Criminal substance abusing adolescents and systemic treatment

Pol, T.M. van der

2

THE EFFECTIVENESS OF MULTIDIMENSIONAL FAMILY THERAPY (MDFT) IN TREATING ADOLESCENTS WITH MULTIPLE BEHAVIOUR PROBLEMS: A META-ANALYSIS

Thimo van der Pol, Machteld Hoeve, Marc Noom, Geert Jan Stams, Theo Doreleijers, Lieke van Domburgh, Robert Vermeiren

Journal of Child Psychology and Psychiatry. 2017;58(5):532-45

ABSTRACT

Introduction

Multidimensional Family Therapy (MDFT) is a well-established treatment for adolescents showing both substance abuse and/or antisocial behaviour.

Method

The effectiveness of MDFT in reducing adolescents' substance abuse, delinquency, externalising, and internalising psychopathology, and family malfunctioning was examined by means of a (three-level) meta-analysis, summarising 61 effect sizes from 19 manuscripts (N = 1,488 participants).

Results

Compared with other therapies, the overall effect size of MDFT was significant, albeit small in magnitude (d = 0.24, p < 0.001), and similar across intervention outcome categories. Moderator analysis revealed that adolescents with high severity problems, including severe substance abuse and disruptive behaviour disorder, benefited more from MDFT than adolescents with less severe conditions.

Conclusions

It can be concluded that MDFT is effective for adolescents with substance abuse, delinquency, and comorbid behaviour problems. Subsequently, it is important to match specific characteristics of the adolescents, such as extent of impairment, with MDFT.

INTRODUCTION

Substance abuse disorders (SUD) in adolescents predispose to a variety of behaviour problems, such as delinquency, externalising, and internalising psychopathology, (Grella, Hser, Joshi, & Rounds-Bryant, 2001; Merikangas et al., 2010) and family malfunctioning (Colins et al., 2011; Cuellar, McReynolds, & Wasserman, 2006; Hoeve, McReynolds, & Wasserman, 2013; McReynolds & Wasserman, 2011). The incidence of SUD related comorbidity is estimated to reach up to 75%, (Grella et al., 2001) which influences treatment outcome substantially. For instance, the presence of externalising psychopathology in combination with SUD increases the likelihood of engaging (Anderson, Ramo, Schulte, Cummins, & Brown, 2007; Anderson, Tapert, Moadab, Crowley, & Brown, 2007; Monahan, 2003) and persisting in delinquent behaviour (Lodewijks et al., 2010; Wasserman, McReynolds, Fisher, & Lucas, 2003). The same pattern has been observed in adolescents with internalising psychopathology (Loeber, Stouthamer-Loeber, & Raskin White, 1999). As such, the presence of multiple behaviour problems in adolescence creates major societal and public health concerns (Johnston & Hauser, 2008; Moffit, 1993). Hence, effective prevention and treatment programmes to address the complex problems of adolescents with SUD are direly needed (Hall et al., 2016; Merikangas et al., 2010).

In the last 30 years, several treatments have been developed to effectively reduce SUD, delinquency and comorbid behaviour problems. Various systematic literature reviews and meta-analyses have concluded that family-based treatments and cognitive behavioural therapy are effective in treating adolescents with SUD, delinquency, and comorbid psychopathology (Carr, 2009; Von Sydow et al., 2013; Holly Barrett Waldron & Charles W Turner, 2008). A promising family-based treatment programme is Multidimensional Family Therapy (MDFT) (Liddle, 2002). The present meta-analysis focuses on the effectiveness of MDFT compared to other treatments in reducing adolescents' substance abuse, delinquency, externalising, and internalising psychopathology, and family malfunctioning.

MDFT

MDFT is a manualised, evidence-based, intensive intervention programme with assessment and treatment modules focusing on four areas: 1) the individual adolescent's issues regarding SUD, delinquency, and comorbid psychopathology, 2) the parents' child-rearing skills and personal functioning, 3) communication and relationship between adolescent and parent(s), and 4) interactions between family members and key social systems (Liddle, 2002). MDFT is based on the family therapy foundation established by Minuchin (1974) and Haley, (1976) and on the ecological systems theory of Bronfenbrenner, (1979) which states that human development is shaped by the interaction of the individual with his or her surrounding social contexts. Within each adolescent's environment there are multiple risk and protective factors that influence and reinforce each other (Brook et al., 1992). Therefore, MDFT was developed to intervene in multiple systems, addressing these risk and strengthening protective factors in the adolescents' environments (Liddle, 1999). MDFT is operational and expanding briskly in Europe and in the United States and targets youth from diverse ethnic and socioeconomic backgrounds in a variety of settings (Liddle, 2002; Rigter et al., 2010).

The effectiveness of MDFT

Three previous meta-analyses (Baldwin, Christian, Berkeljon, & Shadish, 2012; Filges, Andersen, & Jørgensen, 2015; Tanner-Smith, Wilson, & Lipsey, 2013) summarised the results of studies that examined the effectiveness of MDFT alone or together with other family-based treatments. Tanner-Smith et al. (2013) concluded that for substance abuse, family therapy is the treatment with the strongest evidence of comparative effectiveness. The overall effect compared with non-family treatments was small (d = 0.26). Similarly, Baldwin et al. (2012) found family therapies to have a small effect for substance abuse and delinquency compared with treatment as usual (d = 0.21) and alternative treatments, such as group therapy, psychodynamic family therapy, individual therapy, parent groups, and family education (d = 0.26). It must be noted that the Baldwin et al. study did not include any follow up data of the studies they reviewed in their meta-analysis. Filges et al. (2015) concluded that MDFT was successful in reducing adolescents' substance abuse in the short run, but not in the long run (no Cohen's d was reported).

The studies included in the meta-analyses revealed substantial variability in the effectiveness of MDFT, which may be explained by differences in study characteristics. For example, differences in MDFT effectiveness could be related to the severity of substance abuse and/or psychopathology of participants. However, the effect of substance abuse severity, psychopathology, and other potentially important moderators were not considered in previous meta-analyses. The authors of the three meta-analyse mentioned not being able to perform extensive moderator analyses due to a limited number of studies. Therefore, further comprehensive research is needed. Insight into moderating factors of the effectiveness of MDFT is important for identifying which adolescents may benefit most from MDFT; this knowledge is crucial for improvement of assessment and referral practices.

The present meta-analysis

The goal of the present study was to provide a meta-analytic overview of the studies examining the effects of MDFT compared to other interventions for adolescents with SUD and comorbid behaviour problems. First, we examined the overall effectiveness of MDFT regarding substance abuse, delinquency, externalising, and internalising psychopathology, and family functioning. Also, the mean effects of MDFT as compared to cognitive behavioural therapy (CBT), group therapy (GT), and combined treatments (CT) were examined. Second, we conducted moderator analyses in order to investigate whether study characteristics contributed to the effectiveness of MDFT. The most important question to be investigated was if adolescents with severe substance abuse and severe externalising psychopathology benefitted more from MDFT than adolescents with less severe conditions, which is from now on called 'the severity gradient'. To test this severity gradient (C.E. Henderson, Dakof, Greenbaum, & Liddle, 2010) a three-level meta-analysis was utilised. This novel three-level analytic method makes it possible to include more effect sizes per study and account for differences between effect sizes both within and between studies, which prevents important data and information loss, increases statistical power and the number of moderators that can be tested (Assink et al., 2015).

METHOD

Sample of studies

Three criteria guided the selection of studies. First, the study had to examine the effectiveness of MDFT. Second, the study had to report results for one or more of the following outcome measures: substance abuse, delinquency, externalising, and internalising psychopathology, and family functioning, or provide enough details to calculate a bivariate test statistic. Third, in view of study quality, a study had to report the results of a randomised controlled trial (RCT).

Candidate studies meeting the selection criteria with data either published by the 29th of February 2016 or available from primary authors (unpublished manuscripts) were collected as follows. First, the electronic databases PubMed, PsycINFO, Embase, and Web of Science were searched for articles, books, chapters, paper presentations, dissertations, and reviews. Our purpose was to find as many studies as possible, and therefore a variety of terms related to Multidimensional Family Therapy (MDFT) were used. Search terms, such as multidimensional*, famil*, and MDFT, were cross-referenced with therap*, and treat* in English, Dutch, French and German: ((Multidimensional Family Therap*) OR (Multidimensional Family Treat*) OR (MDFT AND (Family OR Therapy OR Multidimensional)). Subsequently, manual searches of references, lists from these publications were conducted to identify relevant studies not found in the electronic databases.

If multiple publications were found that reported on the same study, we only included manuscripts which reported a different outcome measure or a subsample of the original study. Furthermore, we contacted the authors of the publications to check for unpublished materials. Seven manuscripts were received of which one submitted paper (Liddle, 2015) and 4 reports (Grichting, Haug, Nielsen, & Schaub, 2011; Phan, 2011; Tossmann & Jonas, 2010; Verbanck et al., 2010) were eligible to be included in the meta-analysis. In total 210 manuscripts were found, of which we selected 71 on the basis of information in the abstract. After assessing the 71 articles, 19 manuscripts on effects of MDFT met our

criteria and were included in the present meta-analysis. For the purpose of standardisation of the effect sizes and the possibility to examine the influence of severe behaviour problems through moderator analyses, we asked the authors of the manuscripts for supplementary information on substance abuse and psychopathology. The 19 manuscripts together with the retrieved supplementary information yielded 61 effect sizes, resulting from 8 independent studies with a total of 1,488 subjects. Figure 1 presents a flowchart of the selection procedure.

File drawer problem

The tendency of journals to exclude manuscripts reporting non-significant findings, referred to as publication bias, may have implications for the final conclusions of the meta-analysis (Rosenthal, 1991; Van IJzendoorn, 1998). For this, Rosenthal coined the term 'file drawer problem' (1979). Several methods exist to address potential effects of publication bias, but each has its own shortcomings (Rothstein, 2008). The best solution in preventing effects of publication bias is to make extensive efforts to obtain all unpublished materials (Mullen, 2013; Rosenthal, 1991). Following the advice of Rothstein, (2008) three methods addressing publication bias were applied. First, we calculated a fail-safe number, which estimates the number of unretrieved studies reporting null results needed to bring the overall combined effect size to a level at which it would no longer be statistically significant (Rosenthal, 1991). The fail-safe number, 2,554, exceeded Rosenthal's (1995) critical value ($61 * 5 + 10 = 315$). This indicates that the number of unpublished studies with non-significant results that would be required to reduce significant results to non-significant results was sufficient, suggesting no evidence for publication bias.

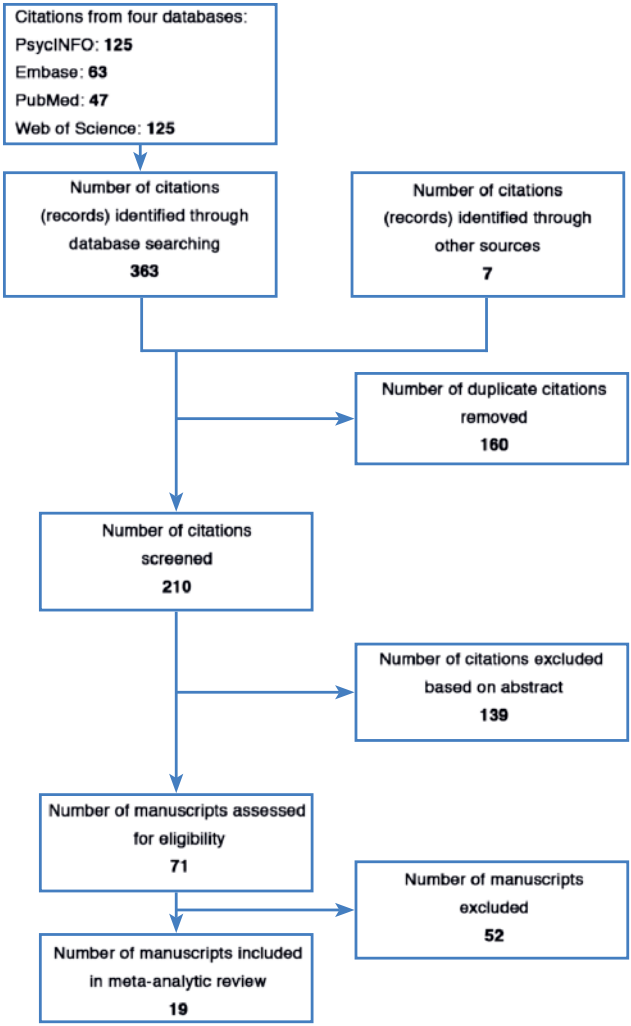


Figure 1. Flowchart of literature search and screening.

A second method of examining publication bias is inspecting the distribution of each individual study's effect size on the horizontal axis

against its sample size and standard error or precision (the reciprocal of the standard error) on the vertical axis. The distribution of effect sizes should form a funnel shape if no publication bias is present, as studies with small sample sizes are expected to show a larger variation in effect size magnitude, whereas studies with large effect sizes are expected to result in effect sizes closer to the overall mean. A violation of funnel plot symmetry reflects publication bias, that is, a selective inclusion of studies showing positive or negative outcomes (Sutton, Duval, Tweedie, Abrams, & Jones, 2000). Figure 2 depicts the funnel plot of effect sizes. In the present study, funnel plot asymmetry was tested by regressing the standard normal deviate, defined as the effect size, divided by its standard error, against the estimate's precision (the inverse of the standard error), which largely depends on sample size (Egger, Smith, Schneider, & Minder, 1997). If there is asymmetry, the regression line does not run through the origin and the intercept significantly deviates from zero. The intercept did not significantly deviate from zero ($z = 1.490$, $p = 0.136$), indicating no publication bias.

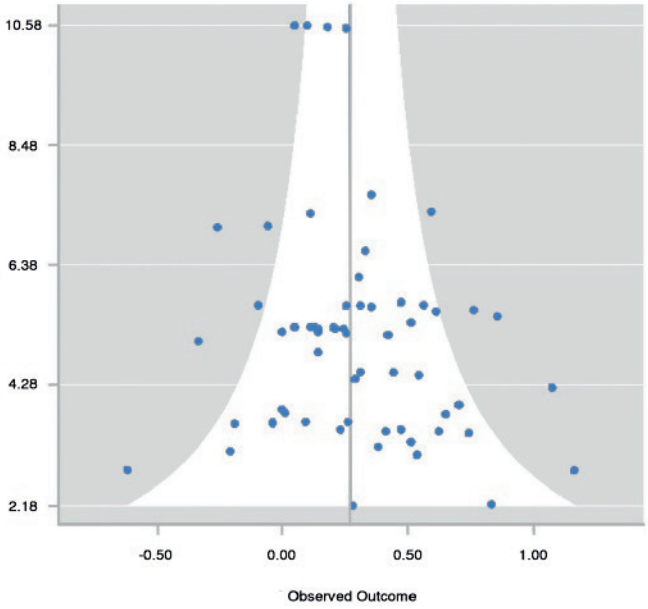


Figure 2. Funnel plot of effect sizes.

Third, we utilised the *P*-curve method, which was recently introduced by Simonsohn et al. (2014). The rationale of the method is that if a set of statistically significant studies contains real evidential value in favour of rejecting a joint null hypothesis, *p*-values extracted from these studies should display a larger share of *p*-values closer to zero as compared to *p*-values in the upper ranges just below the critical value ($p < 0.05$) of statistical significance. Likewise, if there are signs of *p*-hacking, that is, if a non-significant *p*-value is pushed past the critical value for statistical significance, a larger share of the *p*-values should be observed just below the threshold of statistical significance rather than closer to zero. The *P*-curve analyses whether MDFT is being more or less effective than the compared therapies. The *P*-curve test was performed on all of the statistically significant two-tailed *p*-values in our sample. When testing the two-tailed *p*-values the right-skew *p*-value was <0.0001, (Figure 3). The *P*-curve showed statistically significant signs of evidential value and the statistical power estimated was 85%. It can be concluded that the results indicate no evidence of *p*-hacking.

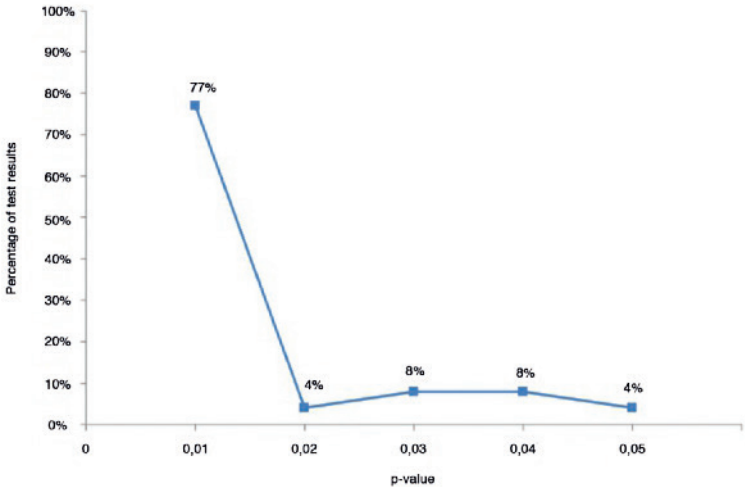


Figure 3. P-curve, testing possible p-hacking.

Note. The Observed p-curve includes 26 statistically significant ($p<0.05$) results, of which 22 are $p<0.025$. There were no non-significant results entered.

Coding of the study outcomes and characteristics

We retrieved the study results (test statistic and value) or data to calculate the effect size from the manuscripts. Next, information on sample descriptors, treatment descriptors, research design, and manuscript characteristics were collected.

For the sample descriptors, we categorised the effect sizes into five primary outcome measures: substance abuse, delinquency, externalising, and internalising psychopathology, and family functioning. We coded the geographical location where the study had been conducted (Europe, United States). As for demographic characteristics, we collected data on age, gender, socioeconomic status (SES), and ethnicity. We coded age of the subjects at the start of treatment. Gender was defined as percentage of males in the sample. The SES was characterised by calculating the mean family income in euros. Furthermore, we defined the percentage of Caucasian, Afro-American, Hispanic, Asian, and other ethnicities (e.g., Caribbean, North-African). The percentage of adolescents in the sample with additional psychiatric disorders was also coded for: conduct disorder, oppositional defiant disorder, and disruptive behaviour disorder (DBD) (i.e., the presence of either CD and/or ODD), attention deficit hyperactivity disorder (ADHD), generalised anxiety disorder (GAD), and depression. Moreover, we collected data on the type of substance abuse and calculated the percentage of cannabis, alcohol, and other drug use in the sample. Finally, we retrieved information on the severity of cannabis use. Using the benchmark established by Hendriks et al. (2011) and also used in Rigter et al., (2013) we retrieved the percentage of adolescents who reported using substances more than 64 of the 90 day intake assessment period.

For the treatment descriptors, we distinguished three treatment comparison groups: cognitive behavioural therapy (CBT), group therapy (GT), and combined treatments (CT). We assigned the comparison group in the Rigter et al. (2013) overarching multi-site trial to the CBT category, because in all sites the comparison group consisted of either CBT alone or CBT complemented with other treatment approaches. CT was coded if more than one treatment module was combined. The following

combinations were found: CBT and motivational enhancement therapy sessions (Dennis et al., 2004), CBT with GT and family interventions (The Adolescent Community Reinforcement Approach, ACRA; Dennis et al., 2004), CBT, motivation enhancement therapy sessions, and family interventions (Family Support Network, FSN; Dennis et al., 2004), and CBT and GT (Residential Substance Abuse Treatment, RST; Liddle, 2015). Finally, treatment duration was collected.

For the research design characteristics, we coded whether studies were conducted by the developers of the treatment or by others (developers, non-developers), to test the assumption that studies carried out by the developers yield higher effect sizes. In this category, overall sample size, treatment group size, comparison group size, and study follow-up duration were analysed as well. For the manuscript characteristics, we coded the year of publication. If the manuscript had not been published, we used the year that the manuscript was written. Finally, the impact factor of the journal in which the manuscript was published was inventoried.

Inter-rater reliability

The first and third author coded the effect sizes and study characteristics. Reliability of the coding scheme was examined by having a subset of the study characteristics coded by two research assistants. Ten manuscripts were randomly selected. Inter-rater agreement was analysed for each of the study outcomes and study characteristics by calculating the percentage of agreement for all study characteristics, Kappa for categorical variables and intraclass correlation for interval and ratio variables. The inter-rater reliability was good, with Kappa's ranging from 0.93 (93% agreement) for comparison group to 1.00 for outcome, geographic location and independence of researchers (100% agreement); intraclass correlations ranged from 0.96 for follow up period (91% agreement) to 1.00 for effect size (91% agreement), SES (91% agreement), average age (100% agreement), and percentage of males (100% agreement).

Analyses

For each study outcome, a Cohen's *d* effect size was coded or calculated. When not provided, formulae provided by Lipsey and Wilson (2001) to transform test statistics into Cohen's *d* or to calculate *d* on the basis of

means and standard deviations, were used. Effect sizes of d = 0.20, d = 0.50 and d = 0.80 were considered as small, medium and large group differences respectively, whereas d = 0.00 would indicate no difference between the experimental and comparison groups (Cohen, 1988). Using standardized z-values larger than 3.29 or smaller than -3.29, (Tabachnick & Fidell, 1989) no outliers were identified. Each continuous moderator variable was centred around its mean. For the categorical variables we made dichotomous dummy variables. The extent of the variation in effect sizes was examined by conducting a test for homogeneity of effect sizes.

Independence of study results is desirable when conducting a meta-analysis in order to prevent a particular study being weighted more strongly than others (Lipsey & Wilson, 2001; Mullen, 2013; Rosenthal, 1991). To deal with dependency of study results, we applied a three-level random effects model (Cheung, 2014; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). This model accounts for three sources of variance: sampling variance (level 1 variance), variance between effect sizes from the same study (level 2 variance), and variance between studies (level 3 variance (Hox, 2002; Van den Noortgate et al., 2013)). A three-level random effects model therefore accounts for the hierarchical structure of the data in which the effect sizes or study results (the lowest level) are nested within studies (the highest level). A likelihood ratio test was used to examine between-study and within-study heterogeneity (Raudenbush & Bryk, 2002).

Moderator analyses were conducted by extending the model with study and effect size characteristics. For these models including moderators, an omnibus test of the fixed-model parameters was conducted, which tests the null hypothesis that the group mean effect sizes are equal. The Knapp and Hartung (2003) adjustment was applied to control for Type I error rates. We used the metafor package (Viechtbauer, 2010) for the R environment (Version 3.2.3; R Development Core Team, 2015) for modelling a three-level random effects model as described by Van den Noortgate et al. (2013). Parameters were estimated using the restricted maximum likelihood procedure.

RESULTS

The 19 manuscripts included in the meta-analysis reported on 8 studies and presented 61 effect sizes. These studies examined 1,488 adolescents in total, of whom 699 received MDFT, and 789 cognitive behavioural therapy (CBT), group therapy (GT), or combined treatments (CT). The effect sizes from the individual studies ranged from d = -0.62 to 1.16. An overview of the characteristics of the 19 manuscripts and the 61 effect sizes is presented in Table 1.

Results indicated that the overall mean effect size for MDFT was beneficial compared to adolescents receiving another form of therapy, $d = 0.24$, $p < 0.01$. For effect sizes, variance between effect sizes within studies (level 2 variance), $\sigma^2 = 0.012$, $\chi^2(1) = 23.00$, $p = 0.14$, was nonsignificant, whereas variation between studies (level 3 variance), $\sigma^2 = 0.048$, $\chi^2(1) = 32.77$, $p < 0.001$ was significant resulting in the examination of the extent to which potential moderators explained effect size variability.

Table 1. Description of Major Characteristics of Studies Used in the Meta-analysis.

no.	study	year	N	N target	N control	comparison condition	study ID	age mean	follow up	% males	% minorities	% conduct disorder	% severe cannabis users	outcome measure	effect size
1	Dakof et al.	2015	112	55	57	Group	1	16.00	18	88	100	52	48	substance abuse	0.05
2	-	2015	54	24	30	Group	1	16.26	18	-	100	-	100	substance abuse	0.09
3	-	2015	58	31	27	Group	1	15.90	18	-	100	-	0	substance abuse	0.01
4	-	2015	112	55	57	Group	1	16.00	18	88	100	52	48	delinquency	0.14
5	-	2015	112	55	57	Group	1	16.00	18	88	100	52	48	externalising	0.21
6	-	2015	112	55	57	Group	1	16.00	18	88	100	52	48	internalising	0.24
7	Liddle et al.	draft	113	57	56	CT	2	15.36	12	75	88	77	70	substance abuse	0.05
8	-	draft	79	43	36	CT	2	15.33	12	-	-	-	100	substance abuse	0.29
9	-	draft	34	14	20	CT	2	15.44	12	-	-	-	0	substance abuse	-0.62
10	-	draft	113	57	56	CT	2	15.36	12	75	88	77	70	delinquency	0.13
11	-	draft	113	57	56	CT	2	15.36	12	75	88	77	70	externalising	0.20
12	-	draft	113	57	56	CT	2	15.36	12	75	88	77	70	internalising	0.11
13	Schaub et al.	2014	450	212	238	CBT	3	16.30	6	85	40	-	48	externalising	0.05
14	-	2014	450	212	238	CBT	3	16.30	6	85	40	-	48	internalising	0.10

no.	study	year	N	N target	N control	comparison condition	study ID	age mean	follow up	% males	% minorities	% conduct disorder	% severe cannabis users	outcome measure	effect size
15	-	2014	450	212	238	CBT	3	16.30	6	85	40	-	48	family	0.18
16	Rigter et al.	2013	450	212	238	CBT	3	16.30	6	85	40	-	48	substance abuse	0.25
17	-	2013	237	108	129	CBT	3	16.39	6	-	-	-	100	substance abuse	0.35
18	-	2013	213	104	109	CBT	3	16.09	6	-	-	-	0	substance abuse	0.11
19	Hendriks et al.	2012	109	55	54	CBT	3	16.80	6	80	28	29	51	externalising	0.25
20	-	2012	109	55	54	CBT	3	16.80	6	80	28	29	51	internalising	0.42
21	Hendriks et al.	2011	109	55	54	CBT	3	16.80	6	80	28	29	51	substance abuse	0.14
22	-	2011	50	28	22	CBT	3	17.03	6	-	-	29	100	substance abuse	0.41
23	-	2011	53	24	29	CBT	3	17.03	6	-	-	29	0	substance abuse	-0.04
24	-	2011	37	21	16	CBT	3	16.80	6	-	-	84	57	substance abuse	1.16
25	-	2011	109	55	54	CBT	3	16.80	6	80	28	29	51	delinquency	0.00
26	Liddle et al.	2011	154	76	78	CT	4	15.40	3	83	84	43	33	delinquency	0.30
27	Phan	2011	101	38	63	CT	3	16.29	6	89	33	-	57	substance abuse	0.14

no.	study	year	N	N target	N control	comparison condition	study ID	age mean	follow up	% males	% minorities	% conduct disorder	% severe cannabis users	outcome measure	effect size
28	-	2011	58	20	38	CT	3	16.48	6	89	33	-	100	substance abuse	-0.19
29	-	2011	43	18	25	CT	3	16.02	6	89	33	-	0	substance abuse	0.38
30	Grichting	2011	60	30	30	CT	3	16.07	6	92	67	-	33	substance abuse	0.00
31	-	2011	20	8	12	CT	3	16.15	6	92	67	-	100	substance abuse	0.28
32	-	2011	40	22	18	CT	3	16.03	6	92	67	-	0	substance abuse	-0.21
33	Tossmann	2010	120	59	61	CT	3	16.21	6	83	30	-	55	substance abuse	0.51
34	-	2010	66	33	33	CT	3	16.42	6	83	30	-	100	substance abuse	0.70
35	-	2010	54	26	28	CT	3	15.96	6	83	30	-	0	substance abuse	0.26
36	Verbanck	2010	60	30	30	CBT	3	16.60	6	93	37	-	65	substance abuse	0.65
37	-	2010	39	19	20	CBT	3	16.67	6	93	37	-	100	substance abuse	0.53
38	-	2010	21	11	10	CBT	3	16.48	6	93	37	-	0	substance abuse	0.83

no.	study	year	N	N target	N control	comparison condition	study ID	age mean	follow up	% males	% minorities	% conduct disorder	% severe cannabis users	outcome measure	effect size
39	Henderson et al. I	2010	45	22	23	CBT	5	15.53	12	-	-	-	100	substance abuse	0.51
40	-	2010	178	90	88	CBT	5	15.36	12	81	-	-	0	substance abuse	0.33
41	Henderson et al. II	2010	50	25	25	CT	4	15.52	3	-	-	-	100	substance abuse	0.23
42	-	2010	104	51	53	CT	4	15.36	3	-	-	-	0	substance abuse	-0.34
43	Henderson et al.	2009	83	40	43	Group	6	13.73	6	74	84	39	2	family	0.44
44	Liddle et al.	2009	83	40	43	Group	6	13.73	6	74	97	39	2	substance abuse	1.07
45	-	2009	83	40	43	Group	6	13.73	6	74	97	39	2	delinquency	0.31
46	-	2009	83	40	43	Group	6	13.73	6	74	97	39	2	internalising	0.54
47	Liddle et al.	2008	224	112	112	CBT	5	15.40	12	81	97	50	21	substance abuse	0.59
48	Hogue et al.	2008	136	74	62	CBT	5	15.50	6	81	80	50	-	substance abuse	0.47
49	-	2008	136	74	62	CBT	5	15.50	6	81	80	50	-	externalising	0.56
50	-	2008	136	74	62	CBT	5	15.50	6	81	80	50	-	internalising	0.76
51	Hogue et al.	2004	51	25	26	CBT	5	15.20	12	67	75	69	-	substance abuse	0.47

no.	study	year	N	N target	N control	comparison condition	study ID	age mean	follow up	% males	% minorities	% conduct disorder	% severe cannabis users	outcome measure	effect size
52	-	2004	51	25	26	CBT	5	15.20	12	67	75	69	-	externalising	0.62
53	-	2004	51	25	26	CBT	5	15.20	12	67	75	69	-	internalising	0.74
54	Dennis et al. I	2004	200	100	100	CT	7	16.00	9	83	39	53	-	substance abuse	-0.06
55	Dennis et al. II	2004	200	100	100	CT	7	16.00	9	83	39	53	-	substance abuse	-0.26
56	Liddle et al. I	2001	152	47	105	Group	8	15.90	12	80	49	-	1	substance abuse	0.25
57	-	2001	152	47	105	Group	8	15.90	12	80	49	-	1	externalising	-0.10
58	-	2001	152	47	105	Group	8	15.90	12	80	49	-	1	family	0.61
59	Liddle et al. II	2001	152	47	105	CT	8	15.90	12	80	49	-	1	substance abuse	0.85
60	-	2001	152	47	105	CT	8	15.90	12	80	49	-	1	externalising	0.35
61	-	2001	152	47	105	CT	8	15.90	12	80	49	-	1	family	0.31

Note. CBT = Cognitive Behavioural Therapy, CT = Combined Treatment, Group = Group Therapy.

Moderator analyses

Table 2 summarises the results of the moderator analyses. Two moderators yielded a positive contribution to effect size. Percentage of severe substance abusers in the study sample was associated with larger effects favouring MDFT, $F(1,45) = 6.150$, $p = 0.017$. This suggests that adolescents with more severe substance abuse benefit more from MDFT than from the comparison treatments. In addition, percentage of DBD was positively related to the effect size, $F(1,5) = 14.072$, $p = 0.013$, indicating that samples with higher percentages of DBD responded better to MDFT. Year of publication yielded a trend, $F(1,59) = 3.638$, $p = 0.061$, showing relatively smaller effects in newer studies.

The effect sizes for the outcome measures substance abuse, delinquency, externalising, and internalising psychopathology, and family functioning, were found to be in the same range, all indicating a small incremental effect over other established treatments with no significant differences between the effect sizes for the five outcome categories. Furthermore, for treatment groups, no significant differences in effect size were found between studies that compared MDFT with CBT and studies that compared MDFT with CT, respectively GT. The geographic location where studies were conducted (i.e., Europe versus United States) had no impact on study results. Studies led by the developers of MDFT had similar outcomes as those led by independent researchers. No moderating effects were found for adolescents' age, gender, SES, ethnic background, duration of therapy, and duration of the follow-up period. Moreover, the rates of depression, GAD, ADHD, CD, and ODD in the sample, and percentage of cannabis, alcohol, and other drugs were not associated with effect size. Finally, study sample sizes and impact factor had no moderating effect.

Table 2. Results for Moderators.

Moderator variables	# Studies	# ES	β0, mean d a (95% CI)	β1 (95% CI)	Omnibus test	Variance level 2b	Variance level 3c
Sample descriptors							
Outcome characteristics					F(4,56) = 0.383	0.015	0.049***
substance abuse (rc)	8	37	0.252 (0.069; 0.436)**				
delinquency	5	5	0.212 (-0.049; 0.473)	0.041 (-0.273; 0.192)			
externalising	5	8	0.168 (-0.058; 0.393)	0.085 (-0.261; 0.092)			
internalising	5	7	0.297 (0.063; 0.531)*	0.044 (-0.143; 0.232)			
family functioning	3	4	0.252 (-0.015; 0.520)	0.000 (-0.230; 0.230)			
Geographic location					F(1,59) = 0.003	0.011	0.061***
United States (rc)	7	36	0.226 (-0.277; 0.728)				
Europe	1	25	0.240 (0.037; 0.443)*	0.015 (-0.527; 0.557)			
Independence researchers					F(1,59) = 0.003	0.011	0.061***
yes (rc)	5	38	0.226 (-0.277; 0.728)				
no	6	23	0.240 (0.037; 0.443)*	0.015 (-0.527; 0.557)			
Age of the subjects	8	61	0.224 (0.052; 0.397)*	0.058 (-0.240; 0.123)	F(1,59) = 0.521	0.012	0.045***
Socioeconomic status (SES)	5	23	0.410 (0.212; 0.609)***	0.000 (-0.000; 0.000)	F(1,21) = 0.626	0.018	0.029***
% Males	8	52	0.262 (0.102; 0.423)**	0.856 (-2.501; 0.789)	F(1,50) = 1.093	0.011	0.039***
% Caucasians	8	54	0.281 (0.080; 0.481)**	0.145 (-0.469; 0.759)	F(1,52) = 0.225	0.008	0.062***
% African-Americans	8	42	0.243 (0.076; 0.411)**	0.429 (-0.299; 1.158)	F(1,40) = 1.419	0.005	0.042***

Moderator variables	# Studies	# ES	β0, mean d a (95% CI)	β1 (95% CI)	Omnibus test	Variance level 2b	Variance level 3c
% Hispanics	8	42	0.262 (0.071; 0.454)**	0.040 (-0.709; 0.790)	F(1,40) = 0.012	0.005	0.058***
% Asians	8	49	0.232 (0.048; 0.416)*	2.676 (-6.430; 11.782)	F(1,47) = 0.349	0.008	0.058***
% Others	8	43	0.226 (0.045; 0.408)*	0.585 (-1.812; 0.641)	F(1,41) = 0.929	0.009	0.053***
% Cultural minority	8	40	0.256 (0.075; 0.437)**	0.139 (-0.555; 0.834)	F(1,38) = 0.165	0.006	0.051***
% CD	7	29	0.272 (0.018; 0.525)*	0.736 (-0.234; 1.705)	F(1,27) = 2.421	0.000	0.096***
% ODD	3	18	0.349 (0.082; 0.616)*	0.271 (-0.879; 1.421)	F(1,16) = 0.250	0.000	0.040
% DBD	2	7	0.494 (0.317; 0.670)***	1.371 (0.432; 2.311)*	F(1,5) = 14.072	0.000	0.000
% ADHD	5	15	0.199 (-0.088; 0.487)	1.118 (-5.008; 2.773)	F(1,13) = 0.385	0.000	0.075**
% GAD	4	14	0.223 (-0.202; 0.648)	0.051 (-3.533; 3.430)	F(1,12) = 0.001	0.000	0.125***
% Depression	6	22	0.273 (-0.008; 0.553)	0.320 (-3.111; 3.751)	F(1,20) = 0.038	0.000	0.090***
% Cannabis	7	54	0.221 (0.040; 0.402)*	0.236 (-0.993; 0.521)	F(1,52) = 0.393	0.005	0.047***
% Alcohol	5	18	0.210 (-0.075; 0.472)	0.490 (-1.899; 0.919)	F(1,16) = 0.543	0.000	0.079***
% Other drugs	5	18	0.200 (-0.108; 0.508)	0.413 (-3.727; 2.902)	F(1,16) = 0.070	0.000	0.092***
% Severe substance abuse	7	47	0.282 (0.067; 0.496)*	0.264 (0.050; 0.479)*	F(1,45) = 6.150	0.001	0.059**
Treatment descriptors							
Comparison condition					F(2,58) = 0.218	0.012	0.060***
Cognitive behavioural therapy (CBT)	2	32	0.281 (0.034; 0.529)*				
Combined treatment (CT) (rc)	4	16	0.257 (0.044; 0.469)	0.025 (-0.222; 0.172)			
Group therapy (GT)	3	13	0.178 (-0.087; 0.444)	0.103 (-0.420; 0.214)			

Moderator variables	# Studies	# ES	β0, mean d a (95% CI)	β1 (95% CI)	Omnibus test	Variance level 2b	Variance level 3c
Duration of treatment in months	8	61	0.263 (0.102; 0.423)**	0.059 (-0.035; 0.154)	F(1,59) = 1.579	0.011	0.039***
Research design							
Total sample size	8	61	0.246 (0.077; 0.416)**	0.000 (-0.001; 0.000)	F(1,59) = 2.448	0.009	0.048***
Sample size treatment group	8	61	0.247 (0.079; 0.414)**	0.001 (-0.002; 0.000)	F(1,59) = 2.374	0.009	0.047***
Sample size comparison group	8	61	0.246 (0.076; 0.417)**	0.001 (-0.002; 0.000)	F(1,59) = 2.491	0.009	0.049***
Follow-up (in months)	8	46	0.237 (0.053; 0.422)*	0.012 (-0.043; 0.019)	F(1,44) = 0.573	0.010	0.057***
Manuscript characteristics							
Year of publication	8	61	0.213 (0.015; 0.410)*	0.031 (-0.063; 0.001)	F(1,59) = 3.638	0.006	0.068***
Impact factor	8	61	0.240 (0.062; 0.417)**	0.006 (-0.048; 0.036)	F(1,59) = 0.084	0.012	0.053***

Note. # studies = number of independent studies; # ES = number of effect sizes; mean d = mean effect size; CI = confidence interval; rc = reference category.

[a] For continuous predictors, the mean effect size indicates the mean effect size of a participant with an average value on the corresponding predictor.

[b] Variance between the effect sizes from the same study.

[c] Variance between studies.

$^{+}$ $p < 0.1$ * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

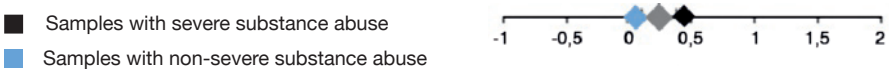
Model with multiple moderators

To examine the unique contribution of each moderator to the variance in effect size, a model with multiple moderators was tested. Variables associated with effect sizes with a $p < 0.20$ in the bivariate moderator analyses reported above were entered in the model. To retain sufficient power in the model with multiple moderators, only the variables for which the number of effect sizes was at least $k = 30$ were included. The following variables were included: percentage of adolescents with severe substance abuse, sample size, and year of publication. The model was found to be significant, $F(3,43) = 5.779$, $p = 0.002$, $k = 47$. Two moderators were significant predictors of effect size: severe substance abuse, $\beta = 0.26$, $p = 0.016$, and year of publication, $\beta = -0.09$, $p = 0.002$. Thus, studies with a larger proportion of subjects with severe substance abuse and older studies yielded larger effect sizes, favouring MDFT. To illustrate the effect of severe substance abuse in samples, Table 3 (Neyeloff, Fuchs, & Moreira, 2012) includes a forest plot that depicts studies with low (0%), moderate (1-99%), and high (100%) severe substance abusers. The forest plot illustrates that in general, MDFT generated larger effect sizes for samples with a higher percentage of severe cannabis users. The computed mean effect sizes for relatively low, moderate, and high severe substance abusers showed that effects of MDFT were non-significant for non-severe substance abusers ($d = 0.09$), small for moderate abusers ($d = 0.28$) and small to moderate for severe substance abusers ($d = 0.38$).

Table 3. Foster plot of individual effect sizes 95% confidence intervals.

No.	Study	Year	Effect Size	95% CI		Forest Plot
1	Dakof et al.	2015	0.05	-0.32	0.42	
2	-	2015	0.09	-0.28	0.46	
3	-	2015	0.01	-0.36	0.38	
4	-	2015	0.14	-0.31	0.59	
5	-	2015	0.21	-0.50	0.92	
6	-	2015	0.24	-0.13	0.61	
7	Liddle et al.	draft	0.05	-0.03	0.13	
8	-	draft	0.29	-0.08	0.66	
9	-	draft	-0.62	-0.99	-0.25	
10	-	draft	0.13	0.05	0.21	
11	-	draft	0.20	0.12	0.28	
12	-	draft	0.11	-0.26	0.48	
13	Schaub et al.	2014	0.05	-0.85	0.95	
14	-	2014	0.10	-0.55	0.75	
15	-	2014	0.18	-0.33	0.69	
16	Rigter et al.	2013	0.25	0.07	0.43	
17	-	2013	0.35	0.17	0.53	
18	-	2013	0.11	-0.07	0.29	
19	Hendriks et al.	2012	0.25	0.07	0.43	
20	-	2012	0.42	0.17	0.67	
21	Hendriks et al.	2011	0.14	-0.13	0.41	
22	-	2011	0.41	0.04	0.78	
23	-	2011	-0.04	-0.41	0.33	
24	-	2011	1.16	0.79	1.53	
25	-	2011	0.00	-0.37	0.37	
26	Liddle et al.	2011	0.30	-0.27	0.87	
27	Phan	2011	0.14	-0.23	0.51	
28	-	2011	-0.19	-0.68	0.30	
29	-	2011	0.38	-0.15	0.91	
30	Grichting	2011	0.00	-0.53	0.53	
31	-	2011	0.28	-0.23	0.79	
32	-	2011	-0.21	-0.58	0.16	
33	Tossmann	2010	0.51	0.00	1.02	

No.	Study	Year	Effect Size	95% CI		Forest Plot
34	-	2010	0.70	-0.20	1.60	
35	-	2010	0.26	-0.37	0.89	
36	Verbanck	2010	0.65	0.24	1.06	
37	-	2010	0.53	-0.02	1.08	
38	-	2010	0.83	0.22	1.44	
39	Henderson et al. I	2010	0.51	-0.06	1.08	
40	-	2010	0.33	-0.22	0.88	
41	Henderson et al. II	2010	0.23	-0.36	0.82	
42	-	2010	-0.34	-0.63	-0.05	
43	Henderson et al.	2009	0.44	-0.01	0.89	
44	Liddle et al.	2009	1.07	0.72	1.42	
45	-	2009	0.31	0.04	0.58	
46	-	2009	0.54	0.27	0.81	
47	Liddle et al.	2008	0.59	0.16	1.02	
48	Hogue et al.	2008	0.47	0.04	0.90	
49	-	2008	0.56	0.13	0.99	
50	-	2008	0.76	0.49	1.03	
51	Hogue et al.	2004	0.47	0.14	0.80	
52	-	2004	0.62	0.27	0.97	
53	-	2004	0.74	0.39	1.09	
54	Dennis et al. I	2004	-0.06	-0.41	0.29	
55	Dennis et al. II	2004	-0.26	-0.61	0.09	
56	Liddle et al. I	2001	0.25	-0.10	0.60	
57	-	2001	-0.10	-0.45	0.25	
58	-	2001	0.61	0.26	0.96	
59	Liddle et al. II	2001	0.85	0.28	1.42	
60	-	2001	0.35	-0.36	1.06	
61	-	2001	0.31	-0.24	0.86	



DISCUSSION

The purpose of this meta-analysis was first to examine the effectiveness of MDFT, compared to other (active) treatments, and second to inventory the effects of severe behaviour problems and other potential moderators. Overall, compared to other treatments and across outcome categories, MDFT showed a significant effect size, d = 0.24, which corresponds to a success rate difference (SRD), (Kraemer & Kupfer, 2006) of approximately 13%. These findings supporting the effectiveness of MDFT are in line with the meta-analyses of other multiple-systems-based treatment, such as multisystemic therapy (MST) (Van der Stouwe, Asscher, Stams, Deković, & Van der Laan, 2014). In addition, MDFT was found to be most effective in adolescents with severe substance abuse and/or disruptive behaviour disorder (DBD).

This 'severity gradient' supported by our finding that MDFT is more effective for those with high severity problems, such as severe substance abuse, is in line with previous research, showing that adolescents with severe cannabis abuse (Rigter et al., 2013) and severe cannabis or substance abusers with comorbid externalising psychopathology benefit most from MDFT (C.E. Henderson et al., 2010; Hendriks, Van der Schee, & Blanken, 2012). This is not surprising, as the treatment goals of MDFT have been designed to serve a broad, heterogeneous group of adolescents with substance use disorders and diverse and complex behaviour problems (C.E. Henderson et al., 2010; Weisz & Kazdin, 2010). Over the years different versions of MDFT have been designed and tested in different countries, in samples with different ages, gender, psychopathology, and in different setting, including clinical and juvenile justice settings. From our findings it seems that MDFT is effective in a variety of settings and for different adolescents, however, the largest effects are found for those with high severity problems. Our finding is consistent with the risk principle of the Risk-Need-Responsivity (RNR)-model, (Andrews et al., 1990; Andrews et al., 2006, 2011) which states that the intensity of interventions should match recidivism risk: those with increased recidivism risk (i.e., with more severe conditions) should receive more intensive treatment. Our findings support the notion that treatment

effectiveness of intensive, comprehensive treatment programmes is better for severely affected youths. Specifically, for MDFT this means that although MDFT is applicable for a broad spectrum of problems, the treatment appears to have surplus value for the most severely impaired youth.

In the model with multiple moderators, an effect of year of publication was found. In early publications, effect sizes for MDFT were larger than in later publications. One possible explanation for this finding would be the "decline effect", a term coined by Ioannidis (2005). He stated that early research is usually small and may be more likely to produce positive results supporting the hypotheses examined than later, larger studies, in which regression to the mean might occur. However, given that we did not find a moderating effect of sample size, this explanation is not likely. It is more likely that confounding moderator, not examined in this meta-analysis, may explain the effect of publication year. Although we have coded many study characteristics, data on features of the intervention, such as different versions of MDFT, or levels of treatment integrity were not available, and therefore we did not examine these potential moderating characteristics.

Further, effects of MDFT on different treatment outcomes, including substance abuse, delinquency, externalising, and internalising psychopathology, and family functioning were about equal in effect size. This suggests that MDFT affects a broad range of domains which may be explained by the multi-focussed approach of MDFT (Liddle, 2002; Liddle & Rigter, 2013). An important finding, enhancing the applicability of MDFT is that this therapy appeared to be similarly effective for boys and girls and for adolescents with different ages, SES and ethnic background, as these were no significant moderators of the effectiveness of MDFT. With regard to age, this is not consistent with an earlier study, which found MDFT to be more effective when the intervention was aimed at younger adolescents, (Hendriks et al., 2011) however, this study has a relatively small sample size, not representative compared to the current meta-analysis. Some studies postulate the development of specific interventions aimed at girls, (e.g. Hipwell & Loeber, 2006) the present meta-analysis found that MDFT is beneficial for a varied group of male and female adolescents from different ethnic backgrounds.

To our knowledge, this is the first meta-analysis on MDFT, using three-level analytic techniques. This novel three-level analytic method makes it possible to study the influence of moderators more extensively and increases statistical power, which allowed us to test the described severity gradient. Another strength of the present meta-analysis is that we only included randomised control trials (RCTs) comparing MDFT with other evidence based, effective therapies, which is considered to be the most robust research design and best equipped to handle threats to a study's internal validity (Weisburd, Lum, & Petrosino, 2001; Welsh, Peel, Farrington, Elffers, & Braga, 2011). Notwithstanding the strength of the present meta-analysis, our findings should be interpreted in the context of some limitations. First, there is a lack of studies that examined family functioning as an outcome measure. Family functioning is considered to be a major focus in the treatment model for MDFT (Dakof, Cohen, & Duarte, 2009). Therefore, more studies regarding family functioning are necessary. Second, although a RCT is considered to be the best research design, there are scholars postulating that due to the selection procedure of RCTs, we should be cautious to generalise the findings in experimental settings to routine youth care (Holly Barrett Waldron & Charles W Turner, 2008). Within clinical samples, there is generally much heterogeneity in adolescent characteristics (e.g., age, substance abuse, delinquency, psychiatric comorbidity). Therefore, adolescent subgroups, within these clinical samples, may differ considerably in treatment outcome (Chan, Dennis, & Funk, 2008; Daudin et al., 2010). Finally, in the current meta-analysis we were unable to examine various types of criminal behaviour, which could generate additional insight. In the five studies that reported delinquency, only one study analysed the influence of MDFT on various types of criminal behaviour (e.g., person crimes, theft, etc.) (Dakof et al., 2015).

For future research we strongly suggest other established treatments addressing substance abusing adolescents with comorbid behaviour problems to test the severity gradient for substance abuse, externalising disorders and possible other important variables, to be able to better match treatment with the characteristics of an adolescent (Bell, Marcus, & Goodlad, 2013; Leijten et al., 2015). Specific for MDFT, one of the directions of future research should be to intensively investigate family

functioning as a moderator of the effectiveness of MDFT. Some studies addressed this quintessential topic for MDFT (Henderson, Rowe, Dakof, Hawes, & Liddle, 2009; Schmidt, Liddle, & Dakof, 1996; Shelef, Diamond, Diamond, & Liddle, 2005). Nevertheless, more research on family functioning is necessary. A further research topic of interest is to study the impact of MDFT on different indices of criminal behaviours (Dakof et al., 2015). This type of research could provide more precise information for which type of adolescents MDFT is the most effective. Moreover, MDFT is an intensive treatment, which is considered to be more expensive than most alternative therapies, and therefore, conducting cost-effectiveness studies carries substantial relevance.

Practical implications of the present meta-analysis are that treatment delivery systems should aim to provide different treatment modules matching the severity of problem behaviours of the youth. MDFT has addressed this issue extensively, by developing diverse modules and researching varied subgroups of adolescents (S. A. Brown & Zucker, 2015; Weisz & Kazdin, 2010); most other treatments targeting this heterogeneous group of adolescents are advised to follow suit. The feasibility of this suggestion can be debated; however, for society the improvement of the quality of care for this group of adolescents is of major importance.

Finally, MDFT, although suitable for a broad spectrum of adolescents with behaviour problems, may be most suitable for adolescents with severe problems, severe substance abuse and disruptive behaviour disorder in particular. Furthermore, this finding could indicate that other less intensive and expensive treatments, for example individual CBT, may be as appropriate for addressing SUD and comorbid psychopathology in adolescents with less severe problem behaviour.

In summary, we conclude that MDFT has an incremental, 13 % advantage over other established treatments. As a unique asset, MDFT can be successfully deployed in male and female adolescents from diverse ethnic backgrounds in a variety of settings, with SUD, delinquency, and diverse comorbid conditions, notwithstanding their age. Furthermore, MDFT was

found to be more effective for adolescents with severe problem behaviour. As such, MDFT can be regarded as a valuable therapy, especially when treating the most challenging group of youth.