



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The effectiveness of youth crime prevention

de Vries, L.A.

Publication date

2016

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

de Vries, L. A. (2016). *The effectiveness of youth crime prevention*. [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.

3.

Effective Ingredients of Prevention Programs for Youth at Risk of Persistent Juvenile Delinquency²

2 Revised version of: De Vries L. A., Hoeve M., Assink, M., Stams, G. J. J. M., & Asscher, J. J. (2015). Practitioner Review: Effective Ingredients of Prevention Programs for Youth at Risk of Persistent Juvenile Delinquency: Recommendations for Clinical Practice. *Journal of Child Psychology and Psychiatry*, 52, 108-121.

Abstract

There is a lack of knowledge about specific effective ingredients of prevention programs for youth at risk for persistent delinquent behavior. The present study combines findings of previous studies by examining the effectiveness of programs in preventing persistent juvenile delinquency and by studying which particular program, sample and study characteristics contribute to the effects. Information on effective ingredients offers specific indications of how programs may be improved in clinical practice. A literature search in PsychINFO, ERIC, PubMed, Sociological Abstracts, Criminal Justice Abstracts, and Google Scholar was performed. Only (quasi)experimental studies and studies that focused on youth at risk for (persistent) delinquent behavior were included. Multilevel meta-analysis was conducted on 39 studies ($N = 9,084$). Participants' ages ranged from 6 to 20 years ($M = 14$ years, $SD = 2.45$). The overall effect size was significant and small in magnitude ($d = .23$). Behavioral-oriented programs, focusing on parenting skills training, behavioral modeling or behavioral contracting yielded the largest effects. Individual, multimodal programs, and programs carried out in the family context proved to be more beneficial than group-based programs. Less intensive programs yielded larger effects. Prevention programs have positive effects on preventing persistent juvenile delinquency. In order to improve program effectiveness, interventions should be behavioral-oriented, delivered in a family or multimodal format, and the intensity of the program should be matched to the level of risk.

3.1 Introduction

Juvenile delinquency is an important societal problem, with negative emotional, physical, and economic consequences for individual victims, local communities and society as a whole. Moreover, juvenile offending is associated with poor health outcomes, and educational, vocational and interpersonal problems in juvenile offenders themselves (Borduin, 1994; Kazdin, 1987). In particular the relatively small group of persistent offenders warrants attention. These youths start committing delinquent acts at an early age, their behavior becomes gradually more disruptive, and offending continues into adulthood (Loeber, Burke, & Pardini, 2009). During early adolescence, these youngsters are exposed to negative peer influences, a starting point for further escalation of problems, they are at high risk for school failure, disengagement from society and involvement in criminal activities in later adolescence and adulthood (Odgers et al., 2008). It is therefore important to establish how juveniles with disruptive behavior problems, who are at risk for becoming a persistent delinquent, can best be prevented from developing a chronic criminal career.

The majority of meta-analytic reviews have focused on a broad range of juvenile offenders, ranging from mild to severe delinquents (e.g., Lipsey 2009; Wilson, Lipsey, & Soydan, 2003), or on severe and chronic juvenile offenders (e.g., James, Stams, Asscher, De Roo, & Van der Laan, 2013; Landenberger & Lipsey, 2005), which limits generalizability to youth at the onset of a criminal career. Therefore, we examine the effectiveness of prevention programs for juveniles at the onset of a criminal trajectory and at risk for persistent offending. These programs usually target youths showing early indications of disruptive behavior problems, who may have committed minor offenses, but who have not yet exhibited a longstanding pattern of severe antisocial and delinquent behavior (Greenwood, 2008; Mulvey, Arthur, & Reppucci, 1993).

A large amount of meta-analyses were directed to one format or type of program targeting antisocial or delinquent behavior (i.e., juvenile offender recidivism), such as cognitive-behavioral therapy (Landenberger & Lipsey, 2005; McCart, Priester, Davies, & Azen, 2006), skills training (Ang & Hughes, 2001; Lösel & Beelmann, 2003), family treatment (Farrington & Welsh, 2003; Latimer, 2001), victim-offender mediation (Bradshaw, Roseborough, & Umbreit, 2006; Nugent, Williams, & Umbreit, 2004), wilderness challenge programs (Wilson & Lipsey, 2000), or specific therapies, such as Multisystemic Therapy (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997). In contrast to these earlier review studies, the present meta-analytic study is focused on identifying specific program components of preventive interventions for delinquent behavior, such as 'conflict resolution' or 'behavioral modeling', that contribute to program effectiveness. Only Landenberger and Lipsey (2005) found that cognitive-behavioral programs including anger control and interpersonal problem solving, but not victim impact or behavior modification components, were related to higher recidivism reductions. However, this meta-analysis focused on severe adult and juvenile offenders.

Although previous research has identified effective programs targeting juvenile delinquency, recidivism in particular, it is still unknown which types or components of

preventive programs are most effective for whom at the onset of a criminal career. Mulvey and colleagues (1993) narratively reviewed the effectiveness of prevention programs for youths with only one or two police contacts, but who had not yet been adjudicated by the juvenile court. Positive effects were found for diversion programs, indicating that well-implemented programs, incorporating behavioral and family-based change strategies, generated reductions in subsequent arrest rates. Other clear evidence of effectiveness was found for behavioral, structural, and multisystemic family therapy. However, these results were based on a narrative review and should therefore be interpreted carefully. Qualitative (narrative) reviews, although informative, lack explicit systematic procedures and detailed analysis of which study characteristics explain differences in study outcomes (Lipsey & Wilson, 2001). The method of quantitative review is especially useful to identify moderator effects, i.e., specific participants and/or program characteristics that may influence the success of an intervention, which are likely to remain invisible in single studies examining effectiveness of preventive programs due to small sample sizes or a lack of variation in these characteristics.

Promising results of family-based programs were confirmed by a recent meta-analysis of Schwalbe and colleagues (2012), revealing that family-based diversion programs resulted in a reduction of recidivism. However, the overall effect of diversion programs on recidivism was non-significant. In contrast, Wilson and Hoge (2012) found that diversion programs were significantly more successful than the traditional justice system, but differences were no longer significant when a successful research design was used (e.g., RCT, or successful matched control design, independency of researchers). Although diversion programs are mainly designed for status and first time offenders diverted from the juvenile justice system, the studies of Schwalbe et al. and Wilson and Hoge also included high risk, chronic or serious offenders. Since juveniles at the onset of their criminal career may have been formally adjudicated by the court, as a result of committing minor offenses, we are also interested in the effectiveness of a broader set of programs for less serious juvenile offenders who have been referred by the juvenile court.

3.2 Review Aim

In sum, previous research has provided information on the effects of curative (judicial) interventions aimed at a broad target group, ranging from mild to severe juvenile offenders. As mentioned in the previous paragraph, most meta-analytic reviews have focused on specific types of programs targeting juvenile offender recidivism. However, to our knowledge, to date there are no meta-analytic reviews that examined to what extent prevention programs in general are effective in preventing juveniles to start or continue a criminal trajectory. Therefore, there is only scant knowledge of which particular program, sample and study characteristics contribute to larger program effects for the target group. For example, it is unknown whether community-based programs are more effective than prevention programs in a court setting, or whether younger juveniles benefit more than older juveniles or young adults. Programs targeting juveniles at risk for delinquency are likely to be more cost-effective

than universal programs that focus on general populations (Greenwood, 1998). Therefore, a systematic review of the effectiveness of prevention programs for youth at the onset of a criminal career is warranted.

The present meta-analysis evaluates *prevention programs* targeting juveniles identified as being at increased risk for a persistent delinquent behavior pattern, allowing an integrated analysis of comparative effectiveness of different programs and approaches (following Lipsey, 2009). The main purpose of this study is to examine the overall effect of prevention programs for persistent juvenile delinquency, and to examine how effectiveness of these programs is influenced by the type and intensity of the program, characteristics of the participants, design of the study, and type of outcome. Identification of effective program ingredients can help improve interventions for the prevention of persistent delinquent behavior in at-risk youths.

3.3 Method

3.3.1 Inclusion Criteria

Studies were selected if they met four main criteria. First, the central outcome measures in this meta-analysis had to be delinquency, criminal offending or recidivism. Studies were included if at least one quantitative outcome measure of delinquency was reported. Studies that focused exclusively on a general category of problem behavior, such as externalizing problems (antisocial or conduct problems) were not included. Delinquency was defined as illegal behavior, prohibited by the law. Recidivism was defined as the second or repeated offense known to the police and court authorities.

Second, studies that involved at-risk youth, with ages 8 to 20 years at the start of the program as treatment and comparison participants, were included. This target group can be described as youths at risk for a persistent delinquent behavior pattern, such as predelinquents with antisocial behavior, first time offenders and delinquents with mainly minor police contacts and offenses (theft, vandalism, menacing). Although rates of delinquency are highest in youths between ages 12-20 years, programs targeting youths from 8 to 12 years were also included, because the present study is focused on prevention programs that could also be designed for school-aged predelinquents. Moreover, it is known that a substantial percentage of these youngsters already come into contact with the police and justice (Snyder, 2001). Studies examining interventions targeting serious, persistent or chronic offenders and incarcerated juveniles (convicted of major offenses, such as violence, murder, forcible rape, armed robbery) were excluded.

Third, we focused on selective and indicated prevention programs that were developed for juveniles at risk for (the progression of) delinquent behavior. The target group of selective prevention consists of juveniles whose risk of developing mental disorders is significantly higher than average. Indicated prevention is focused on high risk juveniles who are identified as having minimal but detectable symptoms of mental disorders (prior to the diagnosis of a disorder). Universal prevention programs, targeting a general population that

has not been identified on the basis of individual risk, were excluded (O'Connell, Boat, & Warner, 2009).

Fourth, in order to maximize research quality, only studies with an experimental (RCT) or quasi-experimental design (in which a treatment condition is compared to a control condition) were included. Nonequivalent comparison designs, in which groups were not randomly assigned to conditions, were included only if a pretest measure of delinquency or antisocial behavior or a variable highly correlated with delinquency (e.g., prior delinquency history) was used. One group pretest-posttest designs were excluded. Finally, studies based on interventions carried out before 1950 were not included.

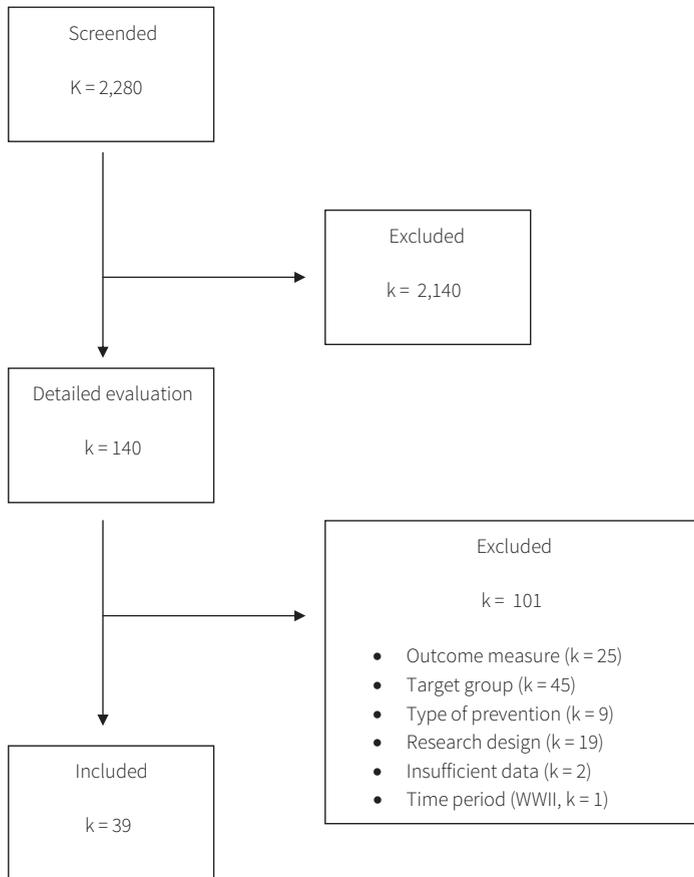
3.3.2 Literature Search Procedures

Electronic databases of PsychINFO, ERIC, PubMed, Sociological Abstracts, Criminal Justice Abstracts, Google Scholar were searched through for articles, books, chapters, dissertations and reports. Until September 2012 studies were collected using keywords regarding research method, program features, study outcomes and respondents in different combinations: (quasi-)experiment, randomized controlled (clinical) trial, program*, intervention*, prevention*, delinquent*, antisocial behavior, crime*, youth at risk, juvenile*, adolescent*, (first time) offender*, and effect*. Next, manual searches of reference sections of articles, reviews and book chapters were conducted. Finally, we contacted authors by email in order to obtain (unpublished material) dissertations, and to receive more information than was provided in the selected articles.

The study selection process is presented in Figure 3.1. Full texts of 140 articles were assessed for eligibility and 101 studies (articles) were excluded because they did not meet the study selection criteria. The final analyses included 39 independent studies (39 samples and 95 effect sizes) written or published between 1973 and 2009.

3.3.3 Coding of Participant, Program, and Study Characteristics

Following the guidelines of Lipsey and Wilson (2001) a coding system was developed. First, with regard to *participant characteristics*, we collected information on mean age at first measurement, gender, ethnicity, SES (based on annual household income, receiving free or reduced school lunch, mean education of parents), level of delinquency, country and degree of urbanization (urban, sub-urban or rural). Second, we retrieved information on the following *program characteristics*: type of referral organization, type of prevention (selective or indicated), setting (home, school, clinic, court, community or ambulant), program format (one-on-one treatment, group, family or mixed/multimodal), type of the program, components of the program, primary target population (juveniles only, parents and juveniles, parents, juveniles and siblings), type of trainer and program drop-out. Regarding time and period of the program, we collected data on total duration in weeks, intensity (number of hours per week and total contact hours), and frequency (number of sessions per week). Finally, we focused on the following *study characteristics*: study design (RCT or

Figure 3.1 Flowchart of Literature Search and Screening

quasi-experimental), method of assignment (random or nonrandom), method of matching, equivalence of groups at pretest, type of control condition, sample size, percentage of drop-out, measurement of program integrity, publication year, and journal impact factor. In order to retrieve specific information on measurement of delinquency we collected information on sources of information (official records or self-reports), length of follow-up period in months, type of delinquency (general, property, violent crime, etc.) and dimension of delinquency and recidivism (participation in delinquency, frequency, seriousness and versatility¹ in offending). The coding form, including a detailed description of the variables, is available on request.

The classification of *type and components of programs* was based on classifications from the Campbell Collaboration, previous research (Bradshaw, Roseborough, & Umbreit, 2006; Farrington & Welsh, 2003; Landenberger & Lipsey, 2005; Latimer, 2001; Lösel & Beelmann, 2003; Nugent, Williams, & Umbreit, 2004; Schwalbe, et al., 2012) and the program descriptions provided in the studies included in the present meta-analysis. This classification resulted in the

following *types of programs*: cognitive skills training, behavioral modification, interpersonal problem solving, social skills training, life skills training, anger management, moral reasoning, mediation and mentoring. In addition, we made a more detailed classification of the following specific *program components*: academic service, employment related service, behavioral modeling, behavioral contracting, victim impact or material and emotional restitution, conflict resolution, community service, parenting skills, communication skills, recreation activities, counseling, rewarding appropriate behavior and self-efficacy.

The coding process started by coding and discussing five randomly selected studies by the two coders (first and third author). Disagreements were resolved through consulting the studies and discussion until consensus was reached. After this process the coding form was refined and all variables of 39 studies were scored by the first and third author. In order to assess inter-rater agreement, 10 studies consisting of 21 analyses were randomly selected and scored by two coders (first and third author). Inter-rater agreement was analyzed by calculating the percentage of agreement for all study characteristics, Kappa for categorical variables and intraclass correlation for continuous variables. The inter-rater reliability for categorical variables proved to be satisfactory, with Kappa's ranging from 0.64 (80% agreement) for program type 'life skills training' to 1.00 (100% agreement) for socio-economic status, program components (modeling, contracting and parenting skills) and primary target population. The inter-rater reliability for continuous variables was very good, with intraclass correlations ranging from 0.99 (90% agreement) for percentage of cultural minority (Hispanic/African American) to 1.00 (100% agreement) for the effect size value, overall mean age of sample, and percentage of males.

3.3.4 Data Analysis

For each study one or more effect sizes were calculated. In order to examine the difference in delinquency scores between the experimental and control group Cohen's d was calculated. Cohen's d was usually calculated on the basis of mean scores and standard deviations or proportions (based on recidivism rates). The reported statistical tests were transformed into Cohen's d with formulas from Lipsey and Wilson (2001) and Mullen (1989). Pretest scores were taken into account by subtracting these scores from the posttest scores of the effect sizes. Each continuous moderator was centered around its mean and dichotomous dummy codes were made for the categorical variables.

Independence of study results is essential when conducting a meta-analysis to prevent that a particular study is weighted more strongly than other studies (Lipsey & Wilson, 2001; Mullen, 1989; Rosenthal, 1991). Following scholars reporting on recent meta-analyses (e.g., Assink et al., 2015; Weisz et al., 2013), in order to take into account dependency of study results, we used a multilevel random effects model for the calculation of combined effect sizes and for conducting moderator analyses (Hox, 2002; Van den Noortgate & Onghena, 2003). In a multilevel meta-analysis all data and effects sizes can be included, which increases the statistical power.

A three-level structure to our meta-analytic models was applied, modeling three types of variance: (1) sampling variance of observed effect sizes on the first level; (2) variance of effect sizes within studies on the second level; and (3) variance between studies on the third level (See also Cheung, 2014; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). The models were extended by including moderator variables to examine whether the variation can be explained by characteristics of studies or effect sizes.

For conducting multilevel analysis we used the user-written function “*rma.mv*” of the metafor package (Viechtbauer, 2010) in the statistical program R (version 3.2.0; R Core Team, 2015). The method of Knapp and Hartung (2003) was applied in order to test individual regression coefficients of the meta-analytic models, meaning that test statistics were based on a *t*-distribution. In models with categorical moderators containing three or more categories, the omnibus test of the null hypothesis that all group mean effect sizes are equal was based on an *F*-distribution. Two separate log-likelihood-ratio-tests were conducted in which the deviance of the full model was compared to the deviance of a model excluding one of the variance parameters. This method was used to determine whether the variance between effect sizes from the same study (Level 2), and the variance between studies (Level 3) were significant (see Assink et al., 2015; Wibbelink & Assink, 2015). The formula of Cheung (2014) was used to assess the sampling variance of observed effect sizes (Level 1). The assessment of parameters was based on the restricted maximum likelihood estimation method. Finally, *p*-values $\leq .05$ (two-tailed) were considered as statistically significant, and *p*-values $< .10$ (two-tailed) were reported as trends.

3.3.5 File Drawer Problem

Publication bias forms a common problem when conducting a meta-analysis. Studies with non-significant results are less likely to be published than those with strong significant results. This tendency, referred to as the *file drawer problem*, may have implications for the final conclusions of the meta-analysis (Rosenthal, 1991).

To investigate whether studies included in the present meta-analysis form a random sample of all studies conducted on the subject, we applied two conventional methods. First, we calculated the fail-safe number, which is the minimum number of additional studies with non-significant results needed to reduce significant meta-analytic results to non-significance. (Durlak & Lipsey, 1991; Rosenthal, 1995). Results of the meta-analysis are considered to be robust if the fail-safe number exceeds the critical value obtained with Rosenthal's (1995) formula of $5 * k + 10$. The number of effect sizes is represented by *k*. Second, we inspected the distribution of each individual study's effect size on the horizontal axis against its sample size, standard error on the vertical axis. If no publication bias is present, the distribution of effect sizes should be shaped as a funnel (Sutton, Duval, Tweedie, Abrams, & Jones, 2000). In the present study, funnel plot asymmetry was tested by regressing the standard normal deviate, defined as the effect size divided by its standard error, against the estimate's precision, which largely depends on sample size (Egger, Smith, Schneider, & Minder, 1997).

3.4 Results

The present meta-analysis included 39 studies, providing data on 9,084 participants ($N = 4,755$ treatment group and $N = 4,329$ control group). Sample sizes ranged from 32 (Augimeri, Farrington, Koegl, & Day, 2007) to 782 participants (McGarrell, & Hipple, 2007), with an average of 229 participants per study. The mean age of the participants was 14.18 ($SD = 2.45$, age range: 6 - 20 years²). An overview of all studies included in the meta-analysis can be found in Appendix 3.A.

The overall mean effect size for the effects of prevention programs was $d = .23$ ($k = 95$ effect sizes), which indicated a small overall mean effect, based on the criteria for interpretation of effect sizes formulated by Cohen (1988)³. The overall mean effect size of .23 corresponds to a significant reduction of 13% in delinquency compared to care as usual or no treatment (based on the success rate difference, SRD, Kraemer & Kupfer, 2006). The fail-safe number, 4,332 ($p < .05$, $k = 95$), exceeded Rosenthal's (1995) critical value ($5 * k + 10 = 485$), which indicated no evidence of publication bias. This outcome was confirmed by testing of funnel plot asymmetry. There was no indication of funnel plot asymmetry, as the intercept did not significantly deviate from zero ($t = .864$, $p = .393$).

The results of the likelihood-ratio tests showed that there was significant variance between effect sizes from the same study (i.e., level 2 variance) and that there was significant variance between studies (i.e., level 3 variance), indicating that the variation across studies might be caused by study, program or participant characteristics (see Tables 1-3). In order to detect if differences between effect sizes have another source than subject-level sampling error, we conducted moderator analyses.

3.4.1 Participant and Program Characteristics

Table 3.1 presents the results concerning the significant categorical moderators (participant and program characteristics). No significant effects of age, sex and ethnicity were found (a table with all moderators, including non-significant results, is available on request). Several program characteristics affected program effectiveness. First, specific program components accounted for a significant proportion of the variation in effect sizes. *Parenting Skills* ($d = .60$, $p = .003$) and *Behavioral Contracting* ($d = .57$, $p = .047$) were significantly associated with better program outcomes, indicating that programs containing these specific components yielded larger effect sizes. Positive trend effects were found for components of *Behavioral Modeling* ($d = .53$, $p = .061$) and *Recreation Activities* ($d = .49$, $p = .073$). The effectiveness of prevention programs was not related to program type.

Further, the composition of the target group was significantly associated with effect size. Programs involving mixed target populations (juveniles, parents and siblings) showed larger effect sizes ($d = .74$, $p < .001$) than programs that targeted only juveniles or juveniles and parents. In addition, the specific setting accounted for significant differences in effect sizes. Programs carried out in court settings ($d = -.31$, $p = .035$) yielded smaller effect sizes than programs carried out in the direct environment of juveniles (home, school, community

and ambulant setting). With regard to the format of the program, programs carried out in a family format ($d = .65, p = .002$) and multimodal format ($d = .37, p = .002$) yielded larger effect sizes than individual ($d = .28, p = .045$) and group-based programs ($d = -.02$). Also, a negative trend effect was found regarding type of trainer, indicating that treatment carried out by peers showed somewhat smaller effect sizes ($d = -.38, p = .091$).

Table 3.2 presents an overview of the continuous moderator variables. First, the intensity of the program (in hours per week) was significant ($d = .16, p = .040$), which indicates that less intensive programs were associated with larger effect sizes. Total contact hours (duration \times intensity) was also significant ($d = .13, p = .043$), indicating that a smaller amount of contact hours of programs was associated with larger effect sizes. The moderator effect of number of sessions was marginally significant ($d = .02, p = .052$), which implies that fewer sessions were related to larger effect sizes.

3.4.2 Study Characteristics

Concerning study moderators, type of matching accounted for a significant proportion of the variation in effect size (see Table 3.3). We found that one study, applying the method of matching on demographics, yielded a larger effect size ($d = 1.65, p = .006$) than studies using other matching methods. A trend was found for dimension of delinquency, indicating that studies measuring seriousness of delinquency ($d = .01, p = .063$) showed smaller effect sizes than studies that measured participation ($d = .28$) and/or frequency ($d = .16$) of delinquent acts. Finally, we found a significant moderating effect of parent reports ($d = .89, p = .046$), indicating that studies using parent reports showed larger effect sizes than studies using self-, teacher- and official reports (resp. $d = .21, d = -.18$, and $d = .23$).

3.4.3 Unique Contribution of Program Characteristics

Several multivariate analyses were conducted to examine the unique contribution of program characteristics to the variance in effect sizes. Because of missing data, we were not able to test all significant moderators simultaneously. First, we tested the combined contribution of significant program characteristics to effect size: setting, format, components and target group of the program. We found a significant effect of components (parenting skills, $p = .009$) and format (one-on-one, $p = .038$; family-based, $p = .096$; see Appendix 3.B). Next, we examined the unique contribution of the other significant moderators, that is, method of matching, intensity of the program, dimension of delinquency, and type of measure, over and above components and program format. We found a significant effect of an indicator of program intensity (number of program sessions, $p = .037$; see Appendix 3.C), dimension of delinquency (seriousness vs. participation, $p = .038$; see Appendix 3.D), and a trend for type of measure (teacher reports vs. official record, $p = .062$), adjusting for components and program format (results of all bivariate and multivariate models are available on request).

Table 3.1 Results for the Overall Mean Effect Size and Categorical Moderators: Participant and Program Characteristics

| Moderator variables | Studies | ES | Mean d (SE) | CI | β (SE) | CI | p | Level-2 variance | Level-3 variance | Omnibus test ^a |
|---|---------|----|----------------|---------------|----------------|---------------|-------|------------------|------------------|---------------------------|
| Overall | 39 | 95 | 0.232 (0.061) | 0.112; 0.353 | | | | 0.044*** | 0.098*** | |
| Program components | | | | | | | | | | |
| Behavioral modeling | | | | | | | | | | |
| No (RC) | 32 | 81 | 0.196 (0.065) | 0.067; 0.326 | | | | 0.043*** | 0.095*** | F(1,92) = 3.611+ |
| Yes | 6 | 13 | 0.532 (0.164) | 0.206; 0.858 | 0.336 (0.177) | -0.015; 0.686 | .061 | | | |
| Parenting skills | | | | | | | | | | |
| No (RC) | 31 | 74 | 0.155 (0.062) | 0.032; 0.279 | | | | 0.043*** | 0.076*** | F(1,92) = 9.526** |
| Yes | 7 | 20 | 0.602 (0.131) | 0.342; 0.861 | 0.447 (0.145) | 0.159; 0.734 | .003 | | | |
| Behavioral contracting | | | | | | | | | | |
| No (RC) | 33 | 84 | 0.198 (0.063) | 0.072; 0.324 | | | | 0.043*** | 0.090*** | F(1,92) = 4.056* |
| Yes | 5 | 10 | 0.573 (0.175) | 0.225; 0.921 | 0.375 (0.186) | 0.005; 0.745 | .047 | | | |
| Recreation activities | | | | | | | | | | |
| No (RC) | 32 | 80 | 0.196 (0.064) | 0.069; 0.322 | | | | 0.045*** | 0.085*** | F(1,92) = 3.284+ |
| Yes | 6 | 14 | 0.494 (0.152) | 0.192; 0.795 | 0.298 (0.165) | -0.029; 0.625 | .073 | | | |
| Type of trainer | | | | | | | | | | |
| Peers | | | | | | | | | | |
| No (RC) | 31 | 73 | 0.240 (0.063) | 0.114; 0.367 | | | | 0.041*** | 0.082*** | F(1,73) = 2.926+ |
| Yes | 1 | 2 | -0.375 (0.354) | -1.081; 0.331 | -0.616 (0.360) | -1.333; 0.102 | .091 | | | |
| Primary target population | | | | | | | | | | |
| Juveniles (RC) | 31 | 78 | 0.142 (0.058) | 0.027; 0.258 | | | | 0.047*** | 0.064*** | F(2,92) = 7.033** |
| Mixed (juveniles and parents) | 2 | 6 | 0.300 (0.238) | -0.173; 0.774 | 0.158 (0.245) | -0.329; 0.645 | .521 | | | |
| Mixed (juveniles, parents and siblings) | 6 | 11 | 0.742 (0.150) | 0.445; 1.039 | 0.599 (0.160) | 0.281; 0.918 | <.001 | | | |
| Program format | | | | | | | | | | |
| Group (RC) | 13 | 40 | -0.021 (0.084) | -0.188; 0.147 | | | | 0.046*** | 0.060** | F(3,88) = 5.405** |
| One-on-one | 8 | 13 | 0.282 (0.123) | 0.038; 0.526 | 0.302 (0.149) | 0.007; 0.598 | .045 | | | |

Table 3.1 (continued)

| Moderator variables | Studies | ES | Mean d (SE) | CI | β (SE) | CI | p | Level-2 variance | Level-3 variance | Omnibus test ^a |
|-----------------------|---------|----|----------------|---------------|----------------|----------------|------|------------------|------------------|---------------------------|
| Family | 4 | 4 | 0.650 (0.189) | 0.275; 1.025 | 0.670 (0.207) | 0.260; 1.081 | .002 | | | |
| Multimodal | 12 | 35 | 0.367 (0.091) | 0.185; 0.549 | 0.388 (0.124) | 0.141; 0.635 | .002 | | | |
| Setting | | | | | | | | 0.029*** | 0.155*** | F(7,47) = 1.177 |
| Ambulant setting (RC) | 4 | 11 | 0.498 (0.221) | 0.053; 0.943 | | | | | | |
| Home | 2 | 2 | 0.566 (0.341) | -0.119; 1.251 | 0.068 (0.406) | -0.749; 0.885 | .868 | | | |
| Public school | 4 | 7 | 0.305 (0.229) | -0.156; 0.766 | -0.193 (0.318) | -0.834; 0.447 | .547 | | | |
| Specialized school | 1 | 2 | -0.236 (0.440) | -1.121; 0.648 | -0.735 (0.492) | -1.724; 0.255 | .142 | | | |
| Clinic | 1 | 5 | -0.182 (0.402) | -0.990; 0.627 | -0.680 (0.459) | -1.602; 0.243 | .145 | | | |
| Court | 2 | 5 | -0.314 (0.302) | -0.922; 0.294 | -0.812 (0.374) | -1.565; -0.059 | .035 | | | |
| Community-based | 4 | 7 | 0.200 (0.225) | -0.252; 0.652 | -0.298 (0.315) | -0.932; 0.336 | .349 | | | |
| Other | 5 | 16 | 0.092 (0.199) | -0.309; 0.493 | -0.406 (0.298) | -1.005; 0.192 | .179 | | | |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean d = mean effect size (d); CI = Confidence interval; β = Estimated regression coefficient; Level 2 variance = variance between effect sizes from the same study; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.

* $p < .10$; ** $p < .05$; *** $p < .001$

Table 3.2 Results for the Continuous Moderators: Program Characteristics

| Moderator variables | Studies | ES | Mean d (SE) | CI | β (SE) | CI | p | Level-2 variance | Level-3 variance | Omnibus test ^a |
|---------------------|---------|----|---------------|---------------|----------------|----------------|------|------------------|------------------|---------------------------|
| Duration in weeks | 23 | 64 | 0.230 (0.090) | 0.050; 0.411 | 0.001 (0.002) | -0.004; 0.006 | .689 | 0.026*** | 0.151*** | F(1,62) = 0.162 |
| Intensity | 12 | 27 | 0.158 (0.121) | -0.092; 0.407 | -0.015 (0.007) | -0.030; -0.001 | .040 | 0.053*** | 0.117* | F(1,25) = 4.675* |
| Total contact hours | 14 | 31 | 0.134 (0.100) | -0.071; 0.340 | -0.001 (0.000) | -0.002; -0.000 | .043 | 0.113*** | 0.058 | F(1,29) = 4.461* |
| Number of sessions | 12 | 28 | 0.019 (0.088) | -0.162; 0.200 | -0.077 (0.038) | -0.155; 0.001 | .052 | 0.106*** | 0.028 | F(1,26) = 4.136+ |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean d = mean effect size (d); CI = Confidence interval; β = Estimated regression coefficient; Level 2 variance = variance between effect sizes from the same study; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.

* $p < .10$; ** $p < .05$; *** $p < .001$

Table 3.3 Results for the Overall Mean Effect Size and Categorical Moderators: Study Characteristics

| Moderator variables | Studies | ES | Mean d (SE) | CI | β (SE) | CI | p | Level-2 variance | Level-3 variance | Omnibus test ^a |
|-------------------------------------|---------|----|----------------|---------------|----------------|---------------|------|------------------|------------------|---------------------------|
| Type of matching | | | | | | | | | | |
| Not applicable (RC) | 24 | 52 | 0.276 (0.076) | 0.125; 0.428 | | | | | | |
| Matched on pre-test measure | 1 | 2 | 0.065 (0.358) | -0.647; 0.778 | -0.211 (0.366) | -0.940; 0.517 | .566 | 0.044*** | 0.093*** | F(6,84) = 1.955+ |
| Matched on personal characteristics | 3 | 13 | 0.178 (0.204) | -0.228; 0.585 | -0.098 (0.218) | -0.532; 0.336 | .654 | | | |
| Matched on demographics | 1 | 1 | 1.653 (0.486) | 0.687; 2.618 | 1.377 (0.491) | 0.399; 2.354 | .006 | | | |
| Matched on two of above | 6 | 15 | 0.104 (0.152) | -0.197; 0.405 | -0.172 (0.170) | -0.510; 0.165 | .312 | | | |
| Matched on three of above | 1 | 5 | -0.182 (0.320) | -0.819; 0.456 | -0.458 (0.329) | -1.113; 0.197 | .168 | | | |
| Equated groupwise | 2 | 3 | 0.126 (0.263) | -0.396; 0.648 | -0.150 (0.274) | -0.694; 0.394 | .584 | | | |
| Delinquency type | | | | | | | | | | |
| General (RC) | 39 | 82 | 0.229 (0.061) | 0.107; 0.351 | | | | 0.047*** | 0.097*** | F(5,89) = 0.685 |
| Violence | 2 | 4 | 0.395 (0.158) | 0.081; 0.710 | 0.166 (0.151) | -0.135; 0.467 | .275 | | | |
| Property crime | 3 | 6 | 0.298 (0.144) | 0.012; 0.584 | 0.069 (0.137) | -0.203; 0.341 | .617 | | | |
| Vandalism | 1 | 1 | 0.383 (0.355) | -0.323; 1.089 | 0.154 (0.353) | -0.548; 0.856 | .664 | | | |
| Drug-related crime | 1 | 1 | 0.012 (0.272) | -0.528; 0.553 | -0.217 (0.269) | -0.751; 0.318 | .423 | | | |
| Other crime | 1 | 1 | -0.005 (0.272) | -0.545; 0.535 | -0.234 (0.269) | -0.769; 0.300 | .386 | | | |
| Delinquency dimension | | | | | | | | | | |
| Participation (RC) | 34 | 64 | 0.283 (0.065) | 0.153; 0.412 | | | | 0.039*** | 0.101*** | F(3,91) = 1.792 |
| Frequency | 10 | 18 | 0.160 (0.101) | -0.040; 0.360 | -0.123 (0.099) | -0.319; 0.073 | .216 | | | |
| Seriousness | 6 | 12 | 0.008 (0.144) | -0.279; 0.295 | -0.274 (0.146) | -0.564; 0.015 | .063 | | | |
| Versatility (number of crime types) | 1 | 1 | -0.010 (0.290) | -0.586; 0.566 | -0.293 (0.287) | -0.863; 0.278 | .311 | | | |
| Type of measure | | | | | | | | | | |
| Official record (RC) | 30 | 74 | 0.230 (0.062) | 0.106; 0.355 | | | | 0.039*** | 0.101*** | F(3,91) = 1.792 |
| Self-report | 7 | 13 | 0.212 (0.134) | -0.053; 0.478 | -0.018 (0.141) | -0.299; 0.263 | .898 | | | |
| Parent report | 1 | 4 | 0.890 (0.323) | -0.248; 1.532 | 0.660 (0.326) | 0.013; 1.307 | .046 | | | |
| Teacher report | 1 | 4 | -0.179 (0.321) | -0.817; 0.459 | -0.409 (0.327) | -1.059; 0.241 | .214 | | | |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean d = mean effect size (d); CI = Confidence interval; β = Estimated regression coefficient
^a Omnibus test of all regression coefficients in the model. + $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

3.5 Discussion

The main purpose of this meta-analysis was to examine the contribution of participant, program and study characteristics to the effectiveness of prevention programs for persistent juvenile delinquency. We found that these programs in general are effective in preventing persistent juvenile criminal behavior. The overall mean effect size ($d = .23$) was significant, but small in magnitude, which corresponds approximately to a 13% reduction in delinquent behavior compared to care as usual or no treatment (Kraemer & Kupfer, 2006). These results suggest that the prevalence of offending could be reduced by about 13% by implementing such programs, irrespective of the base rate of (re)offending, which was estimated to be 50% in a recent meta-analysis by Koehler and colleagues (2013). A 13% reduction in offending against a baseline of 50% would imply an offending rate of 37% in juveniles attending effective prevention programs for persistent juvenile delinquency. However, behavioral-oriented programs, focusing on learning positive behavior through role models, preparing behavior contracts, improving parenting skills, or family-based programs yielded a medium and significant reduction in offending of approximately 30% compared to treatment as usual or no treatment, which amounts to a favorable offending rate of only 20%.

Effect sizes of the present study were somewhat larger than those found in meta-analyses of curative programs (Lipsey, 2009) and aftercare programs following detention of juvenile offenders (James et al., 2013). These studies included programs that were aimed at more severe juvenile offenders, whereas our study was focused on prevention programs targeting juveniles at the onset of their criminal career. Apparently, prevention seems more effective than cure.

3.5.1 Participant Characteristics

Our findings suggest that prevention programs are equally effective for boys and girls, younger and older juveniles, and juveniles from different cultural backgrounds. The finding that boys and girls equally benefit from preventive programs is in line with an earlier review of gender differences in effectiveness of curative interventions for juvenile delinquents (Zahn, Day, Mihalic, & Tichavsky, 2009). Given that we did not find an effect of age on study outcomes, it can be concluded that preventive programs are effective for juveniles with an onset of delinquent behavior from childhood to late adolescence. Although it has been suggested that 'juvenile-onset' juveniles desist from antisocial behavior during early adulthood (e.g., Moffitt, 1993), they have also been documented to continue engaging in criminal behavior beyond adolescence (Fairchild, Van Goozen, Calder, & Goodyer, 2013; Odgers et al., 2007; Wiesner, Kim, & Capaldi, 2005). Finally, our study showed that different ethnic groups respond relatively similar to prevention programs. This is consistent with a meta-analysis of Wilson and colleagues (2003), confirming that mainstream programs for juvenile delinquents were equally effective for minority and white youth.

3.5.2 Program Characteristics

Examining core elements of programs, we found that programs containing behavioral modeling, contracting, or parenting skills yielded larger reductions in delinquency. Studies that focused on these program elements revealed medium effects. These three program components are mainly based on the cognitive social learning theory (SLT) of Bandura (Bandura & Walters, 1963), and are characterized by a behavioral orientation. The positive impact of these components is consistent with findings of Lösel and Beelmann (2003), Lipsey (2009; 2012) and Andrews and colleagues (1990b), indicating that skill building approaches containing a behavioral orientation are most effective. Moreover, earlier studies indicated that multi-faceted programs, including multiple components for parents, youths and their environment (school and community) appear to be more beneficial than narrowly focused programs (McCord, Widom, & Crowell, 2001).

Our study showed relatively large effects for programs with a family and multimodal format (individual, family- and group-based), adjusting for the effects of program components and various other moderators. Although involving the family system seems effective in both preventive and curative interventions (Litschge, Vaughn, & McCrea, 2010), James and colleagues (2013) showed that individual after care programs for severe juvenile offenders were more successful than those focusing on the social (family) system. In accordance with James and colleagues (2013) and earlier studies (Ang & Hughes, 2001; Arnold & Hughes, 1999; Dishion, McCord, & Poulin, 1999; Dishion & Dodge, 2006), we found that individual, family-based and multimodal programs showed larger effects than group-based programs, which proved to be ineffective ($d = -.02$). Group-based programs may include antisocial peers who are negative role models reinforcing one another's delinquent behavior. The ineffectiveness of peer-group programs is confirmed by longitudinal research revealing that "deviancy training" within juvenile friendships predicts increases in delinquency (Dishion, McCord, & Poulin, 1999).

The intensity of the program was related to program effectiveness. The effectiveness of programs reduced when the number of program sessions was relatively high, indicating that highly intensive programs could be counterproductive for less serious offenders, even when adjusting for the influence of other moderators. The finding that less intensive treatment can be effective is consistent with previous research. For example, a meta-analysis on wilderness challenge programs for delinquent youths showed that extended programs (duration over 10 weeks) were related to smaller effects (Wilson & Lipsey, 2000). According to the risk principle of effective judicial interventions, the intensity of an intervention must be adjusted to the juvenile's risk for recidivism (Andrews, Bonta, & Hoge, 1990a; Andrews & Dowden, 2007). This dose-response principle is confirmed in meta-analyses by, among others, Lipsey (2009) and Koehler et al. (2013). For example, diversion programs providing the minimum amount of services proved to be most effective for low-risk youth (Wilson & Hoge, 2012). Notably, the less is more principle has also precedents elsewhere in child psychopathology, for example,

in the domain of (preventive) attachment-based intervention (Bakermans-Kranenburg, Van IJzendoorn, & Juffer, 2003).

3.5.3 Study Characteristics

No differences in magnitude of the effect sizes were found between RCT and quasi-experimental designs. This finding contradicts results from previous reviews indicating that experimental research designs are associated with smaller effects (Latimer, 2001; Lipsey, 2003; Weisburd, Lum, & Petrosino, 2001). However, most included quasi-experimental studies matched groups on different variables prior to assignment of the condition, tested equivalence of groups at pre-test and significant differences between groups were taken into account in the analysis. Moreover, there was no significant difference in sample sizes and drop-out rates between quasi-experimental and experimental studies.

Concerning measurement of delinquency and recidivism, studies that measured participation in and frequency of criminal acts showed larger effect sizes than studies measuring seriousness of criminal behavior. This suggests that reductions in delinquency not necessarily coincide with reductions in seriousness of criminal acts.

3.5.4 Study Limitations

Several limitations of this meta-analysis must be kept in mind. First, an important limitation is that the reported information of the studies included in the meta-analysis was limited. A relatively large amount of studies failed to report important information on program characteristics, such as precise duration and intensity of the program as well as format and setting of the program. Also, it was not possible to examine the specific role of program integrity, as most studies did not report whether the program was adequately implemented (only 6 of 39 studies measured program integrity). Program integrity is an important factor influencing program outcomes (Lipsey, 2009). However, the assessment of program integrity in outcome studies of interventions targeting conduct problems is rare. Likewise, only a few studies use valid and reliable instruments to measure program integrity (see Goense, Boendermaker, Van Yperen, Stams, & Van Laar, 2014). Another limitation is that data of several program descriptors were based on a limited number of studies and effects sizes.

Second, rates of psychopathology are high among juvenile delinquents (e.g., Wasserman, McReynolds, Schwalbe, Keating, & Jones, 2010). Further, psychopathology has been found to be associated with offending (Copeland, Miller-Johnson, Keeler, Angold, & Costello, 2007) and recidivism (Hoeve, McReynolds, McMillan, & Wasserman, 2013). For example, youths in detention (pre-trial) and secure post adjudication facilities report high rates of mental health disorders: 60-65% have one or another disorder (Wasserman, McReynolds, Schwalbe, Keating, & Jones, 2010). Even of those who enter the juvenile justice at system probation or family court intake (pre-trial), 35% have a psychiatric disorder, compared to about 15% in the community. Despite these findings, most studies in this meta-analysis did not report prevalence of mental disorders, and we were therefore not able to test

potential moderating effects of psychopathology. In present meta-analysis only three studies reported specific rates of mental disorders in their samples. Brier (1994) reported that 87% of the experimental group met diagnostic criteria for a learning disability. All participants in the study of Keating and colleagues (2002) were rated in the clinical range of externalizing and internalizing behavior (based on the CBCL parent and teacher ratings). Finally, Vitaro and Tremblay (1994) reported that 73% of the sample scored above 70th percentile on aggressive behavior (measured by the Preschool Behavior Questionnaire).

Although we searched for published and unpublished studies, the present meta-analysis was exclusively based on published studies because unpublished studies did not meet our selection criteria. Although excluding unpublished studies might increase the risk for publication bias, analyses showed that publication bias was unlikely. Finally, it should be kept in mind that the present study was mainly based on Western countries, particularly the USA. Since countries differ in social and political climate, organization of mental health services, ethnic background of clients, etc., it is questionable whether the present results are also representative for nonwestern countries (Deković et al., 2011).

3.5.5 Implications for Policy and Clinical Practice

The present study provided support for the notion that prevention of persistent juvenile delinquency is recommended. Our study shows that prevention programs can be effective in preventing youths from developing a persistent course of criminal behavior and as a result, these programs may prevent a substantial amount of individuals from becoming a future victim of crime. Additionally, the present study provides some important implications for clinical practice.

When implementing best practices, clinical professionals and policy makers should opt for programs that produce the largest effects on preventing delinquency. Regarding the specific approach of crime prevention, it is advised to implement behavioral-oriented programs. Programs should integrate elements of *behavioral contracting*, *modeling* and *parenting skills training*, given that we found the strongest effects of programs with these components. These components are theoretically grounded in the cognitive social learning theory (SLT) of Bandura (Bandura & Walters, 1963). SLT provides clear principles and techniques for practitioners. According to SLT, new patterns of behavior are learned through direct experience or by observing behavior of others. *Modeling* can be perceived as a core technique of SLT: juveniles learn appropriate behavior through observing competent models who demonstrate how the required activities should be performed. In turn, positive behavior is reinforced by *behavior contracts* consisting of valued rewards, which enhance the learning process (Bandura, 1971). The SLT principles offer explicit tools for directly changing inadequate parenting behavior (Scott & Dadds, 2009). Parenting behavioral skill techniques, such as contingency management, are applied in the evidence-based intervention of parent management training (PMT) targeting juveniles with disruptive behavior problems (Michelson, Davenport, Dretzke, Barlow, & Day, 2013).

Given that we found no effects of group-based programs, one should opt for prevention programs that are delivered in a family context or multimodal format. Family interventions focus on altering the interactions among family members and improving the functioning of the family as a unit. Multimodal programs focus on a variety of criminogenic needs instead of a single risk factor. In order to address multiple risk factors, these programs include multiple treatment modalities or distinct intervention elements, such as cognitive behavioral therapy and parenting skills training (Wilson & Lipsey, 2007). Multimodal programs that target multiple needs of delinquent juveniles have been proven effective (Lipsey, 1992; 1995; Lipsey & Wilson, 1998). Finally, the number of sessions in prevention programs for juveniles with low delinquency levels should be kept low (the number of sessions per week in the studied programs ranged from less than one to seven times a week).

Footnotes

- ¹ Number of different crime types measured.
- ² The study is focused on youngsters from 8 to 20 years at the start of the intervention. The time of first measurement in one of the studies was before the start of the intervention (M age at first measurement was 6 years).
- ³ Effect sizes are categorized as small ($d = .20$), medium ($d = .50$), and large ($d = .80$) group differences, whereas the effect sizes of $d = .00$ would indicate that there was no difference between experimental and control groups.

Appendix 3.A Studies Included in the Meta-Analysis

| Study | Year | Main intervention | Treatment format | n EC | n CG | Mean age (in years) | Study design | Condition of CG |
|---------------------|------|------------------------------------|------------------|------|------|---------------------|--------------|-----------------------|
| Alexander & Parsons | 1973 | Short-term behavioral intervention | Family | 46 | 40 | unknown | RCT | CAU |
| Arbuthnot & Gordon | 1986 | Moral reasoning | Multimodal | 24 | 24 | 14.5 | RCT | No treatment |
| Augimeri et al. | 2007 | Community-based program | Multimodal | 16 | 16 | 8.9 | RCT | CAU |
| Baffour | 2006 | Family Group conferencing | Multimodal | 80 | 212 | 13.6 | Quasi | CAU |
| Bergseth & Bouffard | 2007 | Restorative justice program | Multimodal | 164 | 166 | 14.7 | Quasi | No treatment |
| Berry et al. | 2009 | Coaching for communities program | One-on-one | 32 | 31 | unknown | RCT | CAU |
| Blechman et al. | 2000 | Mentoring & skill training | unknown | 55 | 137 | 15.3 | Quasi | CAU |
| Bottcher & Ezell | 2005 | Boot camps | Group | 344 | 277 | 17.5 | RCT | CAU |
| Brier | 1994 | Diversion program | Group | 158 | 34 | 17.7 | Quasi | Minimal contact |
| Bry | 1982 | School-based program | Group | 30 | 30 | 15.5 | RCT | Unknown |
| Byles & Maurice | 1979 | Family therapy | Family | 154 | 151 | 11.6 | RCT | CAU |
| Chernoff & Watson | 2000 | Community-based diversion program | One-on-one | 213 | 300 | unknown | Quasi | CAU |
| Cox | 1999 | Education program | Multimodal | 41 | 42 | 13.0 | RCT | CAU |
| Davidson et al. | 1987 | Diversion program | One-on-one | 76 | 89 | 14.2 | RCT | CAU |
| Davidson & Wolfred | 1977 | Behavior-modification program | Group | 42 | 42 | 14.2 | Quasi | Unknown |
| Fo & O'Donnell | 1975 | Community-based program | One-on-one | 48 | 25 | unknown | RCT | No treatment |
| Gordon et al. | 2007 | Behavioral family therapy | Family | 27 | 27 | 15.4 | Quasi | CAU |
| Gottfredson | 1986 | School-based program | Group | 296 | 249 | unknown | RCT | CAU |
| Hanlon et al. | 2002 | Community-based program | Multimodal | 235 | 193 | 13.3 | Quasi | CAU |
| Hayes | 2005 | Restorative justice conferences | One-on-one | 64 | 158 | 13.7 | RCT | CAU |
| Jones & Ross | 1997 | Boot camps | Group | 331 | 369 | 20.0 | Quasi | CAU |
| Jones et al. | 2004 | Wilderness Adventure program | Group | 24 | 11 | 13.6 | Quasi | Alternative treatment |
| Keating et al. | 2002 | Mentoring | One-on-one | 34 | 34 | 13.1 | Quasi | Waiting list |
| Kelley et al. | 2003 | Probation | Multimodal | 143 | 143 | 15.2 | RCT | Placebo |
| Kirigin et al. | 1982 | Family Group homes | Group | 140 | 52 | 14.6 | Quasi | CAU |
| Lane et al. | 2005 | Probation program | Multimodal | 226 | 236 | 15.0 | RCT | CAU |
| Lee & Haynes | 1978 | Community outreach counseling | Multimodal | 29 | 32 | 15.2 | RCT | CAU |

Appendix 3.A (continued)

| Study | Year | Main intervention | Treatment format | n EC | n CG | Mean age (in years) | Study design | Condition of CG |
|--------------------|------|---------------------------------|------------------|------|------|---------------------|--------------|-----------------|
| Leiber & Mawhorr | 1995 | Skills training | Group | 85 | 141 | 15.8 | Quasi | CAU |
| McGarrell | 2001 | Restorative Justice Conferences | Group | 167 | 168 | unknown | RCT | CAU |
| McGarrell & Hipple | 2007 | Family Group Conferencing | Group | 400 | 382 | 12.6 | RCT | CAU |
| McPherson et al. | 1983 | Counseling | Multimodal | 15 | 60 | 15.0 | RCT | CAU |
| Nugent & Paddock | 1995 | Victim-offender mediation | Multimodal | 111 | 130 | 15.4 | Quasi | CAU |
| Patrick et al. | 2004 | Diversion program | One-on-one | 71 | 68 | unknown | RCT | CAU |
| Poulin et al. | 2001 | Cognitive behavioral program | Group | 57 | 62 | 12.4 | Quasi | Nothing |
| Quinn & Van Dyke | 2004 | Multiple family group program | Family | 360 | 95 | 14.2 | Quasi | CAU |
| Stickler et al. | 2008 | Teen courts | Group | 42 | 33 | 14.9 | RCT | CAU |
| Stuart et al. | 1976 | Behavioral program | One-on-one | 30 | 30 | unknown | Quasi | Placebo |
| Vitaro & Tremblay | 1994 | Multicomponent program | Multimodal | 38 | 38 | 6.1 | RCT | Minimal contact |
| Welsh et al. | 1999 | Community-based program | unknown | 62 | 83 | 13.6 | Quasi | Nothing |

Note: n EC = number of participants in experimental group; n CG = number of participants in control group; RCT = randomized controlled trial; Quasi = quasi-experimental design; CAU = care as usual.

Appendix 3.B Results of the Multivariate Model with Program characteristics: Components, Target population, Format, and Setting

| Moderator variables | Mean <i>d</i> (SE) | CI | <i>t</i> | <i>p</i> | Level-2 variance | Level-3 variance | Omnibus test ^a |
|--|--------------------|---------------|----------|----------|------------------|------------------|---------------------------|
| Intercept | -0.330 (0.295) | -0.926; 0.267 | -1.118 | .270 | 0.028** | 0.041 | F(14,39) = 3.142** |
| Program components | | | | | | | |
| Behavioral modeling | 0.080 (0.321) | -0.570; 0.729 | 0.248 | .805 | | | |
| Parenting skills | 1.038 (0.380) | 0.270; 1.807 | 2.732 | .009 | | | |
| Behavioral contracting | 0.178 (0.256) | -0.341; 0.696 | 0.693 | .492 | | | |
| Primary target population | | | | | | | |
| Mixed (juveniles and parents) (vs. juveniles) | -0.806 (0.518) | -1.854; 0.242 | -1.555 | .128 | | | |
| Mixed (juveniles, parents and siblings) (vs. juveniles) ^b | - | - | - | - | | | |
| Program format | | | | | | | |
| One-on-one (vs. group) | 0.555 (0.258) | 0.032; 1.078 | 2.148 | .038 | | | |
| Family (vs. group) | 1.016 (0.596) | -0.188; 2.221 | 1.707 | .096 | | | |
| Multimodal/mixed (vs. group) | 0.243 (0.227) | -0.215; 0.702 | 1.074 | .289 | | | |
| Setting | | | | | | | |
| Home (vs. ambulant) | -0.329 (0.755) | -1.857; 1.198 | -0.436 | .665 | | | |
| Public school (vs. ambulant) | 0.558 (0.337) | -0.123; 1.240 | 1.657 | .106 | | | |
| Specialized school (vs. ambulant) | -0.150 (0.386) | -0.930; 0.630 | -0.389 | .700 | | | |
| Clinic (vs. ambulant) | 0.148 (0.368) | -0.595; 0.892 | 0.403 | .689 | | | |
| Court (vs. ambulant) | 0.020 (0.348) | -0.684; 0.724 | 0.057 | .954 | | | |
| Community-based (vs. ambulant) | 0.134 (0.257) | -0.385; 0.653 | 0.522 | .605 | | | |
| Other (vs. ambulant) | 0.310 (0.290) | -0.275; 0.896 | 1.072 | .290 | | | |
| # Studies | 22 | | | | | | |
| # ES | 54 | | | | | | |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean *d* = mean effect size (*d*); CI = Confidence interval; Level 2 variance = variance between effect sizes from the same study; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.

^b Redundant predictors were dropped from the model.

+ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Appendix 3.C Results of the Multivariate Model with Program Intensity, Adjusting for Components and Format

| Moderator variables | Mean <i>d</i> (SE) | CI | <i>t</i> | <i>p</i> | Level-2 variance | Level-3 variance | Omnibus test ^a |
|-------------------------------------|--------------------|----------------|----------|----------|------------------|------------------|---------------------------|
| Intercept | 0.042 (0.124) | -0.214; 0.298 | 0.341 | .736 | 0.105*** | 0.037 | F(4,23) = 1.543 |
| Program components | | | | | | | |
| Behavioral modeling | -0.004 (0.263) | -0.548; 0.540 | -0.016 | .987 | | | |
| Parenting skills | -0.398 (0.311) | -1.042; 0.246 | -1.278 | .214 | | | |
| Behavioral contracting ^b | - | - | - | - | | | |
| Program format | | | | | | | |
| One-on-one (vs. group) ^b | - | - | - | - | | | |
| Family (vs. group) ^b | - | - | - | - | | | |
| Multimodal (vs. group) | 0.101 (0.222) | -0.358; 0.560 | 0.454 | .654 | | | |
| Number of sessions | -0.095 (0.043) | -0.184; -0.006 | -2.216 | .037 | | | |
| # Studies | 12 | | | | | | |
| # ES | 28 | | | | | | |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean *d* = mean effect size (*d*); CI = Confidence interval;

Level 2 variance = variance between effect sizes from the same study; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.

^b Redundant predictors were dropped from the model.

+ *p* < .10; * *p* < .05; ** *p* < .01; *** *p* < .001

Appendix 3.D Results of the Multivariate Model with Delinquency Dimension, adjusting for Components and Format

| Moderator variables | Mean <i>d</i> (SE) | CI | <i>t</i> | <i>p</i> | Level-2 variance | Level-3 variance | Omnibus test ^a |
|---------------------------------|--------------------|----------------|----------|----------|------------------|------------------|---------------------------|
| Intercept | 0.002 (0.083) | -0.164; 0.168 | 0.021 | .983 | 0.036*** | 0.048** | F(9,81) = 4.195*** |
| Program components | | | | | | | |
| Behavioral modeling | 0.096 (0.154) | -0.211; 0.403 | 0.621 | .536 | | | |
| Parenting skills | 0.358 (0.150) | 0.059; 0.657 | 2.379 | .020 | | | |
| Behavioral contracting | 0.015 (0.182) | -0.348; 0.377 | 0.080 | .936 | | | |
| Program format | | | | | | | |
| One-on-one (vs. group) | 0.264 (0.147) | -0.028; 0.556 | 1.800 | .076 | | | |
| Family (vs. group) | 0.691 (0.283) | 0.128; 1.254 | 2.444 | .017 | | | |
| Multimodal (vs. group) | 0.292 (0.121) | 0.051; 0.533 | 2.408 | .018 | | | |
| Delinquency dimension | | | | | | | |
| Frequency (vs. participation) | -0.089 (0.096) | -0.279; 0.102 | -0.927 | .357 | | | |
| Seriousness (vs. participation) | -0.294 (0.139) | -0.570; -0.017 | -2.112 | .038 | | | |
| Versatility (vs. participation) | -0.390 (0.271) | -0.930; 0.150 | -1.437 | .155 | | | |
| # Studies | 36 | | | | | | |
| # ES | 91 | | | | | | |

Note. # Studies = Number of studies; # ES = Number of effect sizes; Mean *d* = mean effect size (*d*); CI = Confidence interval; Level 2 variance = variance between effect sizes from the same study; Level 3 variance = variance between studies.

^a Omnibus test of all regression coefficients in the model.

* $p < .10$; ** $p < .05$; *** $p < .01$; **** $p < .001$.