



UvA-DARE (Digital Academic Repository)

To treat or not to treat?

ter Beek, E.

[Link to publication](#)

Citation for published version (APA):

ter Beek, E. (2018). To treat or not to treat? Harmful sexual behavior in adolescence: Needs before risk

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Treatment Effect on Recidivism for Juveniles who have Sexually Offended: a Multilevel Meta-Analysis

Ter Beek, E.
Spruit, A.
Kuiper, C.H.Z.
Van der Rijken, R.E.A.
Hendriks, J.
Stams, G.J.J.M.

Journal of Abnormal Child Psychology.
Advance Online Publication. doi: 10.1007/s10802-017-0308-3



ABSTRACT

The current study investigated the effect on recidivism of treatment aimed at juveniles who have sexually offended. It also assessed the potential moderating effect of type of recidivism, and several treatment, participant and study characteristics. In total, 14 published and unpublished primary studies, making use of a comparison group and reporting on official recidivism rates, were included in a multilevel meta-analysis. This resulted in the use of 77 effect sizes, and 1,726 participants. A three-level meta-analytic model was used to calculate the combined effect sizes (Cohens d) and to perform moderator analyses. Study quality was assessed with the EPHPP Quality Assessment Tool for Quantitative Studies. A moderate effect size was found ($d = 0.37$), indicating that the treatment groups achieved an estimated relative reduction in recidivism of 20.5% as compared to comparison groups. However, after controlling for publication bias, a significant treatment effect was no longer found.

Type of recidivism did not moderate the effect of treatment, indicating that treatment groups were equally effective for all types of recidivism. Also, no moderating effects of participant or treatment characteristics were found. Regarding study characteristics, a shorter follow up time showed a trend for larger effect sizes, and the effect size calculation based on proportions yielded larger effect sizes than calculation via mean frequency of offending. Implications for future research and clinical practice are discussed.

INTRODUCTION

An interest in treatment for juveniles who have sexually offended was first sparked in the late 1970's. Research on adults who had sexually offended showed that they had often started their abusing careers in adolescence (Groth, Longo, & McFaddin, 1982). In accordance with findings from research on adult offenders, it was assumed that dynamic risk factors explaining juvenile sex offending would differ from dynamic risk factors explaining non-sexual juvenile delinquency, and that specific treatment for juveniles who have sexually offended would be needed to target the dynamic risk factors for sexual recidivism (Barbaree & Marshall, 2006). Recent research, however, shows that juveniles who have sexually offended differ psychologically from adult sex offenders (Adjorlolo & Egbenya, 2016). Notably, in general, dynamic risk factors for juvenile sex-offense recidivism do not always differ from risk factors for juvenile general offense recidivism (Carpentier & Proulx, 2011; Christiansen & Vincent, 2013, Worling & Långström, 2003).

Sexual (re)offending by juveniles remains mainly adolescence limited (Beaudry-Cyr, Jennings, Zgoba, & Tewksbury, 2015; Loeber, Hoeve, Slot, & Van Der Laan, 2012; Lussier, Van Den Berg, Bijleveld, & Hendriks, 2012). In 2016, Caldwell conducted a meta-analysis, and found a weighted mean of 5% official recidivism by juveniles. This probably is an underestimation of the true recidivism rate due to the rarity of documented sexual offending (Wittebrood, 2006). Interestingly, the reoffending base rate has significantly declined to a weighted mean of 3% in the last 15 years. Improved quality of treatment was hypothesized to be the most likely explanation (Caldwell, 2016). Research on recidivism reduction through treatment of juveniles who have sexually offended, however, is highly diverse, differing in study design, outcome, treatment and participant characteristics, which is thought to affect study findings (Dopp, Borduin, & Brown, 2015; Hanson, Bourgon, Helmus, & Hodgson, 2009). The current study aims to synthesize the results of earlier studies by testing the hypothesis that specialized treatment for juveniles is effective.

Treatment Components

The Risk-Need-Responsivity (RNR) model (Bonta & Andrews, 2007), prominently states that treatment allocation and delivery should be guided by specific criminogenic treatment needs and recidivism risk levels, and should be delivered responsively (i.e., attuned to personal life goals, learning and motivational style) in order to be effective. As a consequence, aiming treatment at established risk factors for (sexual) reoffending is considered to be imperative (Ter Beek et al., 2017a).

Juveniles who generally display antisocial tendencies, also known as mixed offenders, who have also displayed other forms of offending behavior, seem most at risk for recidivism, including sexual reoffending (Drew, 2013; Hendriks & Bijleveld, 2005; Lawing, Frick, & Cruise, 2010; Van Wijk, Vreugdenhil, Van Horn, Vermeiren, & Doreleijers, 2007). Overall, juveniles who have sexually offended display a fair amount of non-sexual recidivism (Caldwell, 2016; Lobanov-Rostovsky, 2015). Specialized juvenile treatment may therefore need to address risk factors for general delinquency, such as immaturity, impulsiveness, and antisocial beliefs, to effectively reduce sexual as well as general recidivism risk. Seto and Lalumière (2010) and Fanniff and Kimonis (2014), comparing juveniles who have sexually offended with juveniles who have committed non-sexual offenses only, found both similarities (mainly their level of antisocial thoughts and behavior) and differences. Juveniles who had sexually offended showed more atypical sexual interests, more social and emotional problems, and more experiences of sexual victimization, abuse and neglect than did non-sex offenders. These findings suggest that general etiological theories of antisocial behavior might only partially explain juvenile sexual misconduct, and that specific or additional explanations remain relevant.

Christiansen and Vincent's study (2013) of 39,249 juveniles, of whom 695 juveniles had offended sexually, showed prior non-sexual offending, non-contact sexual offenses, offending against a child, early age of onset, poor school performance, and non-school attendance to be predictive of sexual recidivism in juveniles. In 2011, Carpentier and Proulx examined 351 male adolescents who had sexually offended, and they found paternal abandonment, childhood sexual victimization, association with younger children, and having victimized a stranger to be associated with a higher risk for sexual recidivism. Prior to these studies, Worling and Långström (2003) conducted a literature review of juvenile sex offending, which provided empirical evidence for the following risk factors: sexual deviation (e.g., interest in prepubescent children or sexual violence), prior criminal sanctions for sexual assault(s), having two or more victims, victimization of strangers, lack of intimate peer relationships/social isolation, and incomplete offense-specific treatment. Notably, the dynamic (changeable) risk factors, such as sexual deviation, are of most interest for recidivism risk reduction through treatment (Beggs & Grace, 2011; Hanson & Morton-Bourgon, 2005; Pedersen, Rasmussen, & Elsass, 2010).

Although support is found for dynamic risk factors that need to be addressed during treatment, most of the empirically established risk factors for juvenile sexual recidivism (e.g., number of previous victims, prior offending behavior, having offended against a child or a stranger, early age of onset) are considered static (unchangeable). Successful treatment however, may also focus on reduction of negative consequences of static risk

factors (for example, by means of trauma treatment to alleviate the consequences of earlier victimization). Also non-criminogenic (mental health) treatment needs should be addressed (Lord, 2016).

Studies into treatment needs of juveniles who have sexually offended, have substantiated the existence of several psychological typologies. Studies into specific mental illnesses, executive functioning, and personality profiles have, for example, concluded that juveniles with child victims display different treatment needs than juvenile peer abusers (Drew, 2013; Glowacz & Born, 2012; Hart-Kerkhoffs, Doreleijers, Jansen, Van Wijk, & Bullens, 2009; Hendriks, 2006; Kjellgren, Wassberg, Carlberg, Långström, & Svedin, 2006). Also, Butler and Seto (2002), and Hissel, Bijleveld, Hendriks, Jansen, and Collot d'Escury-Koenigs (2006) have found sex-only offenders to differ from mixed offenders (juveniles having also committed non-sexual offenses). Hendriks (2006) also differentiated group offenders from solo offenders. The diversity in the population of juveniles who have sexually offended has become increasingly clear, especially because to date no psychological typology has been found to represent a subgroup at increased risk for sexual recidivism (Cale, Smallbone, Rayment-McHugh, and Dowling, 2016; Lussier, et al., 2012).

Effectiveness of Treatment for Juveniles who have Sexually Offended

Treatments for juveniles who have sexually offended make use of different treatment strategies (i.e., wilderness therapy, cognitive and/or behavioral based group therapy, cognitive and/or behavioral based individual therapy, family therapy and psycho-education) and are delivered in both residential and community settings (Ryan, Lerversee, & Lane, 2010; Veneziano & Veneziano, 2002). Meta-analyses, predominantly including studies that combine adult and juvenile samples, have shown cognitive behavioral based treatments and multisystemic therapy to be most effective in reducing recidivism (Dopp, Borduin, & Brown, 2015; Hanson, Bourgon, Helmus, & Hodgson, 2009; Lösel & Schmucker, 2005; Schmucker & Lösel, 2015; Walker, McGovern, Poey, & Otis, 2004). Treatment types that incorporate these standards are, therefore, generally considered established treatment.

To date, only two meta-analyses have been performed that focus specifically on treatment for juveniles. Walker and colleagues (2004) performed a meta-analysis of ten published and unpublished studies; two studies with and eight without a comparison group. Reitzel and Carbonell (2006) based their findings on nine studies, all using a comparison or control group, and including five unpublished studies. These meta-analyses demonstrated a moderate effect of treatment on sexual recidivism by juveniles. Moderator analyses showed that outcomes measured via official recidivism data (rearrests or reconvictions) resulted in smaller effect sizes compared to measurement via self-report or penile plethysmography (i.e., the measurement of physical arousal to deviant stimuli). Furthermore, random

assignment to treatment and comparison groups yielded larger effect sizes. The hypotheses that cognitive behaviorally based treatments are more effective than other treatment types, was not substantiated in this meta-analysis (Reitzel & Carbonell, 2006).

Reitzel and Carbonell (2006), as well as Walker and colleagues (2004), assessed the effects of treatment on sexual recidivism only, and based their findings on a restricted number of studies. In the last decade more and more robust studies (making use of a comparison or control group and examining the effectiveness of treatment for multiple types of recidivism) have been conducted. In addition, multilevel statistical techniques have become more refined, making it possible to analyze moderators with a restricted number of studies, accounting for both within and between study variability in effect sizes. A multilevel approach enables comprehensive moderator analyses to examine the influence of study, sample, and intervention characteristics (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013; Van den Noortgate & Onghena, 2003). This may be useful in the complex field of juvenile sex offending, where multiple causes may contribute to recidivism. Furthermore, multilevel meta-analytical techniques enable the use of all available effect sizes in the primary studies, thus preserving all information and generating maximum statistical power (Assink et al., 2015). This is especially important when examining the results of studies with small samples, as is the case with many of the primary studies in this field.

STUDY AIM

The aim of this study is to review contemporary research on the effect on officially measured recidivism (re-arrest or reconviction) of treatment for juveniles who have sexually offended. In addition, the potential moderating effects of participant, treatment, and study characteristics were investigated. This provided an opportunity to detect factors that may influence the effect of treatment on the reduction of recidivism.

METHOD

To assess the effect of treatment on recidivism, and the factors moderating this effect, a multilevel meta-analysis was carried out. The term meta-analysis refers to a stepwise procedure and a set of statistical techniques, combining results of independent primary studies into effect sizes, so that overall conclusions can be drawn. An important requirement for traditional univariate meta-analytic approaches is that no dependency between effect sizes is allowed, so that only one effect size per primary study can be included. By stepping away from the traditional univariate approach, it becomes possible to deal with dependency of effect sizes, so all information can be preserved and a maximum of statistical power is

achieved. In this study, applying a three-level structure statistically considers the three different variance components distributed over three levels: differences among all effect sizes or random sampling error (level 1), differences in effect sizes within studies (level 2), and differences in effect sizes between studies (level 3). If there is evidence for heterogeneity in effect sizes, moderator analyses can be conducted to test variables that may explain within-study or between-study heterogeneity. For these analyses, the three-level random effects model can be extended with study and effect size characteristics, making the model a three-level mixed effects model (Assink & Wibbelink, 2016).

Inclusion Criteria

Multiple inclusion criteria were formulated to select the studies. First, the treatment condition had to be aimed at treating sexually offending behavior. Second, the study sample had to exclusively contain juveniles. Therefore, the mean age of the researched group had to lie between 12 and 18 years of age, and/or the study had to specifically report on juveniles or adolescents. Third, the studies had to report treatment and comparison group recidivism rates via officially obtained data for sexual and/or other types of recidivism. And finally, the comparison group had to comprise juveniles who have sexually offended as to ensure basic comparability between groups.

Selection of Studies

All studies published before February 2016 that met the inclusion criteria were to be included in the current meta-analysis. Firstly, several electronic databases were searched, including Campbell library, PubMed, OVID (Medline, PsycINFO, ERIC), and Proquest (Sociological Abstracts, Social Services Abstracts, Proquest Dissertations). Secondly, Google Scholar was searched. The following English search string was used: (sex* offender) AND (juvenile OR adolescent) AND (treatment OR therapy) AND (effect* OR efficacy) AND (recidivism). Finally, the references of other meta-analyses and reviews were checked for eligible studies, and authors of non-published work were contacted. Regrettably the contacted authors did not respond, so a few non-published studies could not be included. A flow chart is presented as Figure 1 in the results section.

The initial search and screening resulted in 23 studies that met the basic criterion of examining the effectiveness of an intervention for juveniles who have sexually offended with the use of a comparison group. After exclusion, mainly because of the use of other outcome variables (see also Table 2), 14 studies remained, with 77 effect sizes, 1,726 participants, and 13 independent samples. Table 1 presents the study characteristics of the included studies. Table 2 specifies the excluded studies, and our reasons for excluding them in italic.

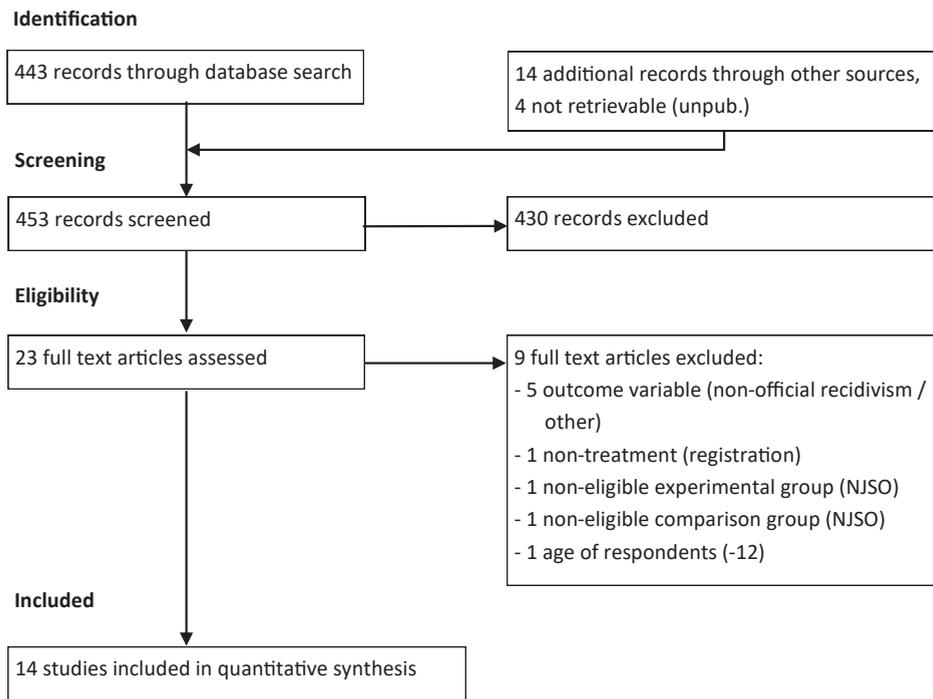


Figure 1. Flowchart of study selection

Coding the Studies

The dependent variable in this meta-analysis was recidivism. Three types of recidivism were distinguished: any recidivism, sexual recidivism, and non-sexual recidivism. In all studies, recidivism was measured by means of official data as to ensure comparability among the effect sizes of the primary studies, because the type of measurement used may influence the effect measured (Reitzel & Carbonell, 2006).

The independent variable was the treatment offered. Type of recidivism, participant, treatment, and study characteristics were coded to assess whether treatment effects varied among the possible moderator variables. In order to reduce the problem of multiple testing (Tabachnik & Fidell, 2013), only moderators of possible theoretical importance were used. Studies with overlapping samples (Worling & Långström, 2000; Worling, Littlejohn, & Bookalam, 2010) were coded with corresponding study identification numbers.

Table 1. Characteristics of Included Studies

Study	Study characteristics					Sample characteristics					Treatment characteristics						
	Treat. coh.	N	# r (M)	Pb	Study quality	Design	IA	Type of rec.	Age	% Male	% Cauc	% Child Vic.	% Off.	Sett.	EXP type	Exp. Cond.	Comparison ^b or Control ^c condition ^a
Barlow 1998	1995	44	2 (328)	N	Weak	OE	Y	1/2	15.2	100	86	-	-	Res	CBT	'level 6' compl.	'level 6' non-completers ^b
Borduin 1990	1983-1985	16	6 (1,130)	Y	Weak	RCT	N	1/2/3	14.0	100	63	50	94	Com	SYS	MST	TAU (individual treatment) ^b
Borduin 2009	1990-2001	48	5 (915)	Y	Mod.	RCT	N	1/2/3	14.0	96	71	-	-	Com	SYS	MST	TAU (Community Services) ^b
Gillis 2010	1990-2005	190	6 (341)	Y	Mod.	OE	Y	1/2/3	-	100	65	-	-	Res	TAU	LEGACY	TAU 1: Dev centers ^b TAU 2: Spec. Treat ^b
Guarino-Ghezzi 1998	1993-1994	75	5 (683)	Y	Mod.	OE	Y	1/2/3	-	100	53	46	83	Res	CBT	Spec. Treatm.	TAU (Nonspecialized residential ther.) ^b
Hendriks 2005	1988-2003	325	3 (256)	Y	Weak	OE	N	1/2/3	14.4	100	62	44	100	Out	CBT	Spec. Treatm.	Assessment only ^c
Holly 2000	1990-1998	64	14 (583)	N	Mod.	OE	Y	1/2/3	14.5	98	-	-	100	Com	CBT	TBASOP compl.	1: Non Completers ^c 2: Assessment only ^c
Johnson 2008	2000-2007	78	4 (367)	N	Strong	OE	Y	1/2/3	15.3	99	78	87	92	Out	SYS	FFT	TAU (individual and group CBT) ^b
Lab 1993	1988-1991	155	2 (248)	Y	Weak	OE	Y	1/2	14.5	98	57	66	24	Out	TAU	SOT	TAU (community based) ^b
Letourneau 2013	2004-2006	131	2 (051)	Y	Strong	RCT	N	1	14.7	100	44	55	86	Com	SYS	MST	TAU ^b
Schram 1991	1984-1991	196	6 (055)	N	Weak	OE	Y	1	14.5	100	89	76	97	Mix	TAU	Mix com. based	TAU (state operated institutions) ^b
Waite 2005	1992-2001	256	4 (316)	Y	Weak	OE	N	1/2/3	16.8	100	44	43	31	Res	CBT	'Selfcon-tained'	TAU/Prescriptive ^b
Wooling 2000	1987-1995	148 ^a	12 (575)	Y	Mod.	OE	N	1/2/3	15.6	94	-	65	98	Mix	CBT	SAFE-T	dropouts ^c / refusers ^c / assessment only ^c
Wooling 2010	1987-1995	148 ^a	6 (352)	Y	Mod.	OE	N	1/2/3	15.6	94	-	55	98	Mix	CBT	SAFE-T	dropouts, refusers & assessment only

Note: Treat. coh. = Years in which treatment was delivered; N = number of participants; # r (M) = number of effect sizes (mean d); Pb = published in peer reviewed article yes/no; Study Quality = strong/moderate/weak according to EPHP quality assessment tool; Design = RCT (randomized controlled trial) or OE (quasi experimental); IA = Independent author, yes/no; Type of recidivism measured: 1 = all recidivism, 2 = sexual, 3 = non-sexual; Age = mean age of sample; % Male = percentage of males in sample; % Caucasian = percentage of Caucasian ethnicity; % Contact offense = percentage contact offense; % Child molest = percentage child molestation; Setting = outpatient, community based, residential or a mix; EXP type = Type of experimental intervention; Experimental Condition = name of experimental intervention; Comparison or Control condition = Type of control intervention(s). ^a = studies using the same participants. ^b = treatment comparison group. ^c = no treatment control group. * the allocation of control group status was based on available information, sometimes control group juveniles may have still have received alternate forms of treatment elsewhere.

Table 2. Excluded Studies and Reasons for Exclusion

Study	First author & year of publication	Treat. cohort	N	Setting	Outcome measure	Experimental treatment	Comparison ^a or Control ^b condition
Brannon 1991	1987-1988	110	Residential	Official Recidivism Rates	Residential State Treatment JSO	Residential State Treatment JSO ^a	
Carpentier 2006	1992-1995	135	Residential	Psychosocial Functioning & Official Recidivism Rates	Play therapy (5 to 12 year olds)*	CBT (5 to 12 year olds) ^a	
Kahn 1991	1984	221	Mixed	Official Recidivism Rates (total group)*	8 Outpatient programs	2 Residential programs ^a	
Seabloom 2003	1977-1986*	122	Outpatient	Official Recidivism Rates	Personal Social Awareness Program completers	Personal Social Awareness Program withdrawers ^b	
Letourneau 2008	1995-2005	222	-	Official Recidivism Rates	Registration*	Non-Registration ^b	
Letourneau 2009	2004-2006	127	Community	Psychosocial Functioning*	JSO Multi Systemic Therapy	JSO Usual Community Services ^a	
Henggeler 2009	2004-2006	127	Community	Psychosocial Functioning*	JSO Multi Systemic Therapy	JSO Usual Community Services ^a	
Van Outsem 2009	-	113	Outpatient	Psychological Functioning*	Treatment completers	Treatment non-completers and 'normal' controls ^b	
Weinrott 1997	NR	69	Outpatient	Phallometric measurement and Selfreport*	Vicarious Sensitization	Cognitive Therapy ^a	

Note. JSO = Juvenile Sex offender; JNSO = Juvenile Non Sex Offender; CBT = Cognitive Behavioral Therapy.

* The characteristic in italic font specifies reason(s) for exclusion. Mostly the outcome variable was other than a measure of recidivism (Letourneau 2009; Henggeler, Van Outsem, and Weinrott). In the Brannon study a comparison group of non-sex offenders was used. In the Carpentier study the respondents were too young to be included in the Kahn study the official recidivism rates were not available for the experimental and control group, and in the Letourneau 2008 study registration was researched which, to our judgement, is not a form of treatment. The Seabloom study included transgressive acts (i.e., homosexuality, transvestitism) that were then considered to be a sexual transgression, but are now considered a variation in normal human sexuality, so their sample was deemed not comparable.

Firstly, **outcome characteristics** were coded as the type of recidivism reported; recidivism in general (including all types of recidivism), sexual recidivism, or non-sexual recidivism.

The following **participant characteristics** were coded. The criminal status of the clients was coded as all adjudicated or as also suspected of a sex crime, as not all studies only included formally adjudicated offenders. The cultural background of the juveniles was coded as the percentage of Caucasians in the researched group. Furthermore, the percentage of juveniles with child victims (< 12 years of age and ≥ 5 years younger than their assailant) was coded, as was the percentage of contact offenders (physical sexual offending, such as rape or sexual assault; not including, for example, exhibitionism). In most reports information on the estimated recidivism risk level of the respondents was available (groups were designated as relatively high at risk or low at risk), so also the mean estimated recidivism risk level of clients was coded as high or low. Finally, the living condition of participants was coded as living at home versus living in a residential facility or mixed (both types of living situation present in the researched group).

Several **treatment characteristics** were coded: firstly, whether the type of treatment was cognitive behavioral, systemic, or another form of (non-established) treatment as usual. Mean treatment duration was coded in months. Finally, the exclusion of clients with a lower IQ was coded as yes or no, for this may influence learnability of the researched group.

As for **study characteristics**, we coded whether the authors were independent researchers or whether they were involved in the development or implementation of the intervention. It was also coded whether the study was peer reviewed. Further, the design of the study (randomized controlled trial versus quasi-experimental) was coded. The type of comparison treatment used was coded as treatment as usual (TAU) or established treatment (incorporating cognitive behavioral treatment and/or systemic therapy). The comparability of the treatment and comparison group(s) was coded as high (including statistically controlling for any potential influential differences found, for example, differences in group recidivism risk levels) or low (when significant differences were reported between groups and this was not controlled for in results). The mean length of follow up was coded in months. The type of effect size was coded as proportion based (the difference in proportion of reoffenders in the treatment and comparison group) or based on mean number of re-offenses in each group. Type of recidivism measured was coded as re-arrest or reconviction. Lastly, study quality was coded by use of the EPHPP Quality Assessment Tool for Quantitative Studies (<http://www.ephpp.ca/tools.html>). This tool assesses the quality of a study as weak, moderate or strong, providing a comprehensive and structured assessment of the concept of study quality (Armijo-Olivo, Stiles, Hagen, Biondo, & Cummings, 2012). The EPHPP was developed by the Effective Public Health Practice Project in Canada. It is a generic tool used

to evaluate a variety of intervention study designs. It has been judged suitable to be used in systematic reviews of effectiveness (Deeks et al., 2003) and has been reported to have content and construct validity (Jackson & Waters, 2005; Thomas, Ciliska, Dobbins, & Micucci, 2004). The tool assesses six domains: selection bias, study design (including appropriateness of the design), confounders, blinding, data collection methods, and withdrawals and dropouts, see Table 1 for the results of the assessment.

The first author of this article coded the included studies according to the suggestions of Lipsey and Wilson (2001). All studies were double coded by the second author. Following the guidelines by Shrout (1989), for the continuous variables single ICCs were calculated for the two-way random effects model, with absolute agreement as a criterion. On all 18 variables single measures showed an ICC of $> .99$ (i.e., substantial reliability). For 23 categorical variables kappa was calculated (Landis & Koch, 1977). Two variables reached substantial interrater reliability ($\text{kappa} = .61-.80$), all others reached almost perfect interrater reliability ($\text{kappa} > .81$).

Calculations and Analyses

Effect sizes were transformed into Cohen's d by using the calculator of Wilson (2013) and formulas of Lipsey and Wilson (2001). Most d -values were calculated based on reported means and percentages of recidivism. A positive effect size indicated that the treatment group benefited more than the comparison group, whereas a negative effect size indicated that the comparison group benefitted more than the treatment group. If a study only mentioned that an effect was not significant (as was the case in 9% of all effect sizes), the effect size was coded as zero (Lipsey & Wilson, 2001). The continuous variables (percentage immigrants, percentage with child victims, percentage contact, percentage mixed offenders, and mean treatment length) were centered around their mean, and all other (categorical) variables were recoded into dummy variables. We checked for the presence of extreme outliers using box plots (Tabachnik & Fidell, 2013); no extreme outliers were identified. Standard errors were estimated using formulas of Lipsey and Wilson (2001).

In all studies, we were able to calculate more than one effect size. Most studies reported on multiple outcome variables (i.e., all types of recidivism, sex only recidivism, non-sex recidivism, and sometimes violent non-sexual recidivism and non-violent non-sexual recidivism). Effect sizes from the same study may prove more alike than effect sizes from different studies. Therefore, the assumption of statistical independency, which underlies classical meta-analytic strategies, was violated (Hox, 2002; Lipsey & Wilson, 2001). In line with recently conducted meta-analyses, we applied a multilevel approach in order to deal with the interdependency of effect sizes (Assink et al., 2015; Houben, Van Den Noortgate, &

Kuppens, 2015; Spruit, Assink, Van Vugt, Van Der Put, & Stams, 2016; Weisz, et.al, 2013). The multilevel approach accounts for the hierarchical structure of the data in which effect sizes are nested within the studies (Van den Noortgate & Onghena, 2003).

We used a three-level meta-analytic model to calculate the combined effect sizes and to perform moderator analyses. Three sources of variance were modelled, including the sampling variance for the observed effect sizes (level 1), the variance between effect sizes from the same study (level 2), and the variance between the studies (level 3) (Cheung, 2014; Van den Noortgate et al., 2013; Wibbelink & Assink, 2015). The sampling variance of observed effect sizes (level 1) was estimated by using the formula of Cheung (2014). Log-likelihood-ratio-tests were performed to compare the deviance of the full model to the deviance of the models excluding one of the variance parameters, making it possible to determine whether significant variance is present at the second and third level (Wibbelink & Assink, 2015). Significant variance at level 2 or 3 indicates a heterogeneous effect size distribution, meaning that the effect sizes cannot be treated as estimates of a common effect size. In that case, we proceed to moderator analyses, because the differences between the effect sizes may be explained by outcome, study, sample, and/or intervention characteristics. Moderator analyses were only performed when each category of the potential moderator was filled with at least three studies. As a result, Study Quality was collapsed into a dichotomous variable (weak versus medium to strong) because only two studies were considered to be of strong quality. Exclusion of IQ could not be tested because only two studies reported on the exclusion of lower IQ respondents.

The multilevel meta-analysis was conducted in R (version 3.2.0) with the metafor-package, using a multilevel random effects model (Viechtbauer, 2010; Wibbelink & Assink, 2015). The restricted maximum likelihood estimate was used to estimate all model parameters, and the Knapp and Hartung-method (2003) was used for testing individual regression coefficients of the meta-analytic models and for calculating the corresponding confidence intervals (see also Assink et al., 2015; Houben et al., 2015; Spruit et al, 2016; Wibbelink & Assink, 2015).

Publication Bias

In systematic reviews, the aim is to include all studies previously conducted that meet the inclusion criteria (Lipsey & Wilson, 2001). However, a common problem is that studies may not have been published due to non-significant or unfavorable findings, and, therefore, are difficult to locate. Not including these studies may lead to an overestimation of the true effect size, the so called "publication bias" (Rosenthal, 1979). In order to check the presence of publication bias in our meta-analysis, we performed a trim and fill procedure (Duval & Tweedie, 2000) by drawing a trim-and-fill plot in R (Version 3.2.0) using the function "trimfill"

of the metafor package (Viechtbauer, 2010). We tested whether effect sizes were missing on the left side of the distribution, since publication bias would only be likely to occur in case of non-significant or unfavorable (i.e., negative) results.

RESULTS

Overall, a significant, small to moderate effect, $d = 0.37$, $p < .001$, of treatment on recidivism was found, indicating that the treatment groups achieved an estimated reduction in recidivism of 20.5% as compared to comparison groups. The likelihood ratio test comparing models with and without between-study variance (level 3) showed that significant variance was present at the between-study level, $\sigma^2_{\text{level } 3} = 0.11$, $\chi^2(1) = 37.12$; $p < .000$. The variance between the effect sizes within studies (level 2) was not significant, $\sigma^2_{\text{level } 2} = 0.01$, $\chi^2(1) = 3.32$; $p = .068$. About 24% of the total effect size variance was accounted for by the sampling variance (level 1), 7% for the variance between effect sizes within studies (level 2), and 69% for the variance between studies (level 3). Because the variance on the third level was significant, we proceeded to moderator analyses to assess factors that could possibly explain variance in treatment effects (see Table 3).

As presented in the last column of Table 3, only the calculation type of effect size moderated the effect of treatment on recidivism in adolescents who have sexually offended. Stronger treatment effects were found for effect sizes that were proportion-based compared to effect sizes based on the mean number of re-offenses. The proportion of recidivism was significantly lower in the treatment groups compared to the comparison groups, but the mean number of offenses for every reoffender in the two groups did not significantly differ. Furthermore, a moderating trend was found for the follow up period, indicating that stronger treatment effects were found in shorter follow up periods. None of the participant characteristics and none of the treatment characteristics moderated the effect of treatment on recidivism in adolescents who have sexually offended.

The trim-and-fill plot imputed estimations of effect sizes of missing studies, indicating the presence of publication bias (see Figure 2). We included the imputed estimations (the open circles) and performed the meta-analysis again to compute an overall effect size that takes the influence of publication bias into account (Duval & Tweedie, 2000). After controlling for publication bias, no significant treatment effect was found ($d = 0.15$, $p = .176$).

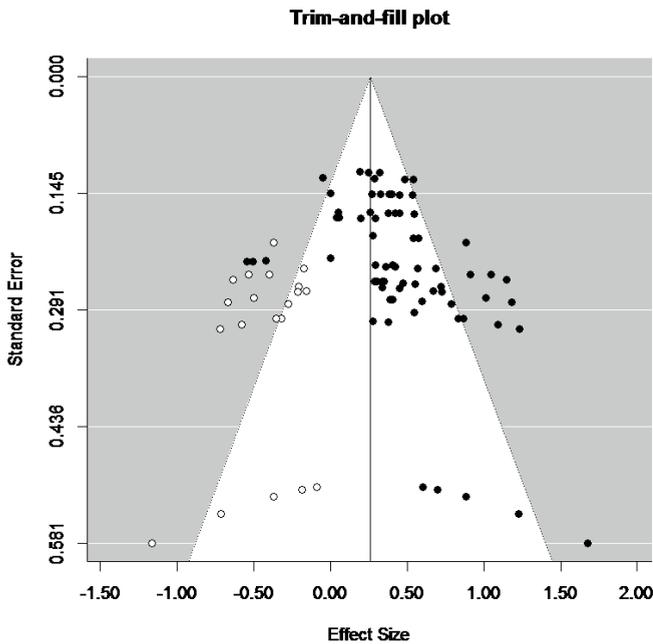


Figure 2. Results of trim-and-fill procedure testing for publication bias.

Note. The closed circles represent the primary studies included. The open circles represent forecasted missing effect sizes, pointing out possible publication bias.

DISCUSSION

A multilevel meta-analysis was performed to assess the strength of the effect on the reduction of recidivism of specific treatment for juveniles who have sexually offended, and to assess what variables have a moderating influence on the effect measured. An overall significant and small to moderate effect ($d = 0.37$) was found, indicating treatment to be more effective in reducing recidivism than comparison or control conditions, a finding similar to that of previous meta-analyses (Reitzel & Carbonell, 2006; Walker et al., 2004). However, indications of publication bias were present, and after statistically controlling for this bias, the effect of treatment on recidivism was no longer significant. Studies with less favorable outcomes may indeed not have been published, and the small to moderate effect found may therefore represent an overestimation of the true overall effect size. Thus, in the current meta-analysis, after correcting for publication bias, an effect of treatment on recidivism, as compared to comparison or control conditions, was not demonstrated. The treatment conditions in the primary studies, however, were mainly compared to other forms of treatment ($N=10$), or to groups of assessment-only juveniles, who may have received

Table 3. Overall Results and Moderator Effects of the Relation between Treatment and Recidivism

Moderator variables	# Studies	# ES	β_0 (mean <i>d</i>)	t_0	β_1	t_1	$F(df_1, df_2)$
Overall relation	13	77	0.367	3.713***			
After trim and fill	18	99	0.154	1.364	(<i>p</i> =.176)		
Outcome characteristics							
Type of Recidivism	13	77					$F(2,74) = 0.543$
Any offenses (RC)	13	29	0.401	3.764***			
Sex offenses	11	19	0.348	3.100**	-0.054	-0.735	
Non-sex offenses	9	28	0.332	3.012**	-0.069	-0.999	
Participant characteristics							
Criminal Status	13	77					$F(1,75) = 1.135$
All convicted (RC)	9	58	0.299	2.506*			
Also suspects	4	19	0.533	2.880**	0.235	1.065	
Percentage Caucasian	11	46	0.333	2.694*	0.406	0.473	$F(1,43) = 0.224$
Proportion Child Molestation	8	44	0.337	2.105*	0.395	0.451	$F(1,42) = 0.203$
Proportion Contact Offenses	10	64	0.332	2.554*	0.084	0.185	$F(1,62) = 0.034$
Recidivism Risk	12	72					$F(1,70) = 1.162$
High risk (RC)	8	49	0.423	3.284**			
Low risk	4	23	0.186	1.044	-0.237	-1.078	
Living Conditions	13	77					$F(2,74) = 0.085$
Residential (RC)	3	12	0.327	1.432			
Living at home	5	31	0.427	2.393*	0.100	0.345	
Mixed	5	34	0.339	1.971 [†]	0.012	0.041	
Treatment characteristics							
Type of Exp. treatment	13	77					$F(2,74) = 0.372$
TAU (RC)	3	14	0.213	0.993			
CBT	6	46	0.438	2.864**	0.226	0.857	
Systemic	4	17	0.384	1.961 [†]	0.171	0.590	
Duration EXP	9	49	0.400*	2.554	-0.002	-0.116	$F(1,47) = 0.014$
Study characteristics							
Authors	13	77					$F(1,75) = 1.262$
Dependent (RC)	6	38	0.488	3.342**			
Independent	7	39	0.265	1.969 [†]	0.223	1.123	
Peer Reviewed	13	77					$F(1,75) = 2.456$
Yes (RC)	9	51	0.463	4.125***			

Moderator variables	# Studies	# ES	β_0 (mean d)	t_0	β_1	t_1	$F(df_1, df_2)$
No	4	26	0.144	0.850	-0.319	-1.567	
Study Design	13	77					$F(1,75) = 2.379$
Randomized (RC)	3	13	0.653	3.131**			
Quasi-experimental	10	64	0.292	2.739**	-0.361	-1.542	
Type of Control Treatment	12	61					$F(1,55) = 0.057$
Treatment as usual (RC)	5	23	0.423	2.895**			
Established treatment	7	34	0.389	3.057**	-0.035	-0.233	
Comparability of Groups	10	56					$F(1,54) = 0.026$
High (RC)	5	35	0.374	2.867**			
Low	5	21	0.351	0.162	-0.023	-0.162	
Length Follow Up	10	68	0.336	2.460*	-0.002	-1.919 ^t	$F(1,66) = 3.681^t$
Type effect	13	77					$F(1,75) = 5.249^*$
Proportion (RC)	13	68	0.396	3.728***			
Means	4	9	0.162	1.167	-0.234	-2.291*	
Recidivism Measure	13	77					$F(1,75) = 1.021$
Reconviction (RC)	3	20	0.178	0.841			
Rearrest	10	57	0.421	3.731***	0.243	1.011	
Study Quality	13	77					$F(1,75) = 0.018$
Strong/Moderate (RC)	7	54	0.380	2.739**			
Weak	6	23	0.353	2.270*	-0.028	-0.134	

Note. # studies = number of independent studies; #ES = number of effect sizes; t_0 = difference in mean r with zero; t_1 = difference in mean d with reference category; mean d = mean effect size (d); $F(df_1, df_2)$ = omnibus test; (RC) = reference category. ^t = trend, significant at a 0.1 level, * = significant at a 0.05 level, ** = significant at a 0.01 level, *** = significant at a 0.001 level.

treatment elsewhere ($N = 5$). Therefore, the difference in effect on recidivism between the treatment conditions and the comparison conditions might have been reduced. Comparing treatment to a control condition (a group that receives no treatment at all), is considered unethical in this population as it requires withholding treatment. Such research, therefore, will remain scarce. Few studies ($N=3$) made use of drop out control groups. These studies, in contrast, may have resulted in an overestimation of the effect of treatment because drop outs are at higher risk for reoffending (Hendriks & Bijleveld, 2005).

Significant variance was present at the between-study level. This indicates that the effect size distribution was heterogeneous, and that moderating variables might explain differences in strength of the effect sizes. Therefore, moderator analyses were performed.

The type of recidivism measured did not moderate the effects significantly; on all types of recidivism a small to moderate effect was achieved. This result could signify that, in juveniles, while treating sexual offending behavior, other offending behavior is also treated. This effect has also previously been established by Lösel and Schmucker (2005) and Schmucker and Lösel (2015). Treatment usually entails a confrontation with the sexual offending behavior, thoughts, and feelings prior to, during, and after the offense, the development of victim empathy, the development of anger management and stress coping skills, social skills training, treatment of the juveniles own victimization experiences, and relapse prevention strategies (Becker & Hunter, 1997; Becker & Johnson, 2001). Moreover, contemporary treatments have been developed to take into account broader issues. The Good Lives Model, for example, was presented as an addition to the relatively risk focused RNR model (Lord, 2016). Contextual treatment forms, such as Multisystemic Therapy-Problem Sexual Behavior, are explicitly holistic in nature (Swenson, Henggeler, Taylor, & Addison, 2005). Thus, risk factors for juvenile delinquency in general also seem to be more and more addressed via treatment aimed at sex offending. This is considered expedient, because while sexual recidivism amongst juveniles is rare, other types of recidivism are relatively frequent (Caldwell, 2010; Caldwell, 2016; Lobanov-Rostovsky, 2015).

None of the tested participant characteristics moderated the effect of treatment on recidivism in juveniles. Thus, the currently coded subgroups within the sample did not account for differences in effects on reducing recidivism between the treatment and comparison or control groups. Notably, few studies defined their participants more specifically than juveniles who have sexually offended, and as a result, in the included studies the heterogeneous group of juveniles who commit sexual offenses was presented as relatively homogenous. This hampered research into moderating effects of participant characteristics. Many studies on specific mental illnesses, traits, executive functioning, and personality profiles of adolescents who have sexually offended have been conducted

(Adjorlolo & Egbenya, 2016; Glowacz & Born, 2012; Hart-Kerhoffs et al., 2009; Lawing et al., 2010; Margari et al., 2015; Purcel, 2010; Seto & Lalumière, 2010). None of the studies included in the current meta-analysis, however, reported on psychological profiles or the specific treatment needs juveniles presented at admission that deemed them eligible or in need of specific treatment (other than their type of offending behavior). Also, several typologies are commonly in use: non-contact versus contact offenders, sex-only offenders versus mixed or sex-plus offenders, peer abusers versus child abusers, and group offenders versus solo offenders (Butler & Seto, 2002; Drew, 2013; Hendriks, 2006; Hissel et al., 2006; Kjellgren et al., 2006), and are described as in need of different treatment approaches. Of these typologies, only two (victim age and type of offending behavior) could be tested in this study. A more extensive moderator analysis could have been performed if studies had reported on treatment needs or typologies of their respondents.

Further, none of the coded treatment characteristics moderated the effect of treatment. The treatment conditions were grouped as predominantly cognitive behavioral based (e.g., SAFE-T, TBASOP), systemic (i.e., Multi Systemic Therapy and Functional Family Therapy), or as treatment as usual for juveniles who have sexually offended (TAU). The first two categories are established treatments, the third category contained a compilation of non-established treatment forms (i.e., adventure or wilderness therapy, individual treatment, or family treatment). The established treatments were expected to achieve better results in reducing recidivism in juveniles based on previous research (Hanson et al., 2009; Lösel & Schmucker, 2005; Schmucker & Lösel, 2015; Walker et al., 2004).

In study characteristics, one moderator significantly explained variance (calculation type of effect size), while for another (follow up time) a moderating trend was found, indicating stronger treatment effects in shorter follow up periods. These results are in line with expectations. Calculation of Cohen's *d* via proportions (the relative number of recidivists per group) yielded a more pronounced effect (difference measured) than calculation via mean number of re-offenses (how many times recidivists reoffended per group). This indicates that specialist treatment showed a stronger reduction in the number of juveniles reoffending than comparison conditions did, but that reoffenders recidivated as many times in both groups. Regarding follow up time, most recidivism by juveniles who have sexually offended was reported in the first years after release (Hendriks & Bijleveld, 2005), while the majority of adolescent offending proves to be adolescence limited (Lussier, et al., 2012; Moffit, 1993). Therefore, differences in percentages of recidivism were expected to be most pronounced in the first years after treatment. Lastly, comparing treatment with a non-established treatment comparison group did not generate larger effects than comparing treatment with an established treatment comparison group. The inclusion of drop-out respondents in an established treatment comparison group (by most of the studies), may

have reduced their effect, since drop-outs were not able to profit fully from the established treatment administered. Drop-outs also may have been more distressed or less motivated than treatment completers; a reason for dropping out and achieving less positive results (Hendriks & Bijleveld, 2005).

Notably, not many moderating influences were detected. However, non-significant or borderline significant results may still prove clinically relevant (Man-Son-Hing, et. al., 2002), especially when an analysis is conducted with a relatively small sample size. We refer to the limitations section for an overview of our borderline significant findings.

LIMITATIONS

This study was limited by only using studies that relied on official measures of recidivism. Thereby, undoubtedly, recidivistic behavior was missed that was not officially reported. Levels of unreported sexual problem behavior ('dark numbers') are known to be quite high (Wittebrood, 2006). Also, official reports involve getting caught, which results in a specific subgroup of treated juveniles who were caught sexually offending, and again getting caught reoffending (Yun & Lee, 2013). The use of a conservative measure, however, to our opinion generated greater comparability between studies and, therefore, more reliable results.

The restrictive inclusion criteria used, and, therefore, the relatively small number of included studies (mostly performed on small samples), presented us with underpowered analyses. However, to assess the effect of treatment the inclusion of studies with a comparison group is imperative (Weisburd, Lum, & Petrosino, 2001). Some results previously detected in other reviews might have been replicated, if a larger sample size would have been available (i.e., the moderating effect of established treatment, high at recidivism risk respondent groups, dependent authors, publication, study design, and re-arrest as outcome measure), because some of these variables did show rather pronounced, yet non-significant, differences in mean effect size (see Table 3).

Furthermore, there are methodological difficulties in the use of trim and fill procedures in multilevel meta-analytic data (Nakagawa & Santos, 2012; Peters, Sutton, Jones, Abrams, & Rushton, 2007; Terrin, Schmid, Lau, & Olkin, 2003). Therefore, the difference between the adjusted and observed mean effect size in this study should be interpreted as indicative of (the effects of) publication bias; the adjusted effect size is not to be considered the true effect size.

In the included studies, most likely due to ethical considerations, one type of treatment was usually compared to another type of treatment, that is, treatment as usual. Only one comparison of a treatment group to an explicit non-treatment (refuser) group was available (Worling & Curwen, 2000). As a result, the moderating effect of no treatment could not be tested and no statement about the effects of therapy as compared to administering no therapy could be made. Also, we could not test whether treatment specifically aimed at juveniles who have sexually offended proved as effective or less effective as non-specific treatment (treatment aimed at juveniles with all types of behavioral problems) in reducing recidivism, for only one study addressed this issue (Guarino-Ghezzi & Kimball, 1998).

Lastly, recidivism is just one aspect relevant to assess the quality of treatment. In addition to reduction of recidivism, treatment aims at improving the quality of life and psychosocial functioning of clients. This dual focus is important to bear in mind, since some treatments may excel in reducing recidivism (the treatment need of society), while others might more primarily serve treatment needs as formulated by their clients (quality of life). A meta-analysis on the effect of treatment on psychosocial improvement would, therefore, complement the findings of the current study.

CONCLUSION

Juveniles who have sexually offended constitute a notoriously heterogeneous group regarding treatment needs and offending behavior patterns. Various treatment forms are aimed at them, although sexual recidivism rates are generally low. Relatively low sexual recidivism rates, however, do not imply that juveniles who have sexually offended are not in need of therapy to reduce recidivism risk, if only because their general recidivism rates are relatively high. To date, recidivism risk factors that are specific to juveniles who have sexually offended mainly indicate at what specific criminogenic treatment needs treatment should be aimed (i.e., childhood sexual victimization, sexual deviation, association with younger children, poor school performance, not attending school and lack of intimate peer relationships/social isolation). Also, mixed offenders, who display more antisocial tendencies, have been shown to be at higher risk for recidivism, and therefore, in greater need of treatment. The current study has shown sex offense specific treatment to have an equal effect on other types of recidivism. Thus, juveniles who have sexually offended and who display specific as well as general criminogenic needs may be best off in treatment aimed at reducing sexual recidivism. Not all juveniles who have sexually offended, however, display these features equally, so not all juveniles who have sexually offended may be in need of sex offense specific treatment as to reduce recidivism.

The authors would like to encourage more robust (independent, high quality) studies on the effect of treatment for juveniles who have sexually offended, possibly making use of a non-specific treatment comparison group. Reporting on specific treatment needs might also shed more light onto what treatment works, for what (type of) sexually offending juvenile. This could improve allocation to treatment and contribute to more positive treatment effects.