State of Hawaii
Department of Health

Child and Adolescent Mental Health Division

Multisystemic Therapy: Treatment Outcome Validity and Predictors of Outcome

January 2004: December 2005

Prepared by
Ryan T. Tolman, B.A., Charles W. Mueller, Ph.D.
Eric L. Daleiden, Ph.D. & Roxanna E. Stumpf, M.A.

For the Period of January 2004 to December 2005
Version 7-11-07

Executive Summary

Background/Context:

- The Child and Adolescent Mental Health Division (CAMHD) has been one of the earliest implementers of Multisystemic Therapy (MST) on a system-wide basis. A considerable research literature exists for MST, which allows CAMHD to benchmark the local implementation to other well-controlled randomized trials.
- Given the CAMHD commitment to ongoing data collection and evaluation, there are numerous datasets that allow for an evaluation of MST in CAMHD. The present report focuses on the validity of ideographic MST goal attainment as measured by therapist's ratings of success ("Successful", "Partially Successful", and "Unsuccessful") and changes in CAFAS and CALOCUS scores from initiation to termination of MST services.

Purposes:

- Examine the validity of the therapist-rated outcome measures (based on goals attained) by comparing these to changes in case manager's ratings of client level of functioning (CAFAS) and level of service needs (CALOCUS) collected around the time of entry and exit from MST.
- Examine client and service characteristics that might predict the extent of goal attainment (therapist-rated outcomes).
- Compare improvement in child functioning observed in MST in Hawaii to rates reported in the MST efficacy literature.

Method:

- This report examines outcomes for 254 children who received MST services between January 1, 2004 and December 13, 2005 and had complete therapist-rated outcome data.
- A sample of 254 youth with MST outcomes was used to examine potential demographic and clinical data that might predict therapist-rated outcomes. These participants comprised the "full MST" sample.
- Of the full MST sample, 122 (48%) had available CAFAS and CALOCUS data near the beginning and ending of MST treatment. These participants made up the "full CAFAS and CALOCUS" sub-sample. Data from this sub-sample was used to validate the MST outcome measures by comparing therapist ratings of success to changes in CAFAS and CALOCUS.
- Data from the "full CAFAS and CALOCUS" sub-sample was used to estimate the amount of improvement and to compare this to rates of improvement reported in the scientific literature.

Key Findings:

- **What percent of MST cases are rated by their therapists as meeting few, some, or all of their MST goals at the end of MST services?** Nearly one half (48%) of all MST cases were rated as "Successful" (all treatment goals met) by therapists at the end of MST services. About 36% of the MST cases were rated as

"Partially Successful" (25% to 99% of goals met) and about 16% were rated as "Unsuccessful" (<25% of goals met).

- **Do CAMHD MST cases show improvement over the course of MST treatment as assessed by other validated measures?** Yes. On average, MST cases showed improvement in level of functional impairment, displaying a significant decrease in CAFAS scores of 12.64 over the course of treatment. MST cases also displayed an improvement in level of service need as evidenced by a decrease in CALOCUS scores of .42 over the course of treatment.
- **Do ratings of "Success" relate to changes in CAFAS and CALOCUS scores over the course of MST services?** Yes. Both CAFAS (r = -0.53) and CALOCUS (r = -0.52) exit scores were significantly related to MST outcome level ("Unsuccessful", "Partially Successful" and "Successful") after controlling for respective entry scores. "Successful" MST cases significantly improved on CAFAS and CALOCUS measures from MST entry to MST exit. "Partially Successful" cases did not show significant change on CAFAS or CALOCUS measures. "Unsuccessful" cases showed significant worsening, as evidenced by higher CAFAS and CALOCUS scores at MST exit compared to entry.
- **How do overall rates of improvement in child functioning compare to rates seen in randomized controlled trials of MST?** The mean pre-post effect size on CAFAS (d = -0.29) and CALOCUS (d = -0.33) scores in MST in Hawaii were somewhat smaller than the mean effect size (d = -.46) observed on similar measures in published randomized controlled trials. However, both effect sizes for MST in Hawaii were within the 95% confidence interval of the mean effect size from the randomized controlled trials on MST.
- **Do client demographic factors predict success in MST (as rated by therapists)?** Males and females did not differ in likelihood of MST success. Age of clients did not predict the likelihood of a "Successful" outcome. Given high rates of missing and multi-ethnic data on ethnicity in addition to changes in the way ethnicity is recorded at CAMHD, no formal analyses were conducted examining ethnicity as a predictor of outcomes.
- **Do clinical factors predict success in MST (as rated by therapists)?** Longer (more days in) MST treatment was the single best predictor of "Successful" outcomes, suggesting a "dose-response" relationship. In addition, youth with one or more comorbid diagnoses were more likely to be rated as "Successful". No other diagnostic variables reliably predicted successful outcomes.
- **Taken together, what do these findings mean?** MST implementation is well underway in Hawaii. Across all cases, there was moderate and reliable improvement in levels of functioning and care needs. About one half of the MST cases were rated as "Successful." Significant positive changes in CAFAS and CALOCUS scores for these MST cases suggest that therapist ratings of success are a valid indicator of improvement. However, cases categorized as "Partially Successful" seem to reflect minimal average change in measures of impairment or care needs; "Unsuccessful" cases, on average, reflect deterioration in functioning and heightened level of care needs. Longer MST treatment predicted increased likelihood of "Success". Consistent with findings from the MST literature more broadly, few client characteristics predict MST success. However youth with

more than one diagnosis were more likely to be rated as "successful" than those with only one diagnosis.

Recommendations:

- **Continue use of therapist measures of outcome.** Therapist ratings of "Success" are associated with improvements on other measures of functional status and service needs.
- **Reconsider the meaning of "Partially Successful".** Many "Partially Successful" cases do not show improvement in level of functioning or service level needs. As such, it might be more accurate to exclude such cases from any overall measure of "Success".
- **Consider other factors that might influence likelihood of "Success".** While few client level factors were predictive of outcome, there remains considerable variability in outcomes that, if explained, could point to ways to improve overall success rates. Such factors might relate to family and ecology characteristics, treatment delivery, or other factors.
- **Consider why youth with comorbid diagnoses are more likely to be rated as "successful".** Youth with comorbid diagnoses might present a treatment profile that differs from youth with only one (most likely a disruptive behavior) diagnosis, which might relate to better outcomes. Given considerable controversy about the effectiveness of evidence-based interventions with youth with complex problems, further work in this area is warranted.
- **Continue to work toward increasing the rate of improvement for MST cases in Hawaii.** While within an acceptable range, average improvement over the course of MST in Hawaii is somewhat smaller than average rates reported in the efficacy literature. Efforts that focus on quality assurance and treatment fidelity could lead to improved outcomes.
- **Consider using the Monthly Treatment Progress Summary Form (MTPS) as a way to examine whether specific treatment targets and/or treatment components relate to MST success.** The MTPS shows promising psychometric properties to date and might prove useful in examining statewide MST services and outcomes. Cases that consistently apply practice elements and focus on treatment targets consistent with the MST treatment approach may well show greater improvement and higher success rates than other cases.
- **Consider using longitudinal data analysis to assess client change in MST.** Longitudinal data analysis based on CAFAS or CALOCUS scores should allow for a powerful examination of factors that possibly influence rates of improvement in MST and provide insight into the pattern of improvement. Such analyses can provide a model for other longitudinal analyses of CAMHD data.
- **Consider MST evaluation of longer-term outcome measures assessed in other MST studies.** Comprehensive evaluation of MST Services in Hawaii should include examination of other outcomes, such as days in out-of-home placement or days of incarceration, which have been used as ultimate outcome measures in other studies (e.g., Henggeler, Melton, Brondino, Scherer, & Hanley, 1997) and are key cost drivers.

Introduction

Over the last two decades, the potential role for evidence-based children's mental health services has been carefully examined and developed. In order to provide scientific rigor to this effort, an emphasis has been placed on research studies that employ randomized controlled trials (RCTs) to test the effects of specific manualized clinical protocols (e.g. Chambless & Hollon, 1998; Chambless & Ollendick, 2001). RCTs are often conducted by treatment developers and include one or more control groups that do not receive the intervention. In many RCTs, researchers pay close attention to treatment fidelity and restrict study participant characteristics in ways that facilitate scientific interpretation (e.g. "this treatment works with this specific population"). These types of studies are typically referred to as efficacy research.

More recently, there has been growing interest in the implementation of evidence-based services in more complex and less controlled settings (e.g., Schoenwald & Hoagwood, 2001) often referred to as effectiveness research. While many effectiveness studies often lack the level of scientific control seen in efficacy research, such studies add to our understanding of the robustness of treatment effects across populations and settings. In addition, effectiveness research (including the present study) allows for an examination of evidence-based interventions within the context of complex mental health service systems with ideographic sets of procedures, policies, staffing, and clientele.

The Child and Adolescent Mental Health Division (CAMHD), Department of Health, State of Hawaii has made a commitment to the incorporation of empirically-based services through its mission to "provide timely and effective mental health prevention, assessment and treatment services to children and youth with emotional and behavioral challenges, and their families" (Child and Adolescent Mental Health Division, 2006). This commitment is reflected in many ways, including careful and ongoing reviews of the empirical literature for children's mental health, systematic tracking of client status and improvement throughout service delivery, the use of a monthly reporting system focused on practices and treatment targets consistent with the empirically-based literature (Child and Adolescent Mental Health Division, 2004; Chorpita, et al., 2002; Daleiden, 2006; Daleiden & Chorpita, 2005; Daleiden, Lee, & Tolman, 2004; Daleiden & Tolman, 2005; Nakamura, Daleiden, & Mueller, 2006) and the implementation of evidence-based interventions.

The present study is focused on the implementation of Multisystemic Therapy (MST) throughout the statewide children's mental health system in Hawaii. MST is a multi-faceted home and community-based intervention that provides intensive services to an entire family for about four to six months. The MST model is based on the extensive literature on the development and maintenance of behavioral problems in youth and focuses on changing social ecology factors that appear causally related to problem behaviors. Clinical features of MST include a comprehensive assessment of the youth, family and broader social ecology, the development of well-defined treatment goals, and the implementation of specific interventions that fit the case assessment and goals.

MST was first implemented in Hawaii to help meet the state's goal of providing an evidence-based service for youth with severe psychosocial and behavioral problems that also embodied CAMHD and Hawaii values of family and community. In general, CAMHD's delivery of MST is congruent with the developers' specification. Specifically, MST in Hawaii is time-limited, lasting approximately 4 – 6 months. MST accesses the services of local treatment teams to help assess youth's problem behaviors and ecology in order to determine treatment goals. MST services can also provide access to psychiatric care and crisis intervention for youth in need. MST therapists have small caseloads and are available to the client 24 hours a day, 7 days a week. MST supervisors oversee a small group of therapists (3 to 4). During the study period, MST in Hawaii differed from MST implemented elsewhere in that it was not under a Network Partnership agreement with MST Services in South Carolina, meaning that Hawaii did not have access to the CQI methodology, technology transfer protocols, and coaching that MST Services provides. Also, CAMHD continues to utilize a trivariate conception of MST success based on percentage of goals met at end of treatment whereas MST Services has shifted to examining instrumental (relatively subjective assessments of family- and youth-functioning) and ultimate (objective measure of youth functioning) goals met at discharge. In addition, during the study period, CAMHD made admission criteria exceptions for younger clients into the MST program, occasionally offered longer treatment episodes, assigned a wider number of treatment goals, and occasionally provided MST booster sessions (brief readmission) for youth who had already undergone full MST treatment.

The present study had three purposes. First, it examined the validity of the therapist-rated outcome measures (based on goals attained) used with the MST protocol. These therapist-rated outcomes are derived from the percentage case-specific goals set at the beginning of MST that are met by the end of the MST service. Such goal setting and evaluation are an integral part of the MST approach. Early in treatment, goals are identified through a mutual consensus process involving the MST therapist (and supervisor), the client and family members. These goals are then tracked over the course of MST. In the present study, we compared these therapist-rated outcomes (goals attained) to other measures of client improvement. Specifically, outcomes were compared to changes in case manager's ratings of client level of functioning and level of service needs collected around the time of entry and exit from MST.

Second, this study examined whether specific client and service characteristics predicted therapist-rated outcomes for MST in Hawaii. The current literature on MST reports few, if any, moderators of treatment effects. Neither gender, age, nor type of population studied (i.e., violent or criminal juvenile offenders, juvenile sexual offenders, substance-abusing youth, psychiatrically disturbed youth, or all other youth populations) has been found to affect the likelihood or extent of positive treatment outcomes from MST (Curtis, Ronan, & Borduin, 2004; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998; Littell, Popa, & Forsythe, 2005). The present study examined the potential impact of client's age, gender, diagnosis, comorbidity and length of MST services on therapist-rated outcomes.

Third, the present study compared the extent of improvement in child functioning observed in MST in Hawaii to rates reported in the MST efficacy literature. While an initial implementation study of an MST continuum of care for youth with Severe Emotional and Behavioral Disturbance in Hawaii included a comparison group (Rowland, et al., 2004), the current statewide implementation of home-based MST for youth with conduct problems does not. As such, the approach of this study was to calculate effect sizes for comparable measures of youth functioning from MST entry to exit found in the randomized controlled trials of MST. These effect sizes were then compared to those found in the present study. A review found four efficacy studies with nine such measures (Borduin, et al., 1995; Henggeler, Melton, & Smith, 1992; Henggeler, et al., 1997; Henggeler, et al., 1999) While effect sizes in most meta-analyses compare improvement rates following treatment to change observed in the randomly assigned control group (e.g. Curtis, et al., 2004; Littell, et al., 2005), an alternative approach was taken so that direct comparisons between the CAMHD MST findings and those from published efficacy studies could be made. One advantage of this approach is that changes in the control group (improvement or deterioration) do not influence effect size estimates. Given that this study did not include long-term follow-up data on MST youth, such data from the literature were excluded in the analyses.

Methods

Description of Measures

MST Therapist-rated Outcomes. During the initial stages of treatment, MST therapists meet with youth and caregivers to frame appropriate MST goals. These goals are typically based on traditional MST goals and the youth's specific needs. Specific goals can be altered during the course of MST, although this does not normally occur. At the end of MST treatment, the MST therapist again meets with the youth and caregivers who determine through a collaborative effort if each of the agreed upon treatment goals has been met. Hawaii currently adopts a three level conceptualization of MST goal-based outcomes in which cases can be considered, "Successful", "Partially Successful", or "Unsuccessful" at discharge from MST. Cases were rated as "Successful" if 100% of assigned goals were completed. "Partially Successful" MST cases achieved 25% or more of their goals, but less than 100%. "Unsuccessful" MST cases were those where less than 25% of treatment goals are met.

Child and Adolescent Functional Assessment Scale (CAFAS; Hodges, 1998). The CAFAS is a 200-item clinician report scale that measures youth's level of functional impairment. Based on their knowledge and experience with the child, raters review behavioral descriptions ordered by level of impairment within eight domains of functioning. The subscales of School Role Performance, Home Role Performance, Community Role Performance, Behavior Toward Others, Mood/Emotions, Mood/Self-Harmful Behavior, Substance Use, and Thinking are calculated by scoring the highest level of impairment (i.e., severe = 30, moderate = 20, mild = 10, no/minimal = 0) endorsed within the respective domain of items. An eight-scale total score is calculated by summing across the eight subscales, whereas a five-scale total is calculated by

summing the raw scores from behavior, substance use, and thinking scales with the maximum score from the school, home, and community role performance scales and with the maximum score from the emotions and self-harm. The CAFAS has been found to have acceptable internal consistency across items, inter-rater reliability across sites, and stability across time (Hodges, 1995; Hodges & Wong, 1996). Studies of concurrent validity have found that CAFAS scores are related to severity of psychiatric diagnosis, intensity of care provided, restrictiveness of living settings, juvenile justice involvement, social relationship difficulties, school-related problems, and risk factors. Studies of predictive validity have found that CAFAS scores from intake assessments predict service utilization and cost for services. Care coordinators serve as the primary raters for the CAFAS and results are entered directly into a networked computer scoring program by care coordinators or statistics clerks. CAFAS scores utilized in the current investigation were calculated using the eight-scale scoring procedure.

Child and Adolescent Level of Care Utilization System (American Academy of Child and Adolescent Psychiatry, 1999). The CALOCUS is a clinician rating form. Clinicians make dimensional ratings on a five-point scale in the domains of risk of harm, functional status, comorbidity, environmental stress, environmental support, resiliency and treatment history, child treatment acceptance and engagement, and parent treatment acceptance and engagement. These ratings may be summed to yield a total score, but are also combined through a detailed algorithm into a level of care judgment into one of seven categories: basic services (Level 0), recovery maintenance and health management (Level 1), outpatient services (Level 2), intensive outpatient services (Level 3), intensive integrated service without 24-hour medical monitoring (Level 4), non-secure, 24-hour, medically monitored services (Level 5), and secure, 24-hour, medically managed services. Evaluation of the instrument's reliability (Fallon et al., 2002) indicated that intra-judge agreement based on clinical vignettes ranged from ICC (2,2) = .57 - .95 across scales with all scales above .70 except for environmental stress and child treatment acceptance and engagement. Preliminary validity analysis found that the CALOCUS total score correlated -.33 with the Child Global Assessment of Scale (CGAS) and .62 with the CAFAS eight-scale total score. Care coordinators serve as the primary raters for the CALOCUS and results are entered directly into a networked computer scoring program by care coordinators or statistics clerks. CALOCUS analyses utilized CALOCUS level of care judgment scores using the method indicated above.

Participants

All MST cases admitted into the CAMHD system on or after January 1, 2004 that also completed MST before December 13, 2005 were reviewed for inclusion in this study. Cases were excluded from analyses if they were classified as MST booster sessions (n = 9), had no closing date or final goal information (n = 89), were open less than one month and were closed as "Other" (n = 6). Cases without gender or diagnostic information, with data entry errors, and with no (zero) treatment goals were also dropped (n = 8). One case longer than 30-days was considered an incomplete case and was also eliminated from analyses. Only the initial MST trial was used for the few children who received more than one trial of MST.

After eliminating cases, the remaining sample included 254 participants. This sample is referred to as the "full MST sample" and was used in the analyses examining predictors of MST outcomes. One hundred twenty two (122) of the full sample of 254 participants had MST entry and exit CAFAS and CALOCUS scores. These 122 participants are referred to as the "full CAFAS and CALOCUS information sub-sample" and were used to examine the validity of therapist-rated MST outcomes by comparing these outcomes to changes in CAFAS and CALOCUS scores over the course of MST and when comparing MST-Hawaii rates of change (CAFAS and CALOCUS) to those reported in the efficacy literature.

Demographic characteristics including age, length of service, gender, and ethnicity for the participants in the full and sub-sample are depicted in Table 1. Participants in the study were on average about 15 years old, ranging from 7 to 18 years. Nearly two-thirds of the participants were males. The largest ethnic category was Multiethnic, with White or Caucasian, Native Hawaiian or Pacific Islander, and Asian comprising the remaining major ethnic categories. Over one third of the participants' ethnicity information was not available. The number of days with MST services ranged from 24 to 244, with a mean around 135 days. There were no significant differences in demographic characteristics or length of treatment between those included and not included in the "full CAFAS and CALOCUS sub-sample" analyses.

Table 2 depicts diagnostic information about youth in the study. Most MST youth carried a primary diagnosis of disruptive behavior disorders, mood disorders, or attention disorders and the rate of these diagnoses did not differ across samples. Since many participants carried more than one diagnosis, Table 2 depicts the number and percent of participants who had *any* diagnosis of these three common disorder categories. The extent of comorbidity among participants is depicted in the bottom of Table 2. More than three-quarter of the youth carried more than one diagnosis across Axis I and II. There were no significant differences in diagnostic characteristics between those included and not included in the sub-sample analyses.

Table 3 depicts MST service data. The length of MST service for CAMHD youth was found to be within the traditional standard, lasting on average approximately 135 days (roughly five months). However there was considerable variability in the length of treatment and this is examined later as a possible predictor of outcomes.

Table 3 also depicts the average number of individual MST goals listed at the end of MST treatment (approximately 4 per case) and the number of these goals that were judged to have been met (slightly less than 3 per case). While the average number of goals assigned is comparable to national norms, the range of goals assigned (one to nine) is greater than that used by MST developers (Henggeler, et al., 1998).

Findings

What percent of therapist-rated MST goals were met at end of treatment?

Percent of goals met was calculated by dividing the number of goals met at end of MST treatment by the number of goals assigned (and multiplying by 100). The mean percent of goals met by CAMHD youth included in this study was around 70% and did not differ across the samples (see Table 4).

Table 1
Mean Age and Frequency of Gender and Ethnicity of MST Youth

	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-sample *N* = 122
Mean age in years (*SD*)	14.83 (2.01)	14.78 (2.11)
Range	7.06 – 18.05	7.06 – 18.05
Gender (% of total)		
Female	91 (35.8%)	50 (41.0%)
Male	163 (64.2%)	72 (59.0%)
Ethnicity (% of available)		
Hispanic, Latino, or Spanish origin		
Cuban	0 (0%)	0 (0%)
Mexican	0 (0%)	0 (0%)
Puerto Rican	0 (0%)	0 (0%)
Asian		
Asian Indian	0 (0%)	0 (0%)
Chinese	0 (0%)	0 (0%)
Filipino	3 (1.2%)	1 (0.8%)
Japanese	3 (1.2%)	2 (1.6%)
Korean	0 (0%)	0 (0%)
Vietnamese	1 (0.4%)	1 (0.8%)
Other Asian	0 (0%)	0 (0%)
Native Hawaiian or other Pacific Islander		
Guamanian or Chamorro	2 (0.8%)	0 (0%)
Micronesian	2 (0.8%)	0 (0%)
Native Hawaiian	5 (2.0%)	2 (1.6%)
Samoan	4 (1.6%)	2 (1.6%)
Other Pacific Islander	1 (0.4%)	0 (0%)
Other		
Alaska Native	0 (0%)	0 (0%)
American Indian	0 (0%)	0 (0%)
Black or African-American	0 (0%)	0 (0%)
Portuguese	0 (0%)	0 (0%)
White or Caucasian	20 (7.9%)	14 (11.5%)
Other race or ethnicity not listed	1 (0.4%)	0 (0%)
Multiethnic	97 (38.2%)	56 (45.9%)
Missing	115 (31.2%)	44 (26.5%)

Note. Mean age in years statistic *n* = 243 due to missing intake date information.

Table 2
Frequency of Primary Diagnoses, Any Diagnoses, and Mean Number of Diagnoses for MST Youth

	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-Sample *N* = 122
Primary diagnosis (% of available)		
Attentional disorders	53 (20.9%)	21 (17.2%)
Adjustment disorders	13 (5.1%)	7 (5.7%)
Anxiety disorders	11 (4.3%)	5 (4.1%)
Mood disorders	70 (27.6%)	38 (31.1%)
Disruptive behavior disorders	84 (33.1%)	41 (33.6%)
Substance related disorders	9 (3.5%)	7 (5.7%)
Pervasive developmental disorders	0 (0.0%)	0 (0.0%)
Mental retardation	4 (1.6%)	1 (0.8%)
Deferred	0 (0.0%)	0 (0.0%)
Miscellaneous disorders	0 (0.0%)	2 (1.6%)
Any disruptive behavior disorder	162 (63.8%)	75 (61.5%)
Any attentional problems	98 (38.6%)	42 (34.4%)
Any mood problems	104 (40.9%)	54 (44.3%)
Mean number of comorbid diagnoses (*SD*)	2.28 (0.88)	2.28 (0.92)
N of clients with comorbidity (%)	200 (78.7%)	93 (76.2%)

Table 3
Mean Length of Service, Number of Goals Assigned, and Number of Goals Met

	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-Sample *N* = 122
Mean length of service in days (*SD*)	134.68 (40.71)	135.82 (41.69)
Range	24 - 244	24 - 244
Mean number of goals at close (*SD*)	4.21 (1.22)	4.05 (1.13)
Range	1 - 9	2 – 7
Mean number of goals met at close (*SD*)	2.94 (1.83)	2.88 (1.79)
Range	0 - 8	0 – 7

Note. Mean length of service is computed on an *N* of 243 due to missing intake date information.

Table 4
Mean Percent of Goals Met at Time of MST Discharge

	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-sample *N* = 122
Mean percent of goals met at close (*SD*)	69.2% (36.5%)	70.02% (37.4%)
Range	0.0% - 100.0%	0.00% - 100.0%

What percent of youth were rated by their therapists as meeting few, some, or all of their goals at the end of MST services?

Hawaii utilizes a trivariate conception of MST outcomes by which "Successful" cases are classified as those youth who complete 100% of stated goals; "Partially Successful" reflects completing 25% to 99% of goals; and "Unsuccessful" cases where less than 25% of MST goals were met. The number and percent of MST cases within each category of outcome are detailed in Table 5. Using these categories, about one half of all MST cases were deemed "Successful", while approximately one third were rated as "Partially Successful", and the remaining were rated as "Unsuccessful". There were no significant differences in rates of "Success" between those included and not included in the sub-sample analyses.

Table 5
Percent of Outcomes in MST Youth

MST Outcomes	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-sample *N* = 122
Successful	122 (48.0%)	62 (50.8%)
Partially successful	92 (36.2%)	40 (32.8%)
Unsuccessful	40 (15.7%)	20 (16.4%)

Do CAMHD cases show improvement over the course of MST treatment as assessed by other validated measures?

CAFAS and CALOCUS scores around MST entry and exit where identified for each participant. CAFAS and CALOCUS ratings within 45 days were recorded. The CAFAS and CALOCUS conducted closest to MST intake and exit was then used for analyses. Cases where one or more CAFAS or CALOCUS evaluation was not administered during either entry and exit time frames are not included in the sub-sample used to assess the validity of the MST outcomes measure.

Average CAFAS and CALOCUS scores for MST youth are reported in Table 6. Average CAFAS scores showed statistically significant improvement ($t = 3.36$, $p < .001$) from MST entry (99.9) to discharge (88.8). Similarly, average CALOCUS scores indicated significant improvement ($t = 3.37$, $p < .001$) from MST entry (3.48) to discharge (3.08). Analyses conducted on the sub-sample with complete CAFAS and CALOCUS data ($n = 122$) also indicated significant improvement in both measures.

Do ratings of "Success" relate to changes in CAFAS and CALOCUS scores over the course of MST services.

The validity of therapist-rated MST outcomes was examined by comparing these ratings to change in CAFAS and CALOCUS scores using the $n = 122$ full CAFAS and CALOCUS sub-sample. Two related sets of analyses were done. In order to assess the size of the relationship between therapist-rated outcomes and change in CAFAS (or CALOCUS) scores, partial correlations between outcomes and exit scores were calculated, after controlling for entry scores. Then, two repeated measures analyses of

Table 6
CAFAS and CALOCUS Entry, Exit, and Difference Scores of MST Youth

	Full MST Sample *N* = 254	Full CAFAS & CALOCUS Sub-sample *N* = 122
CAFAS entry (*SD*; *n*)	99.9 (34.5; *n* = 189)	99.4 (32.4)
CAFAS exit (*SD*; *n*)	88.8 (40.3; *n* = 196)	88.9 (40.8)
CAFAS entry-exit difference scores (*SD*; *n*)	-12.6 (47.4; *n* = 159)	-10.6 (46.0)
CALOCUS entry (*SD*; *n*)	3.48 (1.19; *n* = 176)	3.56 (1.08)
CALOCUS exit (*SD*; *n*)	3.08 (1.52; *n* = 166)	3.13 (1.49)
CALOCUS entry-exit difference scores (*SD*; *n*)	-0.42 (1.44; *n* = 137)	-0.43 (1.48)

variance were conducted to compare the direction and extent of change in CAFAS (or CALOCUS) scores over the course of MST for participants in each of the three outcome categories ("Unsuccessful", "Partially Successful" and "Successful").

To conduct the correlational analyses, levels of success were coded as follows: "Unsuccessful" = 1, "Partial Success" = 2 and "Success" = 3. Since higher scores on CAFAS and CALOCUS reflect greater impairment and higher service needs, inverse (negative) correlations were expected. Simple bivariate correlations examining the relationship between therapist-rated MST outcomes and exit CAFAS and CALOCUS scores were first conducted. Indeed, MST therapist-rated outcomes were negatively correlated with exit CAFAS ($r = -0.52$, $p < .001$) and CALOCUS ($r = -0.51$, $p < .001$) scores indicating that more "Successful" cases as rated by MST outcomes scored lower on level of impairment and service needs at completion of MST. Next, partial correlations between therapist-rated outcomes and exit scores, while controlling for entry scores, were conducted to provide data about change in impairment and services levels. Significant inverse (negative) correlations with CAFAS ($r = -0.53$, $p < .001$) and CALOCUS ($r = -0.52$, $p < .001$) were found, indicating that therapist-rated outcomes were related to these other measures of improvement in the predicted fashion.

The second set of analyses focused on direction and extent of change in functioning and service level need for the three outcome groups of participants. Two 2 × 3 repeated measures ANOVAs (time by outcome category) were conducted to determine specific patterns of change on level of functioning (CAFAS) and service needs (CALOCUS) as a function of therapist-rated outcomes for the full sub-sample. These significant interactions are depicted in Figure 1 and 2 below.

Given these significant interactions, simple main effects tests for time (pre-post MST) were conducted for each of the three outcome groups separately. "Successful" MST cases showed a significant improvement in child functioning and significantly lower service needs at MST exit (CAFAS $t(61) = 6.86$, $p < .001$; CALOCUS $t(61) = 6.76$, $p < .001$). MST entry and exit scores on CAFAS and CALOCUS for "Partially Successful" cases were not significantly different, $t(39) = -0.42$, $p = .68$ and $t(39) = 0.00$, $p = 1.00$ respectively. "Unsuccessful" MST cases showed significant deterioration in client functioning and significant increases on level of care needs, $t(19) = -2.60$, $p < .05$ and $t(19) = -2.90$, $p < .01$.

Figure 1. Mean comparisons between entry and exit CAFAS and MST Outcomes

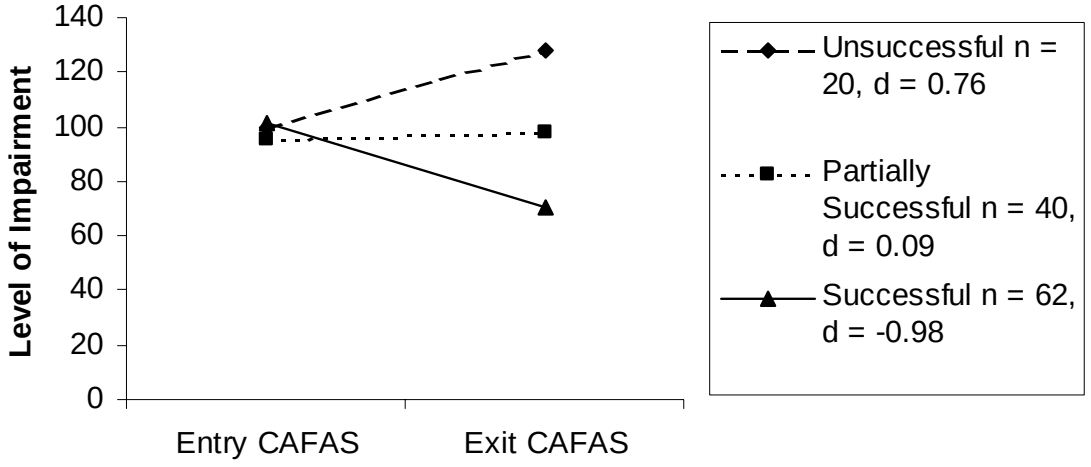
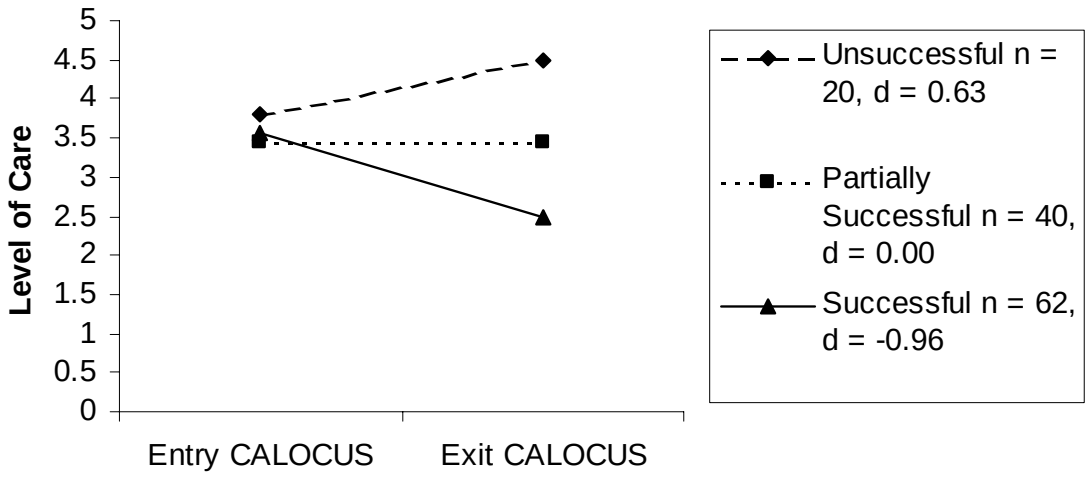


Figure 2. Mean comparisons between entry and exit CALOCUS and MST Outcomes



Overall, findings support the validity of therapist-rated outcomes as meaningful measures of success. These therapist-rated outcome measures were correlated with other validated measures used in the CAMHD system and the pattern of change in functioning and level of care needs differed across the three outcome groups. "Successful" MST cases decreased in CAFAS and CALOCUS scores from MST entry to discharge while

"unsuccessful" showed significant worsening on CAFAS and CALOCUS scores. Of particular note is that the "Partial Success" group did not show reliable change from entry to exit on either CAFAS or CALOCUS scores. This finding suggests that "Partial Success" cases should be interpreted with caution and not categorized together with "Successful" MST cases when conducting program reviews.

How do overall rates of improvement in child functioning compare to rates seen in randomized controlled trials of MST?

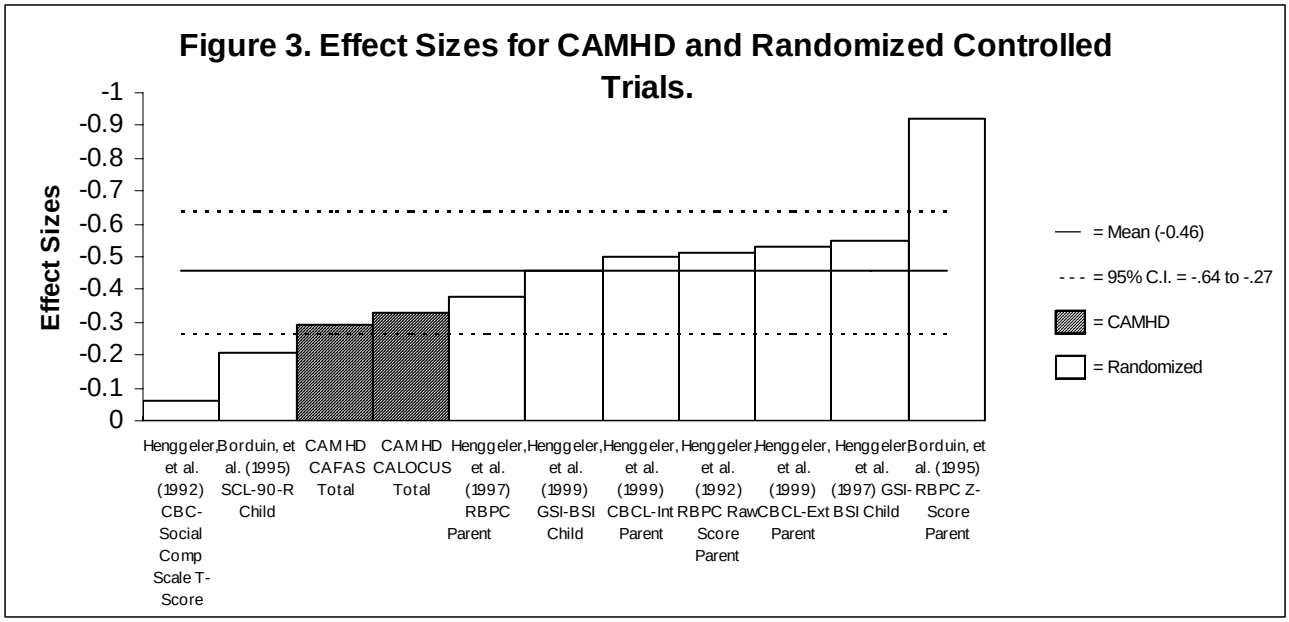
The amount of change in CAFAS and CALOCUS scores allow for rough comparisons of rates of improvement between MST in Hawaii and elsewhere. For these comparisons, all randomized controlled efficacy trials of MST published to date were examined. Four studies that reported the necessary means and standard deviations on one or more measures of child functioning (symptoms, emotional or behavioral problems) both at time of MST entry and exit were found. Pre-post effect sizes for each appropriate measure (n = 9) from each of these published studies were then calculated and compared to the CAFAS and CALOCUS entry-exit effect sizes in the present data.

The mean effect size for the nine published findings was –0.46 with a standard deviation of .25. As can be seen in Figure 3, effect sizes for both CAMHD-MST CAFAS (d = -0.29) and CALOCUS (d = -0.33) scores are below the mean effect size reported in the literature, but are within the range of those reports (-.07 to -.96). Effect size for pre-post CAFAS and CALOCUS outcomes found in MST in Hawaii were also found to be within the 95% confidence intervals calculated for the randomized controlled trials (95% C.I. = -.27 to -.64).

Although outcome improvement effect sizes found in MST in Hawaii are below the mean found in the efficacy literature, this finding is consistent with evaluations of implementation of MST programs that do not receive expert consultation (Henggeler et al., 1997) The diminished effects were significantly related to treatment adherence and point to possible improvement in MST in Hawaii.

What factors, if any, predict therapist-rated levels of success?

All 254 participants in the full MST sample were used to examine predictors of therapist rated success (since complete CAFAS and CALOCUS data were not needed for these analyses). Given that earlier analyses suggested that "Partially Successful" showed little or no average improvement on other measures, these cases were combined with the unsuccessful cases into a "Not Successful" group. In other words, "Successful" cases were ones where youth achieved 100% of goals assigned. Cases that achieved less than 100% of goals were classified as "Non-Successful". This bivariate classification had the added advantage of creating sufficient sample sizes in each condition for these analyses and fit the data well, with one exception. The traditional CAMHD trivariate conception of MST outcomes was retained and reported here when examining length of service effects since this conceptualization of success better fit the data regarding length of MST service.



Note. Effect size comparison analysis was conducted on measures taken from Henggeler, et al. (1992) CBC-Social Comp Scale T-Score* N = 43, Borduin, et al. (1995) SCL-90-R Child N = 70, CAMHD CAFAS Total N = 122, CAMHD CALOCUS N = 122, Henggeler, et al. (1997) RBPC Parent, Henggeler, et al. (1999) GSI-BSI Child N = 56, Henggeler, et al. (1999) CBCL-Int Parent N = 56, Henggeler, et al. (1992) RBPC Raw Score Parent N = 43, Henggeler, et al. (1999) CBCL-Ext Parent N = 56, Henggeler, et al. (1997) GSI-BSI Child, Borduin, et al. (1995) RBPC Z-Score Parent N = 70. Mean and confidence intervals based only on randomized controlled efficacy studies and exclude CAMHD data.

Do client demographic factors predict success in MST (as rated by therapists)?

The likelihood of cases being rated a "Success" did not significantly differ for male or female youth ($\chi^2(1, N = 254) = 1.26$, $p = 0.26$). In the present study, 45.4% of males and 52.7% of females were rated as "Successful" at the end of MST services. Among cases rated as "Non-Successful", 54.6% of youth were males and 47.3% were females. Age of youth was not a significant factor in cases achieving "Successful" outcomes ($t(241) = 1.62$, $p = .11$. Given high missing rates on client ethnicity, no analyses were conducted on this variable.

Do clinical factors predict success in MST (as rated by therapists)?

Next, youth's primary diagnosis and whether the youth carried any of the three most common diagnoses in CAMHD were examined as predictors of MST success. Comparisons were made between MST youth with each of the three most common primary diagnoses seen in CAMHD. Overall, there were no significant differences in rates of successful outcomes as a function of primary diagnosis (top of Table 7) or whether the youth had any of the three common diagnoses. Youth with a primary diagnosis related to disruptive behavior were somewhat less likely to succeed (around 40% compare to about 50% for the other common primary diagnoses), but this difference was not statistically significant, ($\chi^2(2, N = 207) = 2.50$, $p = 0.29$).

Table 7
Number (and Percent) of Successful and Non-Successful Cases by Primary and Any Diagnosis

	Non-Successful (partially or unsuccessful) *N* = 132	Successful *N* = *122*
Primary Diagnosis		
Attentional disorders	26 (49.1%)	27 (50.9%)
Mood disorders	35 (50.0%)	35 (50.0%)
Disruptive behavior disorders	51 (60.7%)	33 (39.3%)
Any Diagnosis		
Attentional disorders	48 (49.0%)	50 (51.0%)
Mood disorders	49 (47.1%)	55 (52.9%)
Disruptive behavior disorders	84 (51.9%)	78 (48.1%)

Next, MST therapist-rated outcomes were compared for youth who carried more than one diagnosis on either DSM Axis I or II to those who carried only one diagnosis. Results from this analysis are presented in Table 8. Surprisingly, youth with more than one diagnosis were more likely to be rated "Successful" than youth with a single diagnosis, $\chi^2(1, N = 254) = 4.53$, $p < 0.05$. This finding stands in direct contrast to a commonly held objection that empirically-based treatments might not be as effective with youth with comorbid disorders. At the same time, this finding is somewhat counterintuitive and should be replicated to determine its generality.

Table 8
Number (and Percent) of Successful and Non-Successful Cases by Comorbidity

Axis I and II any comorbidity	Non-Successful (partially or unsuccessful) *N* = 132	Successful *N* = *122*
No comorbid diagnosis	35 (64.8%)	19 (35.2%)
Comorbid diagnosis	97 (48.5%)	103 (51.5%)

In general, "Successful" cases had longer treatment episodes than "Non-Successful" ones, $t(241) = -2.96$, $p < .01$. When we examined this effect, we saw that it was largely due to shorter treatment periods for the original "Unsuccessful" cases, $F(2) = 13.28$, $p < .001$, so a three-level analysis is depicted in Table 9. The mean length of MST for "Successful" and "Partially Successful" cases was close to the five-month standard for MST ($M = 4.75$ and 4.56 months respectively and did not significantly differ from each other, $t(203) = -1.03$, $p = .31$). However, "Unsuccessful" cases had a significantly shorter mean length of service (3.52 months) compared to both the "Successful" ($t(156) = -5.46$, $p < .001$) and "Partially Successful" cases ($t(121) = -3.75$, $p < .001$).

Table 9
Mean (and Standard Deviation) Number of Days in MST by Trivariate Outcome

Trivariate Outcome	Number of Days of MST
Successful	142.39 (34.18)
Partially Successful	136.85 (43.04)

Unsuccessful	105.47 (42.4)

Note. N = 243. Comparable results were found with the sub-sample of cases with full CAFAS and CALOCUS data.

It is worth noting that comorbidity and length of treatment in MST were not significantly related to each other ($t(254) = -0.92$, $p = 0.36$). As such, criteria for positing a mediational model (Baron & Kenny, 1986), where the effects of comorbidity on therapist-rated outcomes are solely due to longer treatment for comorbid youth, were not met. Given this finding and the more general finding of comorbid youth having a higher likelihood of successful MST outcomes, more research seems indicated.

Summary and Recommendations

This report had three inter-related purposes: to examine the validity of therapist-rated outcome measures (based on goals attained), to examine client and service characteristics that might predict therapist-rated outcomes, and to compare the extent of improvement in child functioning observed in MST in Hawaii to rates reported in the MST randomized controlled trials literature.

In general, therapist-rated MST outcomes seem to be a valid measure of treatment outcome. Level of outcome (successful, partially successful, and unsuccessful) was correlated with both exit CAFAS and CALOCUS scores and amount of change in these scores over the course of MST. The specific validity of the category "partially successful" was less well supported. On average, youth with cases rated as "partially successful" did not show reliable improvement in CAFAS or CALOCUS scores over the course of MST. On average, youth with "successful" cases showed significant improvement and youth with "unsuccessful" cases showed significant declines on these assessment measures.

Comparable to prior findings in the literature, few demographic or clinical level variables predicted therapist-rated outcomes. Therapist-rated outcomes were unrelated to CAFAS and CALOCUS intake scores, youth's gender or primary diagnosis. There was a slight trend suggesting that younger clients were somewhat more likely to be rated more successful than older clients. Longer time in MST predicted more successful outcomes. However this effect was mostly due to shortened treatment lengths for "unsuccessful" cases. Surprisingly, cases with youth who had one or more comorbid diagnoses were more likely to be rated as successful. This effect was not accounted for by longer treatments for youth with comorbid diagnoses and points to interesting questions that can be addressed in further research and evaluation.

The mean pre-post effect size on CAFAS ($d = -0.33$) and CALOCUS ($d = -0.40$) scores in MST in Hawaii are within the range of comparable pre-post changes over the course of MST but somewhat smaller than the mean of these pre-post effect size reported in published randomized controlled efficacy trials.

Given the present findings, the following recommendations are made:

- *Continue use of therapist measures of outcome.* Therapist ratings of "Success" are associated with improvements on other measures of functional status and service needs.
- *Reconsider the meaning of "Partially Successful".* Many "Partially Successful" cases do not show improvement in level of functioning or service level needs. As such, it might be more accurate to exclude such cases from any overall measure of "Success".
- *Consider other factors that might influence likelihood of "Success".* While few client level factors predicted outcomes, there remains considerable variability in outcomes that if explained, could point to ways to improve overall success rates. Such factors might relate to family and ecology characteristics, treatment delivery, or other factors.
- *Continue to work toward increasing the rate of improvement for MST cases in Hawaii.* Average improvement over the course of MST in Hawaii, while within the range found in randomized controlled efficacy studies in the published literature is somewhat smaller than average rates reported in these studies. Efforts that focus on quality assurance and treatment fidelity could lead to improved outcomes (c.f., Henggeler et al., 1997).
- *Consider why youth with comorbid diagnoses are more likely to be rated as "successful".* Youth with comorbid diagnoses might present a treatment profiles that differ from youth with only one (most likely a disruptive behavior) diagnosis, which might relate to better outcomes. Given considerable controversy about the effectiveness of evidence-based interventions with youth with complex problems, further work in this area is warranted.
- *Consider using the Monthly Treatment Progress Summary Form as a way to examine whether specific treatment targets and/or treatment components relate to MST success.* The MTPS shows promising psychometric properties to date and might prove useful in examining state-wide MST services and outcomes. Cases that consistently apply practice elements and focus on treatment targets consistent with MST philosophy and practice may well show greater improvement and higher success rates than other cases. The MTPS might also help elucidate the findings related to comorbidity and therapist-rated outcomes.
- *Consider MST evaluation of ultimate outcome measures assessed in other MST studies.* Comprehensive evaluation of MST Services in Hawaii should include examination of other outcomes, such as days in out-of-home placement or days of incarceration, which have been used as ultimate outcome measures in other studies (e.g., Henggeler et al., 1997) and are key cost drivers.

References

American Academy of Child and Adolescent Psychiatry & American Association of Community Psychiatrists (1999). Child and Adolescent Level of Care Utilization System: User's Manual. Author.

Baron, R. & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Borduin, C.M., Mann, B.J., Cone, L.T., Henggeler, S.W., Fucci, B.R., Blaske, D.M., et. al., (1995). Multisystemic therapy of serious juvenile defenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology, 63*(4), 569-578.

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

Chambless, D.L. & Ollendick, T.H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52,* 685-716.

Child and Adolescent Mental Health Division (2004). *Evidence Based Services Committee: 2004 biennial report, summary of effective interventions for youth with behavioral and emotional needs*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 18, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/ebs/ebs011.pdf

Child and Adolescent Mental Health Division (2006). *Child and Adolescent Mental Health Division strategic plan 2007-2010*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 19, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/camhdplan2.pdf

Chorpita, B.F., Yim, L.M., Donkervoet, J.C., Arensdorf, A., Amundsen, M.J., McGee, C., et al. (2002). Toward large-scale implementation of empirically supported treatments for children: A review of and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice, 9*, 165-190.

Curtis, N.M., Ronan, K.R., & Borduin, C.M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology, 18*(3), 411-419.

Daleiden, E. (2006). *Annual evaluation report: Fiscal year 2006*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 18, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge017.pdf

Daleiden, E.L. & Chorpita, B.F. (2005). From data to wisdom: Quality improvement strategies supporting large-scale implementation of evidence-based services. *Child and Adolescent Psychiatric Clinics of North America, 14*, 329-349.

Daleiden, E., Lee, J., & Tolman, R. (2004). *Annual evaluation report: Fiscal year 2004*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 18, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge011.pdf

Daleiden, E. & Tolman, R. (2005). *Annual evaluation report: Fiscal year 2005*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 18, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge015.pdf

Fallon, T., Pumariega, A., Sowers, W., Klaehn, R., Huffine, C., et al. (2006). A level of care instrument for children's systems of care: Construction, reliability, and validity. *Journal of Child and Family Studies, 15*, 143-145.

Henggeler, S.W., Melton, G.B., & Smith, L.A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60*(6), 953-961.

Henggeler, S.W., Melton, G.B., Brondino, M.J., Scherer, D.G., & Hanley J.H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*(5), 821-833.

Henggeler, S.W., Rowland, M.D., Randall, J., Ward, D.M., Pickrel, S.G., Cunningham, P.B., Meller, S.L., Edwards, J., Zealberg, J.J., Hand, L.D., & Santos, A.B. (1999). Home-based multisystemic therapy as an alternative to the hospitalization of youths in psychiatric crisis: Clinical outcomes. *Journal of the American Academy of Child Adolescent Psychiatry, 38*(11), 1331-1339.

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

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

Littell, J. H., Popa, M., & Forsythe, B. (2005). Multisystemic Therapy for social, emotional, and behavior problems in youth age 10–17. *Cochrane library, 3*, Chichester, UK: John Wiley & Sons, Ltd.

Nakamura, B.J., Daleiden, E.L., & Mueller, C.W. (2006). *Monthly Treatment & Progress Summary Form Report: Progress Rating Validity June 2003: September 2005*. Honolulu, HI: Hawaii Department of Health Child and Adolescent Mental Health Division. Retrieved April 18, 2007, from http://www.hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/mr/mr003.pdf"

Rowland, M.D., Halliday-Boykins, C.A., Henggeler, S.W., Cunningham, P.B., Lee, T.G., Kruesi, M.J.P., & Shapiro, S.B. (2004). A randomized trial of multisystemic therapy with Hawaii's felix class youth. *Journal of Emotional and Behavioral Disorders, 13*(1), 13-23.

Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services, 52,* 1190-1197.