The effectiveness of Multisystemic Therapy (MST): A meta-analysis

Trudy van der Stouwe [a,*], Jessica J. Asscher [a], Geert Jan J.M. Stams [a], Maja Deković [b], Peter H. van der Laan [c,d]

[a] Forensic Child and Youth Care Sciences, University of Amsterdam, Amsterdam, The Netherlands
[b] Clinical Child and Family Sciences, Utrecht University, The Netherlands
[c] Criminology, VU University, Amsterdam, The Netherlands
[d] Netherlands Institute for the Study of Crime and Law Enforcement, Amsterdam, The Netherlands

HIGHLIGHTS

- A multilevel meta-analysis showed that MST has a small effect on delinquency.
- Small effects were found on five of the secondary outcomes.
- MST was most effective with non-ethnic minority, (sex)offending, younger juveniles.
- Larger effects were found with USA studies.
- Larger effects were found when MST was compared to a non-multimodal treatment.

ARTICLE INFO

ABSTRACT

Multisystemic Therapy (MST) is a well-established intervention for juvenile delinquents and/or adolescents showing social, emotional and behavioral problems. A multilevel meta-analysis of $k = 22$ studies, containing 332 effect sizes, consisting of $N = 4066$ juveniles, was conducted to examine the effectiveness of MST. Small but significant treatment effects were found on delinquency (primary outcome) and psychopathology, substance use, family factors, out-of-home placement and peer factors, whereas no significant treatment effect was found for skills and cognitions. Moderator analyses showed that study characteristics (country where the research was conducted, efficacy versus effectiveness, and study quality), treatment characteristics (single versus multiple control treatments and duration of MST treatment), sample characteristics (target population, age, gender and ethnicity) and outcome characteristics (non-specific versus violent/non-violent offending, correction for pre-treatment differences, and informant type) moderated the effectiveness of MST. MST seems most effective with juveniles under the age of 15, with severe starting conditions. Furthermore, the effectiveness of MST may be improved when treatment for older juveniles is focused more on peer relationships and risks and protective factors in the school domain.

© 2014 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author at: Forensic Child and Youth Care Sciences, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands.
E-mail address: t.vanderstouwe@uva.nl (T. van der Stouwe).

1. Introduction

Multisystemic Therapy (MST) is a multi-faceted, short-term, home- and community-based evidence-based intervention for juvenile delinquents and juveniles with social, emotional and behavioral problems, disseminated in fourteen countries (MST Services Inc., 2010). The intervention is considered one of the few empirically supported and evidence-based treatments for conduct problems (see inter alia Littell, 2005). Moreover, it is one of few interventions targeting externalizing behavior problems that intensively monitors treatment integrity (Burns, Hoagwood, & Mrazek, 1999; Goense, Boendermaker, Van Yperen, Stams, & Van Laar, 2014).

MST is based on the premise that adolescent delinquency is associated with an accumulation of criminogenic risk factors (e.g., Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998), in particular individual, family, peer, school and neighborhood characteristics (Henggeler, 2011; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009; Henggeler, Schoenwald, Rowland, & Cunningham, 2002). The idea that these factors should be targeted simultaneously finds its base in Bronfenbrenner's (1979) bio-ecological-system approach, which assumes that human behavior develops within and across contexts. MST mainly focuses on improving family functioning, because it is theorized that improvements in family functioning mediate improvements in peer relationships, school functioning and participation in the community (MST theory of change, Henggeler, 2011). Furthermore, the implementation of MST is highly flexible and designed to address specific individual risk factors. This is in line with the Risk–Need–Responsivity (RNR)-model (e.g., Andrews & Bonta, 2010; Andrews, Bonta, & Hoge, 1990; Andrews, Bonta, & Wormith, 2006), which states that judicial interventions should take into account the recidivism risk, and be matched to the criminogenic needs and learning style and capabilities of the individual.

Therapists visit the juveniles and their families at home and/or in their community to reduce drop-out rates, to provide treatment exactly where and when it is needed, and to increase generalizability of newly acquired skills. Moreover, the therapist is available twenty-four hours a day, seven days a week, and therapeutic sessions may take place up to everyday. MST uses well-established treatment strategies derived from strategic family therapy, structural family therapy, behavioral parent training and cognitive-behavioral therapy (Borduin, 1999). Finally, MST is accompanied by training and supervision, organizational support and adherence measures to monitor treatment integrity (Henggeler, 2011).

Since the first efficacy trial (Henggeler et al., 1986) and subsequent implementation of MST, the treatment has been implemented in a growing number of teams, regions and countries. The target population has expanded from delinquent and antisocial juveniles to abused and neglected juveniles (e.g., Brunk, Henggeler, & Whelan, 1987), sex offenders (e.g., Borduin, Henggeler, Blaske, & Stein, 1990), youth with psychiatric emergencies (e.g., Henggeler et al., 1999), substance-abusing and -dependent juveniles (e.g., Henggeler, Pickrel, & Brondino, 1999), youth with poorly controlled type I diabetes (e.g., Ellis et al., 2004), and juveniles with obesity (e.g., Naar-King et al., 2009). All of these variants have been examined at least once, resulting in a total of 20 published randomized controlled trials in 2012 (MST Services Inc., 2012).

With the growing number of randomized controlled studies, a meta-analysis of the effectiveness of MST became possible. To date, two of such meta-analyses have been conducted: one by dependent researchers, associated with the developers of MST (Curtis, Ronan, & Borduin, 2004), and the other by independent researchers (Littell, Campbell, Green, & Toews, 2005).

The first meta-analysis by Curtis et al. (2004) included published studies with random condition allocation about any available MST variant, resulting in eleven eligible studies, consisting of seven independent (non-overlapping) samples and a total of 708 juveniles, including delinquent juveniles, abused and neglected juveniles and youth at risk for psychiatric hospitalization. The meta-analysis yielded a moderate overall effect of MST compared to the control group ($d = .55$), and larger effect sizes were found specifically for family relationships compared to individual adjustment and peer relationships. Furthermore, optimal conditions of delivery showed larger effects than clinical representative conditions (i.e., efficacy versus effectiveness, Flay et al., 2005).

Shortly after publication of the Curtis et al. review, Littell et al. (2005) published their Cochrane systematic review of MST. Littell and colleagues also included non-published studies, which yielded different outcomes than published studies in the Curtis et al. review. Consequently, the Little et al. meta-analysis consisted of 21 studies from eight independent samples, and a total of 1230 juveniles for whom they found inconclusive evidence for the effectiveness of MST.

More than these contradictive outcomes, an article by Littell (2005) about methods used in systematic reviews raised controversy about the evidence-base of MST. Littell (2005) criticized the prior meta-analysis and previous MST research by questioning methodological quality of the studies and incomplete reports of randomization procedures. Additionally, an important point of criticism was the issue that few studies after MST were conducted independently of the MST developers. Notably, Petrosino and Soydan (2005) reviewed 50 meta-analyses of social interventions and conducted a meta-analysis of 300 randomized field trials of interventions targeting recidivism in order to examine the impact of dependency of researchers on study outcomes. They found that research conducted by dependent researchers yielded consistently and substantially larger effect sizes. Lipsey (1995), however, reported about similar findings in his 1992s meta-analysis that a higher level of treatment integrity, due to closely monitoring researchers, is an alternative explanation for the phenomenon Littell referred to as "program allegiance" or "conflicts of interest".

Since these MST meta-analyses and their responses, the (international) research base for MST has grown. Consequently, it seems time to conduct a new meta-analysis. The main aim of the present multilevel meta-analysis therefore was to examine the impact of MST, specifically with juvenile offenders on delinquency (considered to be the primary outcome, because the initial goal of MST is to prevent delinquency) and on other behavioral and psychosocial outcomes (designated as secondary outcomes). Another aim was to assess moderators that may have an effect on the outcomes. These moderators included post-treatment effects on secondary outcomes, including psychopathology and parenting skills, outcome characteristics, such as informant type, sample characteristics, like age and gender, treatment characteristics, such as control treatment composition and – in line with the objections of Littell (2005) – study characteristics, including publication status and the (in)dependence of researchers.

This meta-analysis differs from both previous meta-analyses with respect to several aspects in order to be able to account for shortcomings of the previous analyses (i.e., exclusion of non-published studies,

exclusion of non-randomized studies and inclusion of other target populations than juvenile delinquents) and to make maximum use of the grown body of research on MST.

First, in contrast to the Curtis et al. (2004) meta-analysis, non-published studies were included to reduce possible publication bias. Published studies on average show larger treatment effects than non-published studies (Hopewell, McDonald, Clarke, & Egger, 2007; McAuley, Pham, Tugwell, & Moher, 2000). Excluding unpublished studies could therefore result in an overestimation of the actual effect size.

Second, in contrast to both meta-analyses, non-randomized (quasi-experimental) studies were included. Although randomized controlled trials are considered a more valid study design enabling causal inference, Shrier et al. (2007) found that non-randomized studies generally produce similar results, and that the advantages of including non-randomized studies outweigh the disadvantages (e.g., increased external validity, more statistical power, less publication bias). Including these studies in the current analyses yields a larger number of studies to be analyzed. The possible influence of study quality (Moher et al., 1998) was accounted for by adding a study quality index.

Third, again in contrast to both previous analyses, only antisocial, conduct disordered and/or delinquent juveniles receiving regular MST were included. Juveniles receiving MST for abuse and neglect, diabetes, psychiatric emergencies and obesity were excluded, because these populations differ from the original target population, which could result in differences in the way MST is carried out and subsequently different treatment effects. Exclusion of these studies is thought to yield results that can better be generalized to the population of juvenile delinquents receiving MST.

Finally, the present study is a multilevel meta-analysis, so that multiple effects within studies can be included to generate more statistical power, and to be able to fully exploit the available research data. The larger number of studies compared to the previous meta-analyses and the multilevel design make it possible to (a) analyze more outcome measures and (b) conduct more comprehensive moderator analyses to assess the influence of study, treatment and sample characteristics on treatment effects and (c) analyze outcome measures over a longer follow-up period. This provides the opportunity to test Littell's (2005) assumptions about publication bias and conflicts of interest not only by conducting moderator analyses after the moderating effects of these aspects, but also by investigating the unique moderating effects that remain through multiple regression analysis. Consequently, the present multilevel meta-analysis can shed a new, more conclusive, light on the effectiveness of MST.

In the current meta-analysis, the following research questions will be addressed: 1) to what extent is MST effective in the prevention of recidivism (primary outcome)? 2) To what extent is MST effective in improving juveniles' functioning on other psychosocial (secondary) outcomes? 3) Which sample, treatment and study characteristics have a moderating effect on (heterogeneous) outcomes? 4) What is the unique contribution of significant moderators when controlling for other significant moderating variables?

2. Method

2.1. Selection of studies

All reports from 1985 until 2012 addressing the effectiveness of MST with antisocial or delinquent juveniles were included. This particular starting year was chosen because the first research on MST – not even named MST yet – was published in 1986 (Henggeler et al., 1986).

First, three electronic databases were searched using the search string "Multisystemic therapy" OR "Multisystemic treatment": Science Direct, Web of Knowledge and Google Scholar. Furthermore, MST reviews and primary studies reference sections were searched for qualifying studies: both Curtis et al. (2004) and Littell et al. (2005) meta-

analyses, the MST services Inc. website including referrals to other websites and databases and MST overview reports (e.g. Henggeler, 2011; MST Services Inc., 2012). This search yielded 112 reports of which 51 studies met the inclusion criteria of our meta-analysis, which resulted in 22 independent (non-overlapping) samples.

2.2. Inclusion criteria

To be included in the current meta-analysis, studies had to meet the following criteria: (1) focus MST, targeting antisocial, conduct disordered and/or delinquent juveniles, (2) assignment of participants took place into MST and one or more control group(s), (3) pre- and post-treatment assessment measures and/or follow-up assessment measures were provided, and (4) statistics suitable for meta-analysis were available.

Two studies were excluded because they provided no usable statistics to extract an effect size for MST versus control treatment (i.e., Glisson et al., 2010; Henggeler et al., 2006). One study (i.e., Mitchell-Herzfeld et al., 2008) reported about research in which assignment methods differed over two inclusion periods. This study was therefore included as two separate samples (pilot and post-pilot respectively). When necessary, the information from the primary outcome study was supplemented with information from the other – secondary – reports. Consequently, the final sample of studies consisted of 22 independent samples from 51 reports.

2.3. Coding the studies

Each study was coded using a detailed coding system for recording outcomes and moderators following the guidelines of Lipsey and Wilson (2001). For every effect size, the follow-up duration (in months) and if the outcome had been corrected for pretreatment measures were coded.

The primary outcome was delinquency, defined as any delinquent or illegal post-treatment activity. This could be reported on a dichotomous scale (any offense) or on a continuous scale (number of offenses). Official record reports as well as self-reported delinquency and all types of delinquency (i.e. violent, non-violent and overall) were included, and the source and type of delinquency were coded as potential moderators.

For delinquency, secondary outcomes reported within the same study were added as continuous moderators, using the effect size closest to post-treatment, because it was hypothesized that a moderating effect could only be achieved when changes showed immediately post-treatment. Thereby, post-treatment effects on psychopathology, internalizing and externalizing behavior, skills and cognitions, substance use, parenting skills, family functioning (e.g., cohesion, emotional connectedness, quality of family relationships), parental mental health, out-of-home placement and peers were included as potential moderators for delinquency.

Secondary outcomes included individual, family and peer outcomes. Individual outcomes were psychopathology, skills and cognitions and substance use. For psychopathology, overall as well as specific psychopathological behavior (i.e., externalizing and internalizing behavior), self-reported as well as other informant reports were included and the type of behavior and informant was coded as potential moderators. Furthermore, because the CBCL was an important source of psychopathological outcomes, a potential moderator discriminating between CBCL and non-CBCL reports was coded as well. For skills and cognitions, a potential moderator discriminating between cognitions, social skills and other skills was included. Again, all informant reports were included and the type of informant was coded as a potential moderator. For substance use, alcohol, soft drugs and other substance use outcomes, from all possible informants were included. Both type of substance and type of informant were included as potential moderators.

Family outcomes were family factors and out-of-home placement. Family factors could be any outcomes reporting on family functioning, parenting skills and parental mental health, reported by the participant,

his parents and other informants. The type of outcome reported (i.e., family functioning, parenting, parental mental health) and the type of informant were included as potential moderators. For out-of-home placement, both incarceration as well as regular (non-criminal) placements and outcomes pertaining to the occurrence (any out-of-home placement) and number of days (frequency) of out-of home placement were included and both characteristics were coded as potential moderators.

Finally, peer outcomes included both peer delinquency as well as peer relation outcomes. For this outcome, again all informants were included and the type of peer outcome and the type of informant were included as potential moderators.

2.4. Moderators

Potential moderators of treatment effects for all studies and outcomes were grouped into sample characteristics, treatment characteristics and study characteristics.

Sample characteristics were target population (i.e., offenders and sex-offenders, both charged, arrested or convicted for an offense, and conduct disordered juveniles, referred from child welfare or mental health services), age, proportion of males, proportion living with biological parents, proportion of migrants (non-Caucasian), proportion previously arrested and number of previous arrests. Because of little variance in age between samples, age was coded into a categorical moderator, discriminating between under and over fifteen years old.

Treatment characteristics were control treatment (one single, specific treatment versus multiple possible treatments), proportion of MST completers and duration of MST treatment (in days).

Study characteristics were (in)dependence of authors (i.e., associated with the developers of MST or not), publication status, country, research design (randomized versus quasi-experimental), research setting (efficacy versus effectiveness) and study quality. Because of little variance in country between samples, the country where the research was conducted was coded into a categorical moderator, discriminating between USA and other countries. Studies were considered efficacy studies when MST treatment therapists were (graduate) students and not (experienced, masters level) clinicians, indicating non-clinical representative conditions (Flay et al., 2005). Study quality was assessed using the Quality Assessment Tool for Quantitative Studies (Thomas, Ciliska, Dobbins, & Micucci, 2004), which classifies study quality based on selection bias, study design, confounders, blinding, data collection method and withdrawals and dropouts as weak, moderate or strong. Only the continuous variable (sum of components) was used.

Inter-rater agreement was based on a random selection of 7 studies that were scored by two coders (first and second author). The inter-rater reliability proved to be good to perfect with 100% agreement for publication and research design characteristics, and 90% agreement for treatment and sample characteristics and effect size calculating statistics.

2.5. Calculation and analysis of effect sizes

For the effect size, the standardized mean difference, or Cohen's d, was calculated using formulas from Lipsey and Wilson (2001), Wilson, (2010) and Mullen (1989). Effect sizes were – if possible – calculated for both pre- and post-treatment and pre-treatment effects were subsequently subtracted from post-treatment effects to correct for pre-treatment differences. Each continuous moderator was centered around its mean and dichotomous dummy codes were made for the categorical variables. Outliers (>3.3 SD; Tabachnik & Fidell, 2007) in follow-up period and effect sizes were winsorized.

To account for dependency of study results in the current analysis design, we used a multilevel random effects model for the calculation of combined effect sizes and for conducting moderator analyses (Hox, 2002; Van den Noortgate & Onghena, 2003). The program MLwiN was used to conduct multilevel analysis, using a multilevel random effects model that allows for the hierarchical structure of the data, in which the effect sizes (the lowest level) are nested within studies (the highest level). The model makes it possible to include multiple outcome effects per study, regarding, for example, different follow-up durations, different measures of the same construct or different informants. The random-effects model is often used, and can be extended by including moderators. The effect size standard error was included in the model as a measure of precision for each effect size. Iterative maximum likelihood procedures were applied to estimate unknown parameters.

Finally, the a priori power was estimated using the given standard errors with a small effect size ($d = .200$) at a significance level of $p < .05$ using the formulas from Snijders and Bosker (1999) and Hox (2010). A priori power based on expected small, medium or large effects at $p < .05$ is generally used as a yard stick to evaluate the statistical power of a study, but post-hoc established statistical power is also informative unless observed effect sizes are extremely small, approaching zero (Hoenig & Heisey, 2001; Tabachnik & Fidell, 2007). Therefore, the observed (post-hoc) statistical power for the overall effect size analyses was also estimated.

After moderator analyses, the significant moderators for each outcome variable were entered into a simple non-hierarchical multiple regression analysis to examine the unique influence of each moderator. First of all correlation analysis of all significant moderators was conducted. Highly correlating (i.e. Pearson's Correlation > .8) moderators were excluded from the multivariate analysis in order to prevent multicollinearity problems. In those cases, the moderators with high correlations with more than one other moderator were excluded.

2.6. File drawer analysis

It is commonly known that studies with no significant or negative results are less likely to be published than studies with positive and significant results. This phenomenon, referred to as the 'file drawer problem' (Rosenthal, 1995), was also highlighted by Littell (2005) with the previous meta-analysis. Although part of this problem should be resolved by the inclusion of non-published studies in the current meta-analytic study, file drawer bias was examined by inspecting the distribution of each individual study's effect size on the horizontal axis against its sample size, standard error or precision (the reciprocal of the standard error on the vertical axis). If no publication bias is present, the distribution of effect sizes should be shaped as a funnel, and therefore a violation of funnel plot symmetry reflects publication bias (Sutton, 2009). Selectivity bias according to the funnel plot was examined using MIX 2.0 (Bax, 2011).

3. Results

The current meta-analysis consists of $k = 22$ studies, reporting on $N = 4066$ juveniles of whom $n = 1890$ received MST treatment and $n = 1835$ constituted the control group. Appendices A, B and C present the included studies and their characteristics.

Cohen's (1988) guideline was used to interpret effect sizes, with effect sizes of $d = .20$ considered small, $d = .50$ medium and $d = .80$ large. The overall distribution of effect sizes is presented in Table 1.

As can be derived from Table 1, all outcome variables, except skills and cognitions, yielded small significant effect sizes. The test for heterogeneity was significant for delinquency, family factors, psychopathology and peer factors. The estimates of the statistical power show that the a priori power was insufficient for substance use and peer factors. Even though the power for peer factors was low, a significant effect was still found. Furthermore, posteriori power testing for substance use showed that the power to detect the observed effect size was sufficient ($power = .782$).

Funnel plot examination suggested that there was publication bias for all outcomes, except for skills and cognition. Table 1 therefore

Table 1
Results for the overall mean effect size, fail-safe numbers and overall mean effect size after trim and fill.

Overall effects	# studies	# ES	Mean d	Z	Heterogeneity	Power	After trim and fill					
							# studies	# ES	Mean d	Z	Heterogeneity	Power
Primary outcome												
Delinquency	20	101	.201	3.7***	4.0***	.980	30	142	.014	0.2	6.1***	.910
Secondary outcomes												
Psychopathology	16	66	.268	5.3***	3.8***	.988	22	77	.166	2.5**	5.1***	.910
Skills and cognitions	7	29	−.016	0.2	1.4	.767	7	29	−.016	0.2	1.4	.767
Substance use	5	18	.291	2.4**	1.6	.508	6	19	.191	1.3	1.4	.393
Family factors	12	70	.143	2.4**	2.1*	.959	13	71	.133	2.0*	2.4*	.910
Out-of-home placement	14	19	.267	3.9***	1.6	.894	18	24	.143	1.6	2.3*	.732
Peer factors	8	29	.213	1.7*	2.0*	.472	11	32	.153	0.6	2.7**	.212

Note. # studies = number of independent studies; # ES = number of effect sizes; Z = difference in mean d with reference category; mean d = mean effect size (d); heterogeneity = within class heterogeneity (Z).

* p < .05.
** p < .01.
*** p < .001.

includes overall effect sizes after a publication bias has been taken into account by means of a trim and fill procedure that is, corrected for publication bias as derived from the funnel plots (as advised by Duval & Tweedie, 2000a,b). Fig. 1 displays the funnel plot for primary outcome delinquency. To apply this with the current multi-level design, effect sizes were filled within studies. Significant small effects only remained for psychopathology and family factors after the trim and fill correction. After trim and fill, the a priori power was insufficient for substance use and peer factors. The analyses yielded insufficient power, even when the power was estimated a posteriori, that is, with the observed effect size (a posteriori power: substance use = .371; peer factors = .159).

3.1. Moderator analysis

Because sets of effect sizes proved to be heterogeneous for primary outcome delinquency, as well as secondary outcomes psychopathology, family factors and peer factors, moderator analyses were conducted for these primary and secondary outcome variables before trim and fill. A moderator is considered to have a significant moderating effect if it significantly improves model fit. Table 2 shows the significant moderators for delinquency with their corresponding statistics. Table 3 shows the same for the secondary outcome measures psychopathology, family factors and peer factors.

Trim and fill plot

Fig. 1. Trim and fill plot for primary outcome delinquency.
Retrieved from: Bax L: MIX 2.0 — professional software for meta-analysis in Excel. Version 2.0.1.4. BiostatXL, 2011. http://www.meta-analysis-made-easy.com.

3.2. Primary outcome: delinquency

As can be derived from Table 2, only significant effects were found if general delinquency was measured and not if specifically violent or non-violent delinquency was assessed. It made no difference if the occurrence or frequency was measured, or if delinquency was measured by means of self-report or archival data. No moderating effects were found for follow-up period or if the outcome was corrected for pretreatment measures.

The effect of MST was moderated by several sample characteristics. Significant effects were only found when the target population consisted of offenders or when the average age of the juveniles under investigation was below 15. Furthermore, larger effects on delinquency were found for studies with a larger proportion of Caucasian or indigenous juveniles. In studies with a larger proportion of previously arrested juveniles, effects were larger as well. No moderating effects were found for the proportion of males and the proportion of juveniles living with biological parents.

Treatment characteristics moderated MST effects as well. Larger effect sizes were found if MST was compared to a single and non-multimodal control treatment type. More MST treatment completers and longer treatment duration also yielded larger effects.

Furthermore, the effect of MST on delinquency was moderated by several study characteristics. Research from dependent researchers, published studies, studies conducted in the USA, efficacy studies and better quality studies yielded larger effect sizes. No moderating effects were found research design (randomized versus quasi-experimental).

Finally, post-treatment effects on several secondary outcomes moderated effects on delinquency. Participants showed less delinquency in studies with larger post-treatment improvement in psychopathology, externalizing behavior and substance use. Studies with more improved parenting skills and parental mental health and less out-of-home placement also yielded larger effects. The magnitude of post-treatment effects on internalizing behavior, skills and cognitions, family functioning and peer factors had no moderating effect on delinquency.

3.3. Secondary outcomes: psychopathology, family factors and peer factors

Only few outcome characteristics moderated the effects of MST. For psychopathology, there was no difference in effects between overall psychopathology and specifically externalizing and internalizing behavior. The type of measure used (CBCL versus other measures) did not moderate outcome effects either. Within family factors, effects were only significant if parental mental health or parenting was measured. No effects were found for family functioning. For peer factors, effects did not differ between peer delinquency and peer relation outcome measures. Effects, however, were only found if other informants than
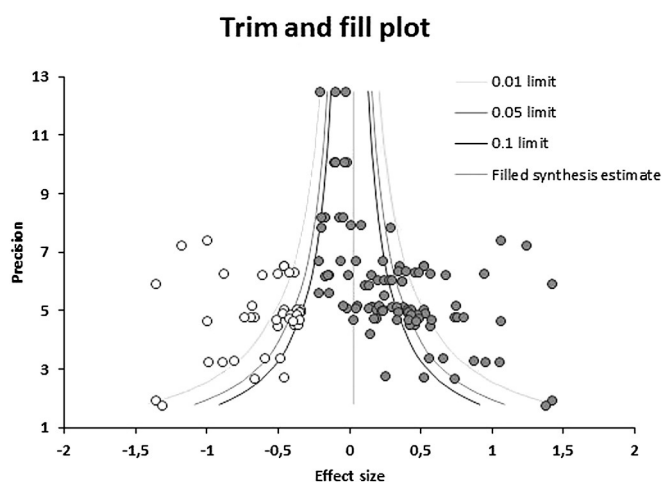
Table 2
The overall mean effect size[a] and significant moderators for delinquency.

Moderator variables	# studies	# ES	β_0, mean d	Z	β_1, RC	Z	Heterogeneity	Δ fit
Delinquency overall	20	101	.201	3.7***			4.0***	–
Outcome characteristics								
Delinquency type							4.0***	7.8**
General (RC)	20	61	.233	3.9***				
Violent	7	21	.115	1.6	−.118	2.2*		
Non-violent	5	19	.082	1.1	−.151	2.6**		
Sample characteristics								
Target population							3.8***	9.3**
Offenders (RC)	11	75	.156	2.7**				
Conduct problems	6	18	.133	1.4	−.023	.2		
Sex offenders	3	8	.700	4.1***	.544	3.0**		
Age							3.8***	8.1***
Under 15 (RC)	8	35	.421	4.7***				
15 years and over	11	65	.105	1.7	−.315	2.9**		
% migrants	18	94	.291		−.745	3.8***	3.5***	17.7***
% previously arrested	7	41	.099		.540	2.2*	2.5*	11.7***
Number of previous arrests	10	76	.341		.074	2.1*	2.4*	5.9*
Treatment characteristics								
Control treatment							4.0***	6.3*
Single treatment (RC)	13	66	.118	2.0*				
Multiple treatments	7	35	.406	4.2***	−.288	2.5*		
Days of MST treatment	14	72	.205		.004	4.0***	2.9**	12.8***
% MST completion	14	64	.162		1.236	2.8**	3.2**	7.2**
Study characteristics								
Authors							3.7***	10.1***
Independent (RC)	12	60	.083	1.4				
Dependent	8	41	.422	5.0***	.339	3.3**		
Publication status							3.8***	11.1***
Non-published (RC)	5	18	−.025	.3				
Published	15	83	.323	5.4***	.348	3.4***		
Country							3.9***	6.5*
USA (RC)	16	83	.275	4.7***				
Non-USA	4	18	−.026	−.3	−.301	2.6*		
Research setting							3.8***	12.8***
Efficacy (RC)	3	22	.727	4.8***				
Effectiveness	17	79	.138	2.7**	−.590	3.7***		
Study quality	20	101	.249		.054	2.6*	3.8***	6.2*
Post-treatment effects on secondary outcomes								
Psychopathology	9	66	.288		.427	2.3*	2.4*	15.1***
Externalizing behavior	11	84	.286		.265	3.3**	2.9**	7.1**
Substance use	5	20	.223		.814	2.5*	0.9	5.3*
Parenting skills	7	64	.238		.998	2.4*	2.5*	18.2***
Parent mental health	6	35	.242		−.044	0.2	2.3*	5.8*
Out-of-home placement	14	86	.233		.623	2.6*	3.4***	6.3*

Note. # studies = number of independent studies; # ES = number of effect sizes; Z = difference in mean d with reference category; mean d = mean effect size (d); heterogeneity = within class heterogeneity (Z); Δ fit = difference with model without moderators (χ^2).

  [a] For trimmed effect sizes see Table 1.
  * $p < .05$.
  ** $p < .01$.
  *** $p < .001$.

the juvenile reported on peer factors, while the informant type did not moderate the effects on psychopathology or family factors. Furthermore, larger effect sizes were found for family factors if a correction for pretreatment measures was applied. No moderating effects were found for follow-up period.

Few sample characteristics had a moderating effect on the secondary outcomes. Significant effects on peer factors were only found for studies conducted with sex offenders, while the target population did not yield different effects on psychopathology or family factors. There were differences in effects on psychopathology for the average sample age: smaller, but still significant effects were found for studies with an average juvenile age of over 15 years. For family factors, only studies with juveniles with an average age under 15 showed improvement, while the average participants' age had no moderating effect on peer factors. Studies with a larger proportion of previously arrested juveniles showed more improvement in only psychopathology after MST. The proportion of males, proportion living with biological parents, proportion of migrants and the average number of previous arrests had no moderating effect on either of the secondary outcomes.

Of all coded treatment characteristics, the proportion of MST completers did not have a moderating effect on any of the outcome measures. The effects on both psychopathology and peer factors were larger in studies with longer treatment duration. Furthermore, effects on family functioning were only found if MST was compared with a single treatment type.

Finally, study characteristics had a moderating effect on all secondary outcomes. Efficacy studies yielded larger effects on psychopathology and family factors, while the research setting did not have a moderating effect on peer factors. Furthermore, lesser quality studies showed larger effect sizes for family factors and peer factors. No moderating effects were found for (in)dependence of authors, publication status, country and research design (randomized versus quasi-experimental).

3.4. Multiple regression analysis

The significant moderators for each outcome variable were entered into a multiple regression analysis to examine the unique influence of each moderator. Before the moderators were included in a multiple

Table 3
The overall mean effect size[a] and significant moderators for secondary outcomes.

Moderator variables	# studies	# ES	β_0, mean d	Z	β_1, RC	Z	Heterogeneity	Δ fit
Psychopathology overall	16	66	.268	5.3***			3.8***	
Sample characteristics								
Age							3.7***	4.4*
Under 15 (RC)	8	24	.400	5.0***				
15 years and over	8	42	.187	3.0**	−.213	2.1*		
% previously arrested	5	22	.195		.458	1.3	2.4*	−8.6**
Treatment characteristics								
Days of MST treatment	11	56	.231		.005	2.5*	3.3***	8.6**
Study characteristics								
Research setting							3.6***	13.1***
Efficacy (RC)	3	5	.900	5.1***				
Effectiveness	13	61	.220	4.7***	−.680	3.7***		
Family factors overall	12	70	.143	2.4**			2.1*	
Outcome characteristics								
Family factor type							2.1*	8.9**
Parent mental health (RC)	6	8	.321	3.7***				
Parenting	8	37	.144	2.3*	−.177	2.2*		
Family functioning	9	25	.076	1.1	−.245	3.0**		
Pretreatment correction							2.3*	5.7*
Not corrected (RC)	2	4	−.137	1.0				
Corrected	12	66	.184	2.7**	−.321	2.5		
Sample characteristics								
Age							2.0*	4.6*
Under 15 (RC)	6	34	.253	3.5***				
15 years and over	6	36	.031	.4	−.222	2.2*		
Treatment characteristics								
Control treatment							2.0*	5.0*
Single treatment (RC)	9	45	.066	1.1				
Multiple treatments	3	25	.328	3.3***	.262	2.3*		
Study characteristics								
Research setting							1.6	13.2***
Efficacy (RC)	3	16	.471	5.3***				
Effectiveness	9	54	.066	1.6	−.405	4.1***		
Study quality	12	70	.126		−.070	2.2*	2.1*	4.7*
Peer factors overall	8	29	.213	1.7*			2.0*	
Outcome characteristics								
Informant type							2.1*	5.7*
Other informants (RC)	4	12	.413	2.5**				
Self-report	6	17	.128	.9	−.286	2.5*		
Sample characteristics								
Target population							1.9	6.7*
Offenders (RC)	4	13	−.054	.3				
Conduct problems	2	4	.169	.9	.223	.9		
Sex offenders	2	12	.596	3.3***	.650	2.7**		
Treatment characteristics								
Days of MST treatment	7	26	.287		.005	2.5*	1.8	4.5*
Study characteristics								
Study quality	8	29	.327		−.220	2.6**	1.9	6.0*

Note. # studies = number of independent studies; # ES = number of effect sizes; Z = difference in mean d with reference category; mean d = mean effect size (d); heterogeneity = within class heterogeneity (Z); Δ fit = difference with model without moderators (χ^2).

  [a] For trimmed effect sizes see Table 1.
   * $p < .05$.
  ** $p < .01$.
 *** $p < .001$.

regression analysis, correlations between all moderators were examined. The occurrence of previous arrest was only reported for non-sex offenders and the number of previous arrests was only reported for published studies. The moderators about previous arrests were therefore excluded. For the primary outcome delinquency, treatment characteristics duration of MST treatment and proportion of MST completers were excluded because they consisted of too little effect sizes among all the other significant moderators.

Furthermore, post-treatment effects on secondary outcomes as moderators could not be included in the analysis with the other moderators, because the different moderators consisted of too little effect sizes to enter all moderators in one single analysis. Post-treatment effects on secondary outcomes were therefore entered into a separate multiple regression analysis. With the latter analysis, post-treatment effects on substance use, parental mental health and out-of-home placement

were available for little studies, and these moderators were excluded from the multiple regression analyses.

With the remaining moderators per outcome variable, multivariate analyses were conducted to examine the unique contribution of each moderator to the overall effect size (Tables 4 and 5). The fit for all outcome variables was improved.

For the ultimate outcome delinquency, larger effects were found when a composite of violent and non-violent delinquency was measured. Of the moderating sample characteristics, uniquely larger effects remained for sex offenders (and not for offenders in general), in studies with an average participants' age of under age 15, and when the sample consisted of a lower proportion of migrants. Larger effects were also found if the control treatment was a single treatment (not consisting of different kinds of treatment). Furthermore, of all study characteristics, only the country in which the study was conducted

Table 4
Results for the multivariate models for primary outcomes.

Moderator variables	β (SD)	Z
Delinquency		
Intercept	−.067 (.307)	.2
Outcome characteristics		
Delinquency type: general	.143 (.047)	3.0**
Sample characteristics		
Target population: offenders	.213 (.150)	1.4
Target population: sex offenders	.477 (.179)	2.7**
Age: under 15	.454 (.186)	2.4*
% migrants	−1.052 (.211)	4.9***
Treatment characteristics		
Control treatment: single treatment	.301 (.132)	2.3*
Study characteristics		
Authors: dependent	.015 (.117)	.1
Publication status: published	.120 (.218)	.6
Country: USA	.417 (.171)	2.4*
Research setting: efficacy	.146 (.201)	.7
Study quality	.056 (.039)	1.4
Δ fit	76.1***	
# ES	93	
Delinquency and post-treatment effects on secondary outcomes		
Intercept	.089 (.120)	.7
Post-treatment effects on secondary outcomes		
Psychopathology	−2.126 (1.133)	1.9
Externalizing behavior	−.203 (.216)	.9
Parenting	2.674 (1.207)	2.2*
Δ fit	34.6***	
# ES	57	

Note. Z = significance of moderator. Δ fit = difference with model without moderators.
* *p* < .05.
** *p* < .01.
*** *p* < .001.

Table 5
Results for the multivariate models for secondary outcomes.

Moderator variables	β (SD)	Z
Psychopathology		
Intercept	0.257 (0.058)	4.4***
Sample characteristics		
Age: under 15	−.171 (.135)	1.3
Treatment characteristics		
Days of MST treatment	.005 (.002)	2.5*
Study characteristics		
Research setting: efficacy	−1.132 (.337)	3.4***
Δ fit	20.6***	
# ES	56	
Family factors		
Intercept	−.262 (.261)	1.0
Outcome characteristics		
Family factor type: parental mental health	.152 (.086)	1.8
Family factor type: parenting	.054 (.053)	1.0
Pretreatment correction: corrected	.358 (.155)	2.3*
Sample characteristics		
Age: under 15	−.086 (.184)	.5
Treatment characteristics		
Control treatment: single treatment	−.217 (.188)	1.2
Study characteristics		
Research setting: efficacy	.112 (.143)	.8
Study quality	−.122 (.051)	2.4*
Δ fit	30.8***	
# ES	70	
Peer factors		
Intercept	.753 (.178)	4.2***
Outcome characteristics		
Informant type: other informants	.316 (.119)	2.7**
Sample characteristics		
Target population: sex offenders	−.353 (.288)	1.2
Treatment characteristics		
Days of MST treatment	.010 (.003)	3.3***
Study characteristics		
Study quality	−.181 (.056)	3.2**
Δ fit	24.6***	
# ES	26	

Note. Z = significance of moderator. Δ fit = difference with model without moderators.
* *p* < .05.
** *p* < .01.
*** *p* < .001.

remained to have a significant effect on delinquency outcomes: larger effects were found with studies conducted in the USA. Finally, of the analyzed post-treatment secondary outcome effects, only studies with a larger post-treatment effect on parenting showed a unique contribution to the primary outcome effect.

Within the secondary outcomes, larger effects were found when someone other than the juvenile reported on peer factors. Furthermore, if the outcome measure could be corrected for pretreatment measures, larger effects were shown on family factors. No unique moderating effects remained for sample characteristics on any of the secondary outcomes. Treatment characteristics had no unique moderating effect on family factors. Larger effects on psychopathology and peer factors, however, were found in studies where MST treatment was of longer duration. Moreover, larger effects were found within study characteristics: efficacy studies showed larger effects on psychopathology. Finally, higher quality studies showed smaller effects on both family as well as peer factors.

4. Discussion

The purpose of the present meta-analysis was to examine the effectiveness of MST and replicate the previous meta-analyses by Curtis et al. (2004) and Littell et al. (2005) with a larger body of studies and controlling for dependency of study results by using multi-level meta-analytic techniques. The larger number of studies made it possible to examine more moderators explaining the effectiveness of MST. Small but significant treatment effects were found on the primary outcome delinquency and on the secondary outcomes psychopathology, substance use, family factors, out-of-home placement and peer factors. The effect on family factors is consistent with the fact that MST aims to improve juvenile functioning by improving parenting and family functioning (e.g., Borduin, 1999; Henggeler, 2011). Moreover, the effects on other outcomes, like peer factors and substance use, underline the multisystemic focus of MST.

According to standard meta-analytic procedures, we conducted a trim and fill analysis to examine the degree to which publication bias might have had an effect on study results by adding 'unknown' file draw studies, which resulted in smaller effect sizes for all outcome variables, but still significant for psychopathology and family factors. MST outcome studies, however, are well tracked and documented by MST Services (see e.g. MST Services Inc., 2012), which makes it very unlikely that outcome studies have been overlooked. Moreover, 6 of the 20 studies in the delinquency effect size outcome were not peer reviewed and five of these were not published. Therefore, the probability seems very small that additional file draw studies with effect sizes of zero exist for delinquency outcomes. Thus, the trim and fill analyses yielded an unrealistic, overly conservative, and biased perspective of MST effects on delinquency. A similar conclusion holds for other outcome variables. We therefore argue that conclusions about the effectiveness of MST should be based on the results without trim and fill correction. Notably, results from moderator analyses were based on the non-filled data.

The current meta-analytic study could show under which conditions MST is the most effective. Therefore, moderator analyses were conducted for outcomes with heterogeneous effect sizes: delinquency, psychopathology, family factors and peer factors. When all significant moderators were entered into the multiple regression model, only few moderators remained significant.

Within the multiple regression model, of all study characteristics, only the country where the research has been conducted retained its unique moderating effect on delinquency. Larger effects were found for studies carried out in the USA, indicating that the transportation of MST to other countries may not be as effective as is generally assumed (see e.g. Ogden, Christensen, Sheidow, & Holth, 2008; Gustle, Hansson, Sundell, & Andree-Löfholm, 2008). Given that Van der Put et al. (2011) found no differences in risk factors for recidivism between Europe and The United States of America, no other treatment outcomes were expected between Europe and the USA, because MST exactly targets those risk factors. Differences in effects of MST could possibly be ascribed to differences in treatment implementation. Furthermore, the fact that country of research was the only study characteristic to retain a unique moderating effect indicates that the bias Littell (2005) had highlighted, might not have the effect she attributed to it. Littell namely reported that dependent researches, published studies, efficacy studies and lesser quality studies showed larger effect sizes, but given the fact that those elements lose their predictive power among other moderators, it seems time to reconsider these findings.

Larger effects were found in studies with an average participants' age of under 15 years and in studies with a larger proportion of Caucasian juveniles, indicating that MST is more effective with younger and non-ethnic minority juveniles. This is especially interesting given that MST studies that tested for moderating effects of ethnicity reported no moderating effects (Borduin et al., 1995; Henggeler, Clingempeel, Brondino, & Pickrel, 2002; Henggeler, Melton, & Smith, 1992; Schaeffer & Borduin, 2005), which is probably due to smaller sample sizes in individual studies (lack of statistical power). Furthermore, these findings are different than those reported by Wilson, Lipsey, and Soydan (2003), who found no differences in treatment effects between Caucasian white and minority juvenile delinquents in spite of the lack of cultural tailoring of treatments. They did state, however, that more research was needed to specifically assess the treatment outcome differences between ethnic minority and Caucasian white (ethnic majority) youth. The findings in the current study endorse that statement once more.

As the target population (i.e., sex-offenders) uniquely predicted study outcomes, variation in the magnitude of effect sizes may depend on the particular population under study. Moreover, as significant effects were only found when a general delinquency outcome measure was used and not when delinquency was measured with a distinction between violent and/or non-violent offenses, type of delinquency explains variation in effect sizes too.

Furthermore, smaller effects of MST were shown if the control treatment was a mix of different kinds of treatments, which may be referred to as multimodal treatment. As Andrews et al. (2006) already emphasized the importance of multimodal treatment approaches, it is well possible that the effects of MST can partly be explained by the multimodal approach of the intervention, because smaller effects were found when MST was compared to a combination of treatments showing some similarity to the multimodal approach.

Finally, in contrast with the MST theory of change (Henggeler, 2011), only studies with larger post-treatment effects on parenting and not on family functioning showed a unique moderating effect on delinquency. This could indicate that mainly short-term better parenting skills (which MST targets) prevent long-term delinquency and recidivism. Interestingly, studies with better parenting did not show improved short-term family functioning (e.g., cohesion, emotional connectedness, quality of family relationships), and studies with better family functioning did not show differences in long-term recidivism too. This is in line with findings by Loeber and Stouthamer-Loeber (1986), who found that parenting practices, such as supervision, had large effects on conduct problems and delinquency. The emphasis of MST therefore should be mainly the improvement of parenting skills, even though this did not have the hypothesized effects on family functioning.

For the other outcomes, mostly study and treatment characteristics had a unique moderating effect. Psychopathology effects were larger for efficacy studies and when MST treatment was of longer duration. The larger effects for efficacy studies indicate that MST has larger effects under optimal (well controlled) treatment conditions. This finding differs from that of Shadish, Matt, Navarro, and Phillips (2000), who in a meta-analysis of psychological therapies found that the difference between efficacy and effectiveness was an artifact of selective assignment in non-randomized trials. Family factors showed more effects when corrected for pretreatment differences and with lesser quality studies. Peer factors showed more improvement with longer MST duration, with lesser quality studies and when not measured by means of self-report. This latter difference between informants is in contrast to the non-existent difference between self-reported and otherwise reported delinquency (mostly through archival data), which is promising for the use of self-reported delinquency measures when examining intervention effects of treatments targeting delinquency.

Several characteristics had a (non-unique) moderating effect on more than one outcome measure. First, in studies where MST was applied for more weeks, larger effects were found on delinquency, psychopathology and peer factors. This finding is difficult to interpret, since Lipsey (2009) found no overall effect for treatment duration in his meta-analysis of effective interventions with juvenile delinquents. It is possible that longer treatment duration is related to the severity of pre-treatment problems, and therefore a higher pre-treatment recidivism risk. Longer, more intensive treatment would then fit the RNR-model through its alignment with the recidivism risk, thereby enhancing effectiveness (i.e., the higher the risk, the more intensive treatment needs to be; e.g., Andrews & Bonta, 2010; Andrews et al., 1990, 2006). Furthermore, larger effects were found on delinquency and psychopathology in studies where more juveniles had a more extensive offending history. These findings are in line with those by Lipsey (2009), who found larger treatment effects in juveniles with higher levels of delinquency risk for a variety of juvenile delinquent treatments.

The findings of the current meta-analytic study differ from previous meta-analyses by Curtis et al. (2004) and Littell et al. (2005). The effect sizes in the present study were smaller than those reported in the Curtis et al. study (i.e., small versus moderate) and larger than those reported by Littell et al. (i.e., no effect versus small effects). This is probably caused by the fact that non-published, non-randomized and more recent studies were added to the analyses in the present study, and child abuse and neglect and psychiatric hospitalization studies were excluded. Larger average effect sizes for family relationships over individual adjustment (Curtis et al., 2004) were not replicated, whereas the effect on peer relationships diminished altogether.

Contrary to previous meta-analyses (Curtis et al., 2004; Littell et al., 2005), no moderating effects were found for research design (i.e., randomized versus non-randomized control group). Less strictly assigned treatment and control groups did not yield larger effects, thereby validating the use of non-randomized studies in the present meta-analysis, and confirming its generalizability (see Shadish et al., 2000).

Additionally, a number of moderators were highly confounded. For instance, the occurrence of previous arrests was not reported for sex offenders, and only published studies reported about the number of previous arrests. Furthermore, although Littell (2006) suggested that some large effects may be found due to conflict of interest of the involved researchers (i.e., researchers having personal or financial stakes in MST), (in)dependence of researchers is linked to other study characteristics. MST developers examined more offenders than juveniles with conduct problems, conducted studies with a lower average age of juveniles and with larger proportions of males, more often conducted efficacy studies, were more often involved with earlier studies, and reported about longer follow-up periods than independent researchers did (see Table 1). Larger effects can thus not simply be attributed to the dependency of the researchers, especially seeing that those factors did not

have a unique moderating effect. Still, findings of the current meta-analytic study were consistent with those of Petrosino and Soydan (2005) in that dependent researchers reported larger effect sizes than independent researchers.

Unfortunately, treatment adherence could not be included in the moderator analyses. Precisely this meta-analysis would have provided the opportunity to assess the assumption that treatment effects are highly dependent of treatment adherence (e.g. Henggeler, Schoenwald, Borduin, & Swenson, 2006). Although treatment adherence measures were more often applied in the more recent studies, the way this adherence was reported proved to be inconsistent, and adherence outcomes could not be coded for a sufficient amount of studies. Some studies reported an actual treatment adherence outcome (without referential information), other studies reported treatment adherence categorically – i.e., low, moderate or high treatment adherence – if adherence had a moderating effect on the outcome variables (without outcome or category of treatment adherence), if adherence measures were administered, or if a measure of treatment adherence was used or not.

There are some limitations of this meta-analytic study that need to be mentioned. First, because of the inclusion of unpublished and non-randomized studies, some studies were of weak study design and therefore had questionable validity. Second, several moderators, mostly study design and publication characteristics, were excluded from the multiple regression analyses, because these were highly correlated with other moderators (multicollinearity). Moreover, the analysis of peer factors and trim and fill analyses of substance use and peer factors were somewhat underpowered, and should therefore be interpreted with caution. Furthermore, the current meta-analysis did not allow examination of more differentiated moderator analyses that may be found in some of the primary studies, such as testing differences between various subgroups of ethnic minority youth. Finally, although the current meta-analytic design allows analyses with more statistical power over individually (potentially) underpowered studies (Lipsey & Wilson, 2001), it does not take into account the (non-)significance of results from individual studies, even though including the standard error in the equation provides for a correction.

In spite of these limitations, this meta-analytic study provides the most up-to-date information on the effectiveness of Multisystemic Therapy, which has been spread over the world since the previous meta-analyses. The current study provides elaborate information about the conditions under which MST could be (more) effective. The present study is the first to control for the multilevel structure of the MST outcomes, and the first to control for multicollinearity in the moderators, shedding new light on both the Curtis et al. (2004) meta-analysis and the Littell (2005) review.

In line with findings of Andrews et al. (2006), especially the multi-modal approach of MST appears effective. It seems that with MST, only the post-treatment improvement of parenting skills and not improvements in family functioning as a whole keep their influence on the long term. A recent study by Van der Put, Deković, Stams, Hoeve, and Van der Laan (2012), however, shows that family risk factors lose their predictive value of recidivism when juveniles get older. Possibly, improved parenting skills better prepare parents for this transition. Furthermore, according to Van der Put et al. (2012) with older juveniles, peer and school factors are more predictive of recidivism. As the current study shows no moderating effect of post-treatment peer factor improvements, the effectiveness of MST could therefore presumably be improved by targeting school and peer related factors more with older juveniles.

Appendix A. Sample, treatment and study characteristics of the studies included in the meta-analysis

| Authors | Year | Sample characteristics | | | | | | | | Treatment characteristics | | | Study characteristics | | | | | |
		N	Pop.	Age	% male	% biol. par.	% migrants	% prev. arr.	No. prev. arr.	Ctrl.	% compl.	Days MST	Indep. authors	Pub. status	Country	Design	Setting	Quality
Asscher	2013	256	CP	16.0	.73		.45	.71	2.23	Mix	.99	174.46	Yes	Yes	NED	RCT	Effect	16
Barnoski	2004	145	OF	15.2	.77					Mix			Yes	No	USA	RCT	Effect	10
Borduin	2009	48	SO	14.0	.96	.98	.29		4.33	Single	1.00	215.6	No	Yes	USA	RCT	Efficacy	12
Borduin	1995	176	OF	14.8	.68	.90	.30	1.00	4.20	Single	.83		No	Yes	USA	RCT	Efficacy	15
Borduin	1990	16	SO	14.0	1.00	1.00	.37			Single	.62	137.25	No	Yes	USA	RCT	Efficacy	11
Butler	2011	108	OF	15.1	.82		.64		2.03	Mix		142.80	Yes	Yes	GBR	RCT	Effect	16
Cunningham	2002	409	OF	14.7	.74			.67		Mix	.81		Yes	No	CAN	RCT	Effect	12
Henggeler	2002b	118	OF	15.7	.79	.90	.53		2.90	Single	.98	130.00	No	Yes	USA	RCT	Effect	15
Henggeler	1997	155	OF	15.2	.82		.81		3.07	Mix		119.60	No	Yes	USA	RCT	Effect	15
Henggeler	1993	84	OF	15.2	.77	.74	.58		3.50	Mix		93.80	No	Yes	USA	RCT	Effect	15
Henggeler	1986	156	OF	14.8	.84				2.10	Mix	.75		No	Yes	USA	Quasi	Efficacy	12
Letourneau	2009	131	SO	14.6	.98	.79	.85			Single	.91	216.55	No	Yes	USA	RCT	Effect	16
Löfholm	2009	156	CP	15.0	.61		.47	.67		Mix	.73	145.80	Yes	Yes	SWE	RCT	Effect	16
Mayfield	2011	252	CP	14.0	.60		.17			Mix			Yes	No	USA	Quasi	Effect	13
Mitchell-Hertzfeld	2008 pilot	269	OF	16.1	.73		.94	.47		Mix	.75	114.94	Yes	No	USA	Quasi	Effect	10
Mitchell-Hertzfeld	2008 post-pilot	629	OF	16.3	.80		.96	.46		Mix	.72	101.92	Yes	No	USA	Quasi	Effect	10
Ogden	2006	75	CP	15.1	.64	.81	.01			Mix	.93	170.10	Yes	Yes	NOR	RCT	Effect	16
Painter	2009	174	CP	11.9	.47		.46	.00	.00	Single		152.50	Yes	Yes	USA	Quasi	Effect	12
Rowland	2005	55	CP	14.5	.58	.90	.90		7.50	Mix	.96		No	Yes	USA	RCT	Effect	16
Stambaugh	2007	265	CP	12.1	.74					Mix		167.75	Yes	Yes	USA	Quasi	Effect	13
Timmons-Mitchell	2006	93	OF	15.1	.78	.95	.22		6.87	Mix	.89	144.84	Yes	Yes	USA	RCT	Effect	14
Timmons-Mitchell	2002	296	OF		.64		.57				.57		Yes	No	USA	Quasi	Effect	10

Note. N = number of participants; pop. = target population; age = average age of participants; % male = proportion of males; % biol. par. = proportion living with biological parents; % migrants = proportion non-Caucasian; % prev. arr. = proportion previously arrested; no. prev. arr. = average number of previous arrests; ctrl. = control treatment composition; % compl. = proportion MST treatment completers; days MST = average duration of MST in days; indep. authors = independent authors yes/no; pub. status = published in a peer reviewed journal yes/no; country = country where the research was conducted; design = research design randomized versus quasi-experimental; setting = research setting effectiveness versus efficacy; quality = study quality assessed using the Quality Assessment Tool for Quantitative Studies (Thomas et al., 2004); CP = conduct problems; SO = sex offenders; OF = offenders; mix = control treatment was a mix of several (possible) treatments; single = control treatment comprised of one particular treatment; NED = The Netherlands, USA = United States of America; GBR = Great Britain; CAN = Canada; SWE = Sweden; NOR = Norway; effect = effectiveness.

Appendix B. Outcome characteristics and post-treatment outcome effects for primary outcome delinquency

Authors	Year	Delinquency # of effect sizes							Post-treatment outcomes d									
		Follow-up (years)	d (M)	Corr	Viol	Nonviol	Any	Selfrep	Psych	Inter	Exter	Skills	Subst	Parent	Fam	ParMH	OOH	Peers
Asscher	2012	0, 2, 3.06	8 (.017)	8	2	2	4	2	.310		.110	−.020		.190	.193			.160
Barnoski	2004	1.5	2 (−.185)	0	1	0	2	0										
Borduin	2009	8.95	5 (.814)	2	2	2	1	2	1.35		1.65				.920	1.000	.650	1.270
Borduin	1995	3.95, 13.7, 21.9	16 (.500)	0	5	5	8	0	.240		1.840			.413	.490	.140	.260	−.260
Borduin	1990	3.09	1 (1.370)	0	0	0	1	0										
Butler	2011	0, .5, 1, 1.5	27 (.415)	27	8	8	12	3	.235	−.060	.047	−.066		.180	−.040		.210	.090
Cunningham	2002	.5, 1, 2, 3	5 (−.074)	0	0	0	4	0		.430	.350	.000		.130	.210	.680	−.050	
Henggeler	2002b	0, .5, 1.5, 4	9 (.196)	3	2	2	4	5		−.120	−.160		.400				.120	
Henggeler	1997	0, 1.7	2 (.115)	0	0	0	0	1	.510		−.090			.260	−.171	.240	.440	−.056
Henggeler	1993	0, 1.9	3 (.476)	1	0	0	2	1			−.030	−.250			.185	−.290	.630	.033
Letourneau	2009	0, .5	2 (.490)	2	0	0	0	2		.400	.470		.460	.105			.270	.116
Löfholm	2009	1, 1.5	4 (−.042)	4	0	0	2	2	−.005	.000	−.020	−.017	.000	−.165		.160	−.010	.360
Mayfield	2011	.5	2 (1.150)	2	1	0	2	0										
Mitchell-Hertzfeld pilot	2008	0, .5, 1.5, 2.5	4 (−.127)	4	0	0	4	0										
Mitchell-Hertzfeld post-pilot	2008	0, .5, 1.5	3 (−.113)	3	0	0	3	0										
Ogden	2006	1.5	1 (.140)	0	0	0	0	1	.663	.520		−.390			−.010			.780
Painter	2009	0	1 (.350)	1	0	0	1	0	.140									
Rowland	2005	0	3 (.500)	2	0	0	0	2		.340	.413		.210		.075		.910	
Timmons-Mitchell	2006	1.5	2 (.525)	0	0	0	1	0	.860			.143	.470		1.070			
Timmons-Mitchell	2002	.5	1 (−.180)	0	0	0	1	0										

Note. d (M) = number of effect sizes (mean); corr = number of effect sizes corrected for pretreatment measure; viol = number of effect sizes for violent offenses; nonviol = number of effect sizes for non-violent offenses; any = number of effect sizes for any (re)offense, not number of (re)offenses; selfrep = number of effect sizes from self-report measures; psych = psychopathology; inter = internalizing; exter = externalizing; skills = skills and cognitions; subst = substance use; parent = parenting skills; fam = family functioning; ParMH = parental mental health; OOH = out-of-home placement; peers = peer factors.

Appendix C. Outcome characteristics for secondary outcomes

Table A
Outcome characteristics for psychopathology and skills and cognitions.

Authors	Year	Psychopathology # of effect sizes								Skills and cognitions # of effect sizes					
		Follow-up (years)	d (M)	Corr	Exter	Inter	Selfrep	Parent rep	CBCL	Follow-up (years)	d (M)	Corr	Social skills	Cognitions	Selfrep
Asscher	2012	0	4 (.210)	4	4	0	1	3	2	0	3 (−.020)	3	0	1	3
Borduin	2009	0	2 (1.500)	2	1	0	1	1	0						
Borduin	1995	0	2 (.945)	2	1	0	1	1	0						
Butler	2011	0	8 (.067)	8	3	2	4	4	6	0	3 (−.066)	3	0	2	3
Cunningham	2002	0	2 (.390)	0	1	1	1	1	2	0	4 (.000)	0	1	2	4
Henggeler	2002b	1.5	2 (−.140)	0	1	1	2	0	2						
Henggeler	1997	0	2 (.210)	2	1	0	1	1	0						
Henggeler	1993	0	1 (−.030)	1	1	0	0	1	0	0	1 (−.250)	1	1	0	1
Henggeler	1986	0	1 (.250)	0	1	0	0	1	0						
Letourneau	2009	0, .5	8 (.341)	8	4	4	4	4	8						
Löfholm	2009	.08, 1.5	12 (−.009)	12	4	4	6	6	11	.08, 1.5	9 (−.061)	9	7	2	7
Ogden	2006	0, 1.5	11 (.646)	11	4	4	3	3	9	0	1 (−.390)	1	1	0	0
Painter	2009	0	2 (.140)	2	0	0	0	2	0						
Rowland	2005	0	5 (.384)	5	3	2	3	2	4						
Stambaugh	2007	1	2 (.095)	0	0	0	0	1	1						
Timmons-Mitchell	2006	0, .5	2 (.895)	2	0	0	0	0	0	0, .5	8 (.408)	8	0	0	0

Note. d (M) = number of effect sizes (mean); corr = number of effect sizes corrected for pretreatment measure; alcohol = number of effect sizes from alcohol measures; soft drugs = number of effect sizes from soft drug measures; selfrep = number of effect sizes from self-report measures; fam. func = number of effect sizes for family functioning; parent. skill = number of effect sizes for parenting skills; parent rep = number of effect sizes from parent-report measures.

Table B
Outcome characteristics for substance use and family factors.

Authors	Year	Substance use # of effect sizes						Family factors # of effect sizes						
		Follow-up (years)	d (M)	Corr	Alcohol	Soft drugs	Selfrep	Follow-up (years)	d (M)	Corr	Fam. func.	Parent. skill	Selfrep	Parent rep.
Asscher	2012							0	10 (.191)	10	3	7	3	4
Borduin	2009							0	4 (.960)	4	2	0	0	2
Borduin	1995							0	9 (.400)	9	2	6	0	1
Butler	2011							0	6 (.033)	6	4	2	3	3
Cunningham	2002							0	4 (.202)	1	2	1	0	2
Henggeler	2002b	0, .5, 4	8 (.191)	4	0	2	6							
Henggeler	1997							0	9 (−.030)	9	6	2	4	5
Henggeler	1993							0	3 (.026)	3	2	0	0	1
Henggeler	1986							0	3 (.190)	2	0	3	0	0
Letourneau	2009	0, .5	2 (.610)	2	0	0	2	0, 0.5	12 (.083)	12	0	12	6	6
Löfholm	2009	.08, 1.5	5 (.018)	5	3	0	5	.08, .5	6 (−.018)	6	0	4	2	4
Ogden	2006							0	2 (−.010)	2	2	0	0	0
Rowland	2005	0	1 (.210)	1	0	0	1	0	2 (.075)	2	2	0	0	0
Timmons-Mitchell	2006	0, .5	2 (.475)	2	0	0	0							

Note. d (M) = number of effect sizes (mean); corr = number of effect sizes corrected for pretreatment measure; exter = number of effect sizes for externalizing behavior; inter = number of effect sizes for internalizing behavior; selfrep = number of effect sizes from self-report measures; parent rep = number of effect sizes from parent-report measures; CBCL = number of effect sizes from CBCL outcome measures; social skills = number of effect sizes for social skills; cognitions = number of effect sizes for cognitions.

Table C
Outcome characteristics for out-of-home placement and peer factors.

Authors	Year	Out-of-home placement # of effect sizes					Peer factors # of effect sizes					
		FU (years)	d (M)	Corr	Incarc.	Any	FU (years)	d (M)	Corr	Delinq	Relation	Selfrep
Asscher	2012						0	2 (.160)	2	1	1	2
Borduin	2009	8.95	1 (.650)	0	1	0	0	6 (1.270)	6	3	3	3
Borduin	1995	13.7	1 (.260)	0	1	1	0	3 (−.260)	3	0	3	0
Butler	2011	1.5	1 (.210)	1	0	1	0	1 (.090)	1	1	0	1
Cunningham	2002	3	2 (−.050)	0	0	1						
Henggeler	2002	.5	1 (.120)	0	0	1						
Henggeler	1997	1.7	1 (.440)	0	1	0	0	6 (−.056)	6	1	5	3
Henggeler	1993	0	1 (.630)	0	1	1	0	3 (.033)	3	0	3	0
Letourneau	2009	.5	1 (.270)	1	0	1	0, .5	6 (.301)	6	3	3	6
Löfholm	2009	.08, 1.5	4 (.117)	0	0	2	.08, .5	2 (.175)	2	2	0	2
Mayfield	2011	6	1 (.000)	1	0	1						
Mitchell-Hertzfeld	2008 pilot	0	1 (.200)	1	0	1						
Mitchell-Hertzfeld	2008 post-pilot	0	1 (.110)	1	0	1						
Ogden	2006	0, 1.5	2 (.685)	1	0	2						
Rowland	2005	0	1 (.910)	0	0	0						

Note. d (M) = number of effect sizes (mean); corr = number of effect sizes corrected for pretreatment measure; incarc. = number of effect sizes for incarceration measures; any = number of effect sizes for any out-of-home placement, not number of days in out-of-home placement; delinq = number of effect sizes from peer delinquency measures; relation = number of effect sizes from peer relations measures; selfrep = number of effect sizes from self-report measures.

References[1]

Andrews, D., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*, 39–55, http://dx.doi.org/10.1037/a0018362.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*(1), 19–52, http://dx.doi.org/10.1177/0093854890017001004.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*(1), 7–27, http://dx.doi.org/10.1177/0011128705281756.

Bax, L. (2011). MIX 2.0. Professional software for meta-analysis in Excel: Version 2.0.1.4. BiostatXL. Retrieved from. http://www.meta-analysis-made-easy.com

Borduin, C. M. (1999). Multisystemic treatment of criminality and violence in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(3), 242–249.

*Borduin, C. M., Henggeler, S. W., Blaske, D.M., & Stein, R. J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 996*(3), 105–113.

*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 & Clinical Psychology, 63*(4), 569–578.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

Brunk, M., Henggeler, S. W., & Whelan, J. P. (1987). Comparison of multisystemic therapy and parent training in the brief treatment of child abuse and neglect. *Journal of Consulting & Clinical Psychology, 55*(2), 171–178.

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*, 199–244.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

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

Duval, S., & Tweedie, R. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89–99.

Duval, S. J., & Tweedie, R. L. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

Ellis, D. A., Naar-King, S., Frey, M., Templin, T., Rowland, M., & Greger, N. (2004). Use of multisystemic therapy to improve regimen adherence among adolescents with type 1 diabetes in poor metabolic control: A pilot investigation. *Journal of Clinical Psychology in Medical Settings, 11*(4), 315–324.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151–175.

Glisson, C., Schoenwald, S. K., Hemmelgarn, A., Green, P., Dukes, D., Armstrong, K. S., et al. (2010). Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *Journal of Consulting & Clinical Psychology, 78*(4), 537–550.

Goense, P., Boendermaker, L., Van Yperen, T., Stams, G. J. J. M., & Van Laar, J. (2014). Implementation of treatment integrity procedures: An analysis of outcome studies of youth interventions targeting externalizing behavioral problems. *Zeitschrift für Psychologie, 222*(1), 12–21, http://dx.doi.org/10.1027/2151-2604/a000161.

Gustle, L. -H., Hansson, K., Sundell, K., & Andree-Löfholm, C. (2008). Implementation of evidence-based models in social work practice: Practitioners' perspectives on an MST trial in Sweden. *Journal of Child & Adolescent Substance Abuse, 17*(3), 111–125, http://dx.doi.org/10.1080/15470650802071713.

Henggeler, S. W. (2011). Efficacy studies to large-scale transport: The development and validation of multisystemic therapy programs. *Annual Review of Clinical Psychology, 7*, 351–381, http://dx.doi.org/10.1146/annurev-clinpsy-032210-104615.

*Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(7), 868–874, http://dx.doi.org/10.1097/00004583-200207000-00021.

Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006). Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting & Clinical Psychology, 74*(1), 42–54.

**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 & Clinical Psychology, 60*(6), 953–961.

**Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and-dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research, 1*(3), 171–184.

*Henggeler, S. W., Rodick, J.D., Borduin, C. M., Hanson, C. L., Watson, S. M., & Urey, J. R. (1986). Multisystemic treatment of juvenile offenders: Effects on adolescent behavior and family interaction. *Developmental Psychology, 22*(1), 132.

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

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M.D., & Cunningham, P. B. (2009). *Multisystemic treatment of antisocial behavior in children and adolescents: Treatment manuals for practitioners* (2nd ed.). New York, NY: The Guilford Press.

Henggeler, S. K., Schoenwald, S. K., Borduin, C. M., & Swenson, C. C. (2006). Methodological critique and meta-analysis as Trojan horse. *Children and Youth Services Review, 28*(4), 447–457, http://dx.doi.org/10.1016/j.childyouth.2005.07.001.

Henggeler, S. W., Schoenwald, S. K., Rowland, M.D., & Cunningham, P. B. (2002). *Serious emotional disturbance in children and adolescents: Multisystemic therapy.* New York, NY: The Guilford Press.

Hoenig, J. M., & Heisey, D.M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*(1), 19–24.

Hopewell, S., McDonald, S., Clarke, M., & Egger, M. (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*(2), http://dx.doi.org/10.1002/14651858.MR000010.pub3 (Art. No.: MR000010).

Hox, J. (2002). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah: Lawrence Erlbaum Associates Publishers.

Hox, J. (2010). *Multilevel analysis: Techniques and applications.* New York: Routledge, Taylor & Francis Group.

Lipsey, M. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents. In J. MacGuire (Ed.), *What works: Reducing offending* (pp. 63–78). Chichester, UK: John and Wiley and Sons.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders, 4*(2), 124–147, http://dx.doi.org/10.1080/15564880802612573.

Lipsey, M. W., & Wilson, D. B. (2001). In L. Bickman, & D. J. Rog (Eds.), *Practical meta-analysis.* Thousand Oaks, CA: SAGE Publications, Incorporated.

Littell, J. H. (2005). Lessons from a systematic review of effects of multisystemic therapy. *Children and Youth Services Review, 27*(4), 445–463, http://dx.doi.org/10.1016/j.childyouth.2004.11.009.

Littell, J. H. (2006). The case for multisystemic therapy: Evidence or orthodoxy? *Children and Youth Services Review, 28*(4), 458–472, http://dx.doi.org/10.1016/j.childyouth.2005.07.002.

Littell, J. H., Campbell, M., Green, S., & Toews, B. (2005). Multisystemic therapy for social, emotional, and behavioral problems in youth aged 10–17. *Cochrane Database of Systematic Reviews, 2005*, 4, http://dx.doi.org/10.1002/14651858.CD004797.pub4.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence.* Mahwah, NJ: Lawrence Erlbaum Associates.

Loeber, R., & Stouthamer-Loeber, M. (1986). Family factors as correlates and predictors of juvenile conduct problems and delinquency. *Crime and Justice, 7*, 29–149.

McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet, 356*(9237), 1228–1231, http://dx.doi.org/10.1016/S0140-6736(00)02786-0.

*Mitchell-Herzfeld, S., Shady, T. A., Mayo, J., Han Kim, D. H., Marsh, K., Dorabawila, V., et al. (2008). *Effects of multisystemic therapy (MST) on recidivism among juvenile delinquents in New York State.* New York: New York State & Office of children and family services.

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A.R., Moher, M., et al. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet, 352*(9128), 609–613, http://dx.doi.org/10.1016/S0140-6736(98)01085-X.

MST Services Inc. (2010). MST multisystemic therapy: Breaking the cycle of criminal behavior by keeping teens at home, in school and out of trouble. Retrieved March 1, 2013, from. http://mstservices.com/

MST Services Inc. (2012). *Multisystemic therapy® (MST®): Research at a glance.* (Research report). Mount Pleasant: MST Services Inc (Retrieved from http://mstservices.com/outcomestudies.pdf).

Mullen, B. (1989). *Advanced BASIC meta-analysis, 1989.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Naar-King, S., Ellis, D., Kolmodin, K., Cunningham, P., Jen, K. C., Saelens, B., et al. (2009). A randomized pilot study of multisystemic therapy targeting obesity in African-American adolescents. *Journal of Adolescent Health, 45*(4), 417–419.

Ogden, T., Christensen, B., Sheidow, A. J., & Holth, P. (2008). Bridging the gap between science and practice: The effective nationwide transport of MST programs in Norway. *Journal of Child & Adolescent Substance Abuse, 17*(3), 93–109, http://dx.doi.org/10.1080/15470650802071689.

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators of criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1*, 435–450.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183.

**Schaeffer, C. M., & Borduin, C. M. (2005). Long-term follow-up to a randomized clinical trial of multisystemic therapy with serious and violent juvenile offenders. *Journal of Consulting & Clinical Psychology, 73*(3), 445–453.

Shadish, W. R., Matt, G. E., Navarro, A.M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126*(4), 512–529.

Shrier, I., Boivin, J., Steele, R. J., Platt, R. W., Furlan, A., Kakuma, R., et al. (2007). Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *American Journal of Epidemiology, 166*(10), 1203–1209, http://dx.doi.org/10.1093/aje/kwm189.

Snijders, T. A. B., & Bosker, R. (1999). *Multilevel analyses. An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435–452) (2nd ed.). New York, NY, US: Russel Sage Foundation.

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing

---

[1] References marked with one or two asterisks indicate primary studies or secondary analysis studies, respectively, in the meta-analysis.

interventions. *Worldviews on Evidence-Based Nursing*, *1*(3), 176–184, http://dx.doi.org/10.1111/j.1524-475X.2004.04006.x.

Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*(5), 765–790, http://dx.doi.org/10.1177/0013164402251027.

Van der Put, C. E., Deković, M., Stams, G. J. J. M., Hoeve, M., & Van der Laan, P. H. (2012). Het belang van vroegtijdig ingrijpen bij jeugdcriminaliteit. *Kind & Adolescent*, *33*(1), 2–22, http://dx.doi.org/10.1007/s12453-012-0001-9.

Van der Put, C. E., Deković, M., Stams, G. J. J. M., Van der Laan, P. H., Hoeve, M., & van Amelsfort, L. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal Justice and Behavior*, *38*, 248–262.

Wilson, D. B. (2010). Practical meta-analysis effect size calculator. Retrieved, 2012/2013, from. http://gemini.gmu.edu/cebcp/EffectSizeCalculator/

Wilson, S. J., Lipsey, M. W., & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. *Research on Social Work Practice*, *13*(1), 3–26, http://dx.doi.org/10.1177/1049731502238754.