DO EMPIRICALLY-SUPPORTED PACKAGES OR THEIR PRACTICES PREDICT SUPERIOR THERAPY OUTCOMES FOR YOUTH WITH CONDUCT PROBLEMS?

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
IN
PSYCHOLOGY
AUGUST 2011

By
Danielle M. Denenny

Thesis Committee:

Charles Mueller, Chairperson
Brad Nakamura
Yiyuan Xu

Keywords: Multisystemic therapy, treatment, common elements, propensity scores

ACKNOWLEDGMENTS

This work would not have been possible without the extraordinary assistance of my committee, the continuous encouragement of my colleagues at CAMHD and UHM CCBT, the generous spirit of RET researchers David and Ryan, and the amazing support of my family and friends.

I am exceptionally fortunate to have partnered on this project with a first-rate committee. My mentor, Dr. Mueller, has encouraged me at every opportunity. From the beginning, he was invaluable in challenging me to refine my ideas, learn new skills, find new ways to use the data at my disposal, and apply my findings in ways that will practically benefit our local families and clinicians. Dr. Nakamura brought a wealth of knowledge and resources concerning evidence-based science and inspired me to ground my project in this important growing field. Dr. Xu was infinitely patient with my questions and emails, and provided insightful comments and direction regarding a variety of methodological issues. These mentors have been exceptionally responsive and kind in teaching me the art and science of productive collaboration. They have made this project a most rewarding and enjoyable learning experience.

For supporting me through this incredible journey, I would also like to personally thank my family and friends. There are no adequate words to express my gratitude to each one of you – Jen, Judy, Yvonne, Yvette, Allison, Henri, Trina and Krista.

ABSTRACT

Treatment response for adolescents with conduct problems is often insufficient, while poor response is associated with a trajectory of negative long-term outcomes. The field has chiefly responded by disseminating and implementing empirically-supported treatment packages (ESTs). An alternative for furthering outcomes is the derived elements approach, which promotes therapist utilization of a high proportion of 'practices derived from ESTs' (PDEs). There has been little research to date demonstrating the real world outcomes of this approach. The primary aim of the present study was to conduct a head-to-head comparison of these dual approaches to delivering evidence-based services for adolescents with conduct problems. Under a rigorous propensity match protocol, this study examined whether level of PDE content or allocation to EST better predicted therapist-rated treatment goal progress in a sample of adolescents with conduct problems. A secondary aim was to investigate the extent to which PDE content mediated the relationship between the application of the EST (versus treatment as usual) and therapy progress. Provision of Multisystemic Therapy (MST), the selected EST package, and PDE content both predicted positive five month treatment outcomes with EST package predicting more variance in outcome. PDE content was a small partial mediator of the effect of treatment type on therapy progress. Results suggest future directions for research on PDEs, and provide further support for the derived elements approach as a complementary strategy to ESTs for bolstering treatment outcomes for this population.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA .....Analysis of Variance

CAFAS.....Child and Adolescent Functional Assessment Scale

CAMHD.....Child and Adolescent Mental Health Division of Hawai‘i Department of Health

CAMHMIS ....Child and Adolescent Mental Health Management Information System of CAMHD

DSM-IV-TR...American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revised

EST .....Empirically-supported treatment

FERPA .....Family Educational Rights and Privacy Act

HIPAA .....Health Insurance Portability and Accountability Act

IIH .....Intensive In-Home Services Therapy

LR .....Linear regression model

MST .....Multisystemic Therapy or Multisystemic Treatment

MTPS .....Monthly Treatment Progress Summary

PDE .....Practices derived from empirically-supported treatment packages

PSM.....Propensity score matching

RET .....Research and Evaluation Team jointly operated by CAMHD and the University of Hawai‘i at Mānoa Department of Psychology

SPSS .....Statistical Package for the Social Sciences software program

CHAPTER 1. INTRODUCTION

OVERVIEW

PURPOSE OF THE CURRENT STUDY

Within children's mental health care today there is unprecedented support for evidence-based services as a means to improve youth outcomes (Garland, Hawley, Brookman-Frazee, & Hurlburt, 2008; Lilienfeld, 2005). The most common evidence-based services implementation strategy has been to develop and disseminate high-fidelity manualized treatment packages (i.e., empirically-supported treatments or 'ESTs') that have strong empirical support from controlled efficacy trials. This solution has been incomplete, however, as effectiveness research has demonstrated that fidelity to EST models is poor in applied settings and low fidelity predicts little or no clinical improvement (Eames et al., 2009; Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002). Moreover, experts disagree over the evidence-based status of several therapy packages (Lilienfeld, 2005; Littell, 2005; Silverman & Hinshaw, 2008) and research on many common therapies has been weak or non-existent (Borntrager, Chorpita, Higa-McMillan, & Weisz, 2009). An alternative 'derived elements approach' to implementation of evidence-based services focuses on optimization of the specific therapy content provided to clients. Therapists are encouraged to maximize their use of 'practices derived from the evidence-base' (PDEs) or case 'PDE content,' defined herein as the proportion of practice elements common to the aggregated body of EST protocols for a particular disorder out of all therapeutic techniques and procedures delivered to a client.

To date, studies have demonstrated substantial overlap in practice content among ESTs for conduct problems, lending initial support for the feasibility of the derived elements approach (Chorpita, Daleiden, & Weisz, 2005; Garland et al., 2008; McLeod & Weisz, 2010). On the other hand, few empirical studies have tested the association between PDE content and treatment outcomes, in part because of the limited attention placed on PDEs within health agencies. More research is needed on the effectiveness of the PDE approach to keep pace with its growing appeal among scientists and practitioners alike. Some putative clinical advantages of a focus on PDE content is its

intrinsic theoretical usefulness for treatment planning, case supervision, and therapist training; there is also good acceptability of the PDE approach among therapists due to greater perceived flexibility in applying empirically-supported practices to youth with complex clinical presentations (Borntrager et al., 2009; Perkins et al., 2007). From a research vantage, focus on PDE content promises to address several important gaps in empirical knowledge such as better understanding of treatment mediators and of similarities and differences between ESTs and usual care therapies. Such knowledge could suggest treatment refinements and have significant implications for agencies looking to invest their limited resources in the evidence-based services strategy with the strongest sustainable benefits.

The present investigation was one of the first to bridge research on both approaches to conceptualizing and delivering evidence-based services. The overarching goals were to compare conventional therapy with a gold-standard EST for conduct problems (i.e., Multisystemic Therapy, MST) on mental health outcomes and level of PDE content, and to examine the explanatory power of PDE content for outcomes. While few systems of care have invested in both the EST and derived elements approaches, the State of Hawai‘i Child and Adolescent Mental Health Division (CAMHD) has maintained standardized longitudinal case records of therapy targets, practices, and psychiatric outcomes for nearly a decade. Therefore, CAMHD was uniquely suited as a partner and has made this project both feasible and compelling.

LITERATURE REVIEW

PSYCHOTHERAPY OUTCOMES FOR YOUTH WITH CONDUCT PROBLEMS

Conduct problems are the primary reason for youth referral to the public mental health system in the United States (Accurso, Taylor, & Garland, 2011). While there is strong empirical support for parent management training and behavior modification for conduct problems in children, treatment response within the adolescent conduct population is often insufficient despite the widespread availability of ESTs for adolescents (Kazdin, 2000). Critically, poor response is associated with great community expense, youth and family distress, and a trajectory of negative long-term outcomes (Copeland, Miller-Johnson, Keeler, Angold, & Costello, 2007). Unfortunately,

researchers have found few treatment-level factors beyond therapist fidelity to the treatment package and clinical supervision characteristics (Accurso et al., 2011; Huey, Henggeler, Brondino, & Pickrel, 2000) to account for the large unexplained variance in treatment response that has plagued psychotherapies including MST (Masi et al., 2011).

PRACTICES DERIVED FROM THE EVIDENCE BASE (PDES)

At the level of clinical techniques, several individual therapy practices have received some support in the conduct recovery literature including parent monitoring, rewards, family engagement, and fostering youth contact with positive peers and adult role models (Lahey et al., 1977; McMahon, Wells, & Kotler, 2006; Robbins et al., 2002; Walker, Ramsey, & Gresham, 2004). However, there is insufficient research to date to link most specific practices to outcomes or to the adolescent population with serious conduct problems. Thus it is premature to code the vast majority of practice elements as either evidence-based or not for adolescents with conduct problems. In the absence of such a list, to move the evidence base on practices forward, a limited number of researchers have turned to PDEs, a proxy variable for evidence-based practices.

In operationalizing the PDE construct, Chorpita and colleagues turned to the youth treatment literature to empirically derive a set of 55 'practice elements' or definable, discrete building blocks of clinical interventions (Chorpita, Becker, & Daleiden, 2007; Chorpita et al., 2005; Garland, Hurlburt, & Hawley, 2006). Chorpita subsequently published the cumulative frequencies of these practice elements within all existing 'winning protocols,' defined as treatments that outperformed another treatment or tied with an established treatment in a randomized clinical trial showing a significant main outcome (Chorpita & Daleiden, 2007, 2009). According to Chorpita & Daleiden's 2007 list, the practice elements with the highest winning frequency counts (all above 30%) for adolescents age 16 and older with serious conduct problems are (in rank order): communication skills, problem solving, social skills training, cognitive coping, parent monitoring, parent praise, maintenance or relapse prevention, modeling, response cost, tangible rewards, therapist praise or rewards, family therapy, and skill building. These PDEs showed the expected level of moderate overlap with the list of derived practices for younger children with disruptive behavior disorders that was generated by Garland and

colleagues (2008) using an approach that combined empirical and expert rankings.

A challenge to research on the PDE construct is how to best translate the existing science into lists of elements that are binary coded as 'PDE - yes/no' for conduct problems. There is no standardized minimum frequency coding cutoff for determining the percentage of ESTs for conduct problems in which an element must appear before it is considered a PDE. Therefore, researcher bias is introduced at some point in coding protocols (Garland et al., 2006).

With the first PDE lists emerging for conduct problems, researchers are beginning to describe the PDE content of therapies and the relationship between PDEs and treatment response. One such study found that most usual care therapy sessions provided at least one PDE, whether directed at children or their caregivers, albeit at relatively low intensity (Brookman-Frazee, Haine, Baker-Ericzen, Zoffness, & Garland, 2010). According to the Brookman-Frazee line of research (2010), the factors associated with higher PDE content were greater youth age and symptom severity, higher parental economic status, lower therapist experience, and therapist cognitive-behavioral or behavioral orientation. There is no comparable research for MST. Furthermore, there is minimal past research to suggest how specific strategies are associated with different child or family outcome trajectories. The few studies conducted suggest the relationship between greater PDE content per case and greater treatment response in youth is somewhat positive but variable. In a study by Mueller and colleagues (2010), the addition of each PDE to the mean monthly load over the first 7.5 months of usual care therapy, controlling for non-PDEs, was estimated to increase the mean rate of global functioning improvement by 13% for youth with disruptive behaviors. However, another analysis found that an increase in the mean monthly load of PDEs provided to young clients had a significant impact on recovery from Attention Deficit Hyperactivity Disorder but not disruptive behavior disorders (Mueller & Tolman, 2009). There are significant limitations on the generalizability of these results, however, due to the small number of studies and complexities of dose-response relationships (Hansen, 2000).

TREATMENT CONDITIONS

Multisystemic Therapy

MST (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998; 2009) was one of two treatments provided in this study. MST is a family-centered, home- and community-based behavioral therapy alternative to restrictive placement for adolescents with serious antisocial or disruptive behaviors. MST was implemented in Hawai‘i in 2001 with a high level of involvement by its developers to facilitate clinical treatment integrity as part of a state-wide effort to implement evidence-based services for youth with conduct problems (CAMHD, 2003). During the entire duration of the present study, CAMHD contracted with Parents and Children Together and The Institute for Family Enrichment service agencies to provide MST to 250 youth per year on average. These agencies continually exceeded fidelity benchmarks and met full MST licensure status, which signifies that their therapy teams met the required model criteria (quality assurance monitoring, program drift monitoring, and agency consultation with MST Services) under the oversight of MST Services, the MST licensing organization affiliated with the Medical University of South Carolina.

Theoretical justification for the MST model is found in Bronfenbrenner’s Social Ecology Theory (1979) and extensive research demonstrating that conduct problems are maintained by interacting factors within the youth, family, and social ecology (Curtis, Ronan, & Borduin, 2004; Haley, 1976; Minuchin, 1974). As detailed in the clinical volumes authored by MST developers (Henggeler et al., 1998; 2009), the model is operationalized through a team structure, supervision guidelines and tools, and nine treatment principles. These principles encourage therapists to conduct an ongoing functional analysis of the problem set, and to remain strengths-based, family-driven, action-oriented, and focused on long-term goals and maintenance of gains (Henggeler et al., 2009). Therapists borrow from a variety of treatment modalities (e.g., strategic family therapy, cognitive behavior therapy) in accordance with their training background, the preferences and needs of families, supervisor recommendations, and evidence-based practices. During sessions, therapists assist families with accessing and managing services, and impart skills to clients. Therapists also provide 24-hour/7-day-a-week

availability, offering crisis and evening support with no limit on family contact hours. This intensive intervention often lasts four to six months and involves 60 hours or more of direct contact between the therapist and the adults in the youth's natural environment (Henggeler et al., 1998; 2009).

The program begins with an ideographic assessment and clear specification of problem behaviors, with the aim of stakeholder consensus on targets. Next, the therapeutic team designs specific interventions for those factors. Common MST targets mentioned in the literature include parent behavior management skill, youth anger management, family functioning, delinquent peer involvement, and family supports (Barth et al., 2007; Borduin, 2007; Henggeler et al., 1998; 2009; Huey et al., 2000). When possible, interventions are integrated to maximize gains (e.g., the youth is trained in relaxation skills while the parent learns to implement a rewards system to shape the adolescent's use of relaxation skills outside session). Suggested targets strongly overlap with proposed MST mechanisms of change (i.e., family relations, increased caregiver follow-through on discipline, and decreased delinquent peer affiliation) and with the wider literature on mechanisms of recovery in conduct populations (Eddy & Chamberlain, 2000; Henggeler et al., 2009; Huey et al., 2000; Schaeffer, 2001). Another emphasis of the MST program structure is on fidelity to the model, which is supported by studies suggesting fidelity mediates MST outcomes of criminal behavior, number of days in out-of-home placement, and family functioning (Huey et al., 2000).

The MST program also specifies staff roles and responsibilities. The manual recommends two to four therapists share caseloads of four to six families, with each team supervised by a master's- or doctoral-level supervisor. MST team members often work for private service organizations under contracts from public criminal justice, child welfare, and mental health agencies. Supervisors are considered to be experts in the protocol, although there is no official certification process. Weekly individual clinical supervision is provided that focuses on case-specific data captured weekly by therapists on standardized report forms such as the Fit Circle Diagram (Henggeler et al., 2008). Supervisor responsibilities include guiding therapists to adhere to the MST model and principles. At the program level, therapy teams are overseen by the agency's Expert

System Supervisor, whose added responsibilities include trainings, quality assurance, and case consultation.

Regarding program effectiveness, MST has sufficient empirical support to be categorized as an EST by leading reviewers (e.g., Kazdin, 1997) and United States policy makers (U.S. Department of Health and Human Services, 2001). MST is also 'well-established' under the American Psychological Association's criteria, with estimated small to medium effect sizes widely reported both for entry-exit effects and short-term effects when compared to individual therapy and restrictive placement. A key meta-analysis of MST that has been used as a benchmark study of between-group effect sizes found that efficacy studies have generally yielded significantly larger mean treatment effects (*Cohen's d* = .81) than effectiveness studies (*d* = .27; Curtis et al., 2004). With respect to results from MST effectiveness studies on the outcomes used in this study, Tolman et al. (2008) found small entry-exit effect sizes for improvements in therapist-rated clinical severity (*d* = 0.33) and global functioning (*d* = 0.29). Curtis et al. (2009) reported a small entry-exit effect for therapy goal progress (*d* = 0.29) and for aggregated functioning for school, criminality, and placement domains (*d* = 0.32). Timmons-Mitchell and colleagues (2006) reported an exceptionally large effect of MST on school and work functioning (e.g., *d* = 1.39).

The scientific literature on MST also addresses questions about instrumental versus ultimate outcomes, and treatment attrition. There have been mixed reports as to whether effect sizes in MST are generally higher for outcomes that are considered instrumental (i.e., targeted therapy goals) or ultimate (e.g., placement, school/community functioning), as reported in a summary table by Curtis and colleagues (Curtis et al., 2009). Furthermore, published rates of completion for MST have generally fallen within the range of 76% to 100% (Curtis et al., 2009; Rowland et al., 2005; Tolman et al., 2008).

While these findings are generally supportive of the EST status of MST, there are methodological weaknesses in the existing literature and few reports on longer-term outcomes (Littell et al., 2005; Littell, Popa, & Forsythe, 2005; Rowland et al., 2005; Tolman et al., 2008). Independent-lab effectiveness studies are needed to examine MST against community-based usual care and to further explain the post-dissemination decrease in effect size.

Intensive In-Home Treatment

IIH is a home- and community-based therapy program in which the therapist works with a family to build youth skills, teach parents strategies to better manage youth behaviors, and improve the quality of parent-youth interactions. Studies of IIH indicate therapists utilize multiple therapy strategies that originate from many different theoretical orientations (Stroul & Goldman, 1990). IIH delivery is not limited to youth with disruptive behavior diagnoses and has been prominent in mental health and child welfare systems. IIH has important antecedents in Intensive Family Preservation Services, an intensive, short-term preventive service whose principal aim is the prevention of out-of-home placement for at-risk families (Kinney, Madsen, Fleming, & Haapala, 1977). IIH is also referred to as home-based services, family-centered services, family-based services or intensive family services (Burns, Hoagwood, & Mrazek, 1999).

The IIH program for conduct problems is neither standardized nor manualized, and is delivered in Hawai‘i as the usual care community-based alternative to packaged programs, with limited program-level support and oversight. Decisions about therapy strategies and targets are made by individual therapists and their supervisors based upon clinical judgment and therapeutic orientation. In the Hawai‘i public health system, IIH therapists are contracted through CAMHD and are required to adhere to the standards of care outlined by the CAMHD policy manual (CAMHD, 2006). CAMHD standards for IIH mandate that therapy is provided by bachelor’s or master’s-level therapists. Service delivery must also be preceded by a thorough assessment of the youth and their family, and include evidenced-based interventions, weekly direct contact hours, coordination with other treatment providers, and 24-hour availability for crisis management. CAMHD policies also require that therapy goals and treatment decisions are made collaboratively by an integrated treatment team of therapists, social workers, physicians, school staff, and caregivers. Within CAMHD, IIH therapists typically have 1.5 hour limits on their weekly contact hours with individual families, higher caseloads, and service episodes lasting twice as long (Hurley, 2008; Denenny & Mueller, 2012). Typically, clinical supervision is less data-driven and provided by a master’s-level professional in a weekly group format.

Effectiveness studies of IIH have yielded mixed results and must be interpreted with caution as IIH treatments vary widely in structure and service intensity. Improvements have been noted within IIH on targeted therapy goals (Levins, 1998), family functioning (Hinckley & Ellis, 1985) and need for out-of-home placement (Miller, 2006; U.S. Department of Health & Human Services, 2000). There is also some support for change in youth behavior and parental effectiveness compared to a waitlist control (e.g., Lewis, 2005). On the other hand, several other studies have found little evidence for IIH effectiveness (e.g., Barton, Baglio, & Braverman, 1994; Dagenais, Begin, Bouchard, & Fortin, 2004; Lindsey, Martin, & Doh, 2002; Littell & Schuerman, 2002; Schuerman, Rzepnicki, & Littell, 1994). Regarding rates of treatment drop-out for individual and family treatments of youth with conduct problems, estimates have been reported as high as 50% (Curtis et al., 2009; Rowland et al., 2005).

Several factors have been suggested to impact the effect of IIH on out-of-home placement and treatment response, including age, gender, antisocial behavior, length of episode, and parental difficulties (Hurley, 2008). Although it remains to be tested, a potential mediator of IIH effect could also be PDEs, which are delivered with low intensity in less structured community-based usual care therapies (Accurso et al., 2011; Garland et al., 2010).

SPECIFIC AIMS AND HYPOTHESES

Specific aims of the present study were threefold. Analysis I was an effectiveness test of MST and IIH under a propensity score match protocol (PSM; Rosenbaum & Rubin, 1983). PSM is a rigorous quasi-experimental design that controls for pre-treatment threats to internal validity and allows for the estimation of causal effects from observational data. The intent of this analysis was to determine whether MST outperforms IIH on therapy goal progress or global functioning after the first months of treatment. The superiority of one treatment over the other was seen as a precondition for the subsequent analyses aimed at helping to explain the performance gap. Although the evidence base for MST is mounting, the extent to which MST outperforms community-based IIH in applied care settings with community therapists remains relatively untested. Nonetheless, based on the EST status of MST and prior uncontrolled comparison studies

of these treatments within CAMHD, the hypothesis for Analysis I was that MST would outperform IIH across both outcome measures.

Analysis II was designed as a head-to-head test of the predictive effect of the type of treatment (MST or IIH) and PDE content on therapy progress. First, the proportion of PDE content in each group was reported. Next, regression modeling was used to estimate the extent to which PDE content explained youth response to treatment. A significant effect was expected for PDE content on progress in both groups given that past research has demonstrated an association between higher PDE content and an increased rate of improvement in youth with disruptive behavior disorders (Mueller, Tolman, Ebesutani, & Bernstein, 2010). Dominance analysis was used to determine the relative strength of PDE content and treatment type as predictors of progress. The partial effect for each predictor was expected to be moderate after controlling for the other predictor. No prior research existed to suggest the relative importance of these predictors. A sizeable decrease in the effect of treatment type was expected when controlling for PDE content, as content is commonly theorized to be the active ingredient of therapy (Chambless & Hollon, 1998, Task Force on Promotion and Dissemination of Psychological Procedures, 1995). It was further predicted that the MST package as a whole would have benefits above and beyond the level of PDE content as MST principles and structure emphasize several characteristics (i.e., early therapy dose, hours of supervision, ideographic assessment) that have been shown to predict positive treatment outcomes (Accurso et al., 2011) and these characteristics were not controlled in this study. An exploratory analysis was also run to assess for differences between treatments in the predictive strength of PDE content for treatment progress, with a larger effect anticipated for IIH given the EST status of MST and the lower variability in scores expected for the MST group compared to the IIH group.

Analysis III was intended as a complementary method for exploring whether PDE content was predominantly responsible for the positive effect of allocation to MST on treatment progress. Specifically, the extent to which PDE content mediates the relationship between treatment type and therapy progress was examined. This is the first known study of PDE content as a potential mediator of treatment. This question is particularly interesting as it builds upon recent work isolating mediators of MST (e.g.,

therapist model adherence on outcomes; e.g., Huey et al., 2000) and could suggest key program refinements for MST and IIH. In the present study, positive partial mediation was expected under the logic outlined for Analysis II; no prediction was made regarding the extent of mediation.

CHAPTER 2. METHODS

SOURCE OF THE DATASET

The present investigation was supported by the CAMHD Research and Evaluation Team (RET), comprised of researchers from CAMHD and the University of Hawai‘i at Mānoa Department of Psychology. RET provides leadership in child and adolescent mental health research and evaluation and creates service-learning opportunities in behavioral health research and evaluation (Chorpita & Mueller, 2008). Staff ensure the systematic collection of case data (e.g., data from monthly clinician-report measures) that is single-entered into the agency’s comprehensive health records database, the Child and Adolescent Mental Health Management Information System (CAMHMIS) under standard operating procedures (CAMHD, 2006; Nakamura, Daleiden, & Mueller, 2007). All data utilized in this study was originally housed in CAMHMIS and extracted per the protocol below.

HUMAN SUBJECTS CONSIDERATIONS

This study was approved and overseen by the University of Hawai‘i at Mānoa Committee on Human Studies Institutional Review Board. Prior to data collection for all participants, youth and their legal guardians underwent standard CAMHD consent procedures pertaining to research use of private client information and also received a standardized notice of rights to privacy. Upon receipt of the formal data request and IRB guidelines, RET staff extracted and electronically transferred the requested de-identified dataset so that no personal public health information was accessed by the researcher. Health Insurance Portability and Accountability Act (HIPAA) and Family Educational Rights and Privacy Act (FERPA) standards were met for the protection of private health information.

PARTICIPANTS

This study focused on intensive community-based services for youth with DSM-IV-TR disruptive behavior diagnoses (conduct, oppositional defiant, disruptive behavior not otherwise specified, or adjustment with conduct). Within the CAMHD service array, the primary options for youth with disruptive behavior disorders were MST, IIH, therapeutic

foster homes, group homes, community-based residential programs, and hospital residential services. All study participants participated in an initial CAMHD eligibility process that included a thorough mental health assessment, verification of DSM-IV-TR diagnoses, and an insurance qualification assessment. Treatment assignments were formalized in a process involving multiple stakeholders, and were based on criteria including the frequency, intensity, and duration of the youth's willful misconduct behaviors; the degree of the youth's impaired functioning; the presence and nature of comorbid diagnoses; and response to prior treatment interventions. Both MST and IIH services required that the youth have an Axis I diagnosis and level of care status described as either exiting out-of-home placement or at imminent risk for out-of-home placement. MST also required clinically-severe conduct behaviors and caregiver willingness to participate in therapy. Within CAMHD, highly impaired youth younger than age 12 were occasionally made eligible for MST services, in a divergence from the program-validated age range of 12 and up. These treatment assignment policies are more fully described in the CAMHD Interagency Performance Standards and Practice Guidelines (Hawai‘i Departments of Education and Health, 2006).

At the beginning of data capture for this study, all cases of youth who began a program of IIH or MST treatment between July 1, 2003 and June 30, 2010 were reviewed. Only youth age 11 or older who were participating in their first episode of MST or IIH (with no prior participation in either therapy) that lasted longer than 30 days were considered. Exclusionary criteria included missing information on diagnosis, gender, and age. Due to youth ineligibility for one or both treatments within CAMHD, youth with any of the following problem sets were automatically excluded at the outset of the study: primary mental disorder due to general medical condition; psychotic diagnosis; moderate to severe mental retardation; a primary autism spectrum diagnosis; or restricted problem of sexual misbehavior.

The remaining sample included 619 cases of MST and 2619 cases of IIH. Data from this sample were subsequently screened based on availability of measurement data and diagnostic fit for the study. Cases were eligible if available data included a baseline Child and Adolescent Functional Assessment Scale (CAFAS; Hodges, 1998, 2003), a follow-up CAFAS, and two or more Monthly Treatment Progress Summary datasets (MTPS;

CAMHD, 2008). This study used listwise deletion for handling missing data, which is the best available approach under the PSM framework (Allison, 2001). The final pre-PSM sample of 195 MST participants and 387 IIH participants contained only youth with a primary or secondary disruptive behavior diagnosis[1].

Next, the need for propensity score matching (PSM; Rosenbaum & Rubin, 1983) was evaluated. PSM is a procedure that controls for threats to internal validity from treatment assignment bias (Guo, Barth, & Gibbons, 2006; Barth et al., 2007). First, we identified several youth variables (age, gender, co-morbidity, substance abuse, baseline level of functioning and primary disruptive behavior diagnosis) that have been linked in the literature with treatment response in this population (Daleiden, Pang, Roberts, Slavin, & Pestle, 2010; Farmer, Mustillo, Burns, & Holden, 2008; Hurley, 2008; Rowland et al., 2005; Tolman et al., 2008). Next, Chi-square and *t*-tests were used to identify those variables that differed between treatment types. Significant differences were noted for baseline functioning (t (580) = -3.96, $p < .001$; MST $M = 103.3$, $SD = 28.5$; IIH $M = 92.7$, $SD = 31.6$); gender ($\chi^2 = 10.20$, $p < .01$; MST $M = .74$, $SD = .44$; IIH $M = .60$, $SD = .49$); and comorbid mental health diagnosis ($\chi^2 = 4.41$, $p < .04$; MST $M = .56$, $SD = .50$; IIH $M = .65$, $SD = .48$).

As a next step in PSM, logistic regression modeling (stepwise backward likelihood ratio method) was performed to determine which variables distinguished between the treatment types, after controlling for their joint dependency. As recommended by Barth and colleagues (2007), a significance level of .30 was required for variable entry into the model, and a significance level of .35 was required for retention. Age ($B = -.13$, $SE_B = .05$, $p = .02$), baseline functioning ($B = .01$, $SE_B = .00$, $p = .00$), gender ($B = -.62$, $SE_B = .20$, $p = .00$), and comorbid mental health diagnosis ($B = .41$, $SE_B = .19$, $p = .03$) were related to treatment type (model $R^2_N = .06$) and thus retained as the conditioning set. Next, the probability of each case receiving MST based on this conditioning set (i.e., propensity score) was estimated using the binary logistic regression propensity module in SPSS 19. The propensity score for MST is defined as the logit of the predicted probability (i.e., log p/ (1 – p)). Finally, an SPSS macro, written by Raynald Levesque

[1] A separate series of analyses were run for youth with any externalizing disorder, with relatively similar results for the separate sample. These data can be requested from the author.

and adapted by John Painter (2004), was used to match each youth in MST to a similar youth in IIH based upon propensity scores using the recommended setting for nearest case without replacement assignment and a conservative caliper (.10 in standard deviation units) to prevent dissimilar matches (Barth et al., 2007).

One hundred seventy one pairs of youth resulted from the propensity match protocol and are hereafter referred to as the matched sample. Data from this sample were used in all of the remaining analyses. Matched sample means for demographics, diagnoses, baseline functioning, and duration of measurement period are reported in Table 1. Chi-square and *t*-tests revealed group differences were not significant with regard to any demographic, diagnostic, or baseline functioning characteristics. Forty-six percent of youth carried a primary or secondary conduct disorder diagnosis and 88% carried at least one comorbid diagnosis. Additionally, most youth carried a primary diagnosis of disruptive behavior, mood, or attentional disorder, while a significant portion of the sample carried a primary or secondary attentional disorder, substance-related disorder, or adjustment disorder. Multiethnic heritage (66%) was most frequently endorsed for ethnicity. According to state census data, most multiethnic youth in Hawai‘i are part-Asian and/or part-Pacific Islander. Scores were similarly high (exceeding 100) for the treatment types on CAFAS global functioning scores nearest to treatment start date, which represented initial global functioning. Overall it can be said that the reported data on demographics, intake functioning and diagnostics are consistent with previous studies of this population within CAMHD (e.g., Rowland et al., 2005; Tolman et al., 2008) and also reflect levels typical of juvenile offender samples (Hodges, 2004; Rogers et al., 2006).

Table 1 also displays the mean duration of the measurement periods for each group. The MST group had a shorter mean measurement period for both CAFAS (t (340) = 2.61, $p < .05$) and MTPS (t (340) = 5.72, $p < .001$). MTPS progress ratings were collected near the mean end of the treatment episode for the MST sample, while CAFAS functioning ratings were collected near the mean end of the treatment episode for the IIH sample.

Table 1. *Matched Sample Characteristics by Treatment Group*

Matched Sample Characteristics		MST	IIH
Gender (% of total)	Male	71%	74%
Mean age in years (*SD*)		15.2 (1.7)	15.1 (1.7)
Intake functioning (CAFAS)	Mean global (*SD*)	105.1 (23.3)	103.8 (25.4)
	Community or general behavior – high impairment	91%	94%
Ethnicity (% of total)	African	3%	1%
	American Indian or Alaska Native	1%	0%
	Asian	8%	11%
	Hispanic or Latino	4%	3%
	Native Hawai‘ian / Other Pacific Islander	9%	11%
	White or Caucasian	9%	12%
	Multiethnic	63%	58%
	Not Available	4%	5%
Primary diagnosis	Disruptive	63%	64%
	Substance-related	3%	4%
	Attentional	11%	8%
	Mood	13%	15%
	Adjustment	2%	3%
Other diagnostic information	Any	45%	48%
	Any Substance-related	28%	29%
	Any Attentional	25%	19%
	Any Mood	28%	39%
	Any Comorbidity	85%	90%
Time variables	Mean duration of treatment	4.3 (1.6)	7.4 (6.5)
	Mean measurement period (MTPS ratings, mos.)	4.3 (1.4)	5.4 (2.1)
	Time from episode start to 2nd CAFAS (months)	7.0 (2.1)	7.5 (1.9)
	% completed treatment episode by MTPS	57%	50%
	% completed treatment episode by CAFAS	90%	67%

MEASURES

Instruments for this study were carefully selected to tap constructs empirically linked to important mental health outcomes for this population, with particular attention to measures that had previously demonstrated adequate or stronger reliability, validity, temporal stability, and sensitivity to treatment. Response to treatment was operationalized using both the CAFAS global composite and MTPS progress ratings. Symptomatic and functional outcomes were included as complementary measures; research shows client scores frequently deviate on these measures as they capture distinct aspects of treatment progress (Bacon, Collins, & Plake, 2002; Nakamura et al., 2007). Previous studies suggest the CAFAS scales and MTPS progress ratings have a significant inverse relationship when used for short-term treatment tracking (r = -.3; Nakamura et al., 2007).

MONTHLY TREATMENT AND PROGRESS SUMMARY

The Monthly Treatment Progress Summary (MTPS; CAMHD, 2008) is a therapist-report measure designed to capture standardized monthly data on 55 practice elements, 48 treatment targets, and clinical progress ratings for each CAMHD treatment case. During the study period, CAMHD provided periodic statewide trainings on the MTPS and ensured therapist access to codebooks that provided rater instructions and item definitions. This study utilized the data on practice elements and clinical progress ratings. Detailed information on the psychometric properties of practice elements and clinical progress ratings has been reported elsewhere in the literature (Orimoto, Higa-McMillan, Mueller, & Tolman, 2009; Nakamura et al., 2007; Weersing, Weisz, & Donenberg, 2002).

PDE Content

PDE content was calculated based on client MTPS practice element data and a published list of practice element frequencies within ESTs for youth with disruptive behavior disorders ages 16 and up (Chorpita & Daleiden, 2007[2]). The first step in the

[2] All analyses were run based upon both the 2007 and 2009 lists, with the 2007 list selected as the main list due to the fact it was derived from studies conducted with adolescent participants closer in age to the study sample. Only the 2007 results are reported herein as 85% of the PDEs on both lists were the same, and results from the 2009 analyses were similar.

protocol involved binary coding of the list of 55 practice elements as PDEs (or not), using a 30% minimum cutoff criterion. This criterion was selected to balance the importance of avoiding a ceiling effect and low variance, while also including within the set many elements supported as best practices for disruptive behavior disorders. As previously noted, there is no universal minimum frequency cutoff for recoding elements as PDEs for conduct problems. To test the sensitivity of the models to the selected cutoff, more restrictive coding criteria and different age-adjusted frequency lists were also explored[3]. Following coding of practices, the PDE content for each case was calculated as the proportion of all practice elements endorsed within the available data that were PDEs for youth with disruptive behavior disorders[4]. Elements were captured throughout the first eight months of the episode, with the mean measurement date at 3.26 months (*SD* = 2.49) into treatment.

Regarding published psychometric data for the MTPS practice elements measure, Chorpita and colleagues (2005) found good preliminary inter-rater reliability (k = .76) for 26 of the 55 practice elements, and acceptable one- and three-month test-retest stability (k = .65 and .50, respectively). Factor analysis of MTPS practice elements has revealed three stable factors: practices associated with behavioral interventions, self-change practices, and family supports (Orimoto et al., 2009). More research is needed on the MTPS practice elements measure to support its validity for capturing treatment process data (Orimoto et al., 2009).

Progress Ratings

Response to treatment was operationalized using the MTPS clinical progress ratings data. From the instrument's checklist of 48 items and two close-ended response fields, clinicians endorsed up to ten treatment targets (i.e., competencies or concerns/problems) that had been addressed during the past reporting month. For each treatment target endorsed, clinicians were asked to provide a rating of the youth's progress on the target

[3] Findings from these alternative models for Analyses II and III were similar to results obtained from the main model. Although not published herein, they can be requested from the author.

[4] An alternative method for calculating PDE content (sum of PDEs divided by number of MTPS datasets completed) was also tested in the subsequent analyses. Results did not change notably.

relative to his or her baseline level of functioning and the team's therapy goal. Ratings were assigned on a 7-point anchored scale ranging from "1 = Deterioration < 0%" to "7 = Complete Improvement 91–100%". From the available MTPS sets, the study captured all data from episode start to eight months, but no more than 45 days after the conclusion of the treatment episode. The ultimate outcome variable was calculated as the grand mean of all of the maximum ratings for each target endorsed during the measurement period.

Psychometric studies indicate the information on progress ratings captured by the MTPS is valid for clinical and research use and sensitive to monthly clinical change (Nakamura et al., 2007). Previous analyses also provide preliminary support for the concurrent validity and reliability of the information on treatment targets captured by the measure (Daleiden & Tolman, 2005).

Child and Adolescent Functional Assessment Scale (CAFAS)

The CAFAS (Hodges, 1998, 2003) is a provider-report scale designed to measure functioning across multiple major life domains in youth ages 5 to 18 with serious emotional disturbance. Within CAMHD and many systems of care, care coordinators serve as raters. The CAFAS global composite score ranges from 0 to 240, with higher scores indicating greater overall functional impairment. The global score is the sum of eight zero to 30-point subscale scores (Role Performance Home, Role Performance School/Work, Role Performance Community, Behavior Toward Others/Self, Moods/Self-Harm, Moods/Emotions, Substance Use, Thinking). To qualify for services under the managed Medicaid program of Hawai‘i, a youth must score 80 or higher.

Repeated measures global composite data were obtained at baseline and follow-up. Baseline CAFAS scores were used in the PSM conditioning set. CAFAS scores were selected closest to the treatment start date and within the window from 90 days prior to treatment entry to 30 days after the start of the episode. To supplement progress rating outcome data, response to treatment was also operationalized using follow-up CAFAS global composite scores captured close to eight months after treatment entry.

Psychometric studies of the CAFAS global scale and subscales indicate sensitivity to quarterly treatment change, temporal stability, and high inter-rater reliability (Hodges, 1995; Hodges & Wong, 1996). Studies have also demonstrated significant associations

between scores and severity of psychiatric diagnosis (Hodges & Gust, 1995; Hodges & Wong, 1996), criminal recidivism (Hoge, 2002), and school absenteeism (Hodges & Kim, 2000). Only small changes in the psychometric properties of the CAFAS have been reported across the many United States systems of care in which it is used for outcome monitoring and effectiveness evaluation (Rosenblatt & Rosenblatt, 2002). Nonetheless, there remain questions about the content and construct validity of existing functional impairment measures, including the CAFAS (Bates, Furlong, & Green, 2006). An additional limitation of using the CAFAS as an outcome measure for this particular study was its quarterly capture within CAMHD.

ANALYTIC APPROACH

DOMINANCE ANALYSIS

Dominance analysis (Budescu, 1993) was utilized to estimate the relative importance of PDE content and treatment type in explaining progress outcomes (Analysis II). Relative importance is defined as the contribution of each predictor to the total predicted criterion variance when the predictor is considered by itself and in combination with the other predictors in the model (Johnson & LeBreton, 2004). Research supporting the validity of dominance analysis for estimating relative importance has been published elsewhere (e.g., Azen & Budescu, 2003). Dominance analysis essentially involves calculation of general dominance weights, which are essentially the relative effect size of each predictor accounting for any collinearity between the predictors. The general dominance weight for each predictor is calculated as the average ΔR^2 obtained by adding that predictor to all possible sets of predictors, which is equivalent to the average squared semi-partial correlation for each predictor taken over all possible regression models. As recommended by Azen and Budescu (2003), rescaled dominance weights were obtained by dividing general dominance weights by the model R^2.

MEDIATION ANALYSIS

Mediation analysis was performed to examine if and how PDE content affects the relationship between treatment type and treatment progress (Analysis III). The analysis involved the causal steps strategy (Baron & Kenny, 1986; Kenny Kashy, & Bolger,

1998), the Sobel test (1982, 1986) and bootstrapping (Preacher & Hayes, 2004, 2008). All of these procedures use information on the unstandardized total, direct, and indirect effects of treatment type (X) on MTPS progress ratings (Y) through PDE content, the proposed mediator (M). The relationship among these variables is represented by the following regression equations:

1. $Y = i_1 + cX + e_1$
2. $Y = i_2 + c'X + bM + e_2$
3. $M = i_3 + aX + e_3$

where i_1, i_2 and i_3 are intercepts and e_2 and e_3 are residuals in the M and Y variables, respectively. While *a* is the path coefficient relating treatment type to PDE content, *b* is the coefficient relating PDE content to treatment progress, controlling for treatment type. Coefficient *c* and *c'* are the respective total effect and direct effect of treatment type on treatment progress.

Under the causal steps strategy, paths *a, b*, and *c* must be significant for mediation. After these significant paths are verified, if path *c'* is not significant then full mediation has occurred, yet if path *c'* is significant, then supplemental tests (i.e., Sobel test and bootstrapping) are needed to address whether partial mediation has occurred. Key limitations of the causal steps approach (e.g., low power, Type I error) have been discussed in further detail elsewhere (Preacher & Hayes, 2004).

The Sobel test is a two-tailed *z*-test of the hypothesis that the mediated effect equals zero in the population. Although Sobel test results are commonly reported for mediation analyses, a notable limitation of the Sobel test is that it relies on confidence limits for the mediated effect that are based on the normal distribution, which may be inaccurate (Preacher & Hayes, 2004).

Bootstrapping is a nonparametric procedure for testing the extent of mediation with greater accuracy (Preacher & Hayes, 2004, 2008). Bootstrapping involves creation of a confidence interval for the indirect effect based on the sampling distribution of the indirect effects for multiple resamples of the data set. One thousand resamples were generated in the present study, as recommended by Preacher and Hayes (2008). Resampling was performed using the INDIRECT SPSS script by Kris Preacher (Preacher & Hayes, 2004, 2008).

CHAPTER 3. RESULTS

CORRELATIONS AMONG STUDY VARIABLES (OTHER THAN TREATMENT TYPE)

Correlations were run among the study variables in order to better understand the matched sample and to screen for initial associations between the many predictors and outcomes. These results are shown in Table 2. A significant bivariate correlation (r = .28) confirmed the expected level of association between therapist-rated progress near five months and CAFAS global functional improvement near seven months. Higher PDE content was associated with higher MTPS progress ratings (r = .28) and greater CAFAS improvement (r = .16). As expected for a repeated measures design, higher CAFAS impairment at entry was associated with higher CAFAS impairment at follow-up (r = .26), while girls (r = -.13) and older adolescents (r = .18) showed more impairment at entry than boys and younger adolescents, respectively. Boys were more likely than girls to carry a comorbid diagnosis in our matched sample (r = .11). Age was not associated with deterioration in functioning (r = -.09), even as developmental challenges can lead to declines. Duration of the measurement period for both CAFAS and MTPS had little association with any variables including outcomes. Overall it can be said that correlations were close in magnitude to expected values and in the anticipated direction.

Table 2. *Correlation Matrix for Study Variables Other than Treatment Type*

Variables	1	2	3	4	5	6	7	8	9	10	*Mean* (*SD*) or %
1. PDE content		.28**	-.18**	.16**	-.03	.02	.02	-.08	.02	-.04	.40 (.10)
2. MTPS PR	.28**		-.28**	.28**	.03	.05	-.01	-.05	.05	-.07	4.35 (1.25)
3. CAFAS 2	- .18**	-.28**		-.80**	.26**	.03	-.07	.01	-.09	.07	87.51 (37.09)
4. CAFAS Δ[5]	.16	.28**	-.80		.38**	.09	-.02	.07	.10	-.09	16.91 (38.84)
5. CAFAS 1	-.03	.03	.26**	.38**		.18**	-.13*	.12*	.02	-.04	104.42 (24.38)
6. Age	.02	.05	.03	.09	.18**		-.04	.00	-.11*	.07	15.12 (1.67)
7. Gender	.02	-.01	-.07	-.02	-.13*	-.04		.11*	.07	-.03	72.5% (boys)
8. Comorbidity	-.08	-.05	.01	.07	.12*	.00	.11*		-.03	.02	87.7%
9. Time PR	.02	.05	-.09	.10	.02	-.11*	.07	-.03		.18**	4.80 (1.87)
10. Time CAF	-.04	-.07	.07	-.09	-.04	.07	-.03	.02	.18**		7.21 (2.02)

[5] CAFAS analyses featured regressions of CAFAS at follow-up controlling for scores at entry instead of difference scores to obviate problems with validity and reliability.

Note: All results are Pearson product-moment or point-biserial correlations. PDE content = MTPS proportion of PDEs out of total elements. MTPS PR= MTPS mean maximum progress rating scores. CAFAS 1, Δ and 2 represent respective entry, difference (post – pre), and follow-up CAFAS global composite scores. Age: unit is years. Gender code: male = 1, female = 0. Comorbidity code: present = 1, absent = 0. Time PR = months from entry to measurement. Time CAF = months from entry to final measurement. $^{*}p < .05$, $^{**}p < .01$, both two-tailed

OVERALL OUTCOMES FOR THE MATCHED SAMPLE

Descriptive statistics for the matched sample are presented in Table 2. Overall the sample contained more boys (72.5%) and more youth with comorbid diagnoses (87.7%). On average, youth were 15.1 years old (*SD* = 1.7). Sample demographic and diagnostic information were similar to data reported elsewhere for the CAMHD system (CAMHD, 2010); thus these data appear to represent an unbiased sample of this population. Additionally, based on mean CAFAS pre-treatment subscale scores in the moderate to severe range for the domains of home functioning (*M* = 21.9, *SD* = 8.8), school functioning (*M* = 23.1, *SD* = 7.8), and behavior towards others (*M* = 18.7, *SD* = 6.0), it may be concluded that this CAMHD sample of adolescents mirrors the population of adolescents with severe conduct disturbance described in the literature. Table 2 also contains the mean PDE content scores (*M* = .40, *SD* = .10) and time factors for the treatment cases under study. Time from treatment entry to follow-up measurement differed across the two outcome variables and was approximately 2.5 months shorter for MTSP progress ratings than for CAFAS scores.

Table 2 displays mean outcomes for the matched sample. For the 240-point CAFAS, paired samples *t*-tests suggested significant mean improvement (16.90 points) in global functioning ($p < .001$), with large variance among participants (*SD* = 38.8). The mean CAFAS score of 87.51 (*SD* = 37.09) did not drop below the CAMHD qualifying entrance threshold of 80 by follow-up measurement. The mean maximum MTPS progress rating was 4.35 (*SD* = 1.25) out of a maximum possible 7-points.

TREATMENT GROUP DIFFERENCES IN PROGRESS OUTCOMES AND PDEs

Mean progress ratings should be considered within the context of the duration of the measurement period and the treatment targets upon which these scores were based. By approximately five months (the mean progress rating measurement period for the

matched sample), 43% of youth had not completed their treatment episode under study. Table 3 describes frequency and recurrence data for the treatment targets. Sixty-nine percent of the most common targets across cases overlapped between groups. Regarding non-overlapping targets, willful misconduct, truancy and running away were endorsed for a high proportion of the MST sample, while positive family functioning, positive thinking, contentment/happiness and depressed mood were endorsed for a high proportion of the IIH sample. Fifty percent of common targets were also the most recurrent targets within cases (top 20% of targets). Target recurrence rates were similar across groups (t (110) = 0.10, p = .92). A few targets were recurrent and common in only one therapy: anger and family functioning (IIH) and running away (MST).

In qualitatively examining the PDEs utilized to address these treatment targets, the five most common intervention strategies used within MST across cases were (in order): communication skills, parent/teacher monitoring, parent praise, problem solving and family therapy. The five most common intervention strategies used within IIH were cognitive, communication skills, problem solving, family therapy, and therapist praise/rewards. Overall, MST therapists were generally more consistent in following through in using the same PDEs month-after-month within their cases (t (340) = -4.04, p < .001).

Table 3. *Treatment Target Frequency & Recurrence by Treatment Type*

MTPS Treatment Target	MST	IIH
Oppositional or Non Compliant Behavior	90%$^{+}$	81%$^{+}$
Positive Peer Interaction	78%$^{+}$	70%$^{+}$
Treatment Engagement	63%	49%
Activity Involvement	60%$^{+}$	65%$^{+}$
Aggression	60%$^{+}$	47%$^{+}$
School Attendance or Truancy	58%	
Willful Misconduct or Delinquency	52%	
Substance Use	49%$^{+}$	35%$^{+}$
Academic Achievement	42%	54%
School Involvement	37%	36%
Anger	35%	65%$^{+}$
Runaway	35%$^{+}$	
Phobia or Fears	26%	
Positive Family Functioning		46%$^{+}$

Positive Thinking or Attitude		43%
Contentment or Happiness		40%
Depressed Mood		39%

Note: Data (%) represents the percentage of youth for whom each target was endorsed.
[+] Denotes recurrently endorsed targets within cases (within the top 20% for the

DATA INSPECTION

The properties of the data from the matched sample were inspected to inform selection of the appropriate formal analyses. As regression analyses were planned, Levene's test was used to identify any significant group difference in variance for outcome variables. For all regressions, Durbin-Watson test statistics indicated the assumption of independent errors was tenable and Variance Inflation Factor and tolerance values indicated low collinearity. Examination of influence residual statistics (Mahalanobis, Cook's distance, Leverage) confirmed the absence of influential cases, with assumptions of linearity and homoscedasticity checked through standardized residual plot inspection. Histogram and normality curves suggested normality of residuals.

ANALYSIS I: EFFECTIVENESS OF MST AND IIH

In Analysis I, outcome differences between the MST and IIH treatments were analyzed with group means comparison and hierarchical linear regression. Comparison of group differences in MTPS progress ratings indicated the MST group experienced a 1.39 point (*SE of Difference* = .11) greater mean improvement than the IIH group (MST M = 5.06, SD = 0.94; IIH M = 3.67, SD = 1.14), which independent t-test results demonstrated was statistically significant (t (340) = -12.30, $p < .001$).

A hierarchical regression analysis of MTPS progress ratings on treatment type, controlling for time, was subsequently conducted to handle treatment group differences in duration of therapy that resulted from a group difference in the MTPS measurement period that was previously noted. As duration of therapy predicted treatment progress (bivariate correlation r = .05) it would otherwise confound the effectiveness test. No youth characteristics were included as covariates as there were no statistically significant group differences subsequent to the propensity match. Table 4 presents the results from

this analysis. When treatment type was added to the model, there was a 36% increase in outcome variance explained (*pr* = .60, *p* < .001). Repeated measures ANCOVAs were also used to test whether any interaction variables should be included within the model. No evidence found was for interactions of treatment type with gender, time, age, comorbidity or primary disruptive behavior diagnosis for progress ratings.

Table 4. *Linear Regression of Progress Ratings on Treatment Type*

Variables in the Model		*B*	*SE B*	*β*
Block 1	(Constant)	4.20	0.19	
	Time	0.00	0.00	.05
Block 2	(Constant)	2.82	0.18	
	Time	0.01	0.00	.24***
	Treatment type	1.57	0.11	.63***

Note: Treatment type was coded MST = 1, IIH = 0. Time unit = days. Dependent variable = MTPS mean maximum progress ratings score
R^2 for Block 1 = .00. ΔR^2 for Block 2 = .36, *p* < .001 for *F* change.
***$p <$ .001, two-tailed

Comparable analyses were performed for the secondary criterion, CAFAS global functioning scores. Although the extent of CAFAS mean change was larger for the MST group (*M* = 18.77, *SD* = 37.34) than for the IIH group (*M* = 15.05, *SD* = 40.30), repeated measures ANOVA testing of the between-subjects main effect of treatment type indicated the size of the group difference was not statistically significant (*F* (1,340) = 0.04, *p* = .84). Paired t-tests revealed within-group change to be significant for both MST (*t* (170) = 6.57, *p* < .001) and IIH (*t* (170) = 4.88, *p* < .001). Next, a regression model was generated for CAFAS outcomes, controlling for baseline CAFAS and duration of the measurement period. Partial correlation values of treatment type and CAFAS outcome scores were small and non-significant (*pr* = -.03), continuing to indicate that treatment type was a poor predictor of change in CAFAS scores. Results from the CAFAS regression are presented in Table 5. Similar models were run for to determine if treatment type predicted any CAFAS subscale outcomes, as a previous MST study found different results across functioning domains (Timmons-Mitchell et al., 2006). Subscale outcomes were also poorly predicted by treatment type[6]. Given treatment type was a poor predictor

[6] Results for subscale data are not published herein but can be requested from the author.

of change in CAFAS scores, and given subsequent analyses examined predictors of treatment response more closely, the MTPS progress rating factor was used as the sole dependent variable for the remaining analyses.

Table 5. *Linear Regression of Global Functioning on Treatment Type*

Variables in the Model		*B*	*SE B*	*β*
Block 1	(Constant)	36.85	10.96	
	Baseline functioning	0.39	0.08	.26***
	Time	0.04	0.02	.08
Block 2	(Constant)	38.33	11.27	
	Baseline functioning	0.40	0.08	.26***
	Time	0.03	0.02	.08
	Treatment type	-2.26	3.91	-.03

Note: Treatment type coded MST = 1, IIH = 0. Time unit = days. Baseline functioning = entry CAFAS global composite. Dependent variable = follow-up CAFAS global composite
R^2 for Block 1 = .07. ΔR^2 for Block 2 = .00, $p > .05$ for F change.
***$p < .001$, two-tailed

ANALYSIS II: RELATIVE IMPORTANCE OF PACKAGE AND PDE CONTENT

In Analysis II, PDE content was explored as a possible explanatory factor for between-group differences in therapist-reported treatment progress. Independent *t*-tests indicated significantly greater PDE content for MST (M =.43, SD =.09) than for IIH (M =.38, SD =.11; t (340) = 4.83, $p < .001$) as hypothesized. To test whether PDE content predicted treatment progress, a hierarchical regression analysis of MTPS progress ratings on PDE content, controlling for time, was generated (Table 6). Results indicated the proportion of PDE content predicted therapist-rated progress ratings across the matched sample (pr = .28). Since both PDE content and treatment type predicted outcomes, their relative importance was examined using hierarchical regression with follow-up dominance analysis. First, progress ratings were regressed on both PDE content and treatment type (entered simultaneously), controlling for time (Table 7[7]). Partial effects were large for treatment type (pr = .57) and small for PDE content (pr = .15).

[7] An interaction between treatment type and PDE content was assessed for in an expanded regression model with the interaction term. No statistically significant interaction effect was found.

Table 6. *Linear Regression of Progress Ratings on PDE content*

Variables in the Model		*B*	*SE B*	*β*
Block 1	(Constant)	4.20	0.19	
	Time	0.00	0.00	.05
Block 2	(Constant)	2.86	0.31	
	Time	0.00	0.00	.05
	PDE content	3.35	0.63	.28***

Note: Time unit = days. PDE content = MTPS proportion of PDEs out of total elements. Dependent Variable = MTPS mean maximum raw progress ratings score
R^2 for Block 1 = .00. ΔR^2 for Block 2 = .08, $p < .001$ for F change.
$^{***}p < .001$, two-tailed

Table 7. *Linear Regression of Progress Ratings on Treatment Type and PDE content*

Variables in the Model		*B*	*SE B*	*β*
Block 1	(Constant)	4.20	0.19	
	Time	0.00	0.01	.05
Block 2	(Constant)	2.30	0.26	
	Time	0.01	0.00	.23***
	Treatment type	1.48	0.12	.59***
	PDE content	1.49	0.54	.12**

Note: Treatment type was coded MST = 1, IIH = 0. Time unit = days.
Dependent variable = MTPS mean maximum raw progress ratings score
R^2 for Block 1 = .00. ΔR^2 for Block 2 = .37, $p < .001$ for F change.
$^{**}p < .01$, $^{***}p < .001$, two-tailed

As linear regression does not adequately handle the question of the relative importance of predictors given any correlation between predictors, a dominance analysis was performed. Rescaled dominance weights indicated the relative percentage of total effect uniquely attributable to treatment type and PDE content was 80.9% and 12.4%, respectively. Both values outweigh the 6.7% relative effect size of the time covariate. As noted in the aforementioned description of the dominance approach, dominance weights are the average squared part correlation for each predictor taken over all possible subset regressions of the outcome variable on the predictors.

An exploratory analysis was done to examine whether there were between-group differences in the predictive strength of PDE content. Given that MST is an EST, it was

anticipated that the level of PDE content within MST would not strongly affect treatment progress. Similarly, given IIH is much less linked to the evidence base, the level of PDE content within IIH was expected to strongly affect treatment progress. To examine this, regressions were run for each treatment type predicting MTPS progress ratings. The partial correlations for outcome with PDE content were small in IIH ($pr = .08$; ΔR^2 for predictor block = .01; $\Delta F = 1.10$, $p = .30$) and moderate in MST ($pr = .23$; ΔR^2 for predictor block = .05; $\Delta F = 9.60$, $p < .01$).

ANALYSIS III: TREATMENT MEDIATION BY PDE CONTENT

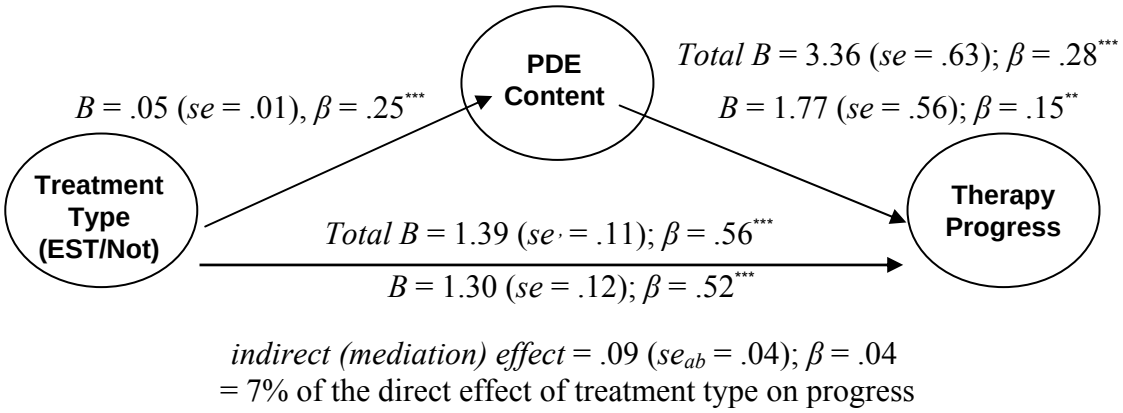
The third question of interest for this study was if and how PDE content might mediate the relationship between treatment type and treatment progress[8]. This analysis was intended to further explore whether the greater PDE content found in MST (Analysis II) could be predominantly responsible for the greater treatment progress in MST versus IIH (Analysis I).

To test for a mediation effect and describe its form, the causal steps approach and bootstrapping protocols were utilized (Baron & Kenny, 1986; Kenny et al., 1998). Results supported the proposed partial mediation hypothesis, although the extent of mediation was smaller than anticipated. Specific findings from the causal steps analysis follow, with results illustrated in Figure 1. As expected and already described, the zero-order relationship between treatment type and PDE content was significant ($\beta = .25$, $p < .001$), as was the relationship between PDE content and therapy progress controlling for treatment type ($\beta = .15$, $p < .01$). Further, significant effects ($p < .001$) were measured for both the total effect between treatment type and therapy progress ($\beta = .56$) and for the direct effect between treatment type and therapy progress controlling for PDE content ($\beta = .52$). The 'indirect (mediation) effect' of treatment type and progress ratings was small (*beta* = .09; $\beta = .04$). Small partial mediation was demonstrated as the direct path of treatment type was significant but smaller than the total path and the indirect (mediation) effect was significantly different from zero, as determined by results from the Sobel test (Sobel, 1982, 1986; $Z = 2.65$ $p < .01$) and bootstrapping (95% bias-corrected and

[8] Mediation of CAFAS outcomes was not examined as PDE content was a poor predictor of CAFAS global and subscale functional improvement.

accelerated bootstrap confidence interval: .03 to .19). Assumptions of the mediation analysis were respected (i.e., independent residuals, no *XM* interaction; MacKinnon, Fairchild, & Fritz, 2007).

Figure 1. *Mediation Model with Path Coefficients*



CHAPTER 4. DISCUSSION

The present study provides an important framework and foundation for future research aimed at furthering the evidence-base on practice elements. Results also provide support for use of the derived elements approach in the adolescent conduct population.

The rigorous propensity-score match effectiveness test yielded a large effect size (Cohen's $d = 1.50$) for MST versus IIH on therapist-reported progress on team-targeted therapy goals over the first five months of therapy. There was a trend towards more global functional improvement in MST compared to IIH. In follow-up analyses aimed at explaining this outcome difference, it was found that youth in MST received a higher mean proportion of PDE content than youth in IIH, and that PDE content was only predictive of progress in MST and not IIH. Furthermore, regression modeling results converged with current theory in suggesting that provision of more PDE content and allocation to an EST (MST) are distinct overlapping predictors of positive response to treatment. A head-to-head test of these predictors indicated that treatment type (MST) was the more powerful predictor of progress on targeted goals. Finally, a mediation analysis aimed at further elucidating the relationships among these study variables suggested that PDE content is a small mediator of the effect of treatment type on progress.

The comparative treatment effect for therapist-rated progress was larger than expected even as it was anticipated that MST would outperform IIH. Prior studies have often reported superior outcomes for MST over usual care therapies yet the effect size in this study (Cohen's $d = 1.50$) was notably larger than a previously reported MST effect size for goal progress (e.g., $d = 0.29$ against usual care; Curtis et al., 2009). This effect was also larger than the mean MST effect size from effectiveness trials ($d = .27$) and efficacy trials ($d = .81$; Curtis et al., 2004). Given the small mediation by PDE content, superior MST therapy progress was more likely attributable to other aspects of therapy. Examination of differences on treatment targets and PDE profiles between treatment groups suggests there may be greater focus within MST (v. IIH) on caregiver parenting practices (use of tangible rewards, monitoring, response cost and praise) and on specific youth behaviors (truancy, aggression, substance use, positive peer interactions), which are proposed mechanisms of change for conduct problems. There is little concern that the

superior progress in MST was due to easier treatment targets as some qualitatively more difficult, objectively measurable behavioral targets were present in the MST profile (runaway, delinquency, truancy) in contrast to a few softer, more subjective targets present in the IIH profile (attitude, happiness, family functioning). The large effect could also be explained by several other therapy features (e.g., hours of therapist training and supervision, focus on ideographic youth assessment and selection of targets/PDEs, clearly defined protocol, handling of treatment barriers) have been examined in previous studies and found to be associated with MST's effectiveness in the community (Henggeler et al., 1997). Some of these factors are not specific to MST, or absent from IIH, but are more likely to occur within MST due to the model. For instance, connecting youth to multiple mentors and caregivers could happen within either MST or IIH, with a presumed beneficial effect. However, the MST therapy team may be more likely to succeed at this labor-intensive target given their higher limit on direct contact hours. Another possible factor that could explain some of this large MST effect is measurement error from systematic differences that might exist between MST and IIH therapists in their self-reporting. Group reporting differences have not yet received much attention in MST studies due to their absence in good experimental design, but could easily originate from the differences in training and supervision between MST and IIH.

Both treatment groups improved on the CAFAS global functioning score and the mean monthly rate of change for MST (approached that of other MST studies measuring this outcome (e.g., Stambaugh et al., 2007). However, there was an unexpected absence of measurable difference between treatment groups on the amount of change in functioning given the sizeable group difference in progress ratings. That MTPS progress rating scores are more closely correlated with youth discharge status (successful versus unsuccessful discharge; $r = .58$) than global functioning scores ($r = .30$) suggests progress rating scores could have greater validity. This is unsurprising given several factors. First, there are some noted problems with the valid measurement and operationalization of functional change (see Bates, 2001 and Bacon et al., 2002 for a more thorough discussion of these issues). In the present study, lower CAFAS score validity may have resulted from the case manager having insufficient information to accurately endorse functioning due to limited contact with the family and therapist, or to the lower sensitivity of the

CAFAS to small and less stable change given its properties. Given the follow-up CAFAS was completed two to three months after the progress ratings on average in this study, it seems less likely that the discrepancy is due to either delay in information transmission from therapist to case manager or delay in generalization of targeted goals to functional gains. Nonetheless, the timing of changes in symptoms and functioning may be quite different (e.g., Howard, Lueger, Maling, & Martinovich, 1993). Because MST yielded similar or greater improvement on the CAFAS mood and home domains compared to IIH, it also seems unlikely that there is a true difference in global functioning rooted in more frequent targeting of mood and general family functioning in IIH. A final challenge for interpreting these findings is that the PSM process increased the odds of Type II error in excluding youth with the highest scores from the matched sample. Given these complexities in interpretation, future researchers might seek to refine these methods of measurement and examine functional change with longitudinal studies.

The IIH level of PDE content was relatively high and did approach that of the MST treatment (95% of youth received between a PDE proportion score between .16 and .60). As yet there are no comparison studies to suggest whether this level of PDE content is typical of IIH and will generalize to IIH services in other systems of care. It seems likely that some of the high level of PDE content reflects the CAMHD system of care's commitment to promoting evidence-based services and practices within usual care. It should be considered that results might be confounded by measurement error, including possible group differences in the quality of provided PDEs. Research suggests there are many challenges in delivering high-fidelity PDEs (Schoenwald et al., 2011). Other potential sources of measurement error include reporting bias from reliance on self-report instruments. Self-report measures in general are known to have lower validity than direct observation measures; it has also been found that therapist self-reports overestimate the fidelity, intensity, and frequency of best practices, particularly in low supervision environments (Horenstein, Houston, & Holmes, 1973; Schoenwald, Sheidow, & Chapman., 2009). As a result, care should be taken in generalizing these results beyond the methods used in this study.

PDE content also had a different effect across the samples. Regression analyses indicated that PDE content is a moderate predictor of progress within MST, but less so in

IIH. The opposite pattern was predicted. It was anticipated that since MST is an EST, cases generally contain multiple interacting features empirically-linked to recovery (i.e., including high-fidelity PDE content but not limited to it) and there could be diminishing returns from adding more PDE content to MST. System of care discussions frequently reflect this assumption that IIH therapists are more likely than MST therapists to benefit from greater focus and trainings on PDE content. The pattern that was found could reflect reporting bias in usual care if the association between PDE content and outcome in IIH is systematically underestimated. Another plausible interpretation is that MST features enhance the impact of each PDE delivered, while IIH does not have such advantage. This would suggest that PDE quantity must be examined within the context of PDE quality and might support the original supposition that IIH therapists are more likely to benefit from trainings with the important caveat that trainers should emphasize quality of PDE delivery.

Next, mediation analysis suggested PDE content was only a small mediator of the effect of treatment type on progress. Graphic inspection of the specific PDE data showed there was significant overlap between treatment groups on frequently used PDEs. In addition, there was low variability within groups. These factors signal a restricted range and could help explain why PDE content mediates only 8% of the variance. This restricted range probably reflects an overstatement of their use by therapists, and raises concerns about the reliability and validity of the PDE data.

Using dominance analysis it was estimated that allocation to MST was a 6.5-fold stronger predictor of goal progress than level of PDE content. That treatment type (MST) emerged as a stronger predictor is reasonable since MST is not only associated with high PDE content, but also with some additional positive outcome predictors. Some of these predictive characteristics were previously noted as qualities that have been associated with MST's effectiveness in applied settings (e.g., level of supervision, treatment fidelity, etc.) Without controlling for PDE delivery variables, and with the level of measurement error in this study, findings on relative rankings are exploratory.

LIMITATIONS

The current study has some analytical limitations. First, research any research involving PDEs or derived elements theory and methods is considered exploratory (Weersing, Rozenman, & Gonzalez, 2009). While our findings suggest future studies in this line of research are warranted, there is clear need to better understand the robustness of these study findings to its specific methods. Methodological refinements can promote greater measurement precision and validity.

The strong external validity of the study, involving a naturalistic sample of clients and their therapists, necessitated a trade-off in internal validity. It is possible that treatment group outcomes differed in part due to uncontrolled variables such as treatment adherence, therapist competency, program operations, or attrition. Most therapists did not overlap between treatment types and MST therapist qualifications were more stringent. The MST program also had built-in quality assurance mechanisms and information available, whereas less information was available on the quality of IIH as it was provided by numerous different agencies and private therapists. As discussed, treatment types also differed on duration of measurement period, due to challenges with missing data and timing of data capture, particularly for the CAFAS. Similarly, there were different levels of nested data (supervisor, therapist, and agency) between treatments and among agencies. All of these challenges are common to studies using retrospective program data from applied settings, and should be carefully considered in future research designs.

Additionally, the study sample was not limited to youth with a conduct disorder diagnosis, although youth with willful and delinquent behavior were the population of interest. The sample of youth with conduct disorder was too small to detect the expected effect sizes, although models were subsequently run using only the conduct disorder participants with similar results. Nonetheless, for purposes of interpretation it is important to keep in mind that differences exist among subtypes that could impact their response to treatment. Similarly, PDE content was calculated based on frequency lists aggregated across disruptive behavior disorders despite subtype differences and a high rate of comorbid diagnoses.

Furthermore, the present study was focused on outcomes for the initial months of a first intensive in-home treatment. This short-term focus precludes much discussion of

lasting effects. While a series of follow-up exploratory regression models were run for CAFAS (i.e., with capture closest to 12 months and treatment exit) and MTPS progress ratings (i.e., with capture closest to seven months and treatment exit), the outcome data overlapped significantly with the outcome data for the short-term measurement period, and therefore cannot be validly interpreted.

FUTURE RESEARCH DIRECTIONS

This study also lays important groundwork that will assist future researchers with refining methods for studying the PDE construct. For example, it may be important to simultaneously measure practice element order, fidelity, intensity, and frequency. Such variables could be a powerful addition to therapist-report measures of practices such as the MTPS or Therapy Process Checklist (Weersing et al., 2002). Studies are also needed to determine standardized cut-off scores for coding practices and methods for calculating treatment content to enable comparison of results across studies and minimize measurement error.

A discussion of PDE fidelity also raises the question of how fidelity at the practice element level is associated with fidelity at the model level, as well as the extent to which existing measures of model fidelity capture PDE dimensions. The TAM-R measure solicits the caregiver's report on the therapist's effectiveness with several strategies (e.g., family engagement, setting of immediate goals) and whether the therapist targeted family and peer interactions, family communication, academics, and parent behavioral management. A more formal study of MST fidelity measures seems warranted at this juncture, focused on the extent to which model fidelity instruments measure outcomes versus model adherence and whether these instruments are useful for measuring fidelity of PDEs. At the model level, therapist adherence to the MST protocol has been associated with improved family relations and decreased negative peer involvement, and, in turn, with decreased delinquent behavior (Huey, Henggeler, Brondino, & Pickrel, 2000). As results from this study suggest, fidelity in delivering PDEs can also be expected to impact mechanisms of change in recovery.

Next, while the focus of this study was on the aggregated effect of the 13 PDEs with the highest frequency within RCT protocols, it will also be important to examine which specific practices contribute to most of the outcome effect. A brief follow-up analysis demonstrated three of these PDEs were more prevalent in MST & also accounted for a medium effect (monitoring, parent praise, and tangible rewards).

Furthermore, while the focus of this study was on MST and usual care as delivered in an applied setting to adolescents with conduct problems, the questions the study addresses and raises

can be asked of other clinical populations as well. Several types of clinical populations seem likely to benefit most from extensions of this type of research, such as those with more limited access to ESTs, those with positive outcomes for usual care that fall short of EST outcomes, and those with EST effect sizes that diminish post-dissemination.

Finally, towards the ultimate goal of furthering the evidence base on supported practices for treatment of adolescents with conduct problems, several additional studies are recommended. Future research could examine practice profiles and how to they might differ across subpopulations or high and low responders. It would also be interesting to examine the relationship of PDE content and package to other population-relevant health outcomes (e.g., youth and family stress, truancy, or days in out-of-home placement) and to examine these relationships using different therapies. Results from such analyses, supplemented by additive and subtractive experiments with practice elements, could assist with the identification of evidence-based practice elements. These studies will be an important next step for validating the derived elements approach in this population, and for translating derived elements research into meaningful recommendations for systems of care.

REFERENCES

Accurso, E. C., Taylor, R. M., & Garland, A. F. (2011). Evidence-based practices addressed in community-based children's mental health clinical supervision. *Training and Education in Professional Psychology, 5(2),* 88-96. doi: 10.1037/a0023537

Allison, P. D. (2001). *Missing data (quantitative applications in the social sciences).* Thousand Oaks, CA: Sage.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8,* 129-148. doi: 10.1037/1082-989X.8.2.129

Bacon, S. F., Collins, M. J., & Plake, E. V. (2002). Does the Global Assessment of Functioning assess functioning? *Journal of Mental Health Counseling, 24(3),* 202-212.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–82. doi: 10.1037/0022-3514.51.6.1173

Barth, R. P., Greeson, J. K, Guo, S., Green, R. L., Hurley, S., & Sisson, J. (2007). Outcomes for youth receiving intensive in-home therapy or residential care: a comparison using propensity scores. *American Journal of Orthopsychiatry, 77(4),* 497-505. doi: 10.1037/0002-9432.77.4.497

Barton, K., Baglio, C., & Braverman, M. (1994). Stress reduction in child-abusing families: Global and specific measures. *Psychological Reports, 75,* 287-304.

Bates, M. P. (2001). *Global functioning within a system of care for youths with serious emotional disturbance: A closer look at the Child and Adolescent Functional Assessment Scale (CAFAS).* Retrieved from Dissertations & Theses (PQD&T) database. (UMI No. 9986996)

Bates, M. P., Furlong, M. J., & Green, J. G. (2006). Are CAFAS subscales and item weights valid? A preliminary investigation of the Child and Adolescent Functional Assessment Scale. *Administration and Policy in Mental Health and Mental Health Services Research, 33(6),* 682-695. doi: 10.1007/s10488-006-0052-4

Borduin, C. M. (2007). Multisystemic treatment of violent youth and their families. In T. A. Cavell, A. Timothy, & K. T. Malcolm (Eds.), *Anger, aggression and interventions for interpersonal violence* (pp. 239-265). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Borntrager, C. F., Chorpita, B. F., Higa-McMillan, C., & Weisz, J. R. (2009). Provider attitudes toward evidence-based practices: Are the concerns with the evidence or with the manuals? *Psychiatric Services, 60,* 677-681. doi: 10.1176/appi.ps.60.5.677

Brookman-Frazee, L. Haine, R. A., Baker-Ericzen, M., Zoffness, R. & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research, 37(3),* 254-269. doi: 10.1007/s10488-009-0244-9

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114,* 542-551. doi: 10.1037/0033-2909.114.3.542

Burns, B. J., Hoagwood, K. & Mrazek, P. (1999). Effective Treatment for Mental Disorders in Children and Adolescents, *Clinical Child and Family Psychology Review, 2(4),* 199-254.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7-18.

Child and Adolescent Mental Health Division (2003). Annual Evaluation Report Fiscal Year 2003: Hawai‘i Department of Health Child and Adolescent Mental Health Division.

Child and Adolescent Mental Health Division (2006). Interagency Performance Standards and Practice Guidelines: Hawai‘i Department of Health Child and Adolescent Mental Health Division. Retrieved from: http://hawaii.gov/health/mental-health/camhd/library/pdf/ipspg/purplebook.pdf

Child and Adolescent Mental Health Division (2008). Instructions and codebook for provider monthly summaries: Hawai‘i Department of Health Child and Adolescent Mental Health Division. Retrieved from: http://hawaii.gov/health/mental-health/camhd/library/pdf/paf/paf-001.pdf.

Child and Adolescent Mental Health Division (2010). Annual Factbook: Hawai‘i Department of Health Child and Adolescent Mental Health Division. Retrieved from: http://hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge030.pdf

Chorpita, B. F., Becker, K. D., & Daleiden E. L. (2007). Understanding the common elements of evidence-based practice: misconceptions and clinical examples. *Journal of the American Academy of Child & Adolescent Psychiatry, 46,* 647-652. doi: 10.1097/chi.0b013e318033ff71

Chorpita, B. F. & Daleiden, E. L. (2007). Biennial report: Effective psychosocial interventions for youth with behavioral and emotional needs. Retrieved from: http://hawaii.gov/health/mental-health/camhd/library/pdf/ebs/ebs012.pdf

Chorpita, B. F. & Daleiden, E. L. (2009). Biennial report: Effective psychosocial interventions for youth with behavioral and emotional needs. Retrieved from: http://hawaii.gov/health/mental-health/camhd/library/pdf/ebs/ebs013.pdf

Chorpita, B. F., Daleiden, E., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research, 7,* 5−20.

Chorpita, B. F., & Mueller, C. W. (2008). Toward new models for research, community, and consumer partnerships: Some guiding principles and an illustration. *Clinical Psychology: Science and Practice, 15,* 144-148.

Copeland, W. E., Miller-Johnson, S., Keeler, G., Angold, A., & Costello, E. J. (2007). Childhood psychiatric disorders and young adult crime: A prospective, population-based study. *American Journal of Psychiatry, 164,* 1668–1675.

Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology, 18,* 411–419. doi: 10.1037/0893-3200.18.3.411

Curtis, N. M., Ronan, K. R., Heiblum, N., & Crellin, K. (2009). Dissemination and effectiveness of multisystemic treatment in New Zealand: A benchmarking study. *Journal of Family Psychology, 23,* 119-129.

Dagenais, C., Begin, J., Bouchard C., & Fortin, D. (2004). Impact of intensive family support programs (IFPS): A synthesis of evaluation studies. *Children and Youth Services Review, 26(3),* 249–263.

Daleiden, E. L., Pang, D., Roberts, D., Slavin, L. A., & Pestle, S. L. (2010). Intensive home based services within a comprehensive system of care for youth. *Journal of Child and Family Studies, 19(3),* 318-325. doi: 10.1007/s10826-009-9300-z

Daleiden, E. L. & Tolman, R. T. (2005). Annual evaluation report Fiscal year 2005: Hawai‘i Department of Health Child and Adolescent Mental Health Division. Retrieved from: http://www.hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge015.pdf.

Denenny, D. M. & Mueller, C. W. (2012). *Multisystemic Therapy.* In R. Levesque (Ed.), Encyclopedia of Adolescence. New York, NY: Springer.

Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Jones, K., Hughes, J. C., et al. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: Care, Health and Development, 35,* 603–612. doi: 10.1111/j.1365-2214.2009.00975.x

Eddy, J. M., & Chamberlain, P. (2000). Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. *Journal of Consulting and Clinical Psychology, 68,* 857–863. doi: 10.1037//0022-0006X.68.5.857

Farmer, E. M. Z., Mustillo, S., Burns, B. J., & Holden, E. W. (2008). Use and predictors of out-of-home placements within systems of care. *Journal of Emotional and Behavioral Disorders, 16,* 5–14. doi:10.1177/1063426607310845.

Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R. J., Haine-Schlagel, R., Ganger, W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services, 61(8)*, 788-795. doi: 10.1176/appi.ps.61.8.788

Garland, A. F., Hawley, K. M., Brookman-Frazee, L., & Hulburt, M. S. (2008). Identifying Common Elements of Evidence-Based Psychosocial Treatments for Children’s Disruptive Behavior Problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 47(5),* 506-515. doi: 10.1097/CHI.0b013e31816765c2

Garland, A. F., Hurlburt, M. S., & Hawley, K. M. (2006). Examining psychotherapy processes in a services research context. *Clinical Psychology: Science and Practice, 13,* 30–46. doi: 10.1111/j.1468-2850.2006.00004.x

Guo, S., Barth, R., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review, 28,* 357–383. doi: 10.1016/j.childyouth.2005.04.012

Haley, J. (1976). *Problem-solving therapy.* San Francisco, CA: Jossey-Bass.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of anti-social behavior in children and adolescents.* New York, NY: Guilford Press.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic Therapy for Antisocial Behavior in Children and Adolescents,* 2nd Edition. New York, NY: The Guildford Press.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child & Adolescent Psychology, 31(2),* 155-167. doi: 10.1207/153744202753604449

Hinckley, E. C. & Ellis, W. F. (1985). An Effective Alternative to Residential Placement: Home-based Services. *Journal of Clinical Child Psychology, 14(3),* 209-213.

Hodges, K. (1995). *CAFAS: Self-training manual.* Ann Arbor, MI: Functional Assessment Systems.

Hodges, K. (1998). *Child and Adolescent Functional Assessment Scale (CAFAS).* Ann Arbor, MI: Functional Assessment Systems.

Hodges, K. (2003). *Child and Adolescent Functional Assessment Scale* (3rd ed.). Ypsilanti, MI: Eastern Michigan University.

Hodges, K., & Gust, J. (1995). Measures of impairment for children and adolescents. *Journal of Mental Health Administration, 22,* 403-413. doi: 10.1007/BF02518634

Hodges, K., & Kim, C. (2000). Psychometric study of the Child and Adolescent Functional Assessment Scale: Prediction of contact with the law and poor school attendance. *Journal of Abnormal Child Psychology, 28,* 287–297.

Hodges, K., & Wong, M. M. (1996). Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale (CAFAS). *Journal of Child and Family Studies, 5,* 445-467. doi: 10.1007/BF02233865

Hodges, K., Xue, Y., & Wotring, J. (2004). Use of the CAFAS to evaluate outcome for youths with severe emotional disturbance served by public mental health. *Journal of Child and Family Studies, 13,* 325–339.

Hoge, R. D. (2002). Standardizing instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29,* 380–396.

Horenstein, D., Houston, B.K. & Holmes, D.S. (1973). Clients', therapists', and judges' evaluations of psychotherapy. *Journal of Counseling Psychology, 20,* 149-153. doi: 10.1037/h0034172

Howard, K. I., Lueger, R., Maling, M., & Martinovich, Z. (1993). A phase model of psychotherapy: Causal mediation of outcome. *Journal of Consulting and Clinical Psychology, 61,* 678-685. doi: 10.1037/0022-006X.61.4.678

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in Multisystemic Therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68,* 451-467. doi: 10.1037//0022-006X.68.3.451

Hurley, S. (2008). *Factors Affecting Long-Term Outcomes Following Intensive In-Home Services.* Retrieved from Dissertations & Theses (PQD&T) database. (UMI No. 3330307)

Kazdin, A. E. (1997). Practitioner review: Psychosocial treatment for conduct disorder in children. *Journal of Child Psychology and Psychiatry & Allied Disciplines, 38,* 161–178.

Kazdin, A. E. (2000). Treatment for aggressive and antisocial children. *Child and Adolescent Psychiatry Clinics of North America, 9,* 841-858.

Kenny, D. A., Kashy, D. A. & Bolger, N. (1998). Data analysis in social psychology. In. D.T. Gilbert, S.T. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology Volume 1* (pp. 233–65). New York, NY: Oxford Univ. Press.

Kinney, J., Madsen, B., Fleming T, & Haapala, D. (1977). Homebuilders: Keeping families together. *Journal of Consulting and Clinical Psychology, 45,* 667-673. doi:10.1037/0022-006X.45.4.667

Lahey, B. B., Gendrich, J. G., Gendrich, S. I., Schnelle, J. F., Gant, D. S. & McNees, M. P. (1977). An Evaluation of Daily Report Cards with Minimal Teacher and Parent

Contacts as an Efficient Method of Classroom Intervention. *Behavior Modification, 1(3),* 381-394. doi: 10.1177/014544557713006

Levins, T. A. (1998). *A comparison of the efficacy of two treatment models for children and adolescents with conduct disorders.* Retrieved from Dissertations & Theses (PQD&T) database. (UMI No. 0421)

Lewis, R. (2005). The effectiveness of Families First services: An experimental study. *Children and Youth Services Review, 27,* 499-509. doi:10.1016/j.childyouth.2004.10.009

Lilienfeld, S. O. (2005). Scientifically Unsupported and Supported Interventions for Childhood Psychopathology: A Summary. *Pediatrics, 115,* 761-764. doi: 10.1542/peds.2004-1713.

Lindsey, D., Martin, S., & Doh, J. (2002). The failure of intensive casework services to reduce foster care placements: An examination of family preservation studies. *Children and Youth Services Review, 24,* 743-775. doi:10.1016/S0190-7409(02)00227-X

Littell, J. (2005). Lessons from a systematic review of effects of Multisystemic Therapy. *Children and Youth Services Review, 27,* 445-463. doi:10.1016/j.childyouth.2004.11.009

Littell, J, Popa, M., & Forsythe, B. (2005). Multisystemic therapy for social, emotional, and behavior problems in youth age 10–17. *Cochrane database of systematic reviews, 3.* doi: 10.1002/14651858.CD004797.pub3.

Littell, J., & Schuerman, J. (2002). What works best for whom? A closer look at intensive family preservation services. *Children and Youth Services Review, 24,* 673-699. doi: 10.1016/S0190-7409(02)00224-4

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology, 58,* 593–614.

Masi, G., Manfredi, A., Milone, A., Muratori, P., Polidori, L., Ruglioni, L., Muratori, F. (2011). Predictors of nonresponse to psychosocial treatment in children and adolescents with disruptive behavior disorders. *Journal of Child and Adolescent Psychopharmacology, 21(1),* 51-55. doi: 10.1089/cap.2010.0039

McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy Strategies Scale. *Journal of Clinical Child and Adolescent Psychology, 39,* 436-443. doi: 10.1080/15374411003691750

McMahon, R. J., Wells, K. C., & Kotler, J. S. (2006). Conduct problems. In E.J. Mash & R.A. Barkley (Eds.), *Treatment of childhood disorders 3rd Edition* (pp. 137-268). New York, NY: Guilford Press.

Miller, M. (2006). *Intensive Family Preservation Programs: Program Fidelity Influences Effectiveness – Revised.* Olympia, WA: Washington State Institute for Public Policy.

Minuchin, S., (1974). *Families & family therapy.* Cambridge, MA: Harvard University Press.

Mueller, C. W. & Tolman, R. T. (2009). Practice elements and Youth Outcomes in Intensive In-Home Services. Unpublished manuscript.

Mueller, C. W., Tolman, R. T., Ebesutani, C., & Bernstein, A. (2010). Improvement for Youth with Disruptive Behaviors Provided Evidence-Based Practices. Unpublished manuscript.

Nakamura, B. J., Daleiden, E. L., & Mueller, C. W. (2007). Validity of treatment target progress ratings as indicators of youth improvement. *Journal of Child and Family Studies, 16(5),* 729-741. doi: 10.1007/s10826-006-9119-9

Orimoto, T.E., Higa-McMillan, C., Mueller, C.W., & Tolman, R. (2009). *Factor structure of therapeutic practices in Hawai'i's youth system of care.* Paper presented at the annual meeting of the Hawai'i Psychological Association, Honolulu, Hawai'i.

Perkins, M. B., Jensen, P. S., Jaccard, J., Gollwitzer, P., Oettingen, G., Pappadopulos, E., & Hoagwood, K. E. (2007). Applying theory-driven approaches to understanding and modifying clinicians' behavior: What do we know? *Psychiatric Services, 58,* 342–348.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879-891.

Robbins, M. S., Szapocznik, J., Santisteban, D. A., Hervis, O. E., Mitrani, V. B., & Schwartz, S. (2002). Brief strategic family therapy for Hispanic youth. In A.E. Kazdin, & J.R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 407-424). New York, NY: Guilford.

Rogers, K. M., Pumariega, A. J., Atkins, D. L., & Cuffe, S. P. (2006). Conditions Associated with Identification of Mentally Ill Youths in Juvenile Detention. *Community Mental Health Journal, 42(1),* 25-40. doi: 10.1007/s10597-005-9001-z

Rosenblatt, A. & Rosenblatt, J. A. (2002). Assessing the effectiveness of care for youth with severe emotional disturbances: Is there agreement between popular outcome measures? *The Journal of Behavioral Health Services & Research, 29(3),* 259-273. doi: 10.1007/BF02287367

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70(1),* 41–55. doi: 10.1093/biomet/70.1.41

Rowland, M. D., Halliday-Boykins, C. A., Henggeler, S. W., Cunningham, P. B., Lee, T. G., Kruesi, M. J. P., & Shapiro, S.B. (2005). A randomized trial of multisystemic therapy with Hawai‘i's felix class youth. *Journal of Emotional and Behavioral Disorders, 13,* 13-23. doi: 10.1177/10634266050130010201

Schaeffer, C. M. (2001). *Moderators and mediators of therapeutic change in multisystemic treatment of serious juvenile offenders.* Retrieved from Dissertations & Theses (PQD&T) database. (AAT 6720)

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38(1),* 32-43. doi: 10.1007/s10488-010-0321-0

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77(3),* 410-421. doi: 10.1037/a0013788

Schuerman, J., Rzepnicki, T., & Littell, J. (1994). *Putting Families First: An Experiment in Family Preservation.* New York, NY: Aldine De Gruyter.

Silverman, W. K. & Hinshaw, S. P. (2008). The second special issue on evidence-based psychosocial treatments for children and adolescents: A ten-year update. *Journal of Clinical Child and Adolescent Psychology, 37,* 1-7. doi: 10.1080/15374410701817725

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco, CA: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological Methodology 1986* (pp. 159-186). Washington, DC: American Sociological Association.

Stambaugh, L. F., Mustillo, S. A., Burns, B. J., Stephens, R. L., Baxter, B., Edwards, D., DeKraai, M. (2007). Outcomes from wraparound and multisystemic therapy in a center for mental health services system of care demonstration site. *Journal of Emotional and Behavioral Disorders, 15(3),* 143-155. doi: 10.1177/10634266070150030201

Stroul, B. A. & Goldman, S. K. (1990). Study of community-based services for children and adolescents who are severely emotionally disturbed. *The Journal of Behavioral Health Services and Research, 17(1),* 61-77. doi: 10.1007/BF02518581

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist, 48(1),* 2-23.

Timmons-Mitchell, J., Bender, M. B., Kishna, M. A., & Mitchell, C. C. (2006). An independent effectiveness trial of Multisystemic Therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology, 35,* 227-236.

Tolman, R. T., Mueller, C. W., Daleiden, E. L., Stumpf, R. E., & Pestle, S. L. (2008). Outcomes from multisystemic therapy in a statewide system of care. *Journal of Child and Family Studies, 17(6),* 894-908. doi: 10.1007/s10826-008-9197-y

U.S. Department of Health & Human Services (2000). *Mental Health: A Report of the Surgeon General.* Washington, DC: National Institute of Mental Health.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practice.* Belmont, CA: Wadsworth/Thomas.

Weersing, V. R., Rozenman, & Gonzalez, A. (2009). Core Components of Therapy in Youth: Do We Know what to Disseminate? *Behavior Modification, 33,* 1. doi: 10.1177/0145445508322629

Weersing, V. R., Weisz, J. R. & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology, 31,* 168-180.