Published in final edited form as:

J Consult Clin Psychol. 2013 December ; 81(6): . doi:10.1037/a0033928.

An Independent Randomized Clinical Trial of Multisystemic Therapy with Non-Court-Referred Adolescents with Serious Conduct Problems

Bahr Weiss, Susan Han, Vicki Harris, and Tom Catron

Vanderbilt University

Victoria K. Ngo

The Rand Corporation

Annalise Caron

American Institute for Cognitive Therapy

Robert Gallop

West Chester University

Carol Guth

Vanderbilt University

Abstract

Objective—Adolescent conduct problems exact serious social as well as personal costs, and effective treatments are essential. One of the most widely disseminated and effective programs for the treatment of serious conduct problems in adolescents is Multisystemic Therapy (MST). However, most evaluations of MST have involved the developers of MST. The purpose of the present study was to conduct an independent evaluation of MST, with non-court-referred adolescents with conduct problems.

Method—Participants were 164 adolescents aged 11 to 18 years who were recruited from self-contained behavior intervention classrooms in public schools. Adolescents and their families were randomly assigned to receive MST or services as usual. Outcome measures assessed conduct problems, school functioning, and court records of criminal behavior. Participants were followed for 18 months after baseline using parent, adolescent, and teacher reports; arrest data were collected for 2.5 years post-baseline.

Results—Two of four primary outcome measures focused on externalizing problems showed significant treatment effects favoring MST. Several secondary and intervention targets pertaining to family functioning and parent psychopathology showed positive effects of MST, and no negative effects were identified.

Conclusions—Results provide some further support for the effectiveness of MST, although smaller effect sizes than previous studies also suggest the complexity of successful dissemination, particularly to non-court-referred populations.

Keywords

multisystemic therapy; MST; delinquency; antisocial behavior; non-court-referred

Adolescent conduct problems exact substantial personal and social costs, making effective treatments essential. One of the most widely disseminated programs for treatment of adolescent conduct problems is Multisystemic Therapy (MST; e.g., Henggeler, Cunningham, Schoenwald, & Borduin, 2009). MST is a family-focused, home-based treatment for adolescents with serious antisocial behavior and emotional disturbance. MST is derived in part from Bronfenbrenner's (1979) theory of social ecology, and views individuals' behavior as embedded in a complex network of interconnected social systems. Treatment is multi-faceted, targeting the behavior of individuals, family, peer, and other key systems such as the school (Henggeler et al., 2009).

In the first MST outcome study of youth with serious antisocial behavior, Henggeler et al. (1986) reported improved family relations and decreased behavior problems for youth receiving MST relative to a comparison group. Subsequent evaluations have led to MST being generally accepted as an efficacious treatment (e.g., U.S. Department of Health and Human Services, 2001; although see Littell, Popa, & Forsythe, 2005, for a contrary opinion) and widely adopted in North America and Europe (Henggeler, 2011).

For a number of years, however, it has been suggested that results of evaluation studies may be unintentionally influenced by investigators' allegiance to the treatment. Several studies (e.g., Houts, Berman, & Abramson, 1994) have found that the magnitude of effects of psychosocial interventions is related to estimates of investigator allegiance even when allegiance is assessed in prior publications (e.g., Robinson, Berman, & Neimeyer, 1990). There have been several evaluations of MST for adolescent conduct problems in which the developers of MST have not been co-authors on study reports or directly involved in the evaluation. In general, these studies have found relatively modest effects. Ogden and Halliday-Boykins' (2004) study is generally considered one of the first independent evaluations of MST, although its second author has published extensively with the developers of MST. Ogden and Halliday-Boykins (2004) was conducted in Norway, with youth with serious antisocial behavior randomly assigned to receive MST or usual child welfare services. Compared to the control group, MST was associated with a significantly greater decrease in internalizing symptoms and a marginally significant greater decrease in externalizing symptoms. The magnitude of effects of Ogden and Halliday-Boykins (2004) as well as that of Sundell et al. (2008), an independent evaluation of MST conducted in Sweden, was more modest compared to other trials of MST (e.g., Borduin et al., 1995). This may have been due to the comprehensive nature of the "usual services" provided by Norway and Sweden's child welfare system.

An independent evaluation of MST was also conducted in the U.K. with court-referred juvenile offenders (Butler, Baruch, Hickey, & Fonagy, 2011). Youth were randomly assigned to receive MST or comprehensive services-as-usual that included substance abuse treatment, anger management training, social skills problem-solving, and other services. Butler et al. (2011) found that MST was associated with a significantly greater reduction in the number of nonviolent offenses at 18 months, and a non-significant trend favoring MST in regard to violent offenses. The MST group also showed significantly greater improvement in delinquent behavior as reported by parents and youth, and for aggressive behavior as reported by parents (but not youth).

In the first independent trial of MST conducted in the U.S., Timmons-Mitchell, Kishna, Bender, and Mitchell (2006) evaluated 93 adolescents involved with the juvenile justice system. Re-arrest rates for adolescents who received MST were significantly lower than for adolescents who received treatment as usual (67% vs. 87%), although time to re-arrest did not differ significantly for the two groups. Post-treatment scores for 4 of 6 subscales on the CAFAS functioning measure (Hodges & Wong, 1996) were significantly better for the MST

group compared to the treatment-as-usual group. However, the magnitude of these effects was lower than found in most previous trials conducted by the developers of MST. Moreover, outcome data were obtained from a single secondary source (i.e., court records coded by research assistants), which limits the generalizability of the results to broader adolescent conduct problems.

The purpose of the present study was to conduct a randomized, independent evaluation of MST in the United States. Strengths of the study included a multi-informant, multi-system assessment focused on psychopathology, interpersonal functioning, the home and school environment, and criminal behavior assessed through court records. Participants were randomly assigned to receive MST or services as usual. The present study is the first randomized, independent assessment of MST conducted in the U.S. involving adolescents with conduct problems who were recruited from a source other than the legal system.

In the present study, treatment fidelity was assessed through expert coding of audiotaped sessions as well as reports from parents and MST supervisors. In most studies of MST, treatment fidelity has relied on parent reports rather than coding of session audiotapes (e.g., Sundell et al., 2008), due to the costs. In the present study, we hypothesized that MST would be associated with significantly greater rates of improvement on our primary outcome measures than the control group, and that arrest rates would be significantly lower among adolescents who received MST than those who did not.

Method

Participants

Participants were 164 adolescents (aged 11 to 18 years) and their parents/guardians. Students in 7th thru 11th grades who were in self-contained behavior intervention classrooms (Moderate Intervention Program; MIP) in junior or senior high public schools were eligible for project participation. The only exclusionary criterion was the lack of a parent or guardian to provide consent. The study was conducted in a large, southeastern U.S. city, and was approved by and conducted in compliance with the overseeing Institutional Review Board.

The developers of MST (who provided training and supervisory consultation) agreed that recruitment of participants from MIP classrooms would produce an appropriate sample for MST. Students are placed in MIP classrooms because the school has deemed that they have serious conduct problems that interfere with their being educated in the general education setting. Based on data provided by the school system, students in MIP classrooms demonstrate significant conduct problems. For example, teacher ratings of MIP students' delinquent behavior on the Teacher Behavior Questionnaire (Catron & Weiss, 1994) are on average 1.8 standard deviations above that for general education students. Moreover, at baseline 68% of project participants had been involved in the juvenile justice system, and the mean *T* score for the Externalizing Problems scale on the parent-report Child Behavior Checklist (Achenbach & Rescorla, 2001) for our participants at baseline was 68, similar to other studies of MST (e.g., Rowland et al., 2005).

Enrollment and assignment to experimental condition—School personnel were paid by the research project to contact the parents/guardians of eligible students. The study was described to families; names and contact information of parents/guardians who provided consent to contact were provided to the research project. The research project then contacted the family, provided additional information, and scheduled a home interview for families interested in participating. During this interview, parent consent and adolescent assent were obtained, and the first assessment was conducted. After the assessment was completed, the

interviewer opened a sealed envelope that contained the family's random assignment to the treatment or control condition.

Participant characteristics—Of 213 families with MIP students in grades 7th thru 11th with a parent or legal guardian to provide consent, 23% declined to participate; the primary reason cited was a lack of time, and more generally a lack of interest. Of the 164 who agreed to participate in the study, 84 were randomly assigned to the treatment group and 80 to the control group. The mean age of the adolescents was 14.6 years ($SD=1.3$), 83% were male, and 60% of the adolescents were African American and 40% were Caucasian. The mean age of the parent or guardian was 40.8 years ($SD=8.8$), 71% of the parents had graduated from high school, and the median reported family income was \$17,500. Table 1 provides demographic information for the groups. Of the 164 participating families, 93% completed project participation (see Figure 1). Based on parent-report, teacher-report, and youth self-report Child Behavior Checklist (Achenbach & Rescorla, 2001), at baseline 87%, 85%, and 52% (respectively) of our sample scored at or above the borderline clinical range for Externalizing Problems, with 96% of the sample scoring at or above the borderline clinical range for Externalizing Problems for one or more informants. Intent-to-treat analyses were used in all analyses.

Treatment conditions

Control—A services-as-usual control group was used. Usual services consisted primarily of a behaviorally-focused classroom management plan provided by the school, with educational instruction occurring in self-contained classrooms. The control group members were assessed on the same schedule as treatment group members.

MST—MST is a principle-based, family-focused treatment program based on research literatures ranging from developmental psychology and child clinical psychology to social work. Treatment is multi-faceted and focuses on multiple systems, targeting disturbance in the behaviors of individuals, family, peers, and dyadic relationships. MST treatment does not involve application of a unique set of techniques but rather interventions for each family are integrated from problem-focused, empirically-validated treatment models that target etiological factors. Application of the interventions is guided by the nine MST treatment principles (e.g., focus on family strengths), the goal of which is to empower parents to manage their adolescent's social and personal life more adaptively (see Henggeler et al., 2009, for program details). In the present study, 96% of the families received family therapy; the three families that did not receive this intervention included two families that withdrew from treatment shortly after enrollment (but were included in the intent-to-treat analyses). In addition, 82% received parent training sessions, 95% received individual parent sessions, 95% received individual adolescent sessions, and 94% of adolescents received school-based interventions from the project that included individualized behavior management plans.

Clinical training and supervision

Initial training—The initial cohort of three clinicians and the on-site supervisor were trained at the research project site by the developers of MST from the Family Services Research Center (FSRC) at the Medical University of South Carolina (MUSC). This training included (a) didactic readings, (b) a 5-day on-site training workshop, and (c) quarterly 1.5-day booster training sessions. Didactic training focused on (a) the major theoretical models of deviant behavior, (b) literature regarding serious emotional disturbance in adolescents, and (c) clinical literature pertaining to the interventions often used in MST. Clinicians who

began involvement in the project after this initial training attended scheduled training sessions offered by FSRC.

Supervision and maintenance of treatment fidelity—The overriding purpose of MST supervision is to maintain treatment fidelity. In the present study, similar to many other MST efficacy and effectiveness studies (e.g., Letourneau et al., 2009; Schoenwald, Sheidow, Letourneau, & Liao, 2003), clinicians received on- as well as off-site supervision. MST clinicians documented in writing the progress of treatment implementation which was provided to the supervisor prior to supervision. Clinicians' contact logs recorded each contact with the family and described steps toward the goals for the family. Following supervision, for each case the on-site supervisor noted the therapist's status with respect to: (a) level of engagement of key members of the social environment; (b) the quality of the therapist's conceptualization of the case; (c) how intermediate goals were linked to overarching outcomes; (d) how interventions would achieve intermediate goals; and (e) identification of factors contributing to barriers to progress and strategies for overcoming barriers. Weekly supervision sessions also were conducted with the off-site consultant at FSRC via telephone. This weekly supervision with the FSRC consultant and additional quarterly booster or as-needed trainings that were conducted on-site were arranged under a consulting agreement with FSRC.

Several mechanisms were used by the FSRC for monitoring treatment fidelity. Most fundamentally, part of the collaborative agreement between the project and FSRC was that the project would be notified if there were concerns on the part of FSRC that the clinical team's adherence to MST was unacceptable, and that whatever clinical support was necessary to rectify the problem would be provided. Several measures were administered to formally monitor adherence (see *Treatment Fidelity Data* section below), including: (a) adherence ratings obtained from parents using the MST Therapy Adherence Measure (TAM; Henggeler & Borduin, 1992) as used in many MST studies (e.g., Letourneau et al., 2009); (b) adherence ratings from audiotapes of MST sessions; and (c) quarterly overall quality ratings by the on-site clinical supervisor and the FSRC consultant.

Clinician characteristics

The clinical team consisted of (a) three clinicians, (b) an on-site MST supervisor who was a licensed clinical social worker with 15 years of clinical experience, and (c) an off-site MST expert consultant at FSRC. During the project's four years of clinical implementation, eight clinicians were employed, the hiring all of whom was approved by FSRC. One had a B.A. in psychology and the others had a master's degree (three in social work, one each in psychiatric nursing, family therapy, divinity, and rehabilitation counseling). Their duration with the project ranged from 4 months to 2.8 years, with an average of 1.4 years. Reasons for the five clinicians' voluntary departure from the project included personal reasons unrelated to the project, or the clinician's perceived non-fit with the demands of the position. This rate of turnover is consistent with a large-scale study of MST clinicians in 45 community settings, in which the average length of time that MST clinicians worked in the study was 1.1 years (Sheidow, Schoenwald, Wagner, Allred, & Burns, 2007).

Assessment

Primary intervention outcomes—Our primary outcome domains were adolescent conduct problems as assessed by parent, adolescent, and teacher reports on the Child Behavior Checklist, and criminal charges that were obtained from court records. Parents completed the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), a broadband measure of children's behavioral and emotional problems. It contains 118 problem items rated on a 0–2 scale, and produces two broadband scales (*Internalizing* and

Externalizing Problems) as well as several narrowband scales. The CBCL scales have an average one-week test-retest reliability of .89 and a correlation of .81 with the Quay and Peterson (1983) Revised Behavior Problem Checklist (Achenbach & Rescorla, 2001). Teachers completed the Teacher Report Form (TRF; Achenbach & Rescorla, 2001) and adolescents completed the Youth Self-Report form (YSR; Achenbach & Rescorla, 2001), teacher- and adolescent-report versions of the CBCL. Like the CBCL, they contain 118 problem items rated on a 0–2 scale, and produce two broadband and eight narrowband scales; their reliability and validity are well-established (Achenbach & Rescorla, 2001). In the present study, given the focus of MST on conduct problems, the *Externalizing Problems* scales were our primary outcome measures.

With parent consent and adolescent assent, data regarding criminal charges were obtained from juvenile court records for one year prior to baseline through 2.5 years post-baseline. For participants who turned 18 during this time span, data were obtained from adult court records. Charges were coded as a status offense, misdemeanor, or felony. The primary analysis included incidents wherein the adolescent was charged with a felony. Secondly, we also analyzed incidents that involved (a) a misdemeanor and/or felony charge; and (b) any incident (status offense, misdemeanor, or felony).

Secondary intervention outcomes—Secondarily, we assessed several specific areas of conduct problems using the Self-Report Delinquency Scale with items covering delinquent behaviors and drug use (SRD; Elliott, Huiziga, & Ageton, 1985). The SRD was used to assess adolescents' reports of specific antisocial behaviors in the last three months and drug use within the last one month. Finally, our outcome analysis analyzed school functioning data, including grades (averaged across core academic subjects), attendance, and number of days suspended during the grading period.

Intervention targets—We assessed several factors explicitly targeted by MST as potential change mechanisms, including (a) family relationships, (b) parenting behavior, and (c) parent mental health problems (Henggeler et al., 2009). To assess family relationships, parents completed the Family Adaptability and Cohesion Evaluation Scales-III (FACES-III; Olson, Portner, & Lavee, 1985) which measures family functioning in regard to instrumental and affective relations. It produces two subscales (*Family Cohesion, Adaptability*), and the measure is well-validated (e.g., Edman, Cole, & Howard, 1990). To assess parenting behavior, the adolescent's primary caregiver completed the Parental Authority Questionnaire (PAQ; Buri, 1991). The PAQ produces three subscales (Authoritarian Parenting, Authoritative Parenting, Permissive Parenting) and has established reliability and validity (e.g., Buri, 1991). To assess parent mental health problems, parents completed the Personality Assessment Inventory (PAI; Morey, 1991), a well-validated (e.g., Ruiz & Edens, 2008) self-administered inventory of adult personality and psychopathology. In the present study, the *Aggressive, Antisocial, Anxiety, Borderline Personality Features, Depression, Paranoia*, and *Positive Impression Management* subscales were used. To reduce the number of subscales for analysis, we used the higher-order *Externalizing (Aggressive, Antisocial, Borderline Personality Features, Paranoia)* and *Internalizing (Anxiety, Depression)* factors identified by Ruiz and Edens (2008).

Treatment fidelity data—Three sources of fidelity data were collected. Following the schedule described in the *Procedures* section below, parents completed the Therapy Adherence Measure (TAM; Henggeler & Borduin, 1992), a 26-item questionnaire measuring family and therapist behaviors related to adherence to MST principles. Second, MST sessions were audiotaped and randomly selected tapes were rated by an MST expert at the FSRC on two dimensions. The first was adherence to the nine MST principles, rated on a

1 (*not at all*) to 7 (*completely*) scale. The second dimension was the Session Checklist items, which are more specific than the MST principles; these items were rated using a 1 (*not at all*) to 5 (*very much*) scale. Initially, one tape per clinician per two weeks was coded, then the frequency was decreased to once per month for each clinician as his/her adherence became well-established. The third source of treatment fidelity data was quarterly global ratings of each clinician's adherence to MST, provided by the on-site MST supervisor and the FSRC-based MST consultant who provided weekly supervision via telephone. A 7-point scale was used: 1=*poor adherence*; 2=*little evidence of adherence*; 3=*unacceptable, needs improvement*; 4=*inconsistent adherence*; 5=*acceptable, some room for improvement*; 6=*good adherence*; and 7=*excellent adherence*.

Service usage—In order to assess services from outside the project that the families and adolescents may have received, parents answered a *Service Use Questionnaire* (SUQ). At the baseline assessment, this questionnaire focused on the past year, and at subsequent assessments it focused on the time period since the previous assessment. The SUQ assessed usage of health-related services in general (e.g., seeing a doctor for a physical injury) as well as mental health-related services specifically. It also assessed whether mental health services were provided by a mental health professional (e.g., a psychologist) or a non-mental health professional (e.g., clergy). In the present study, we used two scores derived from the measure: whether over the course of the study the family received (a) any mental health services, and (b) mental health services provided by a mental health professional.

Satisfaction data—Treatment group parents answered a consumer satisfaction survey at post-treatment and follow-up assessments. This measure contains 16 items that assess various aspects of parents' satisfaction with the program, as well as their global perceptions of improvement in mental health and academic domains attributable to the program. Ratings are on a 1 to 5 scale, with 5 representing more satisfaction or change attributed to the program. Based on results of an exploratory factor analysis, the items were combined to produce an overall satisfaction score.

Procedure

After obtaining parent consent and adolescent assent, adolescents and parents/guardians were interviewed separately by two research assistants either at their home or another convenient location of the participants' choosing (e.g., school). Interviewers administered the assessment battery individually to the adolescent and parent, and read each measure to the adolescent/parent. Teachers were given assessment materials and completed the assessments on their own time. For both treatment and control group families, primary outcome assessments occurred at four timepoints: baseline, 3 months, 6 months, and 18 months. Parents received \$50 per assessment, adolescents \$20, and teachers \$10 per adolescent per timepoint. Treatment group parents completed the fidelity and satisfaction measures at 1, 2, 4, and 5 months after baseline as well as the other timepoints.

Results

Pre-treatment comparability of treatment and control groups

Treatment and control groups were compared on baseline scores for demographics, the primary dependent variables, and rate of project dropout. One-way analyses of variance were used for continuous data and chi-square tests for categorical data. Of the 15 tests one was significant, with the treatment group showing a higher proportion of parents who had graduated from high school ($\chi^2[1] = 5.11, p < .05$, see Table 1). However, this variable was not related to treatment outcome, so it was not included as a covariate in the outcome analyses.

Treatment fidelity

Parent reports using the TAM were completed every 4 weeks, producing 307 completed TAM. On a 5-point scale ranging from 1=*not at all* to 5=*very much*, means and standard deviations for the six TAM factors were: (a) *Adherence* factor mean=4.41 (SD=.51); (b) *Non-productive Sessions* factor mean=1.52 (SD=.66); (c) *Effort to Solve Problems* factor mean=4.55 (SD=.50); (d) *Therapist Attempts to Change Family's Interactions* factor mean=3.29 (SD=1.13); (e) *Therapeutic Lack of Direction* factor mean=1.60 (SD=.66); and (f) *Family-Therapist Consensus* factor mean=4.45 (SD=.56). Thus, overall therapists' adherence to MST principles was moderately high to high, and these means are consistent with those reported in other studies (e.g., Letourneau, Sheidow, & Schoenwald, 2002).

Over the course of the study, 218 audiotaped therapy sessions were coded by the FSRC expert. Mean ratings of clinicians' adherence to the nine MST principles were in the moderate to high adherence range. On a 1 to 7 scale, the lowest mean rating was 4.61 (moderate adherence) for Principle 5 (Interventions should target sequences of behavior within or between multiple systems) and the highest mean rating was 5.92 (high adherence) for Principle 6 (Interventions should be developmentally appropriate and fit the developmental needs of the family/youth). For the Session Checklist items coded by the expert, the mean rating across the 20 items on a 5-point scale was 3.74 (SD=.77) indicating *Pretty Much* (4) adherence to specific therapist behaviors that support MST principles. Means ranged from 2.80 (Discussing how the family can identify and use resources/supports in their natural environment) indicating *Some* (3) adherence, to 4.88 (Using a non-critical tone in addressing family members), with most item means indicating moderate (*Some* to *Pretty Much*) adherence.

In regard to the supervisor's and FSRC consultant's global quarterly ratings, the mean of the 32 ratings (made on a 7-point scale) was 5.08 (SD=1.09), ranging from 4 (*inconsistent adherence*) to 7 (*excellent adherence*), indicating that clinicians were overall implementing MST with acceptable adherence, with some room for improvement. Overall, MST fidelity was acceptable. This is substantiated by the fact that there were no instances where it was necessary for FSRC to intervene or raise concerns regarding the MST implementation during the project.

Overview of outcome analyses

A hierarchical linear model (HLM; Raudenbush & Bryk, 2002) approach was used for our analyses of continuous data (e.g., the CBCL), with the linear and quadratic effects of Time as random coefficients and Treatment as a fixed between-subjects effect. The timepoints consisted of baseline, mid-treatment (3 months), post-treatment (6 months), and follow-up (18 months) assessments. The Time variable was entered as the actual number of months post-baseline that the assessment occurred for each particular family. The linear effect for Time represented the extent to which the combined groups' rate of change differed from zero, whereas the quadratic effect of Time (Time $\times$ Time) represented the extent to which the combined groups' rate of change itself changed across time (i.e., the extent to which the rate of change accelerated or decelerated). The Time (linear) $\times$ Treatment interactions represented the extent to which the two groups' rates of linear change differed, and the Time (quadratic) $\times$ Treatment interactions represented the extent to which the two groups differed in the extent to which their rates of change accelerated or decelerated across time. Independent variables were standardized for ease of interpretation. We used intent-to-treat analyses. To obtain the most complete yet parsimonious model, quadratic effects (Time $\times$ Time, and Time $\times$ Time $\times$ Treatment) were first tested and dropped from the model if non-significant, and the linear effect was then analyzed. Table 2 presents the means and standard deviations for the intervention outcome and target variables.

Outcome results

Primary intervention outcomes—For the parent-report CBCL Externalizing Problems scale, there was a significant Time $\times$ Treatment interaction (see Table 3), in the context of a significant quadratic Time $\times$ Time effect. As Figure 2 illustrates, both groups showed a decelerating decrease in symptoms, with the treatment group's linear rate of decrease significantly greater than that for the control group. The adolescent-report YSR Externalizing Problems scale also showed a significant Time $\times$ Treatment interaction and a significant quadratic Time effect, as well as a significant Time $\times$ Time $\times$ Treatment interaction (see Table 3). As Figure 2 also shows, the treatment group showed a greater linear decrease with a maximal separation from the control group at about 13 months post-baseline.

Treatment effects on the teacher-report TRF Externalizing Problems scale were non-significant (see Table 3). Because some participants after baseline moved to less restrictive educational placements (e.g., general education classrooms) from the MIP classrooms, we examined class placement post-baseline. At Time 2, 4% of the treatment and control participants were in less restrictive placements; at Time 3, 6% of treatment and 1% of control participants were in less restrictive placements; and at Time 4, 10% of both groups were in less restrictive placements. None of these comparisons was significantly different. To examine whether the type of setting (less vs. more restrictive setting) in which the teacher made the ratings influenced the teacher ratings (e.g., because of teacher behavioral standards varying as a function of setting), we conducted follow-up analyses of the TRF, including Placement (MIP vs. less restrictive) as a time-dependent covariate. However, treatment effects on the TRF remained non-significant.

Because the parent-report analyses showed a significant effect but the teacher reports did not, we examined whether social desirability might have influenced parents' ratings of their adolescents' functioning. We first correlated the PAI Positive Impression Management (PIM) scale, which assesses the tendency to present oneself in an excessively favorable light, with parents' reports of child externalizing problems on the CBCL at the four timepoints. The correlation was significant at all four timepoints, ranging from $-.16$ to $-.24$, indicating that higher levels of positive impression management were correlated with lower levels of adolescent behavior problems reported by parents. Second, we re-ran the main HLM analyses for the parent-report CBCL including the PIM scale, first using the PIM baseline scores as a time-invariant covariate, and second using the PIM as a time-variant covariate. In both cases, the intervention effect on parent-report CBCL scores remained significant.

We computed standardized effect size estimates for the primary outcomes involving parent-, teacher-, and youth-reported externalizing problems (see Table 4). We first used the partial correlation coefficient approach for repeated measures structures (Lipsitz, Leong, Ibrahim, & Lipshultz, 2001). This rho coefficient is particularly appropriate for longitudinal data because it is based on the slope rather than comparison of single points, and is less sensitive to pretreatment differences in the outcome variables. The rho coefficient is computed by dividing the parameter estimate for an effect (e.g., Time $\times$ Treatment) by the square root of the variance of the random component of the effect (i.e., in this instance, Time). We also computed d effect size estimates (see Table 4), using the change score approach that others have used (e.g., the MST meta-analysis by Curtis, Ronan, & Borduin, 2004).

We conducted two sets of clinical significance tests on our primary outcome variables (the CBCL scales), following the recommendations of Jacobson, Roberts, Berns, and McGlinchey (1999). First, at each timepoint after baseline we computed separately for the control and treatment groups the percentage of participants who were statistically closer to

the normative population mean versus the clinical population mean; Jacobson et al. (1999) refer to this statistic as Clinically Significant Change (CSC). Second, we computed the Reliable Change Index (RCI; Jacobson et al., 1999), which measures whether change from baseline to subsequent timepoints is reliable; we computed this from baseline to each of the three subsequent timepoints, separately for the control and treatment groups. As Table 4 indicates, the percentage of adolescents in the treatment group considered to be in the normal range of functioning vis-a-vis psychopathology ranged from 34% (parent report Time 2) to 66% (adolescent report Time 3), and in 8 of 9 instances the percentage of adolescents considered to be in the normal range was greater for the treatment group than for the control group. In 9 of 9 instances, the RCI for the treatment group was larger than the RCI for the control group; however, for teacher-reported externalizing problems, none of the RCI exceeded the significance criterion of 1.96.

We used Cox proportional hazards model survival analysis (Cox & Oakes, 1984) to analyze the court data. We tested whether the time to the first felony arrest post-baseline, with participants not arrested within 2.5 years post-baseline considered censored, differed for the treatment and control groups. We included as a covariate whether the participant had been arrested within one year prior to the baseline assessment. The hazard ratio was .90 indicating that across the 2.5 years the treatment group had approximately a 10% less time-adjusted chance of being arrested for a felony than the control group; however, the hazard ratio did not differ significantly from 1 ($p > .25$), indicating that time to being arrested did not differ significantly for the treatment and control groups. In both groups, 45% of participants had a felony arrest within the 2.5 year assessment time frame (i.e., 55% of participants were censored). Secondarily, we also analyzed incidents that involved (a) a misdemeanor and/or felony arrest, and (b) any arrest (status offense, misdemeanor, felony). Results were essentially unchanged (i.e., the effect of treatment did not approach significance, $p > .25$).

Secondary intervention outcomes—In addition to our four primary outcome measures, we also tested the two subscales from the Self-Report Delinquency Scale: *Delinquency* and *Drug Use*. Because these subscales were not interval level data and highly non-normal (e.g., at Time 1, 78% of participants reported no delinquent acts and 74% reported no drug use; kurtosis ranged from 9.41 to 24.04), we analyzed these data as longitudinal ordinal data, using a generalized linear model approach (Lindstrom & Bates, 1990) with maximum likelihood estimation. For the *Delinquency* subscale, there were no significant effects; for the *Drug Use* subscale, the parameter estimate for Time was significant ($t[163] = -2.67$, $p < .01$), with a decreasing slope indicating that the level of reported drug use collapsed across the treatment and control groups decreased. However, no other parameter estimates were significant, indicating that the rates of change for the treatment and control groups did not differ significantly.

To determine whether the treatment and control groups differed in the extent of their placements in general education vs. special education classrooms, we computed the percentage of students who were placed in each of these settings, across the three post-baseline assessment points (all participants started in MIP classrooms at baseline, by virtue of our design). On average across the three assessments, 7% of the treatment group were in general education classrooms as compared to 5% for the control group; these percentages did not differ significantly. With regard to school grades, both groups showed a significant decrease in grades across time ($F[1,148] = 7.27$, $p < .01$), but this rate did not differ for the treatment or control groups. There was, however, a significant quadratic treatment effect for the number of days absent ($F[1,48] = 4.12$, $p < .05$). As shown in Figure 3, the treatment and control groups showed curvilinear effects in opposite directions. The treatment group was approximately .4 standard deviation higher at baseline on absences but approximately .2 standard deviations lower than the control group at the final assessment. There were no

significant effects for the number of days suspended; i.e., the number of days suspended did not change across the timeframe of the study either overall or differentially for the two treatment groups.

Intervention targets—In the next set of analyses, we assessed the effects of treatment on several variables that were intervention targets hypothesized to underlie MST's effects. The structure of these analyses was similar to that for the outcome analyses. For parent-report and adolescent-report FACES Cohesion and Adaptability subscales, no effects involving Time were significant; i.e., there were no significant changes in FACES subscale scores either overall or differentially by group. For parenting behavior, there was a significant curvilinear treatment effect on the PAQ Permissive Behavior subscale, with treatment families reporting a greater rate of decrease than control families (see Table 5 and Figure 3). Finally, there was a significant linear treatment effect for parent internalizing but not externalizing mental health problems as assessed by the PAI (see Table 5), with control parents showing an increase in symptoms whereas treatment parents showed a slight decrease (see Figure 3).

Service usage—The treatment and control groups were first compared in regard to the two variables pertaining to mental health services use from the Service Use Questionnaire (SUQ). Seventy-two percent of the control group and 77% of the treatment group received some form of mental health services from outside the project over the course of the study; these percentages did not differ significantly. Sixty-four percent of the control group and 74% of the treatment group received mental health services provided by a mental health professional from outside the project; these percentages did not differ significantly. We also tested whether these two variables were related to our primary outcomes (i.e., the CBCL variables). The structure of these analyses paralleled that of our main outcome analyses, with the exception that Service Usage replaced Treatment as the categorical independent variable. In all six analyses (2 SUQ variables $\times$ 3 CBCL variables), effects of service usage were non-significant, indicating that service usage external to the project was not significantly related to our primary outcomes.

Satisfaction data—We analyzed treatment families' reports of satisfaction with treatment. On a scale of 1 (*a little satisfied*) to 5 (*a lot satisfied*), parents reported a mean level of satisfaction of 4.62 ($SD=.53$), and adolescents reported a mean level of satisfaction of 4.24 ($SD=.98$). Thus, both parents and adolescents reported being very satisfied with treatment, although adolescents' level of satisfaction was significantly lower than that of parents ($t=3.43$, $p < .005$).

Discussion

The current study is the first fully independent multi-source, multi-informant randomized evaluation of MST in the U.S. with non-court-referred adolescents with serious conduct problems. The study monitored and established MST treatment fidelity using multiple sources of data from parents as well as expert coding of audiotaped sessions and global ratings from the on-site and off-site MST supervisors throughout the implementation period. The fidelity data indicated that therapists' adherence to MST principles was in the moderate to high range, demonstrating acceptable adherence (Letourneau et al., 2002).

Primary intervention outcomes

Two of four primary outcome measures showed significant treatment effects, with MST parents and adolescents reporting on the CBCL significantly greater rates of decrease in externalizing problems than control group parents and adolescents. However, teacher reports

of adolescent externalizing problems and arrest data did not show significant treatment effects.

There are several possible explanations for the inconsistency in findings between parent and adolescent reports, and teacher reports. First, it is likely that parents and adolescents had more emotional investment in the outcome than teachers. Thus, parents and adolescents might have reported the effects of therapy in an overly positive light, with teachers providing a more unbiased perspective. Arguing against this, however, is the fact that when we controlled for positive impression management (PIM) among the parents, effects of treatment as reported by parents remained significant. Moreover, the negative correlations between the PIM scale and parents' reports of their children's externalizing problems on the CBCL suggests that the PIM scale may assess defensiveness that involves one's child as well as oneself. Of course, as is true for all measures of social desirability, interpretation of relations to psychopathology is complex, as this relation could reflect in part that parents who are exceptionally well adjusted (e.g., who are never in a bad mood, and who always keep their promises) have adolescents with fewer behavior problems. However, controlling for the significant negative relation between the PIM and CBCL did not eliminate the significant treatment effect on the CBCL, providing some additional support for the validity of the treatment effect as reported by parents.

Second, it is possible that excessive variability in the teacher data obscured treatment effects. The MST program was not linked to the school year, and assessments extended over more than a year, so in many cases teachers who completed the first assessment were not the teachers who completed the final assessment. This may have created additional error variance that masked treatment effects. In addition, a small number of participants moved to general education classrooms and were assessed by teachers in these classrooms. Such teachers may have had different behavioral standards, thus confounding improved behavior with stricter behavioral standards and increasing variability. However, similar to the parent- and adolescent-report data for which effects were found, the level of variability for the teacher report data did not show large changes across time. In addition, controlling for the type of classroom (MIP vs. general education) in which adolescents were assessed did not change the results from teachers, suggesting that the lack of significant treatment effects for teacher reports was not due to increased variability at follow-up timepoints for teacher report data.

Third, it is possible that teachers reported relatively little change, for either the treatment or control group, because of the persistence of problematic students' reputation among teachers. By the time an adolescent is placed in a special education behavior classroom (in our case, the MIP classroom), his or her behavior has become significantly problematic that these students' reputation likely is well-known among teachers throughout the school. In fact, placement of the student in the behavior intervention classroom essentially labels him or her as having serious behavior problems. Thus, it is possible that teachers did not report change, in either treatment or control students, because persistence of students' reputation overrode perception of any changes in the students' behavior. There is evidence (Smith, Jussim, & Eccles, 1999) that such effects may be larger for African American students, students from low SES backgrounds, and students who are low achievers, which characterizes much of our sample.

Finally, although MST is multi-systemic, the intervention does tend to focus on the home and family more than on the school. Thus, greater changes may have occurred at home and with the family than at school, and teachers may have reported little change because little change occurred at the school. Our therapists did concentrate on the school system as well as the family and home, but given the central importance of school in adolescents' lives, these

null results suggest that it might be useful for MST therapists to develop more targeted interventions for the school setting beyond behavioral plans and contracts to change student behavior. For instance, interventions might involve consultation with teachers to change the way they respond to and manage students' misbehavior, such as ignoring minor misbehavior and focusing on reinforcement of positive behaviors rather than primarily responding to negative behaviors.

MST also did not have a significant treatment effect on arrests. This differs from a number of other randomized trials of MST with adolescents with serious conduct problems (e.g., Borduin et al., 1995; Borduin, Schaeffer, & Heiblum, 2009; Timmons-Mitchell et al., 2006). In these studies, arrest rates were significantly lower for the MST group, whose arrest rates 1 to 9 years post-treatment were 37% to 70% lower than for control group participants. However, in a community-based study of MST implemented without the benefit of weekly expert consultation to monitor treatment fidelity, MST did not produce a significant reduction in arrests at a 1.7-year follow-up (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997).

Secondary intervention outcomes

In contrast to the adolescent-report YSR as well as the parent-report CBCL of externalizing conduct problems, MST had no significant effects on adolescents' reports of delinquent behavior or drug use on the Self-Report Delinquency (SRD) scale, one of our secondary outcome variables. There are several possible reasons why the results may have differed for the SRD. First, almost all SRD items involve very serious behaviors (e.g., involvement in gang fights, selling hard drugs such as heroin), whereas the YSR and CBCL Externalizing Problems scales cover a broader range of behaviors (e.g., from arguing to setting fires). The SRD's focus on highly serious behaviors relative to the CBCL and YSR may have resulted in increased social desirability effects, and decreased treatment effects, for the SRD. However, it also is possible that in this study MST may have been less successful in modifying more serious antisocial behavior, which is reflected more in the SRD than in the CBCL. This possibility is supported by our failure to find a significant treatment effect on arrest record data.

MST had a significant effect on several other secondary measures. First, there was a significant effect on school absences, with the treatment and control groups showing opposite curvilinear patterns. Second, treatment parents reported a greater rate of decrease in permissive parenting than control parents. There was, however, no significant treatment effect on authoritarian or authoritative parenting; the reasons for this discrepancy are not clear. It may be that the behaviors associated with decreases in permissive parenting (potentially, for instance, increases in behavioral limit setting) are easier to learn than the behaviors that would be associated with improvements in authoritarian parenting behavior (e.g., reductions in parental expectations of automatic obedience) or authoritative parenting behavior (e.g., increases in providing clear explanations for rules and discipline); this explanation is speculative. And third, consistent with some other MST outcome studies (e.g., Borduin et al., 1995), MST had a significant effect on parents' mental health problems, with treatment group parents reporting a decrease in their internalizing psychopathology while control group parents reported an increase in symptoms. These findings suggest that the effect of MST on adolescents' conduct problems may in part be mediated by improvement in parents' parenting behaviors and psychopathology, which may enable parents to better monitor their children's behavior and set behavioral limits.

In a review of MST studies, Curtis et al. (2004) found an overall effect size of $d=.55$ for the effects of MST on outcomes, with the d effect size representing point comparisons of the treatment and control groups at a particular timepoint. In our study, the mean d effect size

was .28, averaged across the three primary outcome measures and the three follow-up timepoints. We also computed rho coefficients, which conceptually represent the difference in slopes of change across time for the treatment and control groups (Lipsitz et al., 2001). We believe that the rho effect size is most appropriate for longitudinal data because it reflects trends across time rather than a comparison of specific timepoints, and because it is derived directly from the actual inferential statistical analyses conducted. The mean rho effect size for the linear component (which is most comparable to the d effect size used by Curtis et al., 2004) for the three primary outcome variables was $\rho=.37$.

One possible reason why the effect sizes for the current study were lower than those reported by Curtis et al. (2004), and why MST may have been less successful with more serious antisocial behaviors (as suggested by our failure to find treatment effects for the SRD and arrests), may be because our adolescents were recruited from self-contained behavior intervention classrooms rather than through the court system as has been typically the case in MST evaluations involving delinquent youth. A potential consequence of not having obtained our participants through the court system is that without pressure from the court, families may have been less motivated to energetically participate in therapy. However, the treatment fidelity data in our study do not support the interpretation that families were relatively unmotivated in therapy, as parents' reported adherence and engagement in the intervention was overall in the acceptable range, and comparable to that found in previous studies. But because the fidelity measures were not specifically designed to assess relatively subtle motivation-related behaviors among family members, it is possible that motivation deficits due to not being mandated by court to participate in treatment reduced the efficacy of MST in this study.

Two other independent randomized studies of the effects of MST on adolescents with severe conduct problems that did not recruit participants through the court system were conducted in Scandinavia (Ogden & Halliday-Boykins, 2004; Sundell et al., 2008). In Norway and Sweden, where these studies were conducted, there is no separate juvenile justice system and crimes by juveniles are handled by the child welfare system. Both studies showed relatively modest to no effects, which supports the interpretation that the efficacy of MST was reduced in our study as a consequence of our adolescents being non-court-referred. However, another possible explanation for the results of the Scandinavian studies is that the comprehensive nature of the "usual services" provided by Norway and Sweden's child welfare system provided a higher comparison for MST. In contrast, Butler et al.'s (2011) study, conducted in the U.K. with a comparison group that also received comprehensive services-as-usual (e.g., substance abuse treatment, anger management training, social skills problem-solving), found somewhat stronger effects (e.g., a significant reduction in crimes committed). The findings from Butler et al.'s (2011) study, which did involve court-referred adolescents, suggests that court referral rather than the comparison group receiving comprehensive services may be the key factor influencing our results as well as those of the Scandinavian studies, relative to other MST studies.

Despite the fact that our sample was not court-referred, our sample did have significant conduct problems at baseline, with 96% of the sample scoring at or above the borderline clinical range for Externalizing Problems for one or more informants on the Child Behavior Checklist. Moreover, although these adolescents were selected through the public school system, the majority of youth (68%) had already been involved with the juvenile court system at baseline.

One final explanation for why our results were less strong than those reviewed by Curtis et al. (2004) has to do with the fact that the adolescents in this study spent a significant amount of time in MIP classrooms with other delinquent adolescents. It has been suggested that

when treatments such as that provided by the behaviorally-oriented MIP classrooms aggregate adolescents with conduct problems, the treatments may inadvertently have iatrogenic effects via delinquency training (Dishion, McCord, & Poulin, 1999). Thus, it is possible that the MIP classrooms provided a negative counterbalance to the positive effects of MST, reducing its effects. Actual evidence for such iatrogenic effects is mixed (e.g., Huefner, Handwerk, Ringle, & Field, 2009; Weiss et al., 2005), however, which suggests that such effects may be complex.

Conclusion

Overall, the results of the present study provide some additional support for the efficacy of MST when evaluated independently of its developers. Importantly, they suggest that with sufficient resources to support ongoing MST expert consultation, a moderately high level of treatment fidelity can be established. Despite the diversity of the clinicians' backgrounds, the treatment fidelity data indicated that the clinicians' adherence to MST principles was moderately high to high in this study, similar to that reported in other trials of MST. However, the smaller effect sizes and lack of effect on the arrest data also suggest that fully successful dissemination of MST (i.e., maintaining original effect sizes) is challenging, particularly with adolescents who are not recruited through the juvenile justice system.

Acknowledgments

The research described in this article was supported in part by U.S. NIH grants R01-MH58275 and T32-MH070329. We would like to thank the participating families, and the clinicians who provided the treatment in this study for their support.

References

- Achenbach, T.M.; Rescorla, L.A. Manual for the ASEBA School-Age forms & profiles. University of Vermont Research Center for Children, Youth & Families; Burlington, VT: 2001.
- Bronfenbrenner, U. The ecology of human development: Experiments by nature and design. Harvard University Press; Cambridge, MA: 1979.
- Borduin CM, Mann BJ, Cone LT, Henggeler SW, Fucci BR, Blaske DM, Williams R. Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology*. 1995; 63:569–578. doi: 10.1037/0022-006X.63.4.569. [PubMed: 7673534]
- Borduin CM, Schaeffer CM, Heiblum N. A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting & Clinical Psychology*. 2009; 77:26–37. doi: 10.1037/a0013035. [PubMed: 19170451]
- Buri JR. Parental Authority Questionnaire. *Journal of Personality and Social Assessment*. 1991; 57:110–119. doi: 10.1207/s15327752jpa5701_13.
- Butler S, Baruch G, Hickey N, Fonagy P. A randomized controlled trial of Multisystemic Therapy and a statutory therapeutic intervention for young offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2011; 50:1220–1235. doi:10.1016/j.jaac.2011.09.017. [PubMed: 22115143]
- Catron T, Weiss B. The Vanderbilt School-Based Counseling program. *Journal of Emotional and Behavioral Disorders*. 1994; 2:247–253. doi: 10.1177/106342669400200407.
- Cox, D.R.; Oakes, D. Analysis of Survival Data. Chapman & Hall; London: 1984.
- Curtis NM, Ronan KR, Borduin CM. Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*. 2004; 18:411–419. doi: 10.1037/0893-3200.18.3.411. [PubMed: 15382965]
- Dishion T, J. McCord J, Poulin F. When interventions harm: Peer groups and problem behavior. *American Psychologist*. 1999; 54:755–764. [PubMed: 10510665]

- Edman SO, Cole DA, Howard GS. Convergent and discriminant validity of Faces-III: Family adaptability and cohesion. *Family Process*. 1990; 29:95–103. doi: 10.1111/j.1545-5300.1990.00095.x. [PubMed: 2311756]
- Elliott, DS.; Huizinga, D.; Ageton, SS. Explaining delinquency and drug use. Sage Publications; Beverly Hills, CA: 1985.
- Henggeler SW. Efficacy studies to large-scale transport: The development and validation of multisystemic therapy programs. *Annual Review of Clinical Psychology*. 2011; 7:351–381. doi: 10.1146/annurev-clinpsy-032210-104615.
- Henggeler, SW.; Borduin, CM. Multisystemic Therapy Adherence Scales. Unpublished instrument, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina; 1992.
- Henggeler, SW.; Cunningham, PB.; Schoenwald, SK.; Borduin, CM. Multisystemic therapy for antisocial behavior in children and adolescents. 2nd Ed.. Guilford Press; NYC: 2009.
- Henggeler SW, Melton GB, Brondino MJ, Scherer DG, Hanley JH. 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*. 1997; 65:821–833. doi: 10.1037/0022-006X.65.5.821. [PubMed: 9337501]
- Henggeler SW, Rodick JD, Borduin CM, Hanson CL, Watson SM, Urey JR. Multisystemic treatment of juvenile offenders: Effects on adolescent behavior and family interaction. *Developmental Psychology*. 1986; 22:132–141. doi: 10.1037/0012-1649.22.1.132.
- Hodges K, Wong MM. Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale. *Journal of Child and Family Studies*. 1996; 5:445–467. doi: 10.1007/BF02233865.
- Houts AC, Berman JS, Abramson H. Effectiveness of psychological and pharmacological treatments for nocturnal enuresis. *Journal of Consulting and Clinical Psychology*. 1994; 62:737–745. doi: 10.1037/0022-006X.62.4.737. [PubMed: 7962877]
- Huefner JC, Handwerk ML, Ringle JL, Field CE. Conduct disordered youth in group care: An examination of negative peer influence. *Journal of Child & Family Studies*. 2009; 18:719–730. doi: 10.1007/s10826-009-9278-6.
- Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*. 1999; 67:300–307. doi: 10.1037/0022-006X.67.3.300. [PubMed: 10369050]
- Letourneau EJ, Henggeler SW, Borduin CM, Schewe PA, McCart MR, Chapman JE, Saldana L. Multisystemic therapy for juvenile sexual offenders: 1-year results from a randomized effectiveness trial. *Journal of Family Psychology*. 2009; 23:89–102. doi: 10.1037/a0014352. [PubMed: 19203163]
- Letourneau, EJ.; Sheidow, AJ.; Schoenwald, SK. Structure and reliability of the MST Therapist Adherence Measure scale in a large community sample. Medical University of South Carolina, Family Services Research Center; Charleston, SC: 2002.
- Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46:673–687. doi:10.2307/2532087. [PubMed: 2242409]
- Lipsitz SR, Leong T, Ibrahim JG, Lipshultz S. A partial correlation coefficient and coefficient of determination for multivariate normal repeated measures data. *Journal of the Royal Statistical Society*. 2001; 50:87–95. doi: 10.1111/1467-9884.00263.
- Littell, J.; Poppa, M.; Forsythe, B. Multisystemic Therapy for social, emotional, and behavioral problems in youth aged 10–17. 2005. Campbell Systematic Reviews 2005.1. Retrieved from Campbell Collaboration website: www.campbellcollaboration.org/lib/download/23/1004_R.pdf
- Morey, LC. Personality Assessment Inventory - Professional Manual. Psychological Assessment Resources, Inc; Florida, USA: 1991.
- Ogden T, Halliday-Boykins CA. Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health*. 2004; 9:77–83. doi: 10.1111/j.1475-3588.2004.00085.x.
- Olson, DH.; Portner, J.; Lavee, Y. FACES-III. University of Minnesota; St Paul, MN: 1985.

- Quay, HC.; Peterson, DR. Interim manual for the Revised Behavior Problem Checklist. University of Miami; Coral Gables, FL: 1983.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. Sage Publications; Thousand Oaks, CA: 2002.
- Robinson LA, Berman JS, Neimeyer RA. Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin*. 1990; 108:30–49. doi: 10.1037/0033-2909.108.1.30. [PubMed: 2200072]
- Rowland MD, Halliday-Boykins CA, Henggeler SW, Cunningham PB, Lee TG, Kruesi MJP, Shapiro SB. A randomized trial of multisystemic therapy with Hawaii's Felix Class youths. *Journal of Emotional and Behavioral Disorders*. 2005; 13:13–23. doi: 10.1177/10634266050130010201.
- Ruiz MA, Edens JR. Recovery and replication of internalizing and externalizing dimensions within the Personality Assessment Inventory. *Journal of Personality Assessment*. 2008; 90:585–592. doi: 10.1080/00223890802388574. [PubMed: 18925500]
- Sheidow AJ, Schoenwald SK, Wagner HR, Allred CA, Burns BJ. Predictors of workforce turnover in a transported treatment program. *Administration and Policy in Mental Health and Mental Health Services Research*. 2007; 34:45–56. doi: 10.1007/s10488-006-0061-3. [PubMed: 16767507]
- Smith AE, Jussim L, Eccles J. Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *Journal of Personality and Social Psychology*. 1999; 77:548–565. doi: 10.1037/0022-3514.77.3.548. [PubMed: 10510508]
- Sundell K, Hansson K, Lofholm CA, Olsson T, Gustle LH, Kadesjo C. The transportability of multisystemic therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology*. 2008; 22:550–560. doi: 10.1037/a0012790. [PubMed: 18729669]
- Timmons-Mitchell J, Kishna M, Bender M, Mitchell C. An independent effectiveness trial of multisystemic therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology*. 2006; 35:227–236. doi: 10.1207/s15374424jccp3502_6. [PubMed: 16597218]
- U.S. Department of Health and Human Services. Youth violence: A report of the Surgeon General. Author; Rockville, MD: 2001.
- Weiss B, Caron A, Ball S, Tapp J, Johnson M, Weisz JR. Iatrogenic effects of group treatment for antisocial youths. *Journal of Consulting and Clinical Psychology*. 2005; 73:1036–1044. Doi: 10.1037/0022-006X.73.6.1036. [PubMed: 16392977]

Figure 1.
Participant flow for the study

Figure 2.
Treatment effects on parent-report CBCL and adolescent-report YSR Externalizing Problems

School Absences

Parent PAQ – Permissive

Parent PAI Internalizing problems

Figure 3.
Significant treatment effects on secondary variables

Table 1

Demographic characteristics

	Control Group (<i>n</i> = 80)	Treatment Group (<i>n</i> = 84)
Adolescent		
Mean age (SD)	14.5 (1.4)	14.6 (1.3)
% male	83%	83%
% African-American	64%	56%
% Euro-American	36%	44%
Primary Caregiver		
Mean age (SD)	40.0 (8.0)	41.5 (9.5)
% biological mother	79%	75%
% biological father	9%	4%
% single parent	75%	67%
% high school graduate *	63%	79%
Family		
Median annual income	\$17,500	\$17,500
Mean # adults in household (SD)	1.8 (0.9)	1.8 (0.8)
Mean # children in household (SD)	2.4 (1.4)	2.4 (1.4)

Note.

*
p<.05, for Treatment vs. Control group.

Table 2

Means and standard deviations for intervention outcome and target variables

Primary Outcome Variables				
	Time 1	Time 2	Time 3	Time 4
CBCL Externalizing				
Control Group	23.40 (9.61)	19.54 (10.65)	19.57 (11.50)	19.19 (10.36)
Treatment Group	25.90 (10.63)	20.64 (11.31)	17.68 (10.57)	18.20 (10.82)
YSR Externalizing				
Control Group	17.00 (7.97)	16.38 (8.06)	14.56 (8.04)	14.22 (7.72)
Treatment Group	17.63 (9.03)	15.01 (9.33)	12.39 (7.59)	13.87 (8.53)
TRF Externalizing				
Control Group	22.50 (11.67)	23.21 (13.02)	22.07 (12.99)	20.06 (12.57)
Treatment Group	22.94 (12.50)	19.31 (12.45)	17.66 (12.92)	19.50 (12.39)
Secondary Outcome Variables				
	Time 1	Time 2	Time 3	Time 4
SRD – Delinquency				
Control Group	0.29 (.53)	0.25 (.50)	0.16 (.45)	0.15 (.40)
Treatment Group	0.22 (.50)	0.25 (.55)	0.18 (.42)	0.13 (.42)
SRD – Drug Use				
Control Group	0.43 (.73)	0.26 (.64)	0.21 (.59)	0.15 (.44)
Treatment Group	0.33 (.65)	0.36 (.69)	0.17 (.47)	0.31 (.63)
Core grades ^I				
Control Group	77.20 (15.00)	76.71 (16.47)	72.15 (24.08)	75.73 (18.12)
Treatment Group	76.70 (16.82)	75.16 (19.25)	77.47 (18.10)	72.62 (20.90)
Days absent ^I				
Control Group	4.41 (3.89)	3.72 (3.97)	5.29 (4.83)	5.11 (6.74)
Treatment Group	6.12 (5.85)	3.73 (4.07)	3.73 (5.09)	5.68 (5.43)
Days suspended ^I				
Control Group	1.59 (2.81)	1.21 (2.10)	2.13 (4.05)	1.90 (5.82)
Treatment Group	2.68 (5.01)	0.73 (2.14)	1.13 (2.24)	1.41 (3.28)
Intervention Target Variables				
	Time 1	Time 2	Time 3	Time 4
Parent FACES – Cohesion				
Control Group	57.72 (10.58)	58.00 (9.63)	58.24 (9.67)	57.99 (9.65)
Treatment Group	59.86 (9.82)	58.68 (9.94)	60.11 (8.42)	57.91 (9.46)
Parent FACES – Adaptability				
Control Group	46.89 (6.77)	46.76 (7.07)	46.82 (6.61)	46.62 (6.94)
Treatment Group	47.30 (6.30)	45.84 (5.11)	46.37 (6.28)	46.21 (6.49)
Youth FACES – Cohesion				

Control Group	53.14 (10.46)	53.41 (11.82)	53.11 (8.78)	52.59 (10.06)
Treatment Group	53.80 (11.17)	52.79 (10.19)	55.00 (10.64)	53.38 (12.01)
Youth FACES – Adaptability				
Control Group	44.97 (8.18)	44.67 (8.53)	44.31 (7.96)	43.92 (8.67)
Treatment Group	44.96 (8.62)	42.84 (9.04)	43.53 (9.61)	43.68 (9.07)
Parent PAQ – Authoritarian				
Control Group	34.40 (6.52)	33.86 (5.42)	34.74 (4.88)	33.81 (5.12)
Treatment Group	35.60 (5.66)	34.93 (5.48)	34.51 (5.54)	34.30 (4.53)
Parent PAQ – Authoritative				
Control Group	41.58 (3.40)	40.50 (3.65)	40.65 (3.37)	40.01 (3.45)
Treatment Group	41.28 (4.85)	40.73 (4.06)	39.54 (4.50)	39.83 (3.43)
Parent PAQ – Permissive				
Control Group	20.48 (4.88)	21.13 (5.05)	21.60 (5.68)	20.65 (4.86)
Treatment Group	20.77 (5.46)	19.79 (5.68)	19.34 (4.80)	20.22 (6.36)
Parent PAI Externalizing				
Control Group	79.59 (31.06)	77.81 (35.29)	80.94 (37.72)	80.56 (35.10)
Treatment Group	89.15 (40.14)	75.84 (40.70)	83.13 (41.22)	79.58 (37.11)
Parent PAI Internalizing				
Control Group	44.92 (25.06)	42.65 (26.00)	44.64 (24.76)	49.34 (28.49)
Treatment Group	47.60 (25.56)	44.49 (26.55)	47.41 (24.96)	44.40 (25.34)

Note.

¹ Data for school variables (grades, days absent, days suspended) are for the most recent six-week grading period in relation to when data were collected for the other informants.

Table 3

Results for primary intervention outcomes

Dependent variable	Factor	DF	F
Externalizing Problems on Parent-report CBCL	Time	1, 153	93.89****
	Time $\times$ Treatment	1, 130	5.19*
	Time $\times$ Time	1, 149	47.14****
	Time $\times$ Time $\times$ Treatment	1, 130	2.25
Externalizing Problems on Adolescent-report YSR	Time	1, 148	36.27****
	Time $\times$ Treatment	1, 111	5.55*
	Time $\times$ Time	1, 140	23.45****
	Time $\times$ Time $\times$ Treatment	1, 111	4.64*
Externalizing Problems on Teacher-report TRF	Time	1, 150	1.22
	Time $\times$ Treatment	1, 200	0.24
	Time $\times$ Time	1, 125	2.09
	Time $\times$ Time $\times$ Treatment	1, 74	2.90

Note.

All models included a main effect for Treatment, but it is not reported in this table because it is unrelated to the hypotheses.

*
p<.05,****
p<.0001.

Table 4

Effect sizes, confidence intervals, and clinical significance analyses for primary outcome variables

		Parent Report CBCL	Adolescent Report YSR	Teacher Report TRF
rho	linear	.62	.62	-.13
	quadratic	.57	1.10	.47
d	Time 2	.25 (.08~.41)	.30 (.13~.47)	.28 (.09~.47)
	Time 3	.43 (.26~.59)	.34 (.17~.51)	.27 (.08~.46)
	Time 4	.35 (.19~.52)	.26 (.09~.44)	.03 (-.16~.23)

		Control	Treatment	Control	Treatment	Control	Treatment
CSC	Time 2	35%	34%	45%	56%	29%	38%
	Time 3	42%	47%	51%	66%	33%	49%
	Time 4	41%	43%	52%	63%	36%	41%
RCI	Time 2	4.11	5.83	0.46	3.36	-1.19	1.21
	Time 3	5.26	7.95	3.01	5.15	-0.47	1.31
	Time 4	3.54	6.68	1.92	3.73	0.08	0.34

Note. CSC = Clinically Significant Change, the proportion of participants statistically closer to the mean of the normative group rather than the dysfunctional group. RCI = Reliable Change Index, an absolute value of greater than 1.96 indicates reliable change.

Table 5

Results for significant secondary outcome and intervention target analyses

Dependent variable	Factor	DF	F
Days Absent	Time	1,86	0.06
	Time $\times$ Treatment	1,48	2.87
	Time $\times$ Time	1,74	1.50
	Time $\times$ Time $\times$ Treatment	1,48	4.12*
Parent PAQ – Permissive	Time	1, 153	0.54
	Time $\times$ Treatment	1, 126	5.56*
	Time $\times$ Time	1, 147	0.02
	Time $\times$ Time $\times$ Treatment	1, 126	10.28**
Parent PAI Internalizing problems	Time	1, 153	0.46
	Time $\times$ Treatment	1, 280	5.66*
	Time $\times$ Time	1, 149	1.09
	Time $\times$ Time $\times$ Treatment	1, 130	1.54

Note.

The model included an intercept and main effect for Treatment, but these are not reported in this table because these factors are unrelated to the hypotheses.

*
p<.05,

**
p<.01,

p<.001.