National Institute of Mental Health clinical trials: New opportunities, new expectations. *JAMA Psychiatry, 71*(7), 745-746. doi: 10.1001/jamapsychiatry.2014.426

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. doi: 10.1037/0022-006X.59.1.12

Kaminski, J.W., & Claussen, A.H. (2017). Evidence base update for psychosocial treatments for disruptive behaviors in children. *Journal of Clinical Child and Adolescent Psychology, 46*(4), 477-499. doi: 10.1080/15374416.2017.1310044

Kaminski, J.W., Valle, L.A., Filene, J.H., & Boyle, C.L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*(4), 567-589. doi: 10.1007/s10802-007-9201-9

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P, Landa, R., Murphy, S., & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: Sequential multiple assignment randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(6), 635-646. doi: 10.1016/j.jaac.2014.01.019

Kazantzis, N., Deane, F.P., & Ronan, K.R. (2000). Homework assignments in cognitive and behavioral therapy: A meta-analysis. *Clinical Psychology: Science and Practice, 7*(2), 189-202. doi: 10.1093/clipsy.7.2.189

Kazantzis, N., Whittington, C., Zelencich, L., Kyrios, M., Norton, P.J., & Hofman, S.G. (2016). Quantity and quality of homework compliance: A meta-analysis of

relations with outcome in cognitive behavior therapy. *Behavior Therapy, 47*(5), 755-772. doi: 10.1016/j.beth.2016.05.002

Kazdin, A.E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4-5), 418-428. doi: 10.1080/10503300802448899

Kazdin, A.E., & Mazurick, J.L. (1994). Dropping out of child psychotherapy: Distinguishing early and late dropouts over the course of treatment. *Journal of Consulting and Clinical Psychology, 62*(5), 1069-1074. doi: 10.1037/0022-006X.62.5.1069

Keenan, K., & Wakschlag, L.S. (2002). Can a valid diagnosis of disruptive behavior disorder be made in preschool children? *The American Journal of Psychiatry, 159*(3), 351-358. doi: 10.1176/appi.ajp.159.3.351

Keenan, K., Wakschlag, L.S., Danis, B., Hill, C., Humphries, M., Duax, J. & Donald, R. (2007). Further evidence of the reliability and validity of DSM-IV ODD and CD in preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(4), 457-468. doi: 10.1097/CHI.0b013e31803062d3

Kendall, P.C., Comer, J.S., & Chow, C. (2013). The randomized controlled trial: Basics and beyond. In J.S. Comer & P.C. Kendall (eds.), *Oxford Handbook of Research Strategies for Clinical Psychology.* New York: Oxford University Press.

Kendall, P.C., & Treadwell, K.R.H. (2007). The role of self-statements as a mediator in treatment for youth with anxiety disorders. *Journal of Consulting and Clinical Psycholgy, 75*(3), 380-389. doi: 10.1037/0022-006X.75.3.380

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., Bhattacharya, A., Gimson, A., Brütt, A.L., de Jong, K., & Moore, M. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews, 13*(7), CD011119. doi: 10.1002/14651858.CD011119.pub2

LaHuis, D.M., & Ferguson, M.W. (2009). The accuracy of significance tests for slope variance components in multilevel random coefficient models. *Organizational Research Methods, 12,* 418-443. doi: 10.1177/1094428107308984

Lambert, M.J., Whipple, J.L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520-537. doi: 10.1037/pst0000167

Lantz, B. (2019). Evaluating model performance. In B. Lantz (Ed.), *Machine Learning with R: Expert techniques for predictive modeling*. Packt Publishing Ltd.: Birmingham, UK.

Lavigne, J.V., Lebailly, S.A., Hopkins, J., Gouze, K.R., & Binns, H.J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child & Adolescent Psychology, 38*(3), 315-328. doi: 10.1080/15374410902851382

Lavori, P.W., & Dawson, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clinical Trials, 1*(1), 9-20. doi: 10.1191/1740774504cn002oa

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S.A. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, 8*(1), 21-48. doi: 10.1146/annurev-clinpsy-032511-143152

Lenze, S.N., Pautsch, J., & Luby, J.L. (2011). Parent-child interaction therapy emotion development: A novel treatment for depression in preschool children. *Depression & Anxiety, 28*(2), 153-159. doi: 10.1002/da.20770

Lewis, C.C., Simons, A.D., & Kim, H.K. (2012). The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. *Journal of Consulting and Clinical Psychology, 80*(4), 525-534. doi: 10.1037/a0029131

Lieneman, C.C., Quetsch, L.B., Theodorou, L.L., Newton, K.A., & McNeil, C.B. (2019). Reconceptualizing attrition in Parent-Child Interaction Therapy: "Dropouts" demonstrate impressive improvements. *Psychology Research and Behavior Management, 12*, 543-555. doi: 10.2147/PRBM.S207370

Lindhiem, O. Higa, J., Trentacosta, C.J., Herschell, A.D., Kolko, D.J. (2014). Skill acquisition and utilization during evidence-based psychosocial treatments for childhood disruptive behavior problems: A review and meta-analysis. *Clinical Child and Family Psychology Review, 17*, 41-66. doi: 10.1007/s10567-013-0136-0

Lundahl, B., Risser, H.J., Lovejoy, M.C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review, 26*(1), 86-104. doi: 10.1016/j.cpr.2005.07.004

Lutz, W., Hofmann, S.G., Rubel, J., Boswell, J.F., Shear, M.K., Gorman, J.M., Woods, S.W., & Barlow, D.H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82*(2), 287-297. doi: 10.1037/a0035535

Lutz, W., Schwartz, B., Penedo, J.M.G., Boyle, K., & Deisenhofer, A. (2020). Working towards the development and implementation of precision mental healthcare: An example. *Administration and Policy in Mental Health and Mental Health Services Research, 47*(5), 856-861. doi: 10.1007/s10488-020-01053-y

MacPherson, H.A., Algorta, G.P., Mendenhall, A.N., Fields, B.W., & Fristad, M.A. (2014). Predictors and moderators in the randomized trial of multifamily psychoeducational psychotherapy for childhood mood disorders. *Journal of Clinical Child and Adolescent Psychology, 43,* 459-472. doi: 10.1080/15374416.2013.807735

MacKrill, T., & Sorensen, K.M. (2020). Implementing routine outcome measurement in psychosocial interventions: A systematic review. *European Journal of Social Work, 23*(5), 790-808. doi: 10.1080/13691457.2019.1602029

Magnavita, J.J., & Lilienfeld, S.O. (2016). Clinical expertise and decision-making: An overview of bias in clinical practice. In J.J. Magnavita (Ed.), *Clinical decision-making in mental health practice* (p. 23-60). American Psychological Association: Washington, DC. doi: 10.1037/14711-002

Marker, C.D., Comer, J.S., Abramova, V., & Kendall, P.C. (2013). The reciprocal relationship between alliance and symptom improvement across the treatment of childhood anxiety. *Journal of Clinical Child and Adolescent Psychology, 42*(1), 22-33. doi: 10.1080/15374416.2012.723261

Mausbach, B.T., Moore, R., Roesch, S., Cardenas, V., & Patterson, T.L. (2010). The relationship between homework compliance and therapy outcomes: An updated meta-analysis. *Cognitive Therapy Research, 34*(5), 429-438. doi: 10.1007/s10608-010-9297-z

McDiarmid, M.D., & Bagner, D.M. (2005). Parent-child interaction therapy for children with disruptive behavior and developmental disabilities. *Education and Treatment of Children, 28*(2), 130-141.

McMahon, R.J., & Forehand, R.L. (2005). *Helping the noncompliant child, second edition: Family-based treatment for oppositional behavior.* Guilford Press: New York.

McManus, B., McCormick, M.C., Acevedo-Garcia, D., Ganz, M., & Hauser-Cram, P. (2009). The effect of state early intervention eligibility policy on participation among a cohort of young CSHCN. *Pediatrics, 124*(Supplement 4), S368-S374. doi: 10.1542/peds.2009-1255G

Menting, A.T.A, de Castro, B.O., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: A meta-analytic review. *Clinical Psychology Review, 33*, 901-913. doi: 10.1016/j.cpr.2013.07.006

Merikangas, K.R., He, J., Burstein, M.E., Swendsen, J., Avenevoli, S., Case, B., Georgiades, K., Heaton, L., Swanson, S., & Olfson, M. (2011). Service utilization for lifetime mental disorders in U.S. adolescents: Results of the national comorbidity survey adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adoelscent Psychiatry, 50*(1), 32-45. doi: 10.1016/j.jaac.2010.10.006

Meyer, T.J., Miller, M.L., Metzger, R.L., & Borkovec, T.D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*(6), 487-495. doi: 10.1016/0005-7967(90)90135-6

Murphy, S.A., Collins, L.M., & Rush, A.J. (2007). Customizing treatment to the patient: Adaptive treatment strategies. *Drug and Alcohol Dependence, 88*(Supplement 2), S1-S3. doi: 10.1016/j.drugalcdep.2007.02.001

Naar-King, S., Ellis, D.A., Carcone, A.I., Templin, T., Jacques-Tiura, A.J., Hartlieb, K.B., Cunningham, P., & Jen, K.L.C. (2016). Sequential multiple assignment

randomized trial (SMART) to construct weight loss interventions for African American adolescents. *Journal of Clinical Child and Adolescent Psychology, 45*(4), 428-441. doi: 10.1080/15374416.2014.971459

Nahum-Shani, I., Qian, M., Almirall, D., et al. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods, 17*(4), 457-477. doi: 10.1037/a0029372

National Institute of Mental Health. (2015). NIMH Strategic Plan for Research (NIH Publication No. 02-2650). Retrieved on June 3, 2018 from: http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml

Neece, C.L., Green, S.A., & Baker, B.L. (2012). Parenting stress and child behavior problems: A transactional relationship across time. *American Journal of Intellectual and Developmental Disabilities, 117*(1), 48-66. doi: 10.1352/1944-7558-117.1.48

Niec, L.,N., Barnett, M.L., Prewett, M.S., & Chatham, J.R.S. (2016). Group parent-child interaction therapy: A randomized control trial for the treatment of conduct problems in young children. *Journal of Consulting and Clinical Psychology, 84*(8), 682-698. doi: 10.1037/a0040218

Nix, R.L., Bierman, K.L., McMahon, R.J., & The Conduct Problems Prevention Research Group. (2009). How attendance and quality of participation affect treatment response to parent management training. *Journal of Consulting and Clinical Psychology, 77*(3), 429-438. doi: 10.1037/a0015028

Nock, M.K., & Ferriter, C. (2005). Parent management of attendance and adherence in child and adolescent therapy: A conceptual and empirical review. *Clinical Child and Family Psychology Review, 8*(2), 149-166. doi: 10.1007/s10567-005-4753-0

Owens, E.B., Hinshaw, S.P., Kraemer, H.C., Arnold, L.E., Abikoff, H.B., Cantwell, D.P., Conners, C.K., Elliott, G., Greenhill, L.L., Hechtman, L., Hoza, B., Jensen, P.S., March, J.S., Newcorn, J.H., Pelham, W.E., Severe, J.B., Swanson, J.M., Vitiello, B., Wells, K.C., & Wigal, T. (2003). Which treatment for whom for ADHD? Moderators of treatment response in the MTA. *Journal of Consulting and Clinical Psychology, 71*(3), 540-552. doi: 10.1037/0022-006X.71.3.540

Patterson, G.R. (2002). The early development of coercive family process. In J.B. Reid, G.R. Patterson, and J. Snyder (Eds.), *Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention* (p. 25-44). American Psychological Association: Washington, DC.

Patterson, G.R., Forgatch, M.S., Yoerger, K.L., & Stoolmiller, M. (1998). Variables that initiate and maintain an early-onset trajectory for juvenile offending. *Development and Psychopathology, 10*(3), 531-547. doi: 10.1017/s0954579498001734

Patterson, G.R., Reid, J.B., & Eddy, J.M. (2002). A brief history of the Oregon model. In J.B. Reid, G.R. Patterson, and J. Snyder (Eds.), *Antisocial Behavior in Children*

and Adolescents: A Developmental Analysis and Model for Intervention (p. 3-24). American Psychological Association: Washington, DC.

Paul, G.L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology,* 31, 109-118.

Pelham, W.E. Jr., Fabiano, G.A., Waxmonsky, J.G., Greiner, A.R., Gnagy, E.M., Pelham, W.E. 3rd , Coxe, S., Verley, J., Bhatia, I., Hart, K., Karch, K., Konijnendijk, E., Tresco, K., Nahum-Shani, I., & Murphy, S.A. (2016). Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions. *Journal of Clinical Child and Adolescent Psychology, 45*(4), 396-415. doi: 10.1080/15374416.2015.1105138

Pettit, J.W., Silverman, W.K., Rey, Y., Marin, C., & Jaccard, J. (2016). Moving to second stage treatments faster: Identifying midtreatment tailoring variables for youth with anxiety disorders. *Journal of Clinical Child and Adolescent Psychology, 45*(4), 457-468. doi: 10.1080/15374416.2015.1038824

Perlis, R.H. (2016). Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry, 15*(3), 228-235. doi: 10.1002/wps.20345

Peugh, J.L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112. doi: 10.1016/j.jsp.2009.09.002

Pons, J., Matias-Carrelo, L., Rodriguez, M., Rodriguez, J., Herrans, L.L., Jimenez, M.E., Negron, A., Flores, L., Manon, S., Jimenez, K., Medina, G., Rosario, E., Ortiz Nolasco, N., & Yang, J. (2008). Estudios de validez de la Escala de Inteligencia Wechsler para Adultos Versión III, Puerto Rico (EIWA-III*). Revista Puertorriqueña de Psicología*, *19*, 75-111. doi:

Pfeifer, B.J., & Strunk, D.R. (2015). A primer on psychotherapy process research: A review of cognitive therapy of depression. *Australian Psychologist, 50*(6), 411-415. doi: 10.1111/ap.12147

Rekkas, A., Paulus, J.K., Raman, G., Wong, J.B., Steyerberg, E.W., Rijnbeek, P.R., Kent, D.M., & van Klaveren, D. (2020). Predictive approaches to heterogenous treatment effects: A scoping review. *BMC Medical Research Methodology, 20,* 264. doi: 10.1186/s12874-020-01145-1

Reyno, S.M., & McGrath, P.J. (2006). Predictors of parent training efficacy for child externalizing behavior problems—A meta-anlytic review. *Journal of Child Psychology & Psychiatry, 47*(1), 99-111. doi: 10.1111/j.1469-7610.2005.01544.x

Ros, R., Graziano, P.A., & Hart, K.C. (2017). Parental homework completion and treatment knowledge during group parent-child interaction therapy. *Journal of Early Intervention, 39*(4), 299-320. doi: 10.1177/1053815117718491

Ros, R., Hernandez, J., Graziano, P.A., & Bagner, D.M. (2016). Parent training for children with or at risk for developmental delay: The role of parental homework completion. *Behavior Therapy, 47*(1), 1-13. doi: 10.1016/j.beth.2015.08.004

Rush, A.J., Fava, M., Wisniewski, S.R., Lavori, P.W., Trivedi, M.H., Sackeim, H.A., Thase, M.E., Nierenberg, A.A., Quitkin, F.M., Kashner, T.M., Kupfer, D.J., Rosenbaum, J.F., Alpert, J., Stewart, J.W., McGrath, P.J., Biggs, M.M., Shores-Wilson, K., Lebowitz, B.D., Ritz, L., Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials, 25*(1), 119-142. doi: 10.1016/S0197-2456(03)00112-0

Sachs, G.S., Thase, M.E., Otto, M.W., Bauer, M., Miklowitz, D., Wisniewski, S.R., Lavori, P., Lebowitz, B., Rudorfer, M., Frank, E., Nierenberg, A.A., Fava, M., Bowden, C., Ketter, T., Marangell, L., Calabrese, J., Kupfner, D., & Rosenbaum, J.F. (2003). Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biological Psychiatry, 53*(11), 1028-1042. doi: 10.1016/S0006-3223(03)00165-3

Sanders, M.R. (1999). Triple P—Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review, 2*, 71-90. doi: 10.1023/A:1021843613840

Schueller, S.M., Kwasny, M.J., Dear, B.F., Titov, N., & Mohr, D.C. (2015). Cut points on the Patient Health Questionnaire (PHQ-9) that predict response to cognitive-behavioral treatments for depression. *General Hospital Psychiatry, 37*(5), 470-475. doi: 10.1016/j.genhosppsych.2015.05.009

Schwartz, B., Cohen, Z.D., Rubel, J.A., Zimmermann, D., Wittmann, W.W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research, 31*(1), 33-51. doi: 10.1080/10503307.2020.1769219

Scudder, A., Wong, C., Ober, N., Hoffman, M., Toscolani, J., & Handen, B.L. (2019). Parent-child interaction therapy (PCIT) in young children with autism spectrum disorder. *Child & Family Behavior Therapy, 41*(4), 201-220. doi: 10.1080/07317107.2019.1659542

Shelleby, E.C., & Shaw, D.S. (2014). Outcomes of parenting interventions for child conduct problems: A review of differential effectiveness. *Child Psychiatry & Human Development, 45*(5), 10.1007/s10578-013-0431-5 628-645. doi:

Shih, W., Patterson, S.Y., & Kasari, C. (2016). Developing an adaptive treatment strategy for peer-related social skills for children with autism spectrum disorders. *Journal of Child & Adolescent Psychology, 45*(4), 469-479. doi: 10.1080/15374416.2014.915549

Shimokawa, K., Lambert, M.J., & Smart, D.W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of

psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298-311. doi: 10.1037/a0019247

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press: New York. doi: 10.1093/acprof:oso/9780195152968.001.0001

Southward, M.W., & Sauer-Zavala, S. (2020). Experimental manipulations to test theory-driven mechanisms of cognitive behavior therapy. *Frontiers in Psychiatry, 11*, 603009. doi: 10.3389/fpsyt.2020.603009

Steidtmann, D., Manber, R., Blasey, C., Markowitz, J.C., Klein, D.N., Rothbaum, B.O., Thase, M.E., Kocsis, J.H., & Arnow, B.A. (2013). Detecting critical decision points in psychotherapy and psychotherapy + medication for chronic depression. *Journal of Consulting and Clinical Psychology, 81*(5), 783-792. doi: 10.1037/a0033250

Stokes, J.O., Jent, J.F., Weinstein, A., Davis, E.M., Brown, T.M., Cruz, L., & Wavering, H. (2016). Does practice make perfect? The relationship between self-reported treatment homework completion and parental skill acquisition and child behaviors. *Behavior Therapy, 47*(4), 538-549. doi: 10.1016/j.beth.2016.04.004

Stroup, T.S., McEvoy, J.P., Swartz, M.S., Byerly, M.J., Glick, I.D., Canive, J.M., McGee, M.F., Simpson, G.M., Stevens, M.C., & Lieberman, J.A. (2009). The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophrenia Bulletin, 29*(1), 15-31. doi: 10.1093/oxfordjournals.schbul.a006986

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics (4th ed.)*. Allyn & Bacon: Needham Heights, MA.

Thomas, R., Abell, B., Webb, H.J., Avdagic, E., & Zimmer-Gembeck, M.J. (2017). Parent-Child Interaction Therapy: A meta-analysis. *Pediatrics, 140*(3), e20170352. doi: 10.1542/peds.2017-0352

Thomas, R., & Zimmer-Gembeck, M.J. (2007). Behavioral outcomes of Parent-Child Interaction Therapy and Triple P—Positive Parenting Program: A review and meta-analysis. *Journal of Abnormal Child Psychology, 35*(3), 475-495. doi: 10.1007/s10802-007-9104-9

Timmer, S.G., Urquiza, A.J., Boys, D.K., Forte, L.A., Quick-Abdullah, D., Chan, S., & Gould, W. (2016). Filling potholes on the implementation highway: Evaluating the implementation of parent-child interaction therapy in Los Angeles county. *Child Abuse & Neglect, 53*, 40-50. doi: 10.1016/j.chiabu.2015.11.011

Varadhan, R., Segal, J.B., Boyd, C.M., Wu, A.W., & Weiss, C.O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology, 66*(8), 818-825. doi: 10.1016/j.jclinepi.2013.02.009

Vittengl, J.R., Jarrett, R.B., Weitz, E., Hollon, S.D., Twisk, J. et al. (2016). Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *American Journal of Psychiatry, 173,* 481-490. doi: 10.1176/appi.ajp.2015.15040492

Wakschlag, L.S., Briggs-Gowan, M.J., Carter, A.S., Hill, C., Danis, B., Keenan, K., McCarthy, K.J., & Leventhal, B.L. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *The Journal of Child Psychology and Psychiatry, 48*(10), 976-987. doi: 10.1111/j.1469-7610.2007.01786.x

Ward, M.A., Theule, J., & Cheung, K. (2016). Parent-child interaction therapy for child disruptive behaviour disorders: A meta-analysis. *Child Youth Care Forum, 45*, 675-590. doi: 10.1007/s10566-016-9350-5

Webster-Stratton, C., & Reid, M.J. (2017). The Incredible Years parents, teachers, and children training series: A multifaceted treatment approach for young children with conduct problems. In A.E. Kazdin and J.R. Weisz (Eds.), *Evidence-Based Psychotherapies for Children and Adolescents, 3rd Edition.* Guildford Publications: New York.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)*. NCS Pearson: San Antonio, TX.

Weersing, V.B., & Weisz, J.R. (2002). Mechanisms of action in youth psychotherapy. *Journal of Clinical Psychology & Psychiatry, 43*(1), 3-29. doi: 10.1111/1469-7610.00002

Weis, R., Lovejoy, M.C., & Lundahl, B.W. (2005). Factor structure and discriminative validity of the Eyberg Child Behavior Inventory with young children. *Journal of Psychopathology and Behavioral Assessment, 27*(4), 269-278. doi: 10.1007/s10862-005-2407-7

Weiss, M.J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions, 14*, 3-22. doi: 10.1002/(SICI)1099-078X(199901/03)14:1<3::AID-BIN25>3.0.CO;2-F

Weisz, J.R., Chorpita, B.F., Palinkas, L.A., Schoenwald, S.K., Miranda, J., Bearman, S.K., Daleiden, E.L., Ugueto, A.M., Ho, A., Martin, J., Gray, J., Alleyne, A., Langer, D.A., Southam-Gerow, M.A., & Gibbons, R.D. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry, 69*(3), 274-282. doi: 10.1001/archgenpsychiatry.2011.147

Weisz, J.R., Kuppens, S., Ng, M.Y., Eckshtain, D., Ugueto, A.M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K.M., Krumholz Marchette, L.S., Chu, B.C., Weersing, V.R., & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist, 72*(2), 79-117. doi: 10.1037/a0040360

Werba, B.E., Eyberg, S.M., Boggs, S.R., & Algina, J. (2006). Predicting outcome in parent-child interaction therapy: Success and attrition. *Behavior Modification, 30*(5), 618-646. doi: 10.1177/0145445504272977

Whitney, D.G., & Peterson, M.D. (2019). US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. *JAMA Pediatrics, 173*(4), 389-391. doi: 10.1001/jamapediatrics.2018.5399

Youngstrom, E.A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*, 204-221. doi: 10.1093/jpepsy/jst062

Zisser, A., & Eyberg, S.M. (2010). Parent-child interaction therapy and the treatment of disruptive behavior disorders. In J.R. Weisz & A.E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (p. 179-193). Guilford Press: New York.

VITA

NATALIE HONG

2009 – 2013	B.S., Applied Psychology New York University
2013 – 2016	Program Coordinator Weill Cornell Medicine Center for Autism & the Developing Brain
2016 – 2019	M.S., Psychology Clinical Science in Child and Adolescent Psychology Florida International University
2019 – Present	Doctoral Candidate, Psychology Clinical Science in Child and Adolescent Psychology Florida International University

2018	Academy of Psychological Clinical Science Flexibility Fund Award
2019	Selective Mutism Association Thomas J. Kehle Award
2019-2021	NIH National Research Service Award (NRSA; F31MH122101)
2021	Leadership Education to Advance Diversity Institute Fellow Award

PUBLICATIONS

Conroy, K., Hong, N., Poznanski, B., Hart, K.C., Ginsburg, G.S., Fabiano, G.A., & Comer, J.S. (2021). Harnessing home-school partnerships and school consultation to support youth with anxiety. *Cognitive and Behavioral Practice*. Online ahead of print. doi: 10.1016/j.cbpra.2021.02.007

Comer, J.S., Furr, J.M., del Busto, C., Silva, K., Hong, N., Poznanski, B., Sanchez, A.L., Cornacchio, D., Herrera, A., Coxe, S., Miguel, E., Georgiadis, C., Conroy, K., & Puliafico, A.C. (2021). Therapist-led, internet-delivered treatment for early child social anxiety: A waitlist-controlled evaluation of the iCALM Telehealth Program. *Behavior Therapy*. Online ahead of print. doi: 10.1016/j.beth.2021.01.004

Comer, J.S., Hong, N., Poznanski, B., Silva, K., & Wilson, M. (2019). Evidence base update on the treatment of early childhood anxiety and related problems. *Journal of Clinical Child and Adolescent Psychology, 48*(1), 1-15. doi: 10.1080/15374416.2018.1534208

Cornacchio, D., Furr, J.M., Sanchez, A., Hong, N., Feinberg, L., Tenenbaum, R., Del Busto, C., Bry, L.J., Poznanski, B., Miguel, E., Ollendick T., Kurtz, S., & Comer, J.S. (2019). Intensive group behavioral treatment (IGBT) for children with selective mutism: A preliminary randomized controlled trial. *Journal of Consulting and Clinical Psychology, 87*(8), 720-733. doi: 10.1037/ccp0000422

Fabiano, G.A., Schatz, N.K., Aloe, A.M., Pelham Jr., W.E., Smyth, A.C., Zhao, X., Merrill, B., Macphee, F., Ramos, M., Hong, N., Altszuler, A., Ward, L., Rodgers, D.B., Liu, Z., Ersen, R.K., & Coxe, S. (in press). Comprehensive meta-analysis of

attention-deficit/hyperactivity disorder psychosocial treatments investigated within between group studies. *Review of Educational Research*.

Frost, K.M., Hong, N., & Lord, C. (2017). Correlates of adaptive functioning in minimally verbal children with autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities, 122*(1), 1-10. doi: 10.1352/1944-7558-122.1.1

Furr, J.M., Sanchez, A.L., Hong, N., & Comer, J.S. (2020). Exposure therapy for childhood selective mutism: Principles, practices, and procedures. In Peris, T.S., Storch, E.A, & McGuire, J. (Eds.) *A Clinician's Guide to Exposure Therapy with Children and Adolescents*. New York, NY: Elsevier.

Hong, N. & Comer, J.S. (2018). High-end specificity of the Attention-Deficit/Hyperactivity Problems Scale of the Child Behavior Checklist for ages 1.5-5 in a sample of young children with disruptive behavior disorders. *Child Psychiatry and Human Development, 50*(2), 222-229. doi: 10.1007/s10578-018-0834-4

Hong, N., Cornacchio, D., Pettit, J.W., & Comer, J.S. (2019). Coal mine canaries in clinical psychology: Getting better at identifying early signals of treatment nonresponse. *Clinical Psychological Science, 7*(6), 1207-1221. doi: 10.1177/2167702619858111

Hong, N. & del Busto, C. (2019). Collaboration, scaffolding, and successive approximations: A developmental science approach to training in clinical psychology. *Training and Education in Professional Psychology*. Advance online publication. doi: 10.1037/tep0000260

Hong, N., Feinberg, L., Garcia, D., Comer, J.S., & Bagner, D. (2018). Internet-delivered Parent-Child Interaction Therapy for children with Autism Spectrum Disorder: rationale, considerations, and lessons learned. In McNeil, C.B., Quetsch, L.B., & Anderson, C.M. (Eds.) *Handbook of Parent-Child Interaction Therapy for Children on the Autism Spectrum* (pp. 545-557). New York: Springer. doi: 10.1007/978-3-030-03213-5_30

Hong, N., Sanchez, A.L., & Comer, J.S. (2020). Multi-media field test: Can users strike out OCD with the nOCD app? *Cognitive and Behavioral Practice, 27*, 93-99. doi: 10.1016/j.cbpra.2019.02.005

Schatz, N.K., Aloe, A.M., Fabiano, G.A., Pelham Jr., W.E., Smyth, A., Zhao, X., Merrill, B., Macphee, F., Ramos, M., Hong, N., & Altszuler, A.R. (2020). Psychosocial interventions for ADHD: Systematic review with evidence and gap maps. *Journal of Developmental and Behavioral Pediatrics, 41*(2), S77-S87. doi: 10.1097/DBP.0000000000000778

Zike, I., Xu, T., Hong, N., & Veenstra-VanderWeele, J. (2017). Rodent models of Obsessive Compulsive Disorder: evaluating validity to interpret emerging neurobiology. *Neuroscience, 345*, 256-273. doi: 10.1016/j.neuroscience.2016.09.012