



UvA-DARE (Digital Academic Repository)

One size does not fit all

The need for treatment tailoring in youth and family services

Bijlsma, A.

Publication date

2022

[Link to publication](#)

Citation for published version (APA):

Bijlsma, A. (2022). *One size does not fit all: The need for treatment tailoring in youth and family services*. [Thesis, fully internal, Universiteit van Amsterdam].

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: <https://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.

Chapter 2

A Critical Evaluation of the Risk, Need, and Responsivity Principles in Family Interventions for Delinquent Youth: A Meta-Analysis

This chapter is adapted from:

Bijlsma, A. M. E., Assink, M., & van der Put, C. E. (2022). A critical evaluation of the risk, need, and responsivity principles in family interventions for delinquent youth: A meta-analysis.

Submitted for publication.

Abstract

This meta-analysis aimed to re-examine the available evidence on the effectiveness of the risk, need, and responsivity principles of the RNR model in family interventions for juvenile delinquency. As previous reviews did not examine these principles fully in line with their original definitions, this review aimed to improve the coding of the RNR principles and to re-evaluate their association with intervention effectiveness. A three-level meta-analysis of $k = 31$ studies reporting on 71 effect sizes revealed an overall small and significant intervention effect ($d = 0.382, p < .001$). Although larger effects were found for interventions adhering to any of the RNR principles, none of the RNR principles significantly moderated overall intervention effectiveness. Interventions specifically targeting antisocial recreational activities, and interventions taking into account the youth's age and cultural background did significantly increase overall effectiveness. The results reveal that strong and convincing empirical support for the RNR principles is not yet available, which can mainly be explained by limitations in the design of primary studies on the RNR principles and intervention effectiveness. Suggestions are offered to improve the quality of both primary and secondary research that is needed for establishing a better empirical evidence for the widely acknowledged RNR model.

Keywords: family intervention, juvenile delinquency, meta-analysis, RNR model, RNR principles

Introduction

Recidivism of delinquent youth is a major issue as approximately six in ten prior court referred juveniles in the United States return to court before the age of 18 (Snyder & Sickmund, 2006). Although these numbers reflect a lack of effective programs aimed at reducing youth delinquency (Evans-Chase & Zhou, 2014), promising results have been found in meta-analytic studies that compared family interventions to non-familial responses (Hartnett et al., 2017; Latimer, 2001; Dowden & Andrews, 2003; Van der Stouwe et al., 2014). Following these meta-analyses, family interventions can be defined as in-home and community based interventions aimed at reducing behavior problems of juveniles by improving family functioning. The family risk factors that are targeted in these type of interventions, such as harsh parental discipline and poor parent-child-communication, have been associated with adolescent behavior problems and delinquency (Baldwin et al., 2012). It is theorized that improving family functioning by targeting such family risk factors mediate improvements in other social systems, such as peer relationships, school functioning, and participation in the community (Van der Stouwe et al., 2014). Adolescent delinquency is associated with an accumulation of criminogenic risk factors across such social systems (cf. the ecological systems theory of Bronfenbrenner, 1979). Therefore, many family interventions target criminogenic risks in multiple social systems to create a proper fit between those risks and treatment goals, which is in line with the Risk-Need-Responsivity (RNR) model of Andrews and Bonta (1990) (Dowden & Andrews, 2003; Van der Stouwe et al., 2014).

The RNR model is a theoretical framework that outlines the most important causes of criminal recidivism as well as several principles for effectively reducing criminal engagement. Empirical support for this model has been provided in multiple meta-analyses (e.g., Andrews & Dowden, 2000; 2006; Hanson et al., 2009). However, many of these studies were conducted one or even two decades ago. Therefore, this study aimed to gain knowledge on the effectiveness of applying the RNR principles in family interventions for delinquent youth by replicating and updating the review of Dowden and Andrews (2003). Relative to the work of Dowden and Andrews, the current meta-analysis also synthesized studies on family intervention effectiveness that were published in the past twenty years, and used a more comprehensive coding procedure to examine the moderating effect of the RNR principles on intervention effectiveness. Moreover, an advanced three-level approach to meta-analysis was applied, so that coefficients could be estimated more reliably and more statistical power was achieved than in previous meta-analyses on the RNR principles.

The RNR model as developed by Andrews and Bonta (1990) has become the premier worldwide model for offender assessment and treatment. The most important feature of the RNR model is the focus on applying human services to the justice context instead of relying on deterrence or restoration (Bonta & Andrews, 2017). The model consists of three general principles that guides effective treatment to reduce criminal recidivism: the risk, need, and responsivity principles. The risk principle states that an intervention's intensity should be matched to an offender's risk for recidivism. The need principle indicates that programs should be matched to the unique criminogenic needs of offenders, rather than utilizing a one-size-fits-all approach (Bonta & Andrews, 2017; Gill & Wilson, 2017; Vieira et al., 2009; Wylie et al., 2019). Criminogenic needs are changeable risk factors that are strongly associated with criminal conduct and therefore serve as intervention targets. The needs that are most strongly associated with offending behavior have been labeled as the "Central Eight" by Bonta and Andrews (2017). On the other hand, noncriminogenic needs (e.g., poor self-esteem or depression) are also dynamic attributes of offenders and their circumstances, which, when changed, are not associated with reduced recidivism (Bonta & Andrews, 2017). However, the definition and relevance of non-criminogenic needs in treatment deserve reconsideration, as non-criminogenic needs correspond to responsivity factors that are explicitly important to take into account (Bonta & Andrews, 2003). The *general* responsivity principle prescribes that cognitive social learning methods (e.g., modelling, role-play, or graduated practice) are used to influence behavior. The *specific* responsivity principle states that intervention strategies are aligned with the learning ability, learning style, circumstances, and demographic characteristics – such as gender, age, and ethnicity – of individual offenders (Andrews et al., 2011; Bonta & Andrews, 2007; Bonta & Andrews, 2017).

Adhering to the RNR principles in treatment has been found to produce positive and strong treatment effects across program types, persons, settings, and methodological conditions (e.g., Andrews & Dowden, 2006; Dowden & Andrews, 1999; Hanson et al., 2009). Adhering to the RNR principles may even be the most important explanation for positive program effects, even after accounting for other variables that are assumed to have an effect on treatment effect sizes (Dowden & Andrews, 2003), such as random or nonrandom assignment of participants to experimental and control conditions, and sample size.

The effectiveness of the RNR principles is primarily grounded in findings of multiple meta-analyses conducted by the developers of the RNR model (Bonta & Andrews, 2017). The first meta-analysis synthesized 154 treatment comparisons, and revealed a significantly lower recidivism rate of 35 percent in the treatment conditions that received treatment according to the RNR principles compared to the control conditions that received treatment as usual (Andrews et al., 1990). Later, positive effects of the RNR principles, and particularly the risk principle, were found in meta-analyses on violent reoffending in specifically female and young offender populations (Andrews & Dowden, 2006; Dowden & Andrews, 1999; Dowden & Andrews, 2000). Results of meta-analytic studies by other researchers support the effectiveness of adhering to the RNR principles. For example, sexual offender treatment programs adhering to the RNR principles showed the largest reductions in sexual and general recidivism compared to other treatment programs (Hanson et al., 2009).

Even though meta-analytic research mostly supports the effectiveness of the RNR principles, the findings may be questionable because the coding of the RNR principles was performed inconsistently across reviews (Smith et al., 2009). First, the risk principle is often coded using an aggregate-sample approach instead of a within-study sample approach (Lowenkamp et al., 2006). The former, inspired by Lipsey (1989), means that risk principle adherence is coded whenever a study examines high-risk offenders, defined as offenders with a prior criminal record (Dowden & Andrews, 1999). However, this coding rule does not fully capture the original risk principle, because it only considers the general risk level of the treatment group in terms of prior offenses. It remains ambiguous how this way of coding is related to matching individual offenders to the appropriate level of treatment intensity. Nevertheless, many review authors copied the aggregate-sample approach as available primary studies hardly report on differences in treatment intensity across sampled offenders with different risk levels (Hanson et al., 2009; Koehler et al., 2013; Prendergast et al., 2013).

The within-sample approach to coding the risk principle implies that intervention effects for low- and high-risk groups are reported separately within studies (Andrews and Dowden, 2006). However, this way of coding is also not based on matching risk assessments to treatment goals in individual offenders, but on separating treatment effects for offenders clinically judged as low risk from treatment effects for offenders clinically judged as high risk. The within-sample approach can be more closely related to the risk principle as it is sometimes defined as: “providing intensive interventions

to higher risk offenders and little or no service to low risk offenders” (Hanson et al., 2009, p. 871). However, in practice, almost no primary studies report on differences in treatment intensity (Andrews & Dowden, 2006; Hanson et al., 2009).

Similar to the risk principle, no quantitative reviews seem to exist in which the coding of the need principle aligns with the question how treatment goals matches systematically assessed individual criminogenic needs. Generally, need principle adherence is coded whenever interventions target more criminogenic needs than non-criminogenic needs (Andrews & Dowden, 2006; Dowden & Andrews, 1999; Dowden & Andrews, 2000). In the review of Dowden and Andrews (2003), adherence to the need principle was coded whenever family interventions just targeted family affection/communication or monitoring/supervision even though many other criminogenic needs exist. Studies were coded as not adherent to the need principle when studies were “non-specific” about their targets, which leaves much room for interpretation. In a meta-analysis on the effects of young offender treatment programs in Europe, adherence to the need principle was coded with a three-category item: “low”, “moderate”, and “high”, without further clarification of these levels of adherence (Koehler et al., 2013).

Using such coding rules for examining the effectiveness of the need principle is remarkable, because identifying and addressing the individual criminogenic needs underlying juvenile delinquency is central to the needs principle and seems crucial for the prevention of reoffending (Vieira et al., 2009). The importance of targeting criminogenic needs can be explained by the fact that juvenile delinquency is associated with an accumulation of risk factors across social contexts (i.e., individual, family, peers, school, and neighborhood context) that reflect criminogenic need factors (Henggeler et al., 2009; Van der Stouwe et al., 2014). To examine whether the need principle contributes to program effectiveness, it is important to know whether targeting criminogenic needs, such as the Central Eight factors, is associated with program effectiveness. However, no meta-analyses have yet been performed in which the effectiveness of targeting the most important criminogenic needs - labeled as the “Central Eight” by Bonta and Andrews (2017) - has been examined.

Guidelines for coding the responsivity principle have been more developed for the general responsivity principle than for the specific responsivity principle. In an early study of Andrews et al. (1990), the general responsivity principle was coded as adherent for “all behavioral programs”. Thereafter, adhering to the general responsivity principle

was more specifically coded when programs used social learning or cognitive behavioral techniques (Andrews & Dowden, 2006; Dowden & Andrews, 1999; Dowden & Andrews, 2000; Dowden & Andrews, 2003; Prendergast et al., 2013). Koehler et al. (2013) coded the responsivity principle in a way that refers to the specific responsivity principle: whenever treatment delivery was adapted to offenders' unique learning styles and capabilities, it was coded as adherent to the responsivity principle. Additionally, Hanson et al. (2009) considered the extent to which programs made special efforts to engage offenders in treatment. However, in many studies specific responsivity has not been coded at all (Andrews & Dowden, 2000; Dowden & Andrews, 2003; Andrews & Dowden, 2006).

Overall, it appears that the available meta-analyses have not always properly examined the effectiveness of the risk need and responsivity principles. Also, many of these reviews were conducted more than one or even two decades ago (Andrews & Dowden, 2000; Dowden & Andrews, 2003; Andrews & Dowden, 2006; Hanson et al., 2009). Moreover, more advanced statistical techniques have been developed that allow for a better synthesis of effect sizes in meta-analyses (Assink & Wibbelink, 2016). In recent years, the RNR principles have been implemented in treatment delivered outside the setting of criminal justice, such as child welfare (Brogan et al., 2015; Van der Put et al., 2018). Therefore, renewed insights into the effectiveness of adhering to these principles are important, because the RNR model may promote better allocation of services (Ter Beek, 2018), but will only be effective when applied appropriately (Brogan et al., 2015; Goense et al., 2016; Ter Beek et al., 2018).

In sum, this meta-analysis aimed to re-examine the association between the risk, need, and responsivity principles of the RNR model and the effectiveness of family programs for juvenile recidivism by replicating and updating the review of Dowden and Andrews (2003). To our knowledge, this meta-analysis was the first to examine differences in effect between the aggregate-sample and within-sample approaches to the risk principle, and whether or not intervention effectiveness is dependent on the interaction between a sample's risk level and treatment intensity. Also, the effect of targeting specific criminogenic need and responsivity factors, as well as adherence to the general and specific responsivity principles were examined. Finally, we examined the potential moderating effect of study and sample characteristics, such as study design, treatment duration, gender, and ethnicity.

Method

Study Inclusion

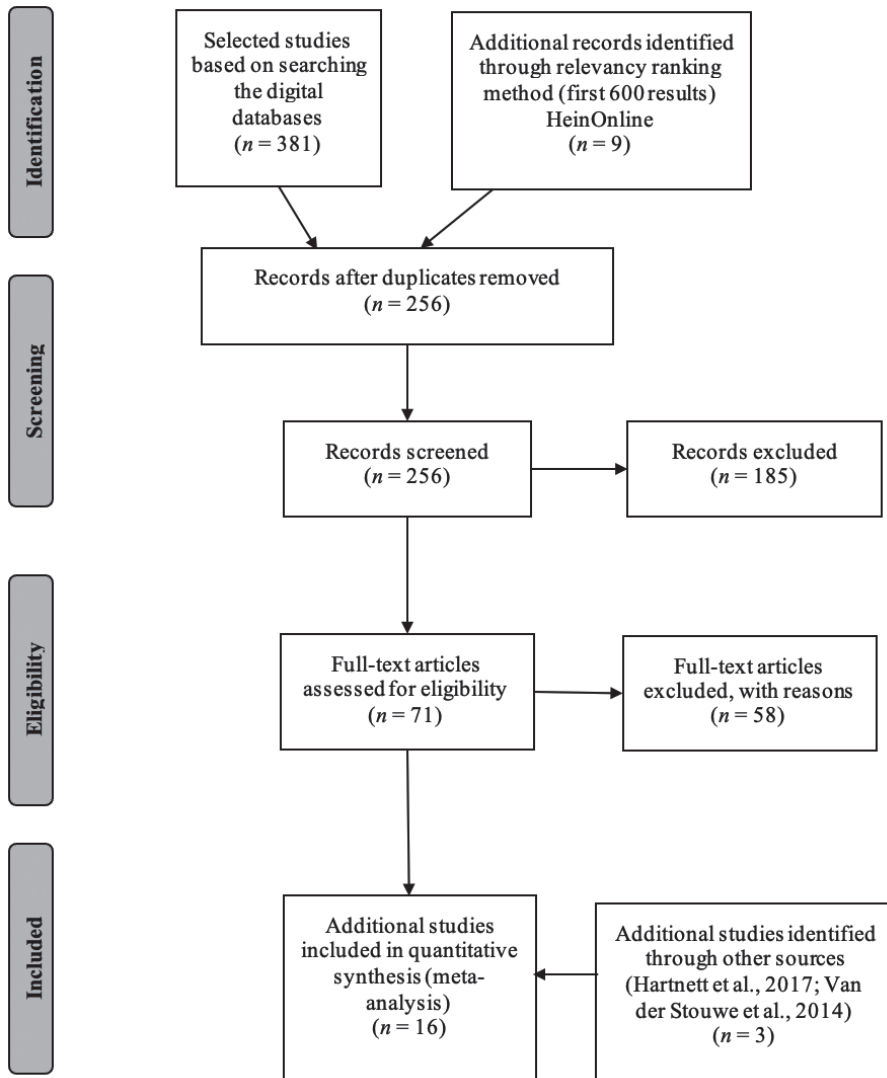
First, we included studies that were synthesized by Dowden and Andrews (2003). The authors did not list the included studies in their meta-analysis, but references to these studies could be derived from a larger meta-analysis of Dowden (1998). However, the study inclusion of Dowden (1998) focused more broadly than on young offenders and family interventions alone. Therefore, we included studies that were synthesized by Latimer (2001) who focused his meta-analysis on the impact of involving families in treatment for young offenders, and in which he included most (89%) of the studies that were also synthesized by Dowden (1998). Of the 35 studies synthesized by Latimer, only 15 could be included in the current meta-analysis given our inclusion criteria that we used in the additional search (see below). Reasons for study exclusion were: absence of a distinct control group (eight studies), no evaluation of a family intervention (three studies), insufficient information for proper study coding (two studies), no results for an intent-to-treat group (one study), and recidivism not being a study outcome (one study). Despite our efforts, such as external library requests and contacting several authors, five studies could not be retrieved.

Second, in an additional search, the following criteria were used to select relevant studies examining the effect of a family intervention on recidivism of juvenile delinquents: (1) The study sample consisted of offenders younger than 18 years old. (2) Each treatment group was offered a family treatment intervention, and each control group received treatment as usual (TAU), another intervention program, or a postponed intervention program. (3) As a form of quality control, only Randomized Controlled Trials (RCT) and quasi-experimental studies (in which a treatment condition is compared to a control condition) were included. (4) Studies had to examine and report post-test and/or follow-up comparisons of recidivism rates, and sufficient statistical information to manually calculate an effect size. (5) Only studies written in Dutch and English of which the full text could be retrieved were included. Also, only studies performed in Western countries (i.e., European countries, Australia, New Zealand, Canada, and the US) were included. (6) Also as form of quality control, studies had to be published in peer-reviewed scientific journals or be (part of) a dissertation that was accessible to the authors of this review. In addition, effect sizes were based on intent-to-treat analyses and not analyses of only intervention completers.

Search Strategy

In searching relevant studies for inclusion, we performed several complementary search strategies. First, we used the backward citation tracking technique for inclusion of studies from Latimer (2001). Second, we searched for studies in the electronic databases PsycINFO, Medline, Web of Science, Scopus, and HeinOnline. In this search, the following keywords were used in varying combinations to retrieve relevant articles, book chapters, dissertations, and reports: family therapy, family, families, intervention, family empowerment, multisystem*, multi-system*, juvenile delinquency, adolescent, young adult*, teen*, youngster*, young people, youth*, minors*, under age*, underage*, juvenile*, girl*, boy*, preadolesc*, adolesc*, recidivism, rearrest*, recidiv*, reconvict, re-convict*, reincarcerat*, re-incarcerat*, reoffen*, re-offen*. In total, the search procedure yielded 981 studies. After thoroughly screening these studies, thirteen studies met the inclusion criteria and were included in the study. Additionally, two recent meta-analyses on Family Functional Therapy and Multisystemic Therapy were screened for eligible studies (i.e., Hartnett et al., 2017; Van der Stouwe et al., 2014), resulting in three additional studies (see Figure 1 for a flow chart of the search results). To determine whether studies were eligible for inclusion, the first author read titles, abstracts, and if necessary, full article texts. Whenever there was doubt about inclusion of a study, the other authors were consulted. The final list of studies eligible for inclusion was discussed and agreed upon by all authors of this study.

Figure 1
Flow Chart of Additional Search Results



Coding of Studies

A coding scheme was developed (Appendix A), partially based on De Vries et al., (2015), and following guidelines of Lipsey and Wilson (2001). First, “general study information” (publication year, study design) was coded. Second, “sample descriptors” (age, gender, percentage of cultural minority) were retrieved. Third, “study and program characteristics” were coded (e.g., sample size, dropout rate, supportive evidence, control group type, treatment duration). Finally, we focused on coding the risk need and responsivity principles. When treatment descriptions in primary studies were insufficient, intervention program manuals or factsheets (when available) were screened. In coding recidivism, four dimensions of delinquency were taken into account: (1) participation (yes/no), (2) frequency, (3) seriousness, and (4) versatility (number of crime types). The measurement type of delinquency was also coded (i.e., official records, self-report, parent report, or other).

For extracting information on study quality, we followed the approach of an integrative study quality coding scheme (Van der Stouwe et al., 2021) that was based on previously validated quality indices, including the Quality Assessment Tools for Quantitative Studies (QATQS, Thomas et al., 2004), the Quality Index (QI, Downs & Black, 1998), and the Cochrane Collaboration’s tool for assessing risk of Bias (Higgins et al., 2011). The coding scheme rates each study on seven features: study design (randomized controlled trial vs. quasi-experimental design), adjustment for pre-test differences (yes vs. no), sample size, (4) attrition (percentage of referred participants that completed the study), (5) reporting of sample characteristics (age, gender, and ethnicity), (6) type of risk assessment used for matching the intervention to the client’s risk level (clinical judgment vs. standardized risk assessment), and (7) type of needs assessment used for matching the intervention to the client’s criminogenic needs (clinical judgment vs. standardized need assessment). Each feature is rated on a two-, three-, or four-point scale (0 indicating lowest quality to 2, 3, or 4 indicating highest quality). The seven item scores were summed resulting in a (total) quality score. The lowest and highest possible scores were 0 and 15, respectively. Across all included studies, the quality index ranged from four to eleven ($M = 8.54$, $SD = 1.93$).

The risk principle was coded following the within-sample approach and the aggregate-sample approach (e.g., Dowden & Andrews, 2000; 2003; Koehler et al., 2013), yet considering the initial definition of the risk principle (Bonta & Andrews, 2017). The within-sample approach was coded as adherent when the intervention intensity was

matched to the individual risk level of offenders, that was assessed by (structured) clinical or actuarial judgment, for example: “The average contact frequency will be determined based on recidivism risk of unacceptable behavior, and therefore may vary.” (Multisystemic Therapy, 2020, p. 5). The aggregate-sample approach was coded as adherent when the general risk level of the sample was matched to the intensity of the intervention. The general risk level of the sample was coded to be “low” when first time offenders were sampled, whereas it was coded “high” when the majority of the sampled participants had formally penetrated the judicial system at the start of the study and had a prior criminal record or showed severe antisocial/violent behavior according to our judgment (e.g., “Inclusion criteria for the study were identical to those used by the MST offering agencies: severe and violent antisocial behavior at home, school or community, sufficiently serious to require treatment”, Asscher et al., 2014, p. 230). The decision whether or not a sample’s general risk level was matched to the intensity of the intervention (coded as yes or no) was based on a description of the treatment intensity in the primary study itself, or a program manual or factsheet. For example, the MST manual reports the following: “MST is an intensive at-home treatment program during three to five months with a 24/7 availability of therapists.” (Multisystemic Therapy, 2020, p. 1). In addition, the interaction between the risk level of the sample (high vs. low) and treatment intensity (total contact hours) was tested to examine how the risk principle affects treatment effectiveness.

The need principle was coded in several ways. First, the two forms of “appropriate forms of family intervention” (i.e., improving the parent-child relationship and parental monitoring/supervision) were coded, in which we followed the original meta-analysis of Dowden and Andrews (2003). Second, the need principle was coded as adherent when primary studies explicitly described that at least one specific need factor was targeted, or when studies described treatment methods that aim to target at least one criminogenic need, which for example is done in the study of Dakof et al. (2015, p. 6): “The goals of stage II are to help teens communicate effectively with their parents and other adults (family intervention: parent-child relationship), develop emotion regulation and coping skills (antisocial personality pattern), and enhance social competence and alternatives to delinquency (criminal involvement; procriminal attitudes) and substance use (substance abuse)”. In cases where none of the Central Eight needs were targeted in an intervention, adherence to the needs principle was coded as non-adherent. In this way, we followed the coding approach in other meta-analyses on the effectiveness of the RNR principles (Andrews & Dowden, 2006;

Dowden & Andrews, 1999; Dowden & Andrews, 2000).

Moreover, in addition to previous meta-analyses, we coded whether or not a criminogenic needs assessment was performed as required by the need principle for treatment tailoring. At first, we also aimed for coding the type of needs assessment (i.e., clinical judgment vs. assessment instrument), but this was not intelligible as a structured instrument for needs assessment was used in only one of the included studies (i.e., Celinska et al., 2018). Thus, we only coded whether or not a needs assessment was performed (0 = no; 1 = yes, with either an instrument or clinical judgment). Furthermore, as Bonta and Andrews (2017) argued that the Central Eight are the most empirically well-supported dynamic risk factors for recidivism, we coded for each of these factors whether or not it was targeted by the intervention examined in a primary study. In this way, targeting each of the Central Eight (criminogenic need) factors could be examined as a moderator of the effectiveness of family interventions. However, the Central Eight factor “family” was not coded, as only studies examining the effectiveness of family interventions were included. Finally, the total number of Central Eight factors targeted by an intervention was coded.

Adherence to the general responsivity principle was coded as adherent when programs used social learning or cognitive behavioral techniques, such as modelling, or role-play (Andrews & Dowden, 2006; Dowden & Andrews, 1999; Dowden & Andrews, 2000; Dowden & Andrews, 2003; Prendergast et al., 2013). For example: “FFT therapists rarely focus on “communication” but instead attempt to change a specific principle that is not evident in the family relational pattern (e.g., brevity, specificity, congruence). This is often accompanied by role-playing within a session.” (Sexton & Alexander, 2004, p. 74). Adherence to the specific responsivity principle was coded as adherent when programs were tailored to at least one of the following factors: client intelligence/cognitive skills level, social support, gender, age/developmental stage, culture/race/ethnicity, psychopathological problems, and motivation (Bonta & Andrews, 2017). For example: “FFT is designed to be provided by therapists of all races and ethnicity, all spiritual systems, and gender/sexual preferences. [...] Matching to the client allows FFT to respect, value, and work within the important cultural, racial, religious, and gender-based values of the client. [...] The therapists at these sites are as diverse as the clients in regard to gender, age, and ethnic origin.” (Sexton & Alexander, 2004, p. 3, p. 52, & p. 96). Finally, the total number of responsivity factors targeted by the intervention were coded.

Determining the inter-rater agreement is an essential aspect of a coding procedure in meta-analytic research, as unreliability in coding procedures may add random variation to the analyses and reduce the reliability of results (e.g., Dieckmann et al., 2009). Therefore, the first, second, and anchor author of this manuscript decided to independently code all studies that could be included. Next, any discrepancy in coding was discussed and resolved until full consensus was established. In this way, a percentual agreement of 100% was reached for all variables that were coded.

Data Analysis

All relevant outcomes reported in the included primary studies were transformed into the standardized difference between two means, also referred to as Cohen's d . In most instances, proportions, means and standard deviations, and odd ratios were transformed into Cohen's d using the methods of Wilson (2019). For calculating each effect size, it was important that the direction of the effect (positive or negative) corresponded with the statistical data reported in the primary study. A positive effect indicated that the intervention group showed less recidivism than the control group, whereas a negative effect indicated that the intervention group showed more recidivism than the control group. If a study provided pretest measurements on recidivism, effect sizes of post-test and follow-up measurements were corrected for these baseline differences by subtracting pre-test effects from post- and follow-up effects. Effect sizes were transformed into z scores and checked for outliers to control for any disproportionate influence of outliers on the results. All standardized effect sizes fell within the range of $-3.29 < z < 3.29$ (Tabachnik & Fidell, 2013), implying that no outliers were identified. A three-level random-effects meta-analytic model was used to synthesize all effect sizes and to model effect size dependency that arose from the fact that more than one relevant effect size could be extracted from individual primary studies (Assink & Wibbelink, 2016). In this model, three levels of variance were accounted for to model the effect size dependency: sampling variance of the observed effect sizes (Level 1), variance of effect sizes within studies (Level 2), and variance of effect sizes between studies (Level 3). The overall effectiveness of family interventions was estimated in an intercept-only model. Potential moderating effects of the coded variables were examined by adding one of the coded variables as a covariate to the intercept-only model.

Publication Bias

A common problem in conducting a meta-analysis is that studies with nonsignificant or negative results are less likely to be published than studies with positive and significant results. This phenomenon is called publication bias and is often referred to as the “file drawer problem” (Rosenthal, 1995). To examine the problem of missing data due to publication bias, we conducted the funnel-plot-based trim and fill method as described by Duval and Tweedie (2000a; 2000b). In case of an asymmetrical distribution of effect sizes, the trim and fill method restores symmetry of the distribution by imputing effect size estimates from “missing” studies. Effect sizes can be imputed either to the left or right side of the estimated mean effect in the funnel plot, depending on whether below average or above average effect sizes are underrepresented in the data. In case of publication bias, below average effect sizes are underrepresented.

Results

Descriptives

In total, $k = 31$ studies published between 1977 and 2020 were included from which $u = 71$ effect sizes were extracted (articles indicated with an asterisk in the references were included in the meta-analysis). The sample size of the included studies ranged from $N = 16$ to $N = 1553$ ($M = 210.58$, $SD = 302$), and the mean age of the young offenders at start of the studies ranged from 11.5 to 16.6 years ($M = 14.86$, $SD = 0.99$). The average percentage of boys in study samples was 78.31%, and the average percentage of youth with a cultural minority background was 41.48%. The average follow-up duration was $M = 34.74$ months ($SD = 54.35$). The studies reported on recidivism data extracted from official records (77.4%), and self-report (22.6%).

Overall Effect Size

The estimated summary effect of family intervention programs on juvenile reoffending of all studies ($k = 31$) and effect sizes ($u = 71$) was $d = 0.382$, $p < .001$, 95% CI [0.228, 0.536] (Table 1). According to Cohen’s criteria (1988), this effect is small in magnitude. This summary effect should be interpreted with caution, as we found significant within-study variance (level 2; 19% of the total variance) and between-study variance (level 3; 65% of the total variance), implying heterogeneity in effect sizes (Table 1). Given this significant variability in effect sizes, testing variables as potential moderators of the summary effect was justified.

Table 1
Overall Effect of Family Interventions for Juvenile Delinquency

	# Studies	# ES	Mean <i>d</i> (SE)	95% CI	Sig. (<i>p</i>)	% Var. level 1	Level 2 variance	% Var. level 2	Level 3 variance	% Var. level 3
Overall effect	31	71	0.382 (0.08) ^{***}	0.228, 0.536	<.001 ^{***}	15.96	.038 ^{***}	18.63	.134 ^{***}	65.41

Note. # Studies = number of studies; # ES = number of effect sizes; Mean *d* = mean effect size (Cohen's *d*); SE = standard error; CI = confidence interval; Sig. = significance; % Var. = percentage of distributed variance; level 1 variance = sampling variance; level 2 variance = variance within studies; level 3 variance = variance between studies.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Analysis of Bias

The results of the trim-and-fill analysis showed that publication bias may have been present in the data, as the effect size distribution was asymmetrical. Appendix C shows that the trim-and-fill algorithm imputed four effect sizes from three studies to the left of the estimated mean effect to restore the symmetry of the effect size distribution. These results reflect an underrepresentation of below average effect sizes, indicating that the estimated mean effect may be an overestimation of the true effect and that publication bias may be present. After the imputed effect sizes were added to the dataset, a re-estimation of the overall effect yielded a smaller but still significant effect ($d = 0.296, p < .01, 95\% \text{ CI } [0.120; 0.473]$). This “corrected” effect was slightly smaller than the initially estimated overall effect ($d = 0.382; \Delta d = 0.086$), implying that publication bias was present to a limited extent.

Moderator Analyses

Table 2 presents the results of all moderator analyses. No significant moderating effects were found for study characteristics including study quality nor sample characteristics. We neither found significant moderating effects for the risk, need, and responsivity principles. However, we did find larger mean effect size estimates for interventions that adhered to each of the RNR principles compared to interventions not adhering to these principles (see Table 2). As for the total number of applied RNR principles, we found once again no significant moderating effect implying there are no significant differences in mean effect between interventions adhering to zero or one principles, two principles, or three principles. We did find larger effect size estimates for interventions adhering to two principles and interventions adhering to three principles (see Table 2). Finally, the moderating effect of the risk principle was also examined by testing the interaction between the risk level of the sample (high vs. low) and treatment intensity (total contact hours). The results (see Table 3) showed no significant moderating effect of the interaction term while controlling for the main effects, meaning that the effect of family interventions does not depend on how treatment intensity is matched to the risk level of juvenile offenders.

As for the other RNR characteristics, three moderating effects were found. First, interventions targeting “Antisocial recreational activities” produced a larger effect size ($d = 0.488$) than interventions not targeting this factor ($d = 0.142$). Second, significant positive moderating effects were found for interventions taking “Age” and “Culture/Ethnicity” into account. Interventions tailored to any of these factors revealed larger effect sizes ($d = 0.521$) than interventions that were not tailored to any of these factors ($d = 0.206$).

Table 2
Results of the Moderator Analyses (Bivariate Models)

Moderator variables	# Studies	# ES	Intercept/Mean d (95% CI)	β_1 (95% CI)	$F(df_1, df_2)^a$	p^b	Level 2 variance	Level 3 variance
Study characteristics								
Design	31	71			0.421 (1, 69)	.519	.038***	.137***
Quasi-experimental (RC)	9	13	0.465 (0.168, 0.762)**					
RCT	22	58	0.351 (0.169, 0.534)***	-0.113 (-0.462, 0.235)	0.623 (1, 69)	.433	.038***	.136***
Type of care control group	31	71						
TAU (RC)	21	44	0.340 (0.151, 0.528)***	0.131 (-0.199, 0.460)	2.567 (1, 69)	.114	.039***	.125***
Other program	10	27	0.470 (0.200, 0.740)***	-0.010 (-0.023, 0.003)	0.212 (1, 64)	.647	.027**	.141***
Publication year	31	71	0.371 (0.220, 0.522)***	-0.037 (-0.200, 0.125)	2.679 (1, 58)	.107	.031**	.131***
Sample age	28	66	0.341 (0.178, 0.504)	-0.006 (-0.012, 0.001)	1.238 (1, 64)	.270	.027**	.137***
Cultural minority (%)	28	60	0.342 (0.177, 0.508)***	-0.006 (-0.017, 0.005)	0.416 (1, 48)	.522	.026*	.110***
Boys (%)	29	66	0.331 (0.166, 0.495)***	-0.003 (-0.014, 0.007)	0.357 (1, 46)	.553	.031*	.153***
Treatment duration (weeks)	20	50	0.384 (0.208, 0.559)***	0.025 (-0.015, 0.066)	1.543 (1, 58)	.219	.034**	.126***
Treatment total contact hours	17	48	0.484 (0.264, 0.703)***					
Number of previous convicted offenses	26	60	0.378 (0.214, 0.542)***					
Study quality	31	71	0.383 (0.228, 0.538)***	-0.010 (-0.085, 0.066)	0.065 (1, 69)	.800	.040***	.134***
Sample risk level	31	71			0.239 (1, 69)	.627	.039***	.137***
Low (RC)	5	15	0.293 (-0.103, 0.689)					
High	26	56	0.399 (0.230, 0.568)***	0.105 (-0.325, 0.536)	2.262 (2, 68)	.112	.040***	.116***
Intervention program	31	71						
MST (RC)	12	29	0.522 (0.291, 0.753)***					
FFT	6	7	0.521 (0.132, 0.910)**	-0.001 (-0.453, 0.451)				
Other	12	35	0.207 (-0.011, 0.426)	-0.315 (-0.632, -0.003)				
Adjusted ES	33	71			0.582 (1, 69)	.448	.037***	.136***
Unadjusted (RC)	26	61	0.361 (0.197, 0.526)***					

Moderator variables	# Studies	# ES	Intercept/Mean <i>d</i> (95% CI)	β_1 (95% CI)	F (df1, df2) ^a	p^b	Level 2 variance	Level 3 variance
Adjusted	7	10	0.466 (0.198, 0.735)***	0.105 (-0.170, 0.380)				
RNR characteristics								
Risk principle (aggregate)	31	71			0.415 (1, 69)	.521	.038***	.137***
No adherence (RC)	6	8	0.273 (-0.100, 0.646)					
Adherence	25	63	0.406 (0.234, 0.577)***	0.133 (-0.278, 0.543)				
Risk principle (within)	31	71			0.278 (1, 69)	.600	.038***	.138***
No adherence (RC)	17	29	0.343 (0.126, 0.560)***					
Adherence	14	42	0.425 (0.201, 0.649)***	0.082 (-0.229, 0.394)				
Needs assessment	31	71			1.123 (1, 69)	.293	.039***	.130***
No assessment (RC)	14	26	0.288 (0.056, 0.521)*					
Assessment	17	45	0.452 (0.250, .655)***	0.164 (-0.145, 0.472)				
Needs principle Central Eight (without family/marital)	31	71			0.558 (1, 69)	.458	.038***	.135
No adherence (RC)	2	2	0.136 (-0.539, 0.811)					
Adherence	29	69	0.396 (0.237, 0.554)***	0.260 (-0.434, .653)				
Central eight: Criminal involvement	31	71			0.534 (1, 69)	.467	.038***	.136***
Not targeted (RC)	10	19	0.297 (0.018, 0.577)*					
Targeted	21	52	0.420 (0.234, 0.607)***	0.123 (-0.213, 0.459)				
Central eight: Antisocial personality	31	71			0.018 (1, 69)	.893	.038***	.139***
Not targeted (RC)	5	12	0.358 (-0.039, 0.755)					
Targeted	26	59	0.387 (0.217, 0.557)***	0.029 (-0.403, 0.461)				
Central eight: Procriminal attitudes	31	71			0.011 (1, 69)	.916	.038***	.139***
Not targeted (RC)	9	18	0.369 (0.070, 0.668)*					

Moderator variables	# Studies	# ES	Intercept/Mean <i>d</i> (95% CI)	β_1 (95% CI)	F (df1, df2) ^a	p^b	Level 2 variance	Level 3 variance
Targeted	22	53	0.388 (0.204, 0.571) ^{***}	0.019 (-0.332, 0.370)				
Central eight: Procriminal associates	31	71			0.778 (1, 69)	.381	.038 ^{***}	.135 ^{***}
Not targeted (RC)	10	15	0.279 (-0.001, 0.559)					
Targeted	21	56	0.428 (0.242, 0.613) ^{***}	0.149 (-0.187, 0.484)				
Central eight: School/work	31	71			3.118 (1, 69)	.082	.037 ^{***}	.123 ^{***}
Not targeted (RC)	8	13	0.147 (-0.156, 0.450)					
Targeted	23	58	0.455 (0.284, 0.626) ^{***}	0.308 (-0.040, 0.656)				
Central eight: Antisocial recreational activities	31	71			4.872 (1, 69)	.031 [*]	.037 ^{***}	.114 ^{***}
Not targeted (RC)	10	22	0.142 (-0.117, 0.401)					
Targeted	21	49	0.488 (0.313, 0.663) ^{***}	0.346 (0.033, 0.658) [*]				
Central eight: Substance abuse	31	71			0.042 (1, 69)	.838	.038 ^{***}	.139 ^{***}
Not targeted (RC)	8	11	0.354 (0.035, 0.673) [*]					
Targeted	23	60	0.392 (0.212, 0.571) ^{***}	0.038 (-0.328, 0.404)				
Number of targeted Central Eight Needs	31	71	0.378 (0.225, 0.532) ^{***}	0.033 (-0.029, 0.095)	1.147 (1, 69)	.288	.038 ^{***}	.132 ^{***}
Family intervention: Parent-child relationship	31	71			2.912 (1, 69)	.092	.038 ^{***}	.123 ^{***}
Not targeted (RC)	6	14	0.119 (-0.221, 0.460)					
Targeted	25	57	0.443 (0.227, 0.610) ^{***}	0.324 (-0.055, 0.703)				
Family intervention: Monitoring/supervision	31	71			1.462 (1, 69)	.231	.038 ^{***}	.131 ^{***}
Not targeted (RC)	9	14	0.230 (-0.062, 0.523)					
Targeted	22	57	0.438 (0.259, 0.618) ^{***}	0.208 (-0.135, 0.551)				

Moderator variables	# Studies	# ES	Intercept/Mean <i>d</i> (95% CI)	β_1 (95% CI)	<i>F</i> (df1, df2) ^a	<i>p</i> ^b	Level 2 variance	Level 3 variance
Appropriate forms of family intervention (PC-relationship and/or monitoring)	31	71			1.550 (1, 69)	.217	.038***	.131***
Not targeted (RC)	5	6	0.157 (-0.234, 0.548)					
Targeted	26	65	0.422 (0.256, 0.589)***	0.265 (-0.160, 0.690)				
General responsivity principle	31	71			0.398 (1, 69)	.530	.038***	.138***
No adherence (RC)	7	9	0.287 (-0.053, 0.627)					
Adherence	24	62	0.408 (0.233, 0.584)***	0.121 (-0.261, 0.503)				
Specific responsivity principle	31	71			3.453 (1, 69)	.067	.037***	.121***
No adherence (RC)	8	18	0.136 (-0.165, 0.437)					
Adherence	23	53	0.458 (0.288, 0.629)***	0.322 (-0.024, 0.668)				
Specific responsivity: Intelligence	31	71			2.443, (1, 69)	.123	.039***	.122***
Not tailored (RC)	19	42	0.284 (0.091, 0.477)**					
Tailored	12	29	0.522 (0.288, 0.756)***	0.238 (-0.066, 0.542)				
Specific responsivity: Social support	31	71			2.743, (1, 69)	.102	.038***	.123***
Not tailored (RC)	10	22	0.197 (-0.071, 0.464)					
Tailored	21	49	0.464 (0.284, 0.645)***	0.268 (-0.055, 0.590)				
Specific responsivity: Gender	31	71			0.004, (1, 69)	.949	.038***	.139***
Not tailored (RC)	14	54	0.385 (0.212, 0.559)***					
Tailored	17	17	0.372 (0.010, 0.734)*	-0.013 (-0.415, 0.388)				
Specific responsivity: Age	31	71			4.693 (1, 69)	.034*	.040***	.111***
Not tailored (RC)	14	35	0.206 (-0.009, 0.421)					
Tailored	17	36	0.521 (0.326, 0.716)***	0.315 (0.025, 0.605)*				

Moderator variables	# Studies	# ES	Intercept/Mean d (95% CI)	β_1 (95% CI)	F (df1, df2) ^a	p^b	Level 2 variance	Level 3 variance
Specific responsivity: Culture/ethnicity	31	71			4.693 (1, 69)	.034*	.040***	.111***
Not tailored (RC)	14	35	0.206 (-0.009, 0.421)					
Tailored	17	36	0.521 (0.326, 0.716)***	0.315 (0.025, 0.605)*	1.691 (1, 69)	.198	.038***	.129***
Specific responsivity: Psychopathology	31	71						
Not tailored (RC)	17	31	0.287 (0.077, 0.497)**					
Tailored	14	40	0.485 (0.265, 0.705)***	0.198 (-0.106, 0.503)	0.967 (1, 69)	.329	.039***	.134***
Specific responsivity: Motivation	21	71						
Not tailored (RC)	11	21	0.276 (0.010, 0.541)*					
Tailored	20	50	0.436 (0.247, 0.626)***	0.161 (-0.165, 0.487)	3.350 (1, 69)	.072	.039***	.120***
Number of targeted responsivity factors	31	71	0.386 (0.237, 0.534)***	0.052 (-0.005, 0.109)				
Adherence to number of RNR Principles (<i>Risk only within</i>)	31	71			0.833 (2, 68)	.439	.037**	.138***
0/1 principles (RC)	6	8	0.168 (-0.199, 0.535)					
2 principles	11	21	0.437 (0.168, 0.706)**	0.269 (-0.186, .724)				
3 principles	14	42	0.425 (0.201, 0.649)***	0.257 (-0.173, .687)				

Note. # Studies = number of studies; # ES = number of effect sizes; mean d = mean effect size Cohen's d referring to the estimated mean effect for a specific category of a variable tested as a moderator (one or more asterisks imply that a mean effect significantly differs from a null effect); CI = confidence interval; β_1 = estimated regression coefficient; df = degrees of freedom; Level 2 variance = variance of effect sizes within studies; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.
^b p value of the omnibus test (a significant value indicates that a coded variable is a moderator of the overall effect of family interventions for juvenile delinquency).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 3
Multiple Moderator Results for the Interaction Between Risk Level (High vs. Low) and Treatment Intensity (in Total Contact Hours)

Moderator variables	β (SE)	95% CI	<i>t</i>
Intercept	0.389	-0.182, 0.960	1.374
Control variables			
Sample risk level (high vs. low)	0.498	-0.198, 1.195	1.443
Treatment intensity (total contact hours)	-0.008	-0.020, 0.003	-1.530
Risk level * treatment intensity	0.033	-0.011, 0.076	0.076
<i>F</i> (df1, df2) ^a	1.032 (3, 44)		
Level 2 variance	0.032		
Level 3 variance	0.140***		

Note. β = estimated regression coefficient; SE = standard error; CI = confidence interval; df = degrees of freedom; Level 2 variance = residual variance in effect sizes extracted from the same study; Level 3 variance = residual variance between studies.

^a Omnibus test of all regression coefficients in the model. *** $p < 0.001$.

Discussion

In line with findings of previous meta-analyses (Dowden & Andrews, 2003; Latimer, 2001; Van der Stouwe et al., 2014), an overall small and significant effect was found for family interventions on juvenile reoffending. Although larger effect sizes were found for interventions adhering to any of the RNR principles, none of the RNR principles significantly moderated the overall intervention effect. In other words, we did not find significant differences in effects between interventions adhering to the RNR principles and interventions not adhering to the RNR principles.

The Risk Principle

In line with Dowden and Andrews (2003), we found no evidence for the effectiveness of adhering to the risk principle in family interventions for delinquent youth. According to Dowden and Andrews, their result was due to a limitation in the coding procedures rather than to any weakness in construct validity of the risk principle. Therefore, we coded the risk principle in two ways, given the original risk principle definition of Bonta and Andrews (2017). However, we still did not find a moderating effect for the within-sample approach (i.e., matching the intervention intensity to the risk level of an individual offender), nor the aggregate-sample approach (i.e., matching the intervention intensity only to the general risk level of the sample instead of individual offenders). We neither found a moderating effect of the interaction between

the risk level of the sample (high vs. low) and treatment intensity (total contact hours), implying that the effect of family interventions does not depend on how treatment intensity is matched to the risk level of juvenile offenders.

These non-significant results may be explained by how the within-sample approach to the risk principle was applied in the included studies. The programs adhering to the within-sample approach matched intervention intensity to the recidivism risk of individual offenders by assessing the recidivism risk on the basis of clinical judgment instead of the outcome of a structured assessment instrument. Multiple studies have shown that statistical methods outperform other methods for risk assessment, and that these methods are crucial for properly identifying juveniles who are at risk for becoming a - life-course persistent - offender (Hanson & Morton-Bourgon, 2009; Schwalbe et al., 2004). Not using the best available methods for risk assessment may be an important reason for not finding a significant effect for adherence to the risk principle in the current meta-analysis.

The Need Principle

Dowden and Andrews (2003) argued that their findings provided strong empirical support for the need principle. However, our results do not confirm this. Although we coded the need principle in multiple ways, the results showed that interventions adhering to the need principle were not significantly more effective than interventions not adhering to the need principle. Further, a remarkable finding was that a standardized instrument for assessing criminological needs of individual offenders was used in only one study (i.e., Strengths and Needs Assessment, Celinska et al., 2018). This indicates that a one-size-fits-all approach to treatment delivery is still often used in studies examining treatment effectiveness rather than matching programs to unique criminogenic needs of individual offenders, which is prescribed by the need principle of the RNR model (Bonta & Andrews, 2017).

However, one of the Central Eight factors (i.e., antisocial recreational activities) was found to positively moderate the effect of family interventions. Risky leisure or recreational routines often involve factors that are predictive of juvenile delinquency, such as the presence of antisocial peers and poorer social control (Hoeben et al., 2016). An example of an intervention that stimulates prosocial activities is Multisystemic Therapy that explicitly considers caregivers as key agents of change in peer relations

and social activities of their children by enhancing parental monitoring and discipline, and facilitating their child's participation in prosocial activities (Henggeler et al., 2009). Further, the intervention Multidimensional Family Therapy (MDFT) aims for a proper social reintegration of adolescents after incarceration, and promotes prosocial peer relations and activities. Another goal of MDFT is to improve all family members' relationships within and between social systems such as school, workplace, and neighborhood (Liddle et al., 2001). Based on the design of the included primary studies and thus also the current meta-analysis, we cannot infer with certainty whether the benefit of prosocial activities is especially gained from programs with an explicit component that targets antisocial recreational activities, or that the same benefit is also gained from programs targeting family and/or peer support that in turn decrease antisocial recreational activities and increase prosocial activities. Future studies in which the potential mediating effect of family and/or peer support is examined are therefore recommended. However, it can be expected that both types of programs contribute to lower recidivism rates, as antisocial recreational activities and low family/peer support have been shown to be well-established and empirically derived dynamic risk factors (Andrews & Bonta, 2006).

In the meta-analysis of Dowden and Andrews (2003), the need principle was defined by only two key elements: "enhancing monitoring/supervision" and "building a positive parent-child relationship". Although the current results revealed larger mean effect sizes for interventions targeting one or both of these elements, targeting these elements did not significantly increase the intervention effects. Based on their results, Dowden and Andrews (2003) highlighted the need for program deliverers to (1) specifically target these two elements, and (2) avoid delivering a nonspecific and unstructured intervention program. As delivering interventions with high treatment integrity (i.e., delivery of the intervention as intended) to young delinquents is indeed vital (Goense et al., 2016), Dowden and Andrews seem right about the importance of avoiding unstructured intervention programs. However, programs involving parents, youth, and their environment (school, peers, and community) appear to be more beneficial than programs only involving parents (De Vries et al., 2015). Therefore, there are more intervention elements (e.g., targeting "Antisocial recreational activities" and "Procriminal associates") than the two mentioned by Dowden and Andrews (2003) that should be part of an effective family intervention for juvenile delinquency.

The Responsivity Principle

Contrary to the findings of Dowden and Andrews (2003), adhering to the general responsivity principle did not significantly increase the effect of family interventions for delinquent youth although we did find larger mean effect sizes for interventions that adhered to the general or the specific responsivity principle compared to interventions that not adhered to these principles. As for specific responsivity factors, we did find that taking “culture/ethnicity” and “age” into account in treatment delivery positively moderates intervention effectiveness. This supports the idea that culturally adapted family-based intervention programs may lead to better intervention outcomes than interventions that are not culturally sensitive (Kumpfer et al., 2002). The positive moderating effect of age may be explained by the existence of different developmental pathways that can lead to delinquency, and the variability in the age of onset for delinquent behavior (Dahlberg & Potter, 2001). The results support the idea that interventions should fit a juvenile’s specific stage of development to reach the best possible outcome. For example, age specific risk factors for delinquent behavior, such as substance use, poor monitoring/supervision, academic failure, and gang membership, are usually more important to consider during adolescence than at a younger age (Van der Put et al., 2012).

Study Limitations and Strengths

This study had several strengths. Our coding of the RNR principles was more in line with the original definitions of the RNR principles of Bonta and Andrews (2017) compared to previous reviews, as ambiguous coding rules were used in for instance the meta-analysis of Dowden and Andrews (2003). We examined the effectiveness of the aggregate-sample and within-sample approaches to the risk principle, as well as the effectiveness of targeting specific criminogenic need factors. We also examined how adhering to the general and specific responsivity principles are associated with intervention effectiveness. Further, studies published after 2001 were also included in the current meta-analysis. Last, we applied a three-level approach to meta-analysis meaning that - contrary to traditional meta-analytic techniques used by Dowden and Andrews (2003) - all information reported in primary studies could be retained, and maximum statistical power could be achieved in the analyses.

However, some limitations must be acknowledged. First, not all primary studies synthesized by Dowden and Andrews (2003) could be included in the current review

for different reasons (see Method section). Second, the results of the trim-and-fill analysis showed that publication bias may have been present in the data, as the distribution of effect sizes was asymmetrical. Third, descriptions of program delivery were limited in most studies, and therefore, intervention manuals or factsheets (when available) had to be retrieved to obtain more information on how program delivery. However, information on treatment fidelity (i.e., accuracy and consistency of intervention delivery) was not reported in most of the included studies meaning that it was uncertain whether the risk, need, and responsivity principles were properly applied in intervention delivery. Adequate assessment of the degree to which a treatment is implemented as intended is still quite uncommon in intervention studies (e.g., Goense et al., 2016). In addition, many studies fail to report essential information on characteristics of the delivered intervention, such as specific intervention practices and delivery techniques (Van der Put et al., 2018). If our study inclusion was restricted to studies that thoroughly assessed program fidelity, the number of studies eligible for inclusion would be too small for testing the moderating effect of the RNR principles. From this perspective, it must be stressed that a thorough assessment of program fidelity in future primary research is vital. Doing so not only improves the knowledge on how specific interventions are delivered and how delivery is associated with intervention effectiveness, but also paves the way for better meta-analytic research in which implemented program elements instead of intended or described program elements can be examined.

Last, none of the included studies used a validated instrument for risk assessment, and only one of the included studies used a structured instrument in assessing the criminogenic needs of individual offenders. This means that despite our intentions, we were unable to thoroughly examine the effectiveness of the risk and need principles according to the original definitions of the risk and need principles as described by Bonta and Andrews (2017); i.e., matching intervention intensity and content to the risk level and need factors of individual offenders. This limitation is probably an important explanation for not finding significant moderating effects of the RNR principles. This finding highlights an important shortcoming in the delivery of treatment in clinical practice and intervention research. We urge future researchers to use validated instruments for risk, need, and responsivity assessment so that the effects of adhering to the RNR principles on program effectiveness can better be assessed. A prerequisite for such studies is the availability of reliable and valid assessment

instruments. Currently, instruments for need assessment and responsivity assessment in particular have not been widely developed and validated. Future research on the development and validation of such instruments is thus also needed.

Conclusion

We did find larger effect size estimates for interventions adhering to any of the RNR principles, but convincing and strong empirical support for the effectiveness of the RNR principles was not obtained. The current results only suggest that adhering to the RNR principles (Andrews & Bonta, 1990) in family interventions is effective, and reveal that strong and convincing empirical support for the principles is not yet available. Suggestions for future primary research are offered to obtain this support. The absence of this support is likely to be driven by limitations in primary research on intervention effectiveness and adherence to the RNR principles. Therefore, we urge future primary researchers to explicitly and thoroughly describe whether and how RNR principles were implemented in studies on treatment effectiveness. Further, it is important that valid and reliable instruments for risk, needs, and responsivity assessment are used so that treatment can be tailored to needs and circumstances of individual offenders. Finally, much more effort should be devoted to adequately assess and report program fidelity, so that study results become driven by actually implemented program elements instead of intended or described program elements. These suggestions for primary research will in turn improve the quality of secondary research that is needed to establish strong empirical evidence for the widely acknowledged value of the RNR model.