

Sustainability of the effects of multisystemic therapy for juvenile delinquents in The Netherlands: effects on delinquency and recidivism

Jessica J. Asscher · Maja Deković · Willeke Manders ·
Peter H. van der Laan · Pier J. M. Prins · Sander van Arum ·
Dutch MST Cost-Effectiveness Study Group

© Springer Science+Business Media Dordrecht 2014

Abstract

Objectives The present study focused on the sustainability of the effects of Multisystemic Therapy (MST) on delinquency and recidivism.

Methods A sample of 256 juveniles with severe and persistent antisocial behavior were randomly assigned to MST (147) and Treatment As Usual (TAU) (109) condition. Pre-test assessment took place before the start of MST/control group treatment. Post-test assessment took place at 6 months after termination of the program. Delinquency (parent and adolescent reported) was assessed 6 months after termination of the treatment. Official judicial data were collected to assess recidivism, with a mean length of follow-up of 3.06 years. ANCOVAs and survival analyses were used to test the effectiveness of MST.

Results The multi-informant data showed that MST is effective in diminishing delinquent behavior as reported by adolescents and parents, with d 's larger than at post-test

Members: D.E.M.C. Jansen, K.M. Vermeulen, E. Buskens, E.J. Knorth, and S.A. Reijneveld

J. J. Asscher (✉)

Forensic Child and Youth Care Sciences, University of Amsterdam, Amsterdam, Netherlands
e-mail: J.J.Asscher@uva.nl

M. Deković

Child and Adolescent Studies, Utrecht University, Utrecht, Netherlands

W. Manders

University of Applied Sciences Arnhem/Nijmegen, Nijmegen, Netherlands

P. H. van der Laan

Free University Amsterdam/NSCR, Amsterdam, Netherlands

P. J. M. Prins

Developmental Psychology, University of Amsterdam, Amsterdam, Netherlands

S. van Arum

De Waag Forensic Outpatient Treatment Center, Utrecht, Netherlands
Developmental Psychology, University of Amsterdam, Amsterdam, Netherlands

Dutch MST Cost-Effectiveness Study Group
Groningen University, Groningen, The Netherlands

assessment immediately after ending of the intervention. The official judicial data, however, suggest that there are no differences between MST and TAU in recidivism. Few and inconsistent moderator effects were found.

Conclusions According to parent and adolescent reports, the beneficial effects of MST were sustained at the follow-up. This was not supported by official data. These results stress the importance of using multi-informant data on delinquency, as each source of information has its own advantages and disadvantages.

Keywords Multisystemic Therapy (MST) · Long-term effects · Official recidivism data

Because of the great personal and societal costs associated with juvenile delinquency, many prevention and intervention programs have been developed during the past decades that aim to decrease juvenile delinquency. Based on the Risk-Needs-Responsivity principle (Andrews and Bonta 2010), these programs target dynamic risk and protective factors, aiming to decrease the risk of the development of persistent delinquency. One home-based program, aiming to decrease delinquent behavior and prevent recidivism is Multisystemic Therapy (MST). Developed by Henggeler and colleagues (e.g., Henggeler et al. 1992; Henggeler 2011), MST focuses on diminishing the risk factors and increasing protective factors in the various systems in which juveniles function, e.g., family, school, peer group, and neighborhood.

Studies in the United States have mainly found positive results in reducing delinquent behavior (Borduin et al. 1995; Henggeler et al. 1992, 1997; Rowland et al. 2005; Timmons-Mitchell et al. 2006). Although in some cases the transportation to other countries and continents seems successful, for example Norway (Ogden and Halliday-Boykins 2004) and the United Kingdom (Butler et al. 2011), Cunningham (2002) found no positive effects of MST in Canada, and neither did Sundell et al. (2008) in Sweden. Moreover, in 2005, Littell et al. (2005) critical review was published, which questioned the randomization procedures and the reporting of findings of the US studies.

In sum, although MST is a promising intervention, which has been extensively studied, the evidence is still not conclusively positive. Despite the broad evidence from the US for the effectiveness of Multisystemic Therapy, the empirical evidence for MST in Europe is less robust. This evidence is especially important now that many European countries focus on ‘evidence-based’ interventions for juvenile delinquents. The evidence-base of MST lies mainly in the US, and has been provided for the largest part by studies in which the program developers were involved. A meta-analysis by Petrosino and Soydan (2005) indicated that effect sizes in evaluations of criminal justice programs conducted by developers of the programs tend to be larger than those conducted by independent researchers, once again stressing with this the importance of independent replication of the results.

There are, to our knowledge, only two European studies so far that have examined long-term follow-up effects. Both of these studies have some limitations. Butler et al. (2011), who reported recidivism findings at 18 months, with results in favor of MST (with 36 % of control group that recidivated vs. 8 % of the MST group, $d=1.03$), have a relatively small sample ($n=53$ MST; $n=48$ control condition). Ogden and Amlund Hagen (2006) conducted a follow-up at 2 years after intake, again with very small subgroups (46 MST vs. 29 TAU) and, moreover, did not use official recidivism data. Consequently, there is a need of

an independent European effectiveness study with outcomes assessed after termination of MST, including official data, and with sufficient power and a relatively large sample.

Recently, a Dutch randomized controlled trial (Asscher et al. 2012) was published, reporting positive outcomes immediately after treatment, for externalizing behavior and delinquency (d 's varied from 0.25 to 0.36), but not for violent delinquency ($d=0.03$). Moreover, mainly positive results were found for parenting, contacts with prosocial peers and some positive results for cognitions (d 's varied from -0.26 to 0.47). Although these results were promising, given that the ultimate goal of MST is preventing long-term delinquency and recidivism, it is important to find out how the juveniles are doing once the treatment has been finished.

The first aim of the present study was to examine sustainability of the effects of MST. We conducted a follow-up using adolescent self-reports and parent reports 6 months after the treatment. Recidivism was assessed 6 months after treatment, 2 years after treatment, and at maximal available follow-up period per participant. Thus, multi-informant data were used. Self-report data only provide information from the perception of the participants. For juvenile delinquents, previous research has also shown that the association between parent and juvenile delinquent self-reports on delinquent behavior is low (Ferdinand et al. 2006; Forehand et al. 1991). For these reasons, we collected both self- and parent-reported data as well as recidivism data.

A second aim was to examine the moderators of MST effectiveness. This is in line with the recent shift in intervention research from the question “does it work?” to the more realistic question “what works for whom (Weisz et al. 2006)”. Previous evaluations of MST have not found conclusive evidence for moderator effects, whereas from the Risk-Needs-Responsivity principle (Andrews and Bonta 2010), it can be expected that programs are likely to work for some participants and not for others. In order to find out whether the program is more effective for specific subgroups, moderator tests were conducted for the following participant characteristics: gender, age, ethnicity, and initial problem severity in terms of number of police contacts at pre-test. Previous research has shown that few programs for juveniles with severe and persistent antisocial behavior are gender-specific and, because of the limited number of girls participating in such programs, the effectiveness of these programs for girls has not yet been proven (Zahn et al. 2009). Age is important to examine, as the effects of interventions may diminish with increasing age (Van der Put et al. 2012). Additionally, previous MST research suggests that native participants show larger improvements than ethnic minority participants (Tran 2010), although other studies found no moderating effects for ethnicity (Henggeler et al. 2002; Sundell et al. 2008). Finally, it is important to determine whether MST is equally effective for more as well as less severe juvenile delinquents. Most interventions are least effective for those with most severe initial conditions (Coatsworth et al. 2001; Stanton and Shadish 1997).

Methods

Recruitment and randomization

All referring agencies (Child Protection Council, juvenile judges, Bureaus Youth Care, local referral institutions) were informed beforehand about the study and gave their consent. The referrers informed the juveniles and their families who were considered to

be eligible for MST that a study was being conducted to investigate the effectiveness of youth care. If the families met the inclusion criteria for MST according to the MST supervisors of the participating institutions, research procedures were explained to the juveniles and their families, and their informed consent to participate in the study was obtained by researchers.

Immediately after referral, participants were randomized with the use of a computerized randomization program. A total of $n=256$ juveniles were assigned to MST ($n=147$) and a 'treatment as usual' (TAU) control group condition ($n=109$). For a more elaborate description of the randomization process, see Asscher et al. (2007) or for procedures followed, see Asscher et al. (2012). Data were collected by research assistants before the start of the treatment and immediately after termination of the treatment (mean=5.72, SD=1.90, months after pretest) and at follow-up (mean=1.08 years, SD=0.247, after pre-test) in the homes of the participants. The complete research protocol can be obtained from the first author.

The design of the study was approved by the institutional review board and the medical ethic committee of Utrecht University.

Participants

Participants were juveniles who were referred to MST between 2006 and 2010 and who met the inclusion criteria for MST (De Waag 2012). A sample size of 64 per treatment condition is sufficient to test the stated hypotheses assuming 0.80 power, an alpha of 0.05, and a medium effect size (Cohen 1992). However, to enable us to conduct moderator analyses taking into account possible attrition, the minimum sample size per treatment condition at pretest was set at a minimum of 100. Inclusion criteria for the study were identical to those used by the MST offering agencies: severe and violent antisocial behavior at home, school or community, sufficiently serious to require treatment, and age between 12 and 18 years. Exclusion criteria were: ongoing treatment by another agency, substance abuse without antisocial behavior, sexual offending, autism, acute psychosis, or imminent risk of suicide, and the presence of the youth in the home posing a serious risk to the youth or to the family.

According to official judicial data, 71 % of the participants had been arrested at least once before treatment. According to the self-reports, 64 % of the adolescents had contact with the police at some point during the year before the baseline assessment. Figure 1 charts the flow of participants from referral to data analyses. Despite extensive tracing efforts, 33 participants were lost to post-intervention assessment, and 59 participants were lost to follow-up. Participants lost to post-intervention assessment and follow-up did not differ significantly from those retained on any assessed variable. Thus, all 256 participants were included in the analyses, and, using LISREL 8.8, the multiple imputation was carried out by the expected maximization algorithm (Graham 2009).

The sample consisted of $n=188$ boys and $n=68$ girls, with an age of mean=16.02, SD=1.31. Fifty-five percent of the adolescents had a Dutch ethnicity. Of the adolescents belonging to ethnic minority groups, most had a Moroccan (34 %) or a Surinamese (32 %) background. Half of the adolescents came from a single-parent family. Fifty percent of the mothers and 36 % of the fathers were unemployed. Forty-five percent of the families experienced financial strains and more than half of the families (56 %) lived below minimum income levels. Independent samples *t* tests for continuous

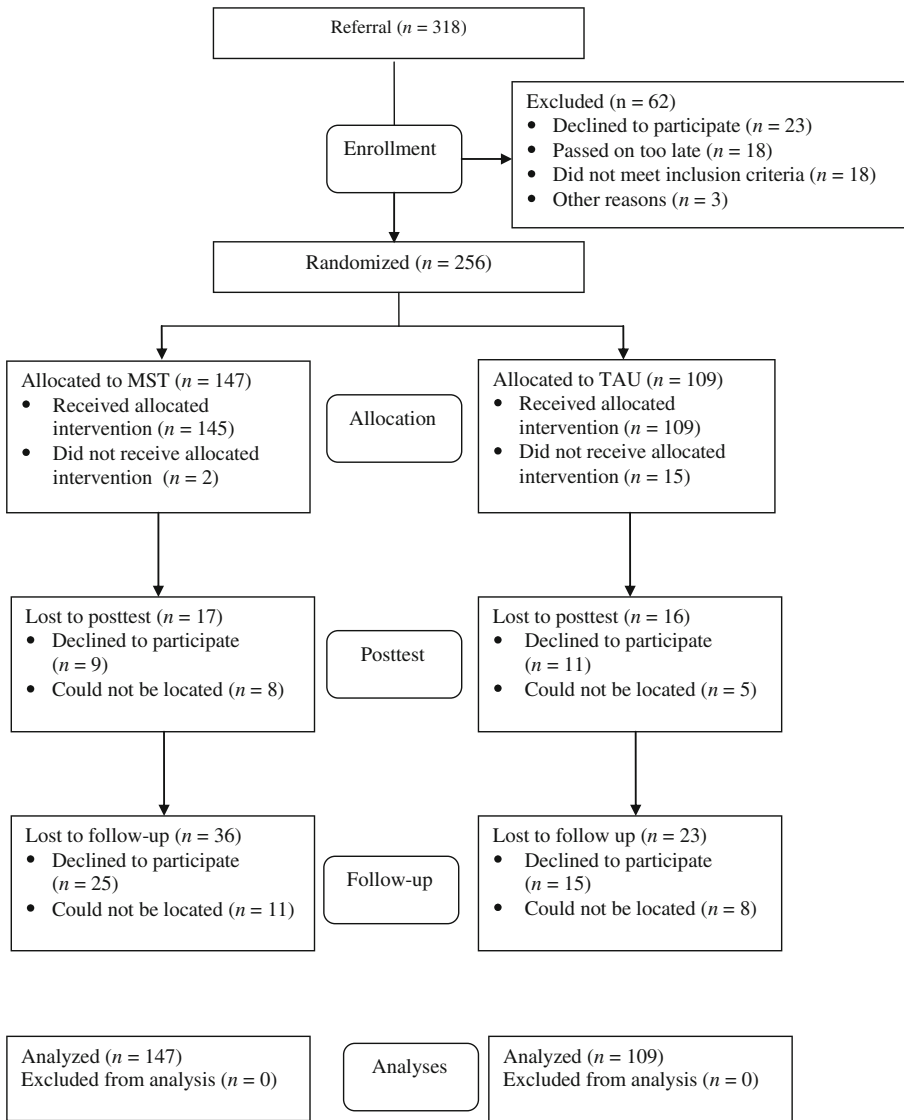


Fig. 1 Flow diagram participants on pre-test, post-test, and follow-up

variables and chi-square analyses for categorical variables were used to examine difference between treatment conditions at T₁ on demographic and the outcome variables. No significant differences were found on any of these variables, suggesting that randomization was successful.

Pathway to MST within the juvenile justice system in The Netherlands

In The Netherlands, there are two referral trajectories to MST within the Juvenile Justice system, and juveniles (or their parents) can also voluntarily participate, for example with a referral from the GP. Within the juvenile justice system, the criminal

law path and the civil law path are distinguished. From the age of 12, juveniles can be sent to MST by referral of the child judge as a special condition within their sentence. This happens when severe antisocial problems are present. Within the civil law path, the child judge may also suggest a treatment when there is severe antisocial behavior, which may lead to delinquency, when the juvenile has not (yet) been arrested. For more information on the referral paths and the consequences of those for randomization, see Asscher et al. (2007).

Conditions

MST MST is based on social ecological and family systems theories, and on research on the causes and correlates of serious antisocial behavior (Henggeler 2011; Henggeler and Schaeffer 2010). It addresses several key systems in which the adolescent is embedded: family, school, peer group, and neighborhood. MST services are often provided in homes at times that are convenient for the families, but meetings are also held in schools, neighborhood settings, or social service agencies. In consultation with family members, the therapist identifies a well-defined set of treatment goals, assigns the tasks required to accomplish these goals, and monitors the progress in regular family sessions at least once a week. The MST therapist training protocol is quite comprehensive, yet the treatment itself is highly individualized to address specific needs of clients.

Treatment as usual (TAU) Participants in the control condition received an alternative treatment that would have been offered had MST not been available. Mostly, these services included individual treatment (individual counseling or supervision by probation officer or case manager, 21 %), and family-based interventions (family therapy, parent counseling, parent groups, or home-based social services, 53 %). Seven percent received a combination of care (e.g., individual treatment and family counseling), and 4 % were placed in a juvenile detention facility. Fifteen percent eventually received no treatment due to various reasons such as moving house or repeated no show at treatment sessions.

Measures

Primary outcomes

Externalizing problems and delinquent behavior In order to assess *externalizing behavior*, several instruments and informants were used. Parents were asked to report on the externalizing behavior of their child with the Child Behavior Checklist (Verhulst et al. 1990). For externalizing behavior problems (aggression and delinquent behavior, 33 items), items had to be answered on a three-point scale, ranging from 0 (*never*) to 2 (*often*). Cronbach's alpha's for externalizing behavior were $\alpha_{T_1}=0.82$, $\alpha_{T_2}=0.81$, and $\alpha_{T_3}=0.93$.

Parents filled out the DSM symptom scales for behavioral problems assessed with the Disruptive Behaviors Disorder rating scales (Oosterlaan et al. 2000). The subscales *Oppositional Defiant Disorder* (9 items) and *Conduct Disorder* (18 items) had to be answered on a four-point scale, ranging from 1 (*not at all*) to 4 (*a lot*). The Oppositional Defiant Disorder (ODD) problems scale (e.g., argues with adults) reached an α of 0.92

(on T_1 , T_2 , and T_3). The Conduct Disorder (CD) subscale (e.g., broke into and entered someone's home or car) had α s of 0.71 (T_1), 0.83 (T_2), and 0.71 (T_3).

Juveniles were asked to report on their own *externalizing behavior* problems by means of the externalizing behavior problems subscale of the Youth Self Report (Achenbach 1991; Verhulst and Van der Ende 1992), which consists of the aggression and delinquency subscale, in total consisting of 30 items, to be answered on a three-point scale, ranging from 0 (*never*) to 2 (*often*). Cronbach's alphas were $\alpha T_1=0.89$; $\alpha T_2=0.90$, and $\alpha T_3=0.74$.

Additionally, two subscales of the Self-Report Delinquency scale (SRD) (Blom and Van der Laan 2006) were used to assess self-report delinquency. Participants were asked to indicate on a list of potential delinquent behaviors whether they engaged in the described behaviors during the past 6 months (*yes* or *no*). The SRD *Violent offending* (5 items, $\alpha T_1=0.74$, $\alpha T_2=0.73$, and $\alpha T_3=0.76$) and *Property offences* (10 items, $\alpha T_1=0.76$, $\alpha T_2=0.82$, and $\alpha T_3=0.82$) scales were used.

Official recidivism data Recidivism data were collected from the official Judicial Registration System. The file containing number of arrests, severity of arrests, and dates of arrests and convictions was provided by the Dutch Ministry of Justice, based on the datafiles of the juveniles who were participating in the RCT.

Recidivism was defined in terms of frequency (dichotomous variable: at least one arrest; and continuous variable: number of arrests), velocity of recidivism (time until first re-arrest) and type of recidivism (categories: violent versus non-violent).

Analytic strategy

Overall effectiveness was examined for all outcome measures by conducting an ANCOVA, with the outcome measures at follow-up as dependent variables, treatment condition as factor and pre-intervention scores of the outcome variables as co-variables. Effect sizes were computed as Cohen's d , based on adjusted means and standard errors, with a positive sign indicating improvement in the MST group relative to the control group.

We examined official delinquency outcomes up to January 2012. Because of the considerable length of the inclusion period, the duration to follow-up was not the same for all participants (on average mean=3.06 years, SD=1.32). Moreover, the time to follow-up was shorter for TAU (mean=2.78 years) than for the MST group (mean=3.27 years), ($t(1, 254)=3.02$, $p<0.05$). To control for the differences in follow-up period between conditions, the official delinquency data were analyzed in two ways. First, We selected those participants for whom recidivism data at 6 months ($n=119$ MST and $n=82$ TAU) and at 2 years post treatment ($n=119$ MST and $n=73$ TAU) were available, (i.e. participants who were followed shorter than 6 months or 2 years were excluded from these analyses). In this way the follow up period was the same for all participants. We compared the two conditions in terms of percentage of participants that recidivated within these periods (6 months and 2 years), frequency of re-arrests, time to re-arrest and violent of non-violent rearrests, using chi-square, t tests and univariate ANOVA tests.

Second, we examined the whole follow-up period for the total sample, thereby correcting for the differences in length of follow-up period between conditions by centering the duration until follow-up in the survival analysis. Cox regression analyses

were used to determine whether the survival curves of MST and TAU were different. Duration to follow-up was centered around the mean and added at step 1 into the Cox regression analysis. Condition (MST or TAU) was added in the second step. A chi-square difference test shows whether MST or TAU condition predicts survival length over duration to follow-up.

For the moderator analyses on the self-report data, the same ANCOVA's were conducted, with the moderators as factor. Post hoc analyses for moderator effects were conducted by splitting the file according to the moderator and again conducting an ANCOVA and calculating effect sizes separately for each group.

Finally, for the moderator tests on the recidivism data, the same regression analysis as above was conducted, but the moderator and the interaction between the moderator and condition were added after adding the centered duration to follow-up and the condition. Following this, chi-square difference tests were conducted to examine whether duration to follow-up was moderated by the interaction condition \times gender, condition \times ethnicity, or condition \times police contacts before start of the treatment.

Results

Intervention effects

In Table 1, the follow-up questionnaire data on externalizing behavior problems and delinquency are presented. To allow for comparison between effectiveness immediately after termination of the program and at follow-up, previously reported (Asscher et al. 2012) effects sizes at post-test are also presented (see last column in Table 1). As can be seen, all effects that were present at post-test are still present at follow-up. Moreover, the effect sizes are even larger at follow-up, when, just like at post-test, there are no significant effects for violent offending.

In order to examine the effectiveness of MST in terms of recidivism, frequency, time to re-arrest, and severity of recidivism were examined for those participants for whom recidivism data were available, both at 6 months and at 2 years post-treatment. The results are presented in Table 2.

The findings indicate that there were no differences between two conditions in frequency or number of arrests at 6 months and at 2 years follow-up, and neither were there differences in time to re-arrest nor in type of re-arrest.

With a Cox regression analysis, the survival curves of the MST and TAU groups were compared regarding the whole sample. At the end of the follow-up period (on average 3.06 years), 63 % of the MST group and 53 % of the TAU group had been rearrested at least once (see also Fig. 2). This difference was not significant: the hazards ratio for condition was 1.136, $p=0.47$, 95 % CI (0.804 to 1.603), which also indicates that there were no significant differences between the groups.

Moderators of effectiveness

To examine whether MST is more beneficial for specific participants, moderator tests were conducted for externalizing behavior problems and recidivism.

Table 1 Intervention effects and effect sizes in *d* at pre-test and follow-up

	Pre-test		6 months follow-up		<i>F</i> for group at follow-up	<i>d</i> 6 months follow-up (95 % CI)	<i>d</i> post-test (95%CI)
	MST	TAU	MST	TAU			
Externalising problems (P)	23.32 (12.60)	22.55 (19.25)	17.02 (10.52)	21.70 (9.57)	16.57***	0.53 (0.28; 0.78)	0.26 (0.01; 0.51)
ODD (P)	2.03 (0.84)	1.93 (0.75)	1.76 (0.59)	2.11 (0.64)	27.77***	0.83 (0.57; 1.09)	0.36 (0.11; 0.61)
CD (P)	1.36 (0.38)	1.34 (0.29)	1.24 (0.26)	1.37 (0.27)	15.76***	0.50 (0.25; 0.75)	0.30 (0.05; 0.55)
Externalising problems (A)	12.40 (9.25)	12.36 (8.32)	10.03 (6.05)	12.20 (6.27)	9.59*	0.39 (0.14; 0.64)	0.26 (0.01; 0.51)
Violent offences (A)	0.38 (0.58)	0.36 (0.57)	0.28 (0.40)	0.28 (0.34)	0.00	0.01 (-0.24; 0.25)	0.03 (0.22; 0.28)
Property offences (A)	0.31 (0.43)	0.29 (0.45)	0.15 (0.22)	0.26 (0.41)	8.23*	0.37 (0.12; 0.62)	0.25 (0.01; 0.50)

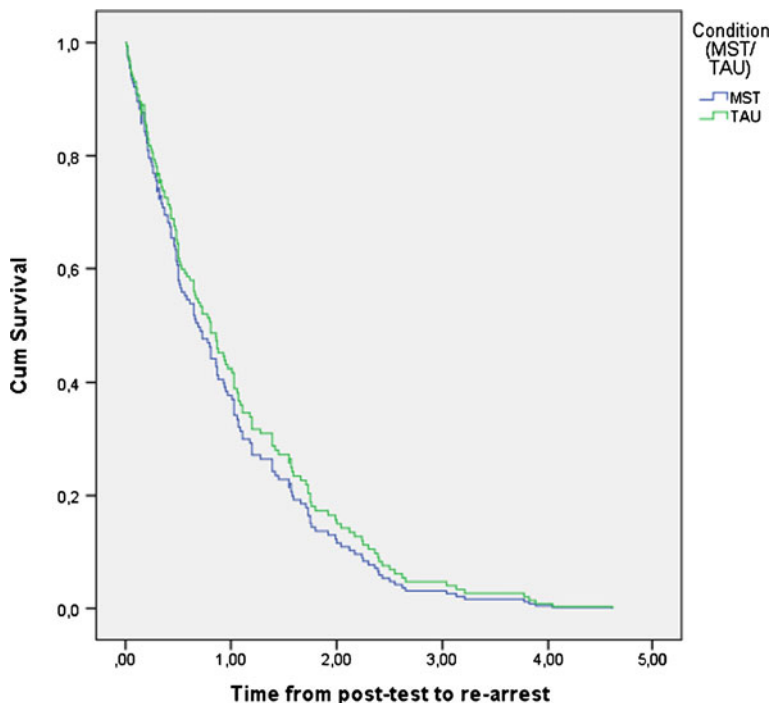
* $p < 0.01$ *** $p < 0.001$

Table 2 Official judicial arrest data at pre-test and at 2 year follow-up

	MST	TAU	F/χ^2	d (95 % CI)
At pre-test	$n=147$	$n=109$		
At least 1 arrest	70.7 %	70.6 %	0.00	
Number of arrests	2.29	2.14	0.26	
Violent offense	54 %	57 %	0.19	
At 6 months post-MST ^a	$n=119$	$n=82$		
At least 1 arrest	34 %	29 %	0.42	0.09 (-0.18; 0.02)
Number of arrests	0.41	0.38	0.35	0.08 (-0.19; 0.02)
Time to re-arrest	2.97 (1.91)	2.58 (2.08)	0.82	0.13 (-0.15; 0.02)
Violent re-arrests ^b	28 %	23 %	0.17	0.05 (-0.22; 0.2)
At 2 years post-MST ^a	$n=119$	$n=73$		
At least 1 arrest	58 %	49 %	1.37	-0.20 (-0.12; 0.52)
Number of re-arrests	1.12	1.22	0.19	0.06 (-0.22; 0.02)
Time to re-arrest	8.28 (6.24)	8.16 (6.69)	0.00	0 (-0.26; 0.02)
Violent re-arrest ^b	50 %	41 %	0.94	0.16 (-0.16; 0.02)

^a Only for those juveniles for whom 6 or 2 years follow-up data were available

^b Only for those juveniles who did recidivate ($n=62$ at 6 months follow-up; $n=151$ at 2 years follow-up)

**Fig. 2** Survival curve for recidivism for MST and TAU group separately

Ethnicity For the questionnaire data, moderator effects were found for ethnicity, with the exception of parent-reports of externalizing behavior at follow-up ($F(4, 251)=4.96, p<0.05$). Posthoc analyses showed stronger effects for native Dutch juveniles ($FT3(1, 138)=16.89, p<0.001, d=0.67$) than for immigrant juveniles ($FT3(1, 112)=1.60, p=0.21, d=0.27$). Additionally, a moderator effect was found for violence ($F(4, 251)=4.60, p<0.05$). Posthoc analyses showed that immigrant juveniles showed more improvement in the MST group ($F(1, 112)=2.65, p=0.11, d=0.27$), whereas native Dutch juveniles performed better in the TAU group ($F(1, 138)=2.00, p=0.16, d=0.24$).

The interaction effect for condition \times ethnicity in Cox regression analysis was not significant. The hazards ratio for condition \times ethnicity was 1.296, $p=0.45$, 95 % CI (0.662 to 2.536), also indicating no differences in recidivism for native Dutch and immigrant juveniles.

Gender There were no moderator effects for gender on questionnaire data at follow-up or regarding recidivism data: hazards ratio=1.081, $p=0.88$, 95 % CI (0.400 to 2.921).

Age In order to examine whether older or younger juveniles would benefit more from MST, the sample was divided in a group that was younger than 16 years of age ($n=113$) and a group of 16 years and older ($n=143$). These analyses revealed no moderator effects for age at follow-up or for recidivism data: hazards ratio=1.058, $p=0.87$, 95 % CI (0.525 to 2.132).

Initial problem severity Various indicators of initial problem severity were examined: both the parent-reported clinical cut-off scores for externalizing behavior problems and severity reported in official arrest data were examined as potential moderators of treatment effectiveness. There were no moderator effects for initial problem severity, for follow-up or recidivism data: hazards ratio=1.171, $p=0.71$, 95 % CI (0.515 to 2.661).

Conclusion and discussion

In the present study, the sustainability of effects of MST was examined by using parent and adolescent reports on externalizing behavior problems 6 months after treatment termination of MST and by using official judicial data at 6 months, 2 years, and, on average, 3.06 years. According to parent and adolescent reports, the follow-up results were promising in the sense that the results that were observed at post-test (for parent and adolescent reported behavior problems, Opposite Defiant Disorder, Conduct Disorder, and self-reported property offenses) were still present at follow-up assessment, and that effect sizes were even larger than at post-test. However, it should be noted that no significant effects were found for violent behavior, either at post-test or at follow-up. The fact that, at follow-up, based on questionnaire outcomes, the effect is larger than immediately at post-test, may indicate an increase of the post-test effects at a later point in time. It is possible that changes that have started during treatment, and continue after treatment, not only changed the adolescent's behavior but also the adolescent's cognitions and the social context of the adolescent (parenting environment, peer

relationships). Possibly the combination of these changes may eventually be important for changing delinquent behavior on the long run. Compared with the US results, the effect sizes in the present study are somewhat smaller (Curtis et al. 2004), even considering the somewhat larger follow-up effect sizes. An explanation for the smaller effect sizes than those reported in the previous meta-analysis is that, in the Netherlands, as in Sweden (see Sundell et al. 2008), the treatment as usual is of high quality and consists of components that are also present in MST. Treatment often takes place outside the clinic, which increases generalizability. A second explanation may be found in terms of implementation of the program. In the study of Petrosino and Soydan (2005), as mentioned in the “Introduction”, it was concluded that effects of criminal justice interventions are likely to be larger when program developers are conducting the study. This may be caused by better focus on program fidelity and treatment integrity by program developers. This phenomenon has been referred to in previous research as the difference between efficacy and effectiveness research or the difference between demonstration programs and clinically driven programs (e.g., Nathan et al. 2000). In these studies, it has been suggested that program effects are likely to be largest when carried out under specific highly controlled circumstances, whereas, when the program is conducted in practice, with a larger heterogeneity in sample and institutions providing the treatment, the effects are likely to drop, simply because the circumstances are less controlled. Eisner (2009) even went one step further by explaining the larger effect sizes when program developers were involved by suggesting conflicts of interest. Based on self- and parent report data, the conclusion that MST is effective in the Netherlands, although with somewhat smaller effect sizes than in the US studies, seems defensible.

The official judicial data, however, revealed no differences between MST and TAU in frequency, timing, and type of re-arrest. The finding that MST is not more effective than TAU in terms of official recidivism data is in line with previous findings reported by Leschied and Cunningham (2002), and contrary to many of the previous MST studies (e.g., Butler et al. 2011; Sawyer and Borduin 2011).

Differences between self-report and official data are not unusual (see, e.g., Petrosino et al. 2003). There are various explanations for the differences in effects when using official judicial data. First, it should be noted that, although the same terms “recidivism” or “official data” are used in studies, the operationalization often differs. Many studies reported several *outcomes* measured in different ways. Criminal offending outcomes are reported as the rates of arrests, convictions, incarcerations, or number of crimes committed. Additionally, results may be inconsistent across the selected outcomes: in most of the studies, positive effects are found for some, but not for all, of these outcomes. For example, Reynolds et al. (2007) reported positive effects for felony arrests, convictions, and incarceration, but not for violent arrest and violent crime conviction. Standard operationalization of official delinquency data is thus lacking in the literature (see also Deković et al. 2011), which makes it difficult to directly compare the findings across the studies.

Second, it should be recognized that each source of information has its own advantages and disadvantages. Self-report data indicate that delinquent behavior is much more widespread than would be expected based on official data (Krohn et al. 2010). Moreover, self- and parent-reported delinquency reports may be subject to social desirability tendencies, and delinquent adolescents show the tendency to play down

their offenses (Tromp and Koot 2010). Both self-reported and official delinquency data may cause bias, and there is only a moderate correlation between these sources of information (Enzmann et al. 2010). Where self-report data may show an underestimation of the delinquent behavior, official judicial data have the risk of being subject to selection bias, caused by policy decisions (Kirk 2006). Moreover, official delinquency data records may not reflect the full volume of crime: not all crimes are detected by the police, and not all crimes will result in registered crime (Kirk 2006). The present study shows that self- and parent reports lead to different conclusions than official delinquency data. The contrary results based on different sources of information stresses the importance of the use of more information sources in order to estimate the value of a program.

Moderator analyses suggest that MST is equally effective for boys and girls, older and younger adolescents, and for juveniles of different levels of initial problems. Some contradicting, moderator effects were found for ethnicity, indicating that, when parent-reported externalizing problems are used, MST seems more effective for native Dutch juveniles, whereas when adolescent-reported violence is examined, effects are larger for migrant juveniles in the MST group. However, one should bear in mind that this moderator effect was found for only one of the outcome measures, so future research should replicate this moderator analysis in order to examine the importance of this finding. The moderator effects must be interpreted cautiously, as only a few effects were found. Previous MST studies did not find moderator effects either (e.g., Henggeler et al. 2002) or only few moderator effects (Sundell et al. 2008), just like the present study.

Limitations

There are several limitations worth mentioning. A first limitation is that we have a large variation in the duration until follow-up assessment because the inclusion of participants in the present study took more than 4 years. It is therefore important to gather recidivism data again later on.

Additionally, despite the fact that our sample is larger than most follow-up RCTs (e.g., Butler et al. 2011; Ogden and Hagen 2006; Sawyer and Borduin 2011), our sample is still relatively small for detecting small effects, especially for subgroups examined, and for testing moderator effects; in order to detect small effects, a sample size of 250 per arm is recommended (Wartna 2005).

Finally, the moderator variables were not blocked in the original design, but examined post hoc. This approach, although common in the field, has recently been criticized as it may add possible bias to the results (see Weisburd and Gill 2013).

Conclusion

Notwithstanding these limitations, the present study presents the follow-up results of the first randomized controlled trial of MST in the Netherlands. Moreover, to date, only two European studies have examined long-term effectiveness of MST and those two studies (Butler et al. 2011; Ogden and Hagen 2006) used very small sample sizes. The sample used in the present study is larger than the samples used in most previous MST trials. The multi-informant data on delinquency provide important insights that, on the one hand, according to parents and adolescents, MST seems to be effective in

diminishing delinquent behavior, whereas, on the other hand, the official judicial data suggest that there are no differences between MST and TAU. These results stress the importance of using official data, participants reports, and/or observational data. A few moderator effects have been found, suggesting that MST to be more effective than TAU for native Dutch and for older juveniles. These findings need to be replicated with larger groups and a longer follow-up period.

Acknowledgments This work was supported by grants from the Netherlands Organisation for Scientific Research (NWO) and the Netherlands Organisation for Health Research and Development (ZonMw). We thank the therapists and staff of the care and treatment organizations ‘De Waag’ and ‘Yorneo’ for their cooperation, and we thank the adolescents and families who participated in the project.

References

- Achenbach, T. M. (1991). *Manual for the Youth Self-Report and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, *16*, 39–55.
- Asscher, J. J., Deković, M., Van der Laan, P. H., Prins, P. J. M., & van Arum, S. (2007). Implementing randomized experiments in criminal justice settings: an evaluation of multi-systemic therapy in the Netherlands. *Journal of Experimental Criminology*, *3*, 113–129.
- Asscher, J. J., Deković, M., Manders, W. A., Van der Laan, P. H., Prins, P. J. M., & Dutch MST Cost-Effectiveness Research Group (2012). A randomized clinical trial of Multisystemic Therapy in The Netherlands. *Journal of Experimental Criminology*. doi: 10.1007/s11292-012-9165-9.
- Blom, M., & Van der Laan, A. M. (2006). *Monitor Jeugd Terecht*. [Monitor Juveniles and the Law]. Retrieved November, 24, 2009, from: <http://www.wodc.nl/onderzoeksdatabase/monitor-jeugd-terecht-2006.aspx>.
- Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., & Williams, R. A. (1995). Multisystemic treatment of serious juvenile offenders: long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology*, *63*, 569–578.
- Butler, S., Baruch, G., Hickey, N., & Fonagy, P. (2011). A randomized controlled trial of Multisystemic Therapy and a Statutory therapeutic intervention for young offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*, 1220–1235.
- Coatsworth, J. D., Santisteban, D. A., McBride, C. K., & Szapocznik, J. (2001). Brief strategic family therapy versus community control: Engagement, retention, and an exploration of the moderating role of adolescent problem severity. *Family Process*, *40*, 313–332.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- 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.
- Cunningham, A. J. (2002). *One step forward: Lessons learned from a randomized study of Multisystemic Therapy in Canada*. London, Ontario: Praxis: Research from the Centre for Children & Families in the Justice System, June 2002.
- Deković, M., Slagt, M. I., Asscher, J. J., Boendermaker, L., Eichelsheim, V. I., & Prinzie, P. (2011). Effects of early prevention programs on adult criminal offending: A meta-analysis. *Clinical Psychology Review*, *31*, 532–544.
- De Waag. (2012). *Inclusie Criteria MST*. [Inclusion criteria MST]. Retrieved from: <http://www.dewaagnederland.nl/docs/folders-en-flyers/multi-systeem-therapie.pdf?sfvrsn=2>
- Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, *5*, 163–183.
- Enzmann, D., Haen Marshall, I., Kilias, M., Junger-Tas, J., Stekete, M., & Gruszczynska, B. (2010). Self-reported youth delinquency in Europe and beyond: First results of the second international self-report delinquency study in the context of police and victimization data. *European Journal of Criminology*, *7*, 159–183.

- Ferdinand, R. F., Van der Ende, J., & Verhulst, F. C. (2006). Prognostic value of parent-adolescent disagreement in a referred sample. *European Child and Adolescent Psychiatry*, *15*, 156–162.
- Forehand, R., Frame, C. L., Wierson, M., Armistead, L., & Kempton, T. (1991). Assessment of incarcerated juvenile delinquents: Agreement across raters and approaches to psychopathology. *Journal of Psychopathology and Behavioral Assessment*, *13*, 17–25.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.
- Henggeler, S. W. (2011). Efficacy studies to large-scale transport: The development and validation of Multisystemic Therapy programs. *The Annual Review of Clinical Psychology*, *7*, 351–381.
- 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 and Adolescent Psychiatry*, *41*, 868–874.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic Therapy with violent 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., & Schaeffer, C. M. (2010). Treating serious emotional and behavioral problems using Multisystemic Therapy. *The Australian and New Zealand Journal of Family Therapy*, *31*, 149–164.
- Kirk, D. S. (2006). Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *Journal of Quantitative Criminology*, *22*, 107–129.
- Krohn, M. D., Thornberry, T. P., Gibson, C. L., & Baldwin, J. M. (2010). The development and impact of self-report measures of crime and delinquency. *Journal of Quantitative Criminology*, *26*, 509–525.
- Leschied, A. W., & Cunningham, A. (2002). *Seeking effective interventions for serious young offenders. Interim results of a four-year randomized study of Multisystemic Therapy in Ontario, Canada*. London (CA): Centre for children and families in the justice system of the London Family Court Clinic.
- Littell, J. H., Popa, M., & Forsythe, B. (Eds.). (2005). *Multisystemic Therapy for social, emotional, and behavior problems in youth age 10-17. Cochrane Library, issue 3*. Chichester: John Wiley & Sons Ltd.
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Between psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, *126*, 964–981.
- Ogden, T., & Amlund Hagen, K. (2006). Multisystemic Treatment of serious behaviour problems in youth: Sustainability of treatment effectiveness two years after intake. *Child and Adolescent Mental Health*, *11*, 142–149.
- Ogden, T., & Halliday-Boykins, C. A. (2004). Multisystemic Treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health*, *9*, 77–83.
- Oosterlaan, J., Scheres, A., Antrop, I., Roeyers, H., & Sergeant, J. A. (2000). *Vragenlijst voor Gedragsproblemen bij Kinderen (VvGK)*. Handleiding. Lisse: Swets & Zeitlinger.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, *1*, 435–450.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *The Annals of the American Academy of Political and Social Science*, *589*, 41–61. doi:10.1177/0002716203254693.
- Reynolds, A. J., Temple, J. D., Ou, S. R., Robertson, D. L., Mersky, J. P., Topitzes, J. W., & Niles, M. D. (2007). Effects of a school-based, early childhood intervention on adult health and well-being. *Archives of Pediatric Adolescent Medicine*, *161*, 730–739.
- Rowland, M. D., Halliday-Boykins, C. A., Henggeler, S. W., Cunningham, P. B., Lee, T. G., Kruesi, M. J. P., et al. (2005). A randomized trial of Multisystemic Therapy with Hawaii's Felix Class Youths. *Journal of Emotional and Behavioral Disorders*, *13*, 13–23.
- Sawyer, A. M., & Borduin, C. M. (2011). Effects of Multisystemic Therapy through midlife: A 21.9 year follow-up to a randomized clinical trial with serious and violent juvenile offenders. *Journal of Consulting and Clinical Psychology*, *79*, 643–652.
- Stanton, M. D., & Shadish, W. R. (1997). Outcome, attrition, and family couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. *Psychological Bulletin*, *122*, 170–191.

- Sundell, K., Hansson, K., Löfholm, C. A., Olsson, T., Gustle, L. H., & Kadesjö, C. (2008). The transportability of Multisystemic Therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology*, *22*, 550–560.
- Timmons-Mitchell, J., Bender, M. B., Kishna, M. A., & Mitchell, C. C. (2006). An independent effectiveness trial of Multisystemic Therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology*, *35*, 227–236.
- Tran, N. T. (2010). *Moderators of Multisystemic Therapy outcome for children with Conduct Disorders*. Nashville, Tennessee: Vanderbilt University.
- Tromp, N. B., & Koot, H. M. (2010). Self- and parent report of personality pathology: informant agreement and relations to dysfunction. *Journal of Personality Disorders*, *24*, 151–170.
- Van der Put, C. E., Stams, G. J. J. M., Hoeve, M., Deković, M., Spanjaard, H. J. M., Van der Laan, P. H., & Barnoski, R. P. (2012). Changes in the relative importance of dynamic risk factors for recidivism during adolescence. *International Journal of Offender Therapy and Comparative Criminology*, *56*, 296–316.
- Verhulst, F., Koot, J. M., Akkerhuis, G. W., & Veerman, J. W. (1990). *Praktische handleiding voor de CBCL [CBCL manual]*. Assen/Maastricht: Van Gorcum.
- Verhulst, F., & Van der Ende, J. (1992). Agreement between parents' and adolescents' self-reports of problem behavior. *Journal of Child Psychology and Psychiatry*, *33*, 1011–1023.
- Wartna, B. S. J. (2005). *Evaluatie van daderprogramma's. Een wegwijzer voor onderzoek naar de effecten van strafrechtelijke interventies speciaal gericht op het terugdringen van recidive* (Evaluation of offender programmes. A guide to research into effects of penal interventions aimed at reducing recidivism). The Hague: WODC.
- Weisburd, D., & Gill, C. (2013). Block Randomized Trials at Places: Rethinking the Limitations of Small N Experiments. *Journal of Quantitative Criminology*. doi:10.1007/s10940-013-9196-z.
- Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*, *61*, 671–689.
- Zahn, M. A., Day, J. C., Mihalic, S. F., & Tichavsky, L. (2009). Determining what works for girls in the Juvenile Justice System: A summary of evaluation evidence. *Crime and Delinquency*, *55*, 266–293.

Jessica Asscher, developmental psychologist, is an assistant professor Forensic Child and Youth Care Sciences at the University of Amsterdam. Her research interests concern the effectiveness of (juvenile justice) interventions and the development of behavior problems.

Maja Deković developmental psychologist, is a full professor at Department of Child and Adolescent Studies, Utrecht University and leader of the research programme “Development and treatment of psychosocial problemst”. Her research interests include the development of problem behaviour, parent–adolescent relationships, family interaction and effects of family-based interventions.

Willeke Manders studied Psychology at Tilburg University. She obtained her PhD at the Radboud University for a study on families and adolescent problem behaviour. Currently she is conducting a study on the implementation and effectiveness of Multisystemic Therapy in the Netherlands as a postdoctoral fellow at Utrecht University. Her research interests are family assessment, the development of problem behavior, and the implementation and effectiveness of interventions.

Peter H. van der Laan, PhD., is Professor of Probation and Parole at the VU University Amsterdam, The Netherlands, and is Senior Researcher at the Netherlands Institute for the Study of Crime and Law enforcement (NSCR), Amsterdam. His research interests regard juvenile delinquency, criminal justice issues and the effectiveness of criminal justice interventions.

Prof dr. Pier J.M. Prins is a full professor of child clinical psychology at the Department of Developmental Psychology at the University of Amsterdam. Prof. Prins is director of the research program “Mechanisms of Change and Psychosocial Treatments for Behaviorally and Emotionally Disordered Youth” at the University of Amsterdam. Currently, he is involved in the multisite Amsterdam-Groningen Study evaluating moderators and mediators of CBT with anxiety disordered children and adolescents. He is also involved in the development and evaluation of a computer game to remediate executive functions in children with ADHD. Pier Prins is a licensed GZ-psychologist and also a licensed behavior therapist.

Sander van Arum is chief clinical director at De Waag - a center for out patient forensic psychiatry with locations in 8 cities in The Netherlands. He has a Master of Science in Special Education from the University of Utrecht. He is a fully licensed psychotherapist and family therapy supervisor and chairs the forensic systems therapy section of the Dutch Association for Relational - and Family Therapy. Focus in his clinical work in forensic psychiatry is disruptive behaviour disorders, relational violence, child abuse, family trauma and family resilience.