

Dissemination and Effectiveness of Multisystemic Treatment in New Zealand: A Benchmarking Study

Nicola M. Curtis
Massey University

Kevin R. Ronan
CQUniversity Australia

Naamith Heiblum
MST Services

Kylie Crellin
CQUniversity Australia

The transportability of Multisystemic Therapy (MST) for the treatment of juvenile offenders in a community-based context was examined in the current study. Results of this New Zealand study showed that significant pre- to posttreatment improvements occurred on most indicators of ultimate (i.e., offending behavior) and instrumental (i.e., youth compliance, family relations) treatment outcomes. Reductions in offending frequency and severity continued to improve across the 6- and 12-month follow-up intervals. In comparison to benchmarked studies, the current study demonstrated a more successful treatment completion rate. Additionally, overall treatment effect sizes were found to be clinically equivalent with the results of previous MST outcome studies with juvenile offenders and significantly greater than the effect sizes found in the control conditions. The findings of this evaluation add to the growing body of evidence that supports MST as an effective treatment for antisocial youth.

Keywords: MST, juvenile offenders, conduct disorder, benchmarking, effectiveness

Antisocial behavior in youth represents a complex and pervasive clinical problem. The prevalence of conduct disorder has been found to be as high as 11% in 11 to 15 year olds (Fergusson, Horwood, & Lynskey, 1993; McGee, Feehan, Williams, & Anderson, 1992; see also Fergusson, Horwood, & Ridder, 2005). Antisocial behavior in youth has been found to have significant consequences for individuals, peers, families, and communities. For example, a prospective study conducted in New Zealand found significant relationships between conduct problems in middle childhood (7–9 years) and unfavorable outcomes in young adulthood (21–25 years) (Fergusson et al., 2005). These outcomes included increased criminal behavior, substance abuse, depression, anxiety disorders, antisocial personality disorders, suicide attempts, teenage pregnancy, domestic violence, unemployment, and welfare dependence, as well as decreased educational achievement. Given the prevalence of antisocial behaviors and the damaging consequences

of these behaviors for youth, families, and the broader community, the need for effective and readily available treatments is apparent.

Reviews of empirically supported child and adolescent treatments have identified Multisystemic Therapy (MST) as a treatment program of choice for antisocial behavior in youth (Burns, Hoagwood, & Mrazek, 1999; Kazdin & Weisz, 1998). MST is an intensive, strengths-based, time-limited, family- and community-based treatment approach that targets the individual, family, peer, school, and community elements that contribute to and maintain problematic externalizing behavior in youth (Henggeler, 1989; Henggeler & Borduin, 1990; Ronan & Curtis, 2008). In particular, MST is focused on empowering parents and other important members of the youth's ecology to develop the necessary skills and competencies to help the youth reduce problematic behavior and function more effectively.

A recently conducted meta-analysis examined the effects of MST for youth with antisocial behavior, including juvenile offenders (Curtis, Ronan, & Borduin, 2004). The results, across seven randomized controlled trials (RCTs) conducted in the United States, reflected an overall effect size (*ES*; Cohen's *d*) of .55. Specifically, MST-treated youth experienced a decrease in the number of arrests (and seriousness of arrests), symptomatology, deviant peer relations, and drug use. Youth also experienced increases in positive family relations, supportive peer relations, school attendance, and parental monitoring (Curtis et al.). Further, MST was found to have a high "successful completion" rate (Hunsley & Lee, 2007). Compared to an average early dropout rate of approximately 50% in child and family

Nicola Curtis, School of Psychology, Massey University, Palmerston North, New Zealand; Kevin Ronan and Kylie Crellin, Healthy Communities and Department of Behavioural and Social Sciences, CQUniversity Australia, Rockhampton, Australia; Naamith Heiblum, MST Services, Rehovot, Israel.

The authors would like to acknowledge and thank Takuya Minami for feedback on our benchmarking strategy. The authors would also like to acknowledge and thank all of the families, therapists, and supervisors that took part in this study.

Correspondence concerning this article should be addressed to Kevin R. Ronan, Healthy Communities, Centre for Social Science Research, CQUniversity Australia, Bruce Highway, Rockhampton QLD 4702, Australia. E-mail: k.ronan@cqu.edu.au

treatments (Nock & Ferriter, 2005), MST has an average completion rate of 86% (with a range of 76–100% across seven studies; Curtis et al.).

All studies included in the meta-analysis were RCTs. A key element of efficacy research is the RCT. However, given the nature of MST (i.e., conducted in clients' homes), certain elements are unable to be controlled. Accordingly, the design of research used to examine the outcomes of MST are generally "hybrids of efficacy and effectiveness research" (Schoenwald, Sheidow, Letourneau, & Liao, 2003, p. 224). Despite this, some studies in the meta-analysis were more characteristic of efficacy research and others were more characteristic of effectiveness research. One major distinction between these studies was that therapy was carried out by highly trained graduate student therapists closely supervised by one of the MST developers in one set of studies (i.e., efficacy studies). The other set of studies employed community based therapists and lacked some of the supervisory features evident in the efficacy research, most notably the contextual and close supervision features (Curtis et al., 2004). This distinction appeared to be important as the efficacy studies were found to yield significantly greater treatment effects ($ES = .81$) when compared to the effectiveness studies ($ES = .27$; Curtis et al.). Thus, it is quite possible that efficacy-based successes seen for MST may be different when disseminated in a community context (e.g., Henggeler, 2004; Norcross, Beutler, & Levant, 2006).

The current study reflects conditions characteristic of the effectiveness studies, with therapists working in community mental health settings receiving supervision from community-based supervisors. Multiple methods have been developed to evaluate outcomes based on effectiveness research methodologies (Hunsley & Lee, 2007; Minami et al., 2008). For example, one method involves a direct comparison with a treatment readily available within the community (i.e., treatment-as-usual; TAU). This methodology has previously been employed by studies examining MST (e.g., Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Henggeler, Melton, & Smith, 1992). In the absence of an available control group, an alternative method to direct comparisons that is being used with greater frequency is benchmarking (e.g., Hunsley & Lee, 2007; Minami et al., 2008).

Benchmarking allows researchers to compare the results of treatment conducted in natural settings to best practice standards (Hunsley & Lee, 2007; Minami et al., 2008). The implementation of benchmarking involves four stages: (a) identification of the population and treatment; (b) selection of a 'best practice' benchmark; (c) measurement of outcomes comparable to the benchmark; and (d) comparison of outcomes to the benchmark (Weersing, 2005). Hunsley and Lee (2007) have recently recommended a benchmarking methodology to those who are transporting efficacious treatments to community settings. More specifically, Hunsley and Lee have identified several populations and accompanying treatments and have identified best practice benchmarks. The two benchmarks recommended by Hunsley and Lee are treatment completion rates as well as effect size

estimates of outcomes relevant to a particular area. For youth with antisocial behavior, Hunsley and Lee identified the MST meta-analysis conducted by Curtis et al. (2004) as the benchmark study with the relevant best practice outcomes being treatment completion and effect sizes reflecting (a) ultimate, (b) instrumental, and (c) overall outcomes. Another benchmarking methodology has been provided by Minami and colleagues (Minami et al., 2008; Minami, Serlin, Wampold, Kircher, & Brown, 2006). This methodology extends Hunsley and Lee's framework (Hunsley & Lee, 2007) by providing a specific means for both calculating and comparing *ESs* derived for an effectiveness study with benchmark *ESs*. Taking into account standard effect sizes ranges (i.e., Cohen, 1977), this method allows one to determine statistically whether results of a current study are clinically equivalent to those of best practice benchmarks (see Minami et al., 2008).

Accordingly, this study was developed with the following aims: (a) to evaluate the effectiveness of MST in assisting families to engage and finish treatment; (b) to evaluate the effectiveness of MST in terms of ultimate outcomes including reduced youth offending and recidivism, days in formal out-of-home placements, increased school and/or employment attendance; (c) to evaluate the effectiveness of MST in terms of instrumental outcomes including improved youth psychosocial functioning and family relations; and (d) to benchmark these and aggregate findings against RCT findings using a combination of recent benchmarking methodologies reflecting both successful treatment completion rates and magnitude of treatment effects comparisons (Hunsley & Lee, 2007; Minami et al., 2006, 2008).

Method

Overview of the Design

The current study used a one-group pretest/posttest design and included follow-up intervals to evaluate the preliminary effectiveness of MST with youth and their families in New Zealand. To accommodate for the limitations of this design (Cook & Campbell, 1979), a benchmarking procedure was employed (described in detail below). The research was conducted in accordance with the ethical standards for the treatment of human participants as outlined by the New Zealand Psychological Society and was carried out following ethical approval from the Massey University human ethics committee as well as the Ministry of Health's national human ethics committee.

Participants

Inclusion and exclusion criteria. Inclusion criteria were: (a) youths ranging in age from 8 to 18 years, (b) the manifestation of significant antisocial and externalizing behavior problems, including having engaged in documented criminal behavior or been documented as at risk of offending, (c) risk for out-of-home placement, and (d) having a parent or caregiver who was sufficiently motivated to enrol in the program. Exclusion criteria were (a) youth for whom

a primary caregiver could not be identified, (b) youth in need of crisis stabilization because of active suicidal, homicidal, or psychotic behavior, (c) youth with pervasive developmental disorder (e.g., autism), and (d) youth with an $IQ < 70$.

Participant characteristics. Sixty-five of the 68 youth and their families who were referred to an MST program between January 2002 and December 2003 gave their consent to participate in this study. Youth ranged in age from 8.6 to 17.0 years ($M = 13.83$, $SD = 1.88$), and 71% ($n = 46$) were male. Ethnic composition of the sample was as follows: New Zealand Pakeha (White/European descent) (83%, $n = 54$); Maori (9%, $n = 6$); Samoan (3%, $n = 2$); Tongan (1.5%, $n = 1$); Ethiopian (1.5%, $n = 1$); and Other (1.5%, $n = 1$). The Deprivation Index (Crampton, Salmond, & Kirkpatrick, 2004) was used to estimate the average socio-economic status of participants. Sixty-nine percent of participants ($n = 45$) lived in the most deprived areas of New Zealand where household

incomes averaged \$17,700. Youth referred also had an average offending frequency of more than two documented offending incidents in the 6 months before treatment (see Table 1).

Research Procedures

Referrals and referral reasons. Youth were referred to the program for a range of behavioral and mental health problems. Youth were referred to the program by a number of agencies including the Child, Youth, and Family service (CYF; 37%, $n = 24$), Police Youth Aid, schools, medical practitioners (35%, $n = 23$), or a Child and Adolescent Mental Health Service (23%, $n = 15$). Details of referral agencies were not available for three of the referred youth (5%). Primary referral reasons included verbal/physical aggression at home, school, or in the community (60%, $n = 39$), truancy (14%, $n = 9$), substance abuse (8%, $n = 5$), non-compliance and family conflict (6%, $n = 4$), suicide/

Table 1
Demographic Characteristics of Youth and Their Families

Demographic characteristics	<i>n</i>	Total sample (%)	<i>M</i>	<i>SD</i>
Youth gender				
Female	19	29		
Male	46	71		
Age			13.85	1.99
Ethnicity				
Pakeha ^a	54	83		
Maori	6	9		
Samoan	2	3		
Tongan	1	1.50		
Ethiopian	1	1.50		
Other	1	1.50		
Primary referral reason				
Verbal/physical aggression	39	60		
Truancy	9	14		
Substance Abuse	5	8		
Noncompliance	4	6		
Suicide/homicide threats	3	4		
Running away	2	3		
Burglary/stealing	2	3		
Sexual assault	1	2		
History of involvement with agencies	3–5		4.09	0.69
Frequency of offending in six months before treatment			2.12	3.12
Family size	65		4.70	1.50
Family composition				
Single-parent	32	49		
Biological mother plus another adult	18	28		
Two biological parents	11	17		
Foster parents	4	6		
Employment status of primary caregiver				
Full-time	17	26		
Not working	33	51		
Part-time	15	23		
Custody status				
Parent	44	68		
Child, youth, and family service ^b	21	32		

^a New Zealand term for those of European descent. ^b Participation criteria for youth included having a primary caregiver. Although the Child Youth and Family Service held legal custody for 32% of the youth, each youth lived with at least one caregiver.

homicide threats (5%, $n = 3$), running away from home (3%, $n = 2$), burglary/stealing (3%, $n = 2$), and sexual assault/sexually inappropriate behavior (1%, $n = 1$). Referred youth had a number of primary diagnoses, including Conduct Disorder (CD; 36%, $n = 23$), ADHD (23%, $n = 15$), Oppositional Defiant Disorder (ODD; 17%, $n = 11$), mood disorders (9%, $n = 6$), anxiety disorders (3%, $n = 2$), learning disorders (3%, $n = 2$), and substance abuse disorders (3%, $n = 2$). Four (6%) youth did not meet diagnostic criteria for any disorder. Multiple problems (i.e., co-morbid conditions) were noted in 51% ($n = 33$) of the youth clinical histories. Twenty-nine percent ($n = 19$) previously had received care from seven or more mental health, social, educational, or judicial services. A further 51% ($n = 33$) previously had received care from between five and six agencies (see also Table 1). Seventy-two percent ($n = 47$) of youth had been experiencing difficulties for more than 5 years; 25% ($n = 16$) for between 3 and 4 years; and 3% ($n = 2$) for less than 3 years.

MST treatment. MST was implemented following established training and supervision procedures as detailed in the treatment manual (see Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998) and was based on the nine MST treatment principles. Treatment practices followed the social-ecological model whereby problem behaviors are considered to (a) develop in response to complex and varied contributing factors linked with the multiple systems in which youth and families are located and (b) require a sustainable and ecologically valid treatment delivery mechanism. Accordingly, MST therapists worked in the youths' homes at times that were convenient for families. Meetings also were held in community settings including schools, social service agencies, or other settings (e.g., marae).¹ Services were delivered to the family as a whole (rather than solely to the identified youth), and tailored to the individual needs and goals of family members, particularly parents. Interventions were planned in the context of a family's values, beliefs, and culture. Additionally, in line with standard MST practice, fidelity was monitored on a regular basis (i.e., monthly) by asking parents to fill out the Therapist Adherence Measure (Henggeler & Borduin, 1992), a 26-item measure developed by expert consensus to reflect the treatment principles and assess family and therapist behaviors specific to the implementation of MST. Overall treatment adherence in this study was found to be high (see Curtis, 2004).

MST teams and clinicians. There were three MST teams in total, located in Christchurch (Site 1), Hamilton (Site 2), and Wellington (Site 3). MST teams were funded by the Ministry of Health and established to provide treatment to youth with high and complex needs in publicly funded community mental health settings. Over the course of the study, 14 therapists and 5 supervisors were trained in MST. Two withdrew their consent to participate in this study following their resignation as therapists. Of the remaining therapists and supervisors, 53% ($n = 9$) had a Bachelor's or BA (Hons) degree; 23% ($n = 4$) had a Masters degree, 12% ($n = 2$) had some other graduate qualification, and 12% ($n = 2$) held a recognized profes-

sional qualification in another field (i.e., registered nurse, registered social worker). MST therapists/supervisors had between 1 and 28 years ($M = 12$) of clinical experience in social work (59%, $n = 10$), psychology (17%, $n = 3$), teaching (6%, $n = 1$), counseling (6%, $n = 1$), occupational therapy, (6%, $n = 1$), and family therapy (6%, $n = 1$). Seventy-one percent ($n = 12$) of MST therapists/supervisors were female, 94% ($n = 16$) described themselves as European (i.e., "Pakeha"), and 6% ($n = 1$) as Samoan/Maori. During the project, 6 of the 14 therapists and 2 of the 5 supervisors who trained in and delivered MST treatment resigned from their respective teams (42% attrition). Thus, although 14 therapist started the study, 8 finished; similarly, 5 supervisors started, 3 finished. Although this attrition needed to be managed to ensure continuity of services, it did not appear to impact on an overall high level of fidelity (Curtis, 2004).

Data Collection

In this study, ultimate outcome data (i.e., details of frequency and severity of offending behavior, days in out of home placements, days absent from school) were collected systematically from agencies (i.e., schools, judicial, and social welfare agencies) by an independent evaluation coordinator at pre- and posttreatment, and again at 6- and 12-month follow-up. A range of instrumental measures were also administered to the main caregiver in each family by the evaluation coordinator to measure change in youth behavior, parent, and family functioning at pre- and post-treatment, at various intervals throughout treatment, and at follow-up periods. Based on an initial agreement with MST treatment teams, family engagement was prioritized over obtaining instrumental outcome data. Thus, to reduce obstacles to engagement (e.g., Kazdin, Holland, & Crowley, 1997), parents were reassured separately by both therapists and the independent evaluator that completion of the assessment measures was entirely voluntary and separate from treatment related efforts and successful completion of the MST program.

Measures of Treatment Outcome

Ultimate Outcomes

Frequency and severity of offending behavior. Details of all offending behavior that occurred in the 6 months before commencing MST treatment were obtained directly from Police Youth Aid or CYF Youth Justice officers by the evaluation coordinator. The date and nature of each contact with a judicial agency (Youth Aid or CYF) was recorded (defined as a warning, charge, or conviction from a judicial agency). In conjunction with the New Zealand Police, the 17-point Seriousness Scale used in previous MST RCTs (i.e., Hanson, Henggeler, Haefele, & Rodick, 1984) was adapted to correspond to New Zealand judicial offence codes. Low scores (1–4) were characterized by status of-

¹ A marae is an established meeting place for New Zealand Maori including whanau (family), hapu, (sub tribe), and iwi (tribe).

fences (truancy, missing person, disorderly behavior); mid range values (5–10) were characterized by crimes such as assault, breaking and entering, and carrying a dangerous weapon; and high scores (11–17) were associated with violent crimes including armed robbery, criminal sexual conduct, and murder. After the completion of treatment and at 6- and 12-month follow-up, the same details were collected again from Police Youth Aid or CYF Youth Justice.

Days in out-of-home placements. Details of all formal out-of-home placements mandated by the court or CYF that occurred in the 6 months before commencing MST treatment were obtained directly from CYF caseworkers by the evaluation coordinator. After the completion of treatment and at 6- and 12-month follow-up, the same details were obtained.

School and vocational attendance. Details of attendance at an educational or employment setting were obtained from the relevant organization for the 6 months before commencing MST treatment. An attendance metric was created by dividing the total number of possible half days (the measure of attendance in New Zealand schools) the youth could potentially have attended by the number of half days actually attended. A wide range of school, work-skill, or tertiary training options were included in measures of attendance (i.e., primary, intermediate or secondary school, alternative education courses, trade skill courses, tertiary institute courses, apprenticeships, part- or full-time employment). After the completion of treatment and at 6- and 12-month follow-up, the same details were obtained.

Instrumental Outcomes

Individual adjustment and behavioral change. Nine items of the Therapist Adherence Measure (TAM items 27–38; Henggeler & Borduin, 1992) that related specifically to youth adjustment were administered to the primary caregiver as part of the complete TAM. These TAM items (TAM-B) assessed aspects of youth adjustment including anxiety, depression, aggression, incidence of self-harm and deviant peer association. Items are rated on a 5-point Likert scale ranging from *almost always* (1) to *never* (5). Internal reliability of these items in the present study was calculated using Cronbach's alpha. At pre- and posttreatment, the alpha coefficients were both found to be 0.70.

Multisystemic behavioral rating scale. The Multisystemic Behavioral Rating Scale (MST-BRS) was designed specifically for this study to assess progress in targeted areas of MST treatment. Administered to the primary caregiver by phone, the MST-BRS is a brief 11-item measure that assesses aspects of youth compliance, family communication, and family relations. All items were rated on a 5-point Likert scale ranging from *not at all* (1) to *very much* (5) and summed to create a rating scale total index. This measure was administered at pre- and posttreatment assessment, and at 6- and 12-month follow-up. Cronbach's alphas for the MST-BRS at pretreatment, posttreatment, 6- and 12-month follow-up were found to be 0.85, 0.95, 0.96, and 0.87, respectively.

Parental monitoring. The two-item Parental Supervision Index (PSI; Jang & Smith, 1997) was adapted for use as a self-report rating of parental monitoring. Ratings for both items were made on a 5-point Likert scale ranging from *never* (1) to *almost always* (5) and summed to create a total score. Cronbach's alphas for the PSI at pretreatment, posttreatment, 6- and 12-month follow-up were found to be 0.69, 0.86, 0.94, and 0.81, respectively.

Client Satisfaction

The CSQ-8 (Attkisson & Zwick, 1982) is an eight-item version of the Consumer Satisfaction Questionnaire (Larsen, Attkisson, Hargreaves, & Nguyen, 1979). The level of satisfaction experienced by families in this study was assessed using a revised version of the CSQ-8. Four items were modified and an additional item was developed to connect family satisfaction more directly with specific elements of the MST program (i.e., therapist availability, home-based services, treatment planning). The nine items were rated on a 5-point Likert scale ranging from very dissatisfied (1) to very satisfied (5). At posttreatment, Cronbach's alpha for the nine items used in this sample was found to be 0.79.

Statistical Analysis

Descriptive and Pre-Post Comparisons

Descriptive statistics were calculated for both sample demographics and outcome measures (i.e., ultimate outcomes including offending frequency and severity, school attendance, out-of-home placement; instrumental outcomes including TAM-B, MST-BRS, and PSI). Repeated measures ANOVA and Chi Square analyses were also conducted to assess pre- to post- and follow-up treatment effects.² Additionally, descriptive statistics were calculated for client satisfaction (CSQ-8) measured at posttreatment.

Benchmarking Comparisons

Comparison outcomes and benchmarks. In following a benchmarking framework developed by Hunsley and Lee (2007), this methodology involves comparisons of two major variables: participant completion rates and major study outcomes (in this case, ultimate, instrumental, and aggregated outcomes) with a 'best practice benchmark'. Given that Hunsley and Lee specified our meta-analysis (Curtis et al., 2004) as the benchmark for adolescents with antisocial behavior, this study was chosen as the benchmark for participant completion rates. However, as Curtis et al. examined between-group effect sizes, and this study used a within-group design more typical of effectiveness studies, the findings from our meta-analysis were seen as an inappropriate benchmark for the other outcomes of interest. Therefore, the RCTs ($n = 3$) reflecting the characteristics of

² These analyses were conducted using the SPSS for Windows program, Standard Version 11.0.

participants within the current study (i.e., chronic juvenile offenders) were identified as benchmarks (Borduin et al., 1995; Henggeler et al., 1992, 1997).

Statistical comparisons. As recommended by Hunsley & Lee (2007), a two-sample t test for proportions was conducted to examine the difference in participant completion rate between the current study and the benchmark. While a range was also provided for the benchmark (76–100%), an average value was used (86%) in line with recommendations (see Hunsley & Lee).

For the second comparative outcome, Hunsley and Lee (2007) recommended the calculation of ESs for key variables (i.e., ultimate, instrumental, and aggregated outcomes). However, they did not specify a statistic that could be used to compare quantitatively the results of the current study to the benchmarks. As a result, we extended the Hunsley and Lee approach through a methodology developed by Minami et al. (2006; see also Minami et al., 2008). Briefly, Minami et al. (2006) recommended a procedure for using meta-analytic techniques to develop benchmarks (i.e., treatment efficacy and natural history/control benchmarks) that can then be statistically compared (using the “good-enough principle” and the non-central t statistic; see below for more detail) to the results of an effectiveness study.

Based on this methodology, the first step involved is the calculation of an ES index for both the current study and the selected benchmarks. For present purposes, the d index is defined as the mean within group gain score divided by the standard deviation of the control time (pretreatment) corrected for a small sample size.³ Consistent with Minami et al. (2006) recommendations, the formulas for $d_{(i)}$ and the variance ($\sigma^2_{d(i)}$) of $d_{(i)}$ are as follows:

$$d_{(i)} = \left(1 - \frac{3}{4n - 5}\right) \frac{M_{Post} - M_{Pre}}{SD_{Pre}}$$

$$\sigma^2_{d(i)} = \frac{2(1 - r)}{n} + \frac{d^2}{2n}$$

Where r is defined as the correlation between scores at pre- and posttreatment and is estimated as follows (see Lipsey & Wilson, 2001):

$$r = \frac{\sqrt{d^2}}{\sqrt{d^2 + 4}}$$

Following the calculation of an ES for each of the identified benchmark studies, the results were aggregated to create a single ES index for each condition (i.e., treatment and control; see Minami et al., 2006). The equation used was as follows:

$$d_{(B)} = \frac{\sum d_i}{\sqrt{\sum \sigma^2_{d(i)}}} \bigg/ \sqrt{\sum \frac{1}{\sigma^2_{d(i)}}}$$

Following the calculation of an aggregated ES for the treatment and control benchmarks, the final step involved a statistical comparison to determine if the current study was either: (a) clinically equivalent to the treatment benchmark; (b) superior to the control benchmark but inferior to the treatment benchmark; or (c) inferior to the control benchmark (Minami et al., 2006). To establish clinical equivalence, Minami et al. set a statistical criterion of $\Delta = 0.2$ that deemed “any difference between the benchmark and the population represented by the sample that is under $\Delta = 0.2$ to be clinically trivial” (Minami et al., 2006, p. 7). Based on the good enough principle, the $\Delta = 0.2$ was chosen based on it being the upper limit of Cohen’s small effect size range (Minami et al., 2006; see also Cohen, 1977).

To determine if the ES of a current study condition is clinically equivalent to benchmark conditions, a $t_{(Treatment)v,\lambda;.95}$ statistic is calculated. For this purpose, the degrees of freedom are defined as $\nu = N - 1$, the noncentrality parameter as $\lambda = \sqrt{N}(d_{B(Treatment)} - \Delta)$, and t is set at the 95th percentile of the noncentral t distribution.⁴ The $t_{(Treatment)v,\lambda;.95}$ statistic in turn allows a critical value ($d_{CV(Treatment)}$) to be calculated. The critical value is the value that the ES from the current study must exceed to be classified as clinically equivalent (Minami et al., 2006):

$$d_{CV(Treatment)} = t_{(Treatment)v,\lambda;.95} / \sqrt{N}$$

Finally, it is also necessary to determine if the current study exceeded the results found in the control condition benchmark. For this purpose, a $t_{(Control)v,\lambda;.95}$ statistic is calculated. For this statistic, the degrees of freedom are defined as $\nu = N - 1$, the noncentrality parameter as $\lambda = \sqrt{N}(d_{B(Control)} + \Delta)$, and t is set at the 95th percentile of the noncentral t distribution. As above, the $t_{(Control)v,\lambda;.95}$ statistic allows a critical value to be calculated. In this instance, the critical value is the value that the ES from the current study must exceed to claim effectiveness greater than that obtained within the comparison conditions (e.g., time alone or treatment as usual; Minami et al., 2006).

Results

Pretreatment Status on Ultimate Outcomes

Thirty-five percent ($n = 29$) of the youth had been in contact with the youth justice system in the pretreatment assessment period (6 months before commencing MST treatment). Across the whole sample, the average severity rating of pretreatment incidents was 3.1 out of a total of 17

³ The method of effect size calculation that uses the pretreatment standard deviation is recommended by Minami et al. (2008) as preferable to the method that uses the pooled standard. It has been found to be a better estimate of the true effect as it is not biased by factors such as repeated measures and treatment (see Morris, 2000; see also Minami et al., 2006, 2008).

⁴ These analyses were conducted using Exploratory Software for Confidence Intervals, which is a program developed by Geoff Cumming (see www.psy.latrobe.edu.au/esci).

($SD = 3.53$). In addition to offending behaviors, 58% ($n = 36$) of the youth had lived in a mandated out-of-home placement during the pretreatment period. These placements ranged in length from 1 to 183 days ($M = 33.62$, $SD = 50.96$). The average pretreatment school attendance was 55% of available half-days (ranging from 108 to 292 possible half-days, $SD = 35.17$). Forty percent of youth ($n = 27$) had been withdrawn from school during the pretreatment period; 22% ($n = 15$) of youth had been temporarily excluded from school or suspended (for periods ranging between 3 and 15 days); and 18% ($n = 12$) were permanently excluded from a school.

Treatment and Assessment: Completion Rates and Length of Treatment

The current study had a high rate of treatment completion ($n = 64$; 98%). The majority of participants did complete measures at both pre- and posttreatment (85%, 83%, 55% completion rates for MST-BRS, TAM-B, and PSI, respectively). For ultimate outcome indicators, data for 100% of the completed sample ($n = 64$) were collected at both pretreatment and posttreatment, between 91% and 98% (ns ranged from 58 to 63) at 6 month follow-up, and between 81% and 97% (ns ranged from 52 to 62) at 12 month follow-up. The average length of MST treatment was 155 days ($SD = 39.22$) with the range spanning 61 to 253 days.

Pretreatment, Posttreatment, 6- and 12-Month Follow-up Data

A series of repeated measures analyses (ANOVA) were conducted to compare participants' pre- and posttreatment and 6- and 12-month follow-up scores in the areas of: (a) frequency and severity of offending behavior, (b) attendance at school/vocational training and, (c) days in mandated out-of-home placements.

Frequency of offending behavior. A significant difference was found in the frequency of offending behavior between pre- and posttreatment, $F(1, 63) = 3.96$, $p = .05$; see Table 2. These gains were maintained at 6-, $F(1, 62) = .34$, $p > .05$, and 12-month follow-up, $F(1, 61) = 1.64$, $p > .05$. The mean number of offences across the sample reduced from 2.25 at pretreatment, to 1.52 at posttreatment, to 1.22 at 6-month follow-up, and to 0.95 at 12-month follow-up. The number of participants who engaged in offending behaviors between assessment intervals (and, for pretreatment, in the 6 months before that assessment) also decreased after treatment and continued to decrease across the follow-up period. The proportions of participants who offended were as follows: 51% ($n = 33$) at pretreatment; 41% ($n = 26$) at posttreatment; 35% ($n = 22$) at 6-month follow-up; and 27% ($n = 17$) at 12-month follow-up.

Severity of offending behavior. A significant difference in the severity of offending behavior was not found from pre- to posttreatment, $F(1, 63) = 1.62$, $p > .05$ or between pretreatment and 6-month follow-up, $F(1, 62) = 1.46$, $p > .05$. However, the severity of offences from pretreatment to 12 month follow-up was significantly decreased, $F(1, 61) = 4.24$, $p < .05$. The average severity of offending behavior reduced across intervals from 3.27 to 2.70 (17%) between pre- and posttreatment, from 2.70 to 2.54 (6%) between posttreatment and 6-month follow-up, and from to 2.54 to 1.92 (24%) between 6- and 12-month follow-up.

School attendance. On average, school attendance increased by 15% during treatment, $F(1, 63) = 17.93$, $p < .001$. At the 6-, $F(1, 57) = 9.72$, $p < .01$ and 12-month follow-up, $F(1, 51) = 5.28$, $p < .05$, school attendance percentage was significantly lower in comparison to posttreatment levels, and had almost returned to pretreatment levels. Average school attendance was 53% at pretreatment, 67% at posttreatment, and 55% at both 6-month follow-up and 12-month follow-up intervals, respectively.

Table 2
Means, Standard Deviations, and F Values, and Effect Sizes for Ultimate and Instrumental Outcomes at Pre- and Posttreatment

Outcome variable	Treatment period				<i>F</i> ^d	<i>d</i>
	Pretreatment		Posttreatment			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Ultimate outcomes						0.32
Offending						
Frequency	2.25	3.14	1.52	2.54	3.96*	0.23
Severity	3.27	3.48	2.70	3.45	1.62	0.16
OHPs (days)	38.48	54.88	13.50	28.48	16.13***	0.45
School attendance ^{a,b}	51%	34.29	67%	29.20	14.93***	0.44
Instrumental outcomes						0.75
MST-BRS	5.11	2.13	6.12	2.43	5.34*	0.29
PSI	4.74	1.36	6.80	2.69	42.85**	0.48
TAM-B	3.48	0.64	3.67	0.62	3.69 [‡]	0.47

Note. OHPs = Out-of-home placements; MST = MST-Behavioral Rating Scale; PSI = Parental Supervision Index; TAM-B = Therapist Adherence Measure – Behavior (youth adjustment scale).

^a School attendance reflects percentage of attendances (possible half days attended/actual half days attended). ^b This was a chi-square analysis.

† $p = .06$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Days in formal out-of-home placements. Significant pre- to posttreatment differences in the number of days youth spent in formal out-of-home placements were found, $F(1, 63) = 16.13, p < .001$. On average, days spent out-of-home reduced from 38 days pretreatment to 13 days posttreatment. This gain was not maintained at 6-month, $F(1, 57) = 5.61, p < .05$ or 12-month follow-up, $F(1, 56) = 4.15, p < .05$, as average days out-of-home increased to 31 at both 6-month follow-up and 12-month follow-up intervals, respectively.

Instrumental Outcomes

Individual adjustment and behavioral change. Based on parent/caregiver responses to the TAM-B, youth demonstrated behavioral improvements across treatment reflecting a trend towards significance, $F(1, 52) = 3.69, p = .06$. The MST-BRS did reflect significant improvements in youth and family functioning between pre- and posttreatment, $F(1, 55) = 5.34, p < .05$. Significant increases continued at 6-month, $F(1, 28) = 41.04, p < .001$ and 12-month follow-up, $F(1, 12) = 15.73, p < .01$. Similarly with parental monitoring, parent/caregivers reported a significant increase from pre- to posttreatment, $F(1, 34) = 42.85, p < .01$. Significant increases continued at 6-month, $F(1, 31) = 6.42, p < .05$ and were maintained at 12-month follow-up, $F(1, 15) = 2.74, p > .05$.

Client Satisfaction

On average, families reported satisfaction with the program, with responses ranging from *satisfied* to *very satisfied* ($M = 34.5$ [out of 45], $SD = 6.77$). The relationships between the CSQ (i.e., the measure of family satisfaction) and (a) gender, (b) ethnicity, (c) family composition (single or two parent family), and (d) caregiver employment status were tested using *t*-tests. No significant differences were found (all $ps > .05$).

Effect Sizes for Ultimate and Instrumental Outcomes

The aggregate pre- to posttreatment effect size for the sample across ultimate and instrumental measures was $d = 0.53$ (see Table 3 for a breakdown of effect sizes).

Benchmarking

To assess comparability of offending across samples, we report offending history data from previous RCTs. For comparison studies, we report the descriptive reports of lifetime arrests before treatment for each study. Before treatment, the average number of lifetime arrests across study samples ranged from 3.07 ($SD = 2.07$; Henggeler et al., 1997) to 4.20 ($SD = 1.40$; Borduin et al., 1995); Henggeler et al. (1992) reported an average of 3.50 arrests ($SD = 2.80$). In comparison, our own study assessed offending in the 6 months before treatment and reported an average of 2.12 ($SD = 3.12$) offences.

In terms of treatment comparisons, successful completion rates were first compared. Participants in the current study followed the treatment through to completion significantly

more often than the average number of participants in benchmarked studies, $t(179) = 2.33, p = .02$ (see Table 3).

In terms of ultimate and instrumental outcomes, the current study ($d = .53$) was found to be clinically equivalent to the treatment benchmark ($d_B = .32$), $t(63) = 2.66, \lambda = .93, p = .05$.⁵ Additionally, the current study was found to significantly exceed the control benchmark ($d_B = .00$), $t(63) = 3.38, \lambda = 1.60, p < .05$. Thus, the effectiveness of MST provided to NZ antisocial youth delivered via a community based clinic was comparable in magnitude to MST delivered in RCTs conducted in the United States and was superior to TAU conditions.

Discussion

This study evaluated the effectiveness of MST in terms of (a) successful treatment completion, (b) reducing youth offending and recidivism, (c) reducing days in formal out-of-home placements, (d) increasing school or vocational attendance, and (e) improving youth psychosocial functioning and family relations.

Ninety-eight percent of the families in this study successfully completed treatment. The high treatment completion rate may well reflect an explicit MST focus on motivation and engagement strategies. It may also reflect an explicit emphasis in this study of favoring engagement over data collection from families. In addition to a favorable treatment completion rate, the current study found that after completion of MST, youth and family relations were improved, youth were attending school more often, youth were removed from their home less often, and the frequency of offending behavior was reduced. Results show significantly improved levels of parental monitoring and significant reductions in antisocial and related behaviors in youth.

One of the most promising findings of this study was that MST conducted through New Zealand publicly funded community mental health clinics with a sample of juvenile offenders was found to be clinically equivalent to findings from MST studies conducted in the United States with juvenile offenders. Additionally, MST was superior to the comparison condition benchmarks. Although benchmarking cannot overcome various threats to internal validity produced by a single group design (Cook & Campbell, 1979), similarities in benchmarked indicators strengthens confidence that MST was responsible for changes across treatment. Importantly, this also suggests that positive treatment outcomes identified in RCTs may be able to be transferred to publicly funded not for profit clinics. This is particularly

⁵ The Minami et al. (2006, 2008) methodology does not allow for the possibility that the results of an effectiveness study may exceed an efficacy benchmark. As our calculated *ES* was greater than the treatment benchmark, this was of interest. Based on Minami and colleagues rationale of clinical equivalence, we followed the methodology they outlined for establishing superiority of effectiveness study findings in relation to benchmarked natural history/control conditions. Based on this calculation, it was concluded that the aggregated *ES* of the current study was not clinically superior to the aggregated *ES* of the treatment benchmark.

Table 3

Benchmarking: Comparison Completion Rate Percentages and Effect Sizes

	Current study	Curtis et al., 2004	Aggregated benchmark <i>ES</i> s		Borduin et al., 1995		Henggeler et al., 1992		Henggeler et al., 1997	
			MST	IT/US	MST	IT	MST	US	MST	US
Completion rate	98%*	86%								
Ultimate outcome <i>ES</i>										
Objective indicators	.32									
Subjective indicators							.54 ^a	.29 ^a	.53 ^a	.26 ^a
Instrumental outcome <i>ES</i>	.75				.36	-.18	.14	.03	.04	-.03
Aggregate <i>ES</i>	.53		.32	.00	.36	-.18	.34	.18	.28	.12

Note. *ES* = Effect Size; MST = Multisystemic Therapy; IT = Individual Therapy; US = Usual Services.

^a Self-reported and parent-reported ultimate outcome measures (i.e., Self-Report Delinquency Scale; Global Severity Index of the Brief Symptom Inventory; Revised Problem Behavior Checklist). Ultimate outcomes similar to those assessed in the current study (e.g., offending frequency and severity) were captured at posttreatment but not at pretreatment, thus preventing prepost comparisons and a resultant within-group effect size calculation.

* $p < .05$.

encouraging in light of recent evidence that transporting evidence-based interventions into publicly funded, not for profit settings, may be less successful than transporting them into private agencies (Glisson et al., 2008; Schoenwald, Kelleher, & Weisz, 2008).

One area of concern with the benchmarking procedure was that the participants within the current sample may not have reflected as severe an antisocial profile as in one or more of the benchmarking studies. Although the benchmark studies did report a higher average arrest rate than the current study, it is important to note that the metrics were not directly comparable. The index used at pretreatment in the comparison studies was lifetime arrests whereas in our study they were documented contacts (defined as a warning, diversion, arrest, or conviction from a judicial agency) in the 6 months immediately before treatment. Nevertheless, despite these variations of definition, it is clear that the current sample and the samples in the benchmarking studies all reflected significant problems related to offending behavior.

In comparison to the pre-to-post outcomes, the follow-up data reflected some mixed findings. In particular, gains in school attendance and out-of-home placements deteriorated across follow-up, whereas offending indices continued to improve. In terms of school attendance, an incongruity may be reflected in the variations in school attendance across the assessment periods. Specifically, given the average age of the youth at 12-month follow-up ($M = 15.65$ years), it is possible that the reduced school attendance evident at 6- and 12-month follow-up is a reflection of the difficulties associated with the transition from school to training programs or employment, as 15 to 16 years is a common age in New Zealand for such a transition. Indeed, the reduction in school/vocational attendance at 6- and 12-month follow-up may be a reflection of the disproportionate difficulty that youth within this age bracket have securing employment or employment-related skills training in their communities (Statistics New Zealand, 2004).

Initial gains in out-of-home placements also deteriorated across follow-up. As with school attendance, this finding may be at least partially attributed to the increasing age of participants and the number of youth that left home during

the follow-up periods. By contrast, indices associated with offending demonstrated continuing improvement across all intervals. The number of participants who committed an offence was almost halved and the frequency and severity across the study period (i.e., through 12 month follow-up) reduced across the sample by over 50% and 40%, respectively. Given the deterioration on the other ultimate indicators, these combined findings may reflect a disproportionate focus on relapse prevention for offending-related behavior.

An area of concern within the current study relates to the high level of therapist and supervisor attrition (42%). This rate of attrition may well have impacted negatively on treatment effectiveness by causing disruptions in treatment and by creating problems with team stability and continuity. One reason for the high attrition may be that many therapists reportedly found it taxing to provide intensive support to families outside normal working hours. Another cause might have been the extensive travel required of some therapists to cover large geographical areas. In addition, several therapists and supervisors noted that it was difficult to adjust to the rigorous quality assurance aspects of the program that demanded significant changes to their working routine (i.e., preparing for and attending weekly case supervision, therapy adherence, working primarily in the homes of families, ongoing and intensive outreach related to MST services amongst professionals and associated agencies). The attrition rate combined with the anecdotal findings highlight the importance of assessing reasons for therapist attrition as well as systematic efforts designed to reduce turnover.

Other limitations in the current study include the number of parents who reported difficulties completing all of the assessment measures because of persistent competing demands (e.g., domestic responsibilities, work commitments, the youth's disruptive and challenging behaviors). These difficulties raise the possibility that those parents who completed all of the assessment measures may have been less stressed and/or derived more benefit from the MST program than those parents who did not complete the measures. Future research might address the possibility of variable

outcomes as a function of level of engagement and/or degree of stress in the family system.

Subsequent to this initial evaluation of MST in New Zealand, a randomized controlled study is now recommended to explore the effectiveness of MST in comparison to usual services and other treatment approaches available in New Zealand. In designing a future RCT for MST in New Zealand, care should be taken to ensure that appropriate comparison groups are chosen. This recommendation follows the finding of the current benchmarking exercise that suggests that the type of control group may contribute to between-group results. For example, the meta-analysis conducted by Curtis et al. (2004) found that efficacy studies had a significantly greater effect on treatment outcomes ($ES = .81$) when compared to the effectiveness studies ($ES = .27$). Initially, this finding was attributed to a difference in the level of supervision provided to therapists; that is, the therapists in the efficacy studies were graduate students being supervised by a developer of an MST treatment manual whereas the therapists in the effectiveness studies were community-based therapists not supervised by an MST developer (Curtis et al., 2004). However, an additional difference between the efficacy and effectiveness studies that was noted in Curtis et al. was the assigned control condition. The current study lends support to the idea that the assigned control condition (i.e., usual services versus a more specific intervention modality) may have contributed to the significant between-group ES differences found by Curtis et al. In fact, the benchmarked MST conditions within-group ES s were relatively comparable ranging from .28 to .36. By contrast, ES s for the three comparison conditions indicated that the efficacy study (Borduin et al., 1995) control ES was negative ($ES = -.18$) whereas the two from the effectiveness studies (Henggeler et al., 1992, 1997) were both positive ($ES = .18$ and $.12$, respectively; see Table 3). These findings also highlight the value of calculating both between- and within-group ES 's in studies where this is possible (i.e., those that use mixed factorial designs and those that use a meta-analytic procedure).

Taken together, the findings of this evaluation add to the growing body of evidence that supports MST as an effective treatment for antisocial behavior in youth. Further, parent-reported satisfaction rates indicate that the clients perceived a benefit from their involvement with MST and appreciated receiving this type of help for their youth. Based on current findings, MST appears to be a valuable addition to existing health, judicial, and social services in New Zealand. As MST is more widely disseminated, efforts will be needed to ensure that support for and research on the ongoing dissemination of the treatment model continues, while taking into account the social, cultural, and ethnic factors that are unique to a particular country or context.

References

- Attkisson, C. C., & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Planning, 5*, 233–237.
- Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., et al. (1995). Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology, 63*, 569–578.
- Burns, B. J., Hoagwood, K., & Mrazek, P. J. (1999). Effective treatment for mental disorders in children and adolescents. *Clinical Child and Family Psychology Review, 2*, 199–254.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company.
- Crampton, P., Salmond, C., & Kirkpatrick, R. (2004). *Degrees of deprivation in New Zealand: An atlas of socioeconomic difference*. Auckland, New Zealand: David Bateman Ltd.
- Curtis, N. M. (2004). *Multisystemic Treatment (MST) of antisocial behaviour in youth: The role of the family in facilitating change*. Unpublished doctoral thesis, Massey University, New Zealand.
- Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology, 18*, 411–419.
- Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1993). Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15-year-olds. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 1127–1134.
- Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry, 46*, 837–849.
- Glisson, C., Schoenwald, S. K., Kelleher, K., Landsverk, J., Hoagwood, K. E., Mayberg, S., et al. (2008). Therapist turnover and new program sustainability in mental health clinics as a function of organization culture, climate, and service structure. *Administration and Policy in Mental Health, 35*, 124–133.
- Hanson, C. L., Henggeler, S. W., Haeefe, W. F., & Rodick, J. D. (1984). Demographic, individual, and family relationship correlates of serious and repeated crime among adolescents and their siblings. *Journal of Consulting & Clinical Psychology, 52*, 528–538.
- Henggeler, S. W. (1989). *Delinquency in adolescence*. Newbury Park: Sage Publications.
- Henggeler, S. W. (2004). Decreasing effect sizes for effectiveness studies – implications for transport of evidence-based treatments: Comment on Curtis, Ronan, & Borduin (2004). *Journal of Family Psychology, 18*, 420–423.
- Henggeler, S. W., & Borduin, C. M. (1990). *Family therapy and beyond: A multisystemic approach to treating the behavior problems of children adolescents*. Pacific Grove, CA: Brooks/Cole.
- Henggeler, S. W., & Borduin, C. M. (1992). *Multisystemic therapy adherence scales*. Unpublished instrument. Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.
- 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*, 821–833.
- 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*, 953–961.
- Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of*

- antisocial behavior in children and adolescents: Treatment manuals for practitioners*. New York: Guilford Press.
- Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice*, 38, 21–33.
- Jang, S. J., & Smith, C. A. (1997). A test of reciprocal causal relationships among parental supervision, affective ties, and delinquency. *Journal of Research in Crime & Delinquency*, 34, 307–336.
- Kazdin, A. E., Holland, L., & Crowley, M. (1997). Family experience of barriers to treatment and premature termination from child therapy. *Journal of Consulting and Clinical Psychology*, 65, 453–463.
- Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, 66, 19–36.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, 2, 197–207.
- McGee, R., Feehan, M., Williams, S., & Anderson, J. (1992). DSM-III disorders from age 11 to age 15 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 50–59.
- Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2006). Using clinical trials to benchmark effects produced in clinical practice. *Quality & Quantity*. Retrieved June 1, 2008, from <http://www.springerlink.com/content/u11t4p0740u11557/fulltext.pdf>
- Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology*, 76, 16–124.
- Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology*, 53, 17–29.
- Nock, M. K., & Ferriter, C. (2005). Parent management of attendance and adherence in child and adolescent therapy: A conceptual and empirical review. *Clinical Child and Family Psychology Review*, 8, 149–166.
- Norcross, J. C., Beutler, L. E., & Levant, R. F. (Eds.) (2006). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC: American Psychological Association.
- Ronan, K. R., & Curtis, N. M. (2008). Treatment with antisocial youth and families. In L. VandeCreek & J. B. Allen (Eds.), *Innovations in clinical practice: Focus on group, couples, & family therapy*. *Innovations in clinical practice* (pp. 5–27). Sarasota, FL: Professional Resource Press/Professional Resource Exchange.
- Schoenwald, S. K., Kelleher, K., & Weisz, J. R. (2008). Building bridges to evidence-based practice: The MacArthur foundation child system and treatment enhancement projects (Child STEPs). *Administration and Policy of Mental Health*, 35, 66–72.
- Schoenwald, S. K., Sheidow, A. S., Letourneau, E. J., & Liao, J. G. (2003). Transportability of evidence-based treatments: Evidence for multi-level influences. *Mental Health Services Research*, 5, 223–239.
- Statistics New Zealand. (2004). Household labour force survey: March 2004 quarter. Retrieved May 28, 2004, from <http://www.stats.govt.nz/domino/external/pasfull.nsf/web/Hot+Off+Press+Household+Labour+Force+Survey+March+2004+quarter>
- Washington State Institute for Public Policy. (2001). *The comparative costs and benefits of programs to reduce crime* (4th ed.). Olympia, WA: Washington State Institute for Public Policy.
- Weersing, V. R. (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 1058–1062.

Received February 5, 2008

Revision received October 10, 2008

Accepted October 16, 2008 ■