



# A Comprehensive Meta-Analysis of Randomized Evaluations of the Effect of Child Social Skills Training on Antisocial Development

Andreas Beelmann<sup>1</sup>  · Friedrich Lösel<sup>2,3</sup>

Received: 8 July 2019 / Revised: 14 April 2020 / Accepted: 22 April 2020 /  
Published online: 5 May 2020  
© The Author(s) 2020

## Abstract

**Purpose** Developmental and life course criminology (DLCC) engages not only in correlational longitudinal research but also in programs of developmental prevention. Within this context, child training on social skills plays an important role. The present article contains a comprehensive meta-analysis of randomized controlled trials (RCTs) on the effects of this type of intervention.

**Method** We updated a meta-analysis on this topic Lösel & Beelman (Annals of the American Academy of Political and Social Science 587:84–109, 2003) to cover more recent studies while focusing specifically on aggression, delinquency, and related antisocial outcomes. From a systematic search of 1133 reports, we found 113 studies with 130 eligible RCT comparisons between a program and control group. Overall, 31,114 children and youths were included in these evaluations. Most interventions were based on a cognitive-behavioral approach.

**Results** Overall, the mean effect was positive, but rather small ( $d = 0.25$  using the random effect model). There were similar effects on aggression, delinquency, and other outcomes, but a tendency to somewhat stronger effects in behavior observations and official records than in rating scales. Most outcome measurements were assessed within 3 months or up to 1 year after training. Only a minority ( $k = 14$ ) had follow-up assessments after more than 1 year. In the latter studies, mean effects were no longer significant. Indicated prevention for youngsters who already showed some antisocial behavior had better effects than universal approaches and (partially related to this) older youngsters benefited more than preschool children. There was much heterogeneity in the findings. Evaluations performed since our previous meta-analysis in 2003 did not reveal larger effects, but training format, intensity, and other moderators were relevant.

**Conclusions** Mean results are promising, but more long-term evaluations, replications, booster approaches, and combinations with other types of interventions are necessary to ensure a substantial impact on antisocial development in the life course.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40865-020-00142-8>) contains supplementary material, which is available to authorized users.

**Keywords** Child social skills training · Aggression · Delinquency · Prevention · Meta-analysis

## Introduction

Many of the advances in developmental and life course criminology that led to the foundation of this journal have come from prospective longitudinal research. These designs have substantially increased our knowledge on the onset, continuity, and aggravation of as well as the desistance from criminal behavior along with our understanding of risk and protective factors and potential causal mechanisms. However, there is also a second strongly expanding field of developmental and life course criminology that has applied findings on risk and protective factors in order to prevent undesirable trajectories. Accordingly, programs aiming to avoid the onset and persistence of antisocial behavior in young people became a key topic of practice, policy, and research (e.g., [1–3]). Based on the findings of longitudinal research, numerous programs have been implemented in families, kindergartens, (pre)schools, family education centers, child guidance clinics, neighborhoods, youth justice institutions, and other contexts to prevent children from embarking on a life of crime [1, 4]. The mean effects of these early developmental prevention programs have been positive [3, 5–8]. Indeed, in a comprehensive review of meta-analyses, Farrington et al. [2] found desirable average effects for family-based, school-based, individually focused, and general prevention programs. However, typical effect sizes were small in statistical terms and contained much variation between and within various types of programs and implementations. Accordingly, the authors emphasized the need for more sound primary studies and research integrations. Reviews also showed that only a small proportion of prevention programs based on rigorous empirical evaluation, and these showed far fewer replicated and very long-term effects [9–11]. This article updates our earlier systematic review of international evaluations of child social skills trainings designed to prevent aggression, delinquency, and related antisocial behavior in young people that had covered findings up to 2000 [12, 13].

Child social skills' training is a widely used approach for preventing antisocial behavior and crime in the young. Originally developed in the 1970s within a social-behavioral framework [14], there are now numerous models aiming to promote healthy prosocial development (see [15]). Social skills training is disseminated across the globe in numerous manualized versions such as the PATHS curriculum [16] or the Dina Dinosaur school program [17]. At first glance, the different brand names suggest differences in the aims, conception, and methods, but a closer look at the concrete models shows much similarity. For example, modern concepts of social skills training involve not only behavioral skill components (e.g., verbal and nonverbal communication and interaction skills) but also training in social-cognitive (e.g., adequate social information processing and cognitions about self and others) and social-emotional skills (e.g., emotional regulation and expressions of emotions).

Therefore, modern social skills training relates to the broader concept of social competence [18]. Social competence comprises interpersonal and communication abilities that lead to positive social interactions, particularly in situations of social conflict [19–21]. Taking this broader view, recent training programs share two basic aims. The first is to promote certain social and social-cognitive competencies for a

healthy social development that will enable a successful coping with social developmental tasks and the establishment of mutually satisfactory social relationships with peers and adults. The promotion of social competencies should enable youngsters to realize their own social interests while simultaneously considering the social rights of significant others. This balance can vary depending on age, context, and situation. It should help to build up satisfactory social relations and friendships, promote interaction with peers, and facilitate coping with family and school problems. The second aim of social training programs is to prevent externalizing behavior problems such as aggression, delinquency, or violence, as well as internalizing problems such as anxiety or depressive symptoms. As research in developmental and life course criminology has shown, a number of different pathways and even many early starting and relatively persistent delinquents desist over time [22, 23]. However, one can consistently find the small group of relatively early starting and seriously antisocial youngsters postulated by Moffitt [24]. Moreover, early aggression and conduct problems are major risk factors for other behavioral, mental health, and social problems in later life [26–28]. Therefore, successful developmental prevention based on child skills training can be a highly important intervention.

In the field of developmental crime and delinquency prevention, child social competence and skills training typically use structured manuals. These often contain up to 20 training sessions in groups of about 10 participants. However, there are also more intensive programs (see [1, 18]). The programs target either all children of a school or neighborhood (universal prevention), selected children at enhanced risk (e.g., from poor single-parent backgrounds), or indicated groups such as youngsters who have already shown some behavior problems. Programs focus on key risk and protective factors for antisocial development such as social problem solving, nonaggressive social information processing, self-control, perspective taking, prosocial attitudes, alternative thinking in social conflict situations, or adequate evaluation of behavioral outcomes [29, 30]. Typical methods include age-appropriate instructions, group discussions, role-playing, singing and drawing, feedback, videos, individual guidance, and homework. Methods are primarily educational but can include therapeutic elements. Social skills training programs are widely used because they can reach both the general population and at-risk groups in schools and other institutional contexts. Universal approaches avoid selection and potential stigmatization, but for practical and financial reasons, they cannot be as intensive as indicated programs that require a more intensive intervention. Child skills training are quite easy to implement in kindergarten, pre-school, or school contexts, and therefore reach those groups that are most in need of support more often than family-based programs [4].

Social skills and social competence trainings have been evaluated intensively over the last decades, and several meta-analyses have been conducted. Some addressed a broad range of outcomes (e.g., [31, 32]); others were designed specifically to prevent antisocial behavior and crime (e.g., [13, 33, 36]). Evaluations have shown that—on average—child social skills trainings are able not only to promote socially competent behavior and social-cognitive skills but also to prevent early antisocial development. For example, our previous analysis [13] found mean post-intervention effect sizes of  $d = 0.39$  based on 127 studies, thereby indicating a moderate success. Effects were somewhat higher when programs had a cognitive-behavioral orientation, when treatment was more intensive, and when they addressed at-risk or

indicated groups (i.e., targeted instead of universal prevention) as well as older children (> 13 years; see [12]). There are also other issues that require more differentiated information on the effectiveness of child skills training to prevent antisocial development. First, effects on measures of social competence (e.g., prosocial behavior and social-cognitive problem solving) were larger than those on antisocial behavior ( $d = 0.43$  compared with 0.29; see [12]). This may have been due to “bottom effects,” because the prevalence of antisocial behavior in low-risk groups (i.e., in universal prevention) is rather low. Second, follow-up and long-term effects were very rare and showed decreasing effect sizes compared with post-intervention effects. The lack of sound long-term evaluations in prevention experiments has been emphasized repeatedly (e.g., [9, 10, 34]). This is particularly relevant, because child skills and other developmental prevention programs aim (more or less explicitly) to reduce antisocial behavior in the long term and not just for a few months (although any positive influence is welcome).

Against this background, the present article contains an update of our previous review and meta-analysis of international randomized evaluations of the effects of child social skills training on young people [12, 13]. Since our previous review, more RCTs on the effects of social skills training programs have appeared. In contrast to our previous meta-analysis, we now wanted to focus on antisocial outcomes such as aggression, delinquency, and violence. In general, we aimed to assess the current state of effectiveness in this field. More specifically, we pursued the following goals: First, we asked whether the more recent studies replicate our previous findings and endorse the potential of child skills training in developmental and life course criminology. Second, we examined whether improvements in theoretical and practical approaches have led to an increase in effectiveness. Third, we aimed to provide more differentiated knowledge about practically relevant moderators of effect size. Fourth, we searched for “blind spots” in research that need to be reduced if we are to improve developmental crime prevention. Based on our previous meta-analysis, the present review could apply proven methods while integrating more recent primary studies. Nonetheless, whereas our previous meta-analysis also contained cognitive and psychological outcomes, we now concentrate on aggression, delinquency, and other outcomes relating to antisocial behavior.

## Method

### Eligibility Criteria

As mentioned, this article contains an update of a previous meta-analysis [12, 13]. With one exception (see below), we used similar selection criteria and methods but extended the publication year of primary studies from 2000 to 2015. Included studies had to meet the following criteria:

- (1) Eligible studies had to contain a systematic evaluation of a *social skills training program* specially designed for the prevention of antisocial behavior and crime in children and youth. We excluded all studies evaluating a social skills training combined with additional programs or program components (e.g., programs

including parent training, teacher training, or home visits). Likewise, we excluded programs focusing on other areas of problem behavior (e.g., prevention of internalizing problems, substance use, or suicide) because these programs contain contents and exercises (e.g., speaking in front of groups, resist group pressure, promoting self-esteem) not relevant or even contradict to the prevention of antisocial behavior.

- (2) Studies had to compare an intervention and control group in a randomized controlled design (RCT). Stratified modes of randomization (e.g., randomized field trial, randomized block design, matching plus randomization, randomization at the level of clusters such as schools and classes) were included. Pre- and post-intervention data had to be available. Not only untreated and wait-list but also attention groups and treatment as usual groups were accepted as control groups. However, designs comparing two different programs without an untreated control condition were excluded.
- (3) The *age range* of the treated youngsters had to be between 0 and 18 years.
- (4) The program had to be preventive in the narrow sense. Studies on universal prevention and on targeted prevention in at-risk groups (selective or indicated prevention) were included. Treatment programs for already adjudicated delinquents or other clinical groups (e.g., internalizing disorders) were excluded. The content of the included programs may be similar to treatments for adjudicated juvenile delinquents. However, the legal context is different and there are already various meta-analyses that had shown mean positive effects for treatment of young offenders (e.g., [25, 35, 36]; for overviews see [37, 38]). Programs for youngsters with conduct disorders or oppositional defiant disorders were included because these were insofar preventive as they targeted at-risk groups for later offending.
- (5) Different from our prior meta-analysis on this topic, the studies had to assess at least one outcome for antisocial behavior or crime (e.g., parent report, teacher report, self-report, observational data, or official records). Outcome data had to be provided in sufficient detail to permit a calculation or reliable estimation of effect sizes.
- (6) We included all retrievable published or unpublished reports in English or German language up to the year 2015. Although some relevant studies may have been published in other languages, our search suggested that the vast majority of sound quality studies was in English. Insofar, we do not assume a serious language bias in the present review.

## Search Strategy

We used the same strategy to locate relevant literature as Lösel and Beelmann [13]. The current extension of eligible primary studies from 2000 to 2015 enabled an updated picture of findings in recent research. We used the following search procedure:

- (1). Electronic searches were conducted in the following most relevant databases: PsycINFO, PsycARTICLES, PubPsych, ProQuest Dissertations & Theses Full Text, ERIC, Web of Science, and PubMed. A combination of search terms was

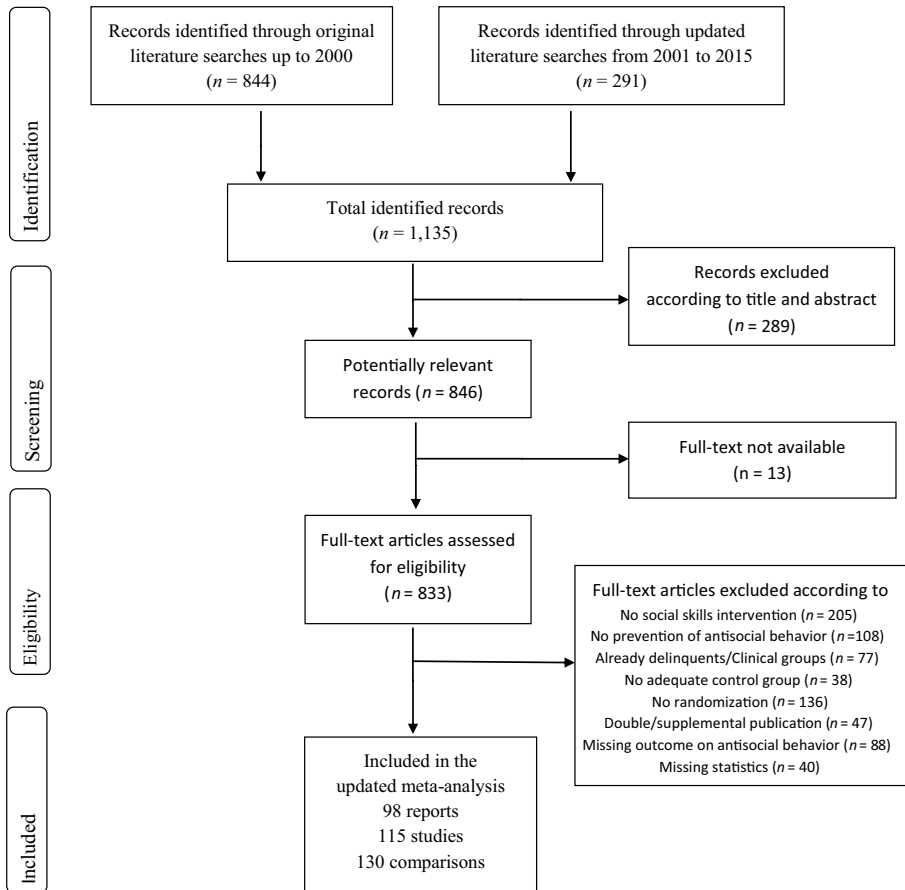
- used referring to the type of intervention (e.g., *prevention, social skills training*), the intervention field (e.g., *antisocial behavior, delinquency, crime*), age of the target group (e.g., *children, adolescents*), and the study design (e.g., *randomization, control group*, see Supplement A). The exact search syntax varied slightly for each database.
- (2). As in our prior searches, the reference lists of existing reviews or meta-analyses published since 2000 were screened for more recent evaluation studies (e.g., [2, 3, 32, 39, 40]).
  - (3). In addition, the references of all eligible reports were screened for further relevant studies.

A total of 851 reports were identified within the first literature search up to the year 2000. These resulted in 135 intervention–control contrast comparisons (see Lösel & Beelmann [13] for details of the selection process). From this pool, we now had to delete 27 reports with 50 intervention–control comparisons because they did not use an outcome measure for antisocial behavior (see selection criterion 5). Our updated literature searches since 2001 based on the title and the keywords produced 291 additional reports. From these, we finally had to exclude 250 reports because they evaluated combined or different interventions (e.g., parent and teacher training), were not primarily conducted for the prevention of antisocial behavior or crime, or the target groups were not appropriate (e.g., adjudicated offenders). Further reports had no RCT design or no untreated control group or only an alternative intervention condition, were totally or partially double publications of already included studies (additional follow-up studies of included studies were considered), did not report sufficient statistical data, or had not assessed antisocial behavior as an outcome (for details see Fig. 1).

In sum, we ended up with a study pool of 98 reports (see Supplement B, including additional reports from the respective studies). Some reports contained more than one independent study or separate data for two or more distinct subgroups (e.g., two age or risk groups). These were treated as separate data sets. In addition, several studies had more than one intervention group, thereby allowing more than one comparison between an intervention and a control condition. Therefore, the 98 reports resulted in a total of 113 studies and 130 RCT comparisons between an intervention and a control group. These comparisons were the final basic units of our analyses.

## Coding Procedures

The coding of the updated reports, studies, and comparisons was carried out as in our previous study. The first author and two trained students coded all new comparisons according to the characteristics of the report (e.g., year, country of origin), methods (e.g., design, follow-up assessments), intervention programs (e.g., type, intensity, setting), and the children and adolescents trained (e.g., age, gender, risk factors). Selections of these variables are displayed in Table 1 below. On the effect size level, we coded outcome categories of antisocial behavior (aggression, delinquency, oppositional-defiant behavior, and general measures of antisocial behavior), the assessment method (self-report, parent and teacher ratings, official data), the time between the termination of the intervention and the outcome assessment, and the effect sizes. According to our previous experience, simple categories like publication characteristics



**Fig. 1** Prisma flow chart for the identification of relevant research reports, studies, and comparisons

or age of participants reached very high inter-rater agreement, whereas complicated categories such as type of program were rather difficult to rate (see also [41]). Therefore, to ensure reliability, the coding team met weekly for a period of about 1 year to discuss unclear issues until all inter-coder reliabilities attained Cohen's  $\kappa > 0.8$ , which corresponded to our previous coding reliabilities.

## Statistical Analyses

Unbiased standardized mean differences were calculated from means and standard deviations of treatment and control groups or from test statistics as recommended by Lipsey and Wilson [42]. Pooled pretest standard deviations were used for effect size calculation as proposed by Morris [43]. If the reports mentioned non-significant results with no further details, we counted these as zero effects. Since many studies contained several relevant outcome measures, we computed at first 385 individual effect sizes for the various antisocial behavior outcomes. Of these, 246 (63.9%) were calculated from means and standard deviations, 49 (12.7%) were estimated as zero effects, 57 (14.8%)



**Table 1** Overall effect size statistics

Assessment	FEM		REM		<i>k</i>	%I	OR
	$d_+$	$\pm CI$	$d_+$	$\pm CI$			
Period							
Post-Intervention	0.11***	$\pm 0.02$	0.25***	$\pm 0.09$	119	12.5	1.57
	Qt (df = 118) = 454.86, $p < 0.001$ ; $I^2 = 89.6\%$						
Follow-up 1	0.12***	$\pm 0.06$	0.15**	$\pm 0.10$	20	7.5	1.31
	Qt (df = 19) = 45.40, $p < 0.001$ ; $I^2 = 58.1\%$						
Follow-up 2	0.04	$\pm 0.07$	0.12	$\pm 0.16$	14	6.0	1.24
	Qt (df = 13) = 47.41, $p < 0.001$ ; $I^2 = 72.6\%$						

*FEM*, fixed effect size model; *REM*, random effect size model; *ES*, effect size;  $d_+$ , weighted mean effect size; *CI*, Limits of the 95% confidence interval; *k*, number of comparisons; %I, Improvement in percent (REM); *OR*, odds ratio (REM)

\*\* $p < .01$

\*\*\* $p < .001$

were recalculated from reported test statistics, and 33 (8.6%) were taken directly from the reports.

These individual effect sizes were then aggregated at the comparison level by averaging them within each comparison between an intervention and a control condition according to the coded post-intervention and follow-up time intervals. However, as comparisons applied not the same assessment intervals (ranged from immediately up to 72 months after the training), it was necessary to build unified post-intervention and follow-up time periods. Therefore, we classified all assessments up to 3 months as Post-intervention, all assessments between 3 and up to 12 months as Follow-up I, and all assessments at 12 and more months after the termination of the intervention as Follow-up II assessments. In total, within the 130 comparisons, 173 outcome assessments had been conducted (124 at Post-intervention, 26 at Follow-up I, and 23 at Follow-up II). To obtain only one effect size for each time interval for each comparison, all within comparison effect sizes were averaged for the respective periods (e.g., two assessments that were conducted immediately after the training and 1 month later were aggregated within our post-intervention time interval up to 3 months). This procedure resulted in 119 Post-intervention comparisons, 20 Follow-up I, and 14 Follow-up II comparisons (see Table 1).

Comparison effect sizes were then integrated across comparisons following Lipsey and Wilson's [42] recommendations on weighting by the inverse of the squared standard error using methods developed by Hedges and Olkin [44]. For our analyses, we used SPSS 24 with macros developed by Wilson [45] and the procedure metafor for R.

## Results

### Description of Study Characteristics

The included reports were published since 1971. Forty-one (41.8%) appeared since 2000, the last year of the search in our previous review [13]. Most reports stemmed



from investigations in the US (74 reports or 75.5%). Further reports originated from Germany (8), Canada (6), Israel and the Netherlands (each 2), and England, Spain, China, Switzerland, Italy, and Austria (one report each). On the comparison level, sample sizes had a huge range and varied between 13 and 6733 with a mean of 262.6 ( $SD = 795.9$ ,  $k = 130$ ). Sixty-six comparisons (50.8%) had sample sizes up to 50, 47 comparisons (35.8%) more than 100, and 14 (10.8%) more than 500 participants. In total, 31,140 children and adolescents participated in the comparisons (16,864 in the training groups).

## Overall Intervention Effects and Stability of Outcomes

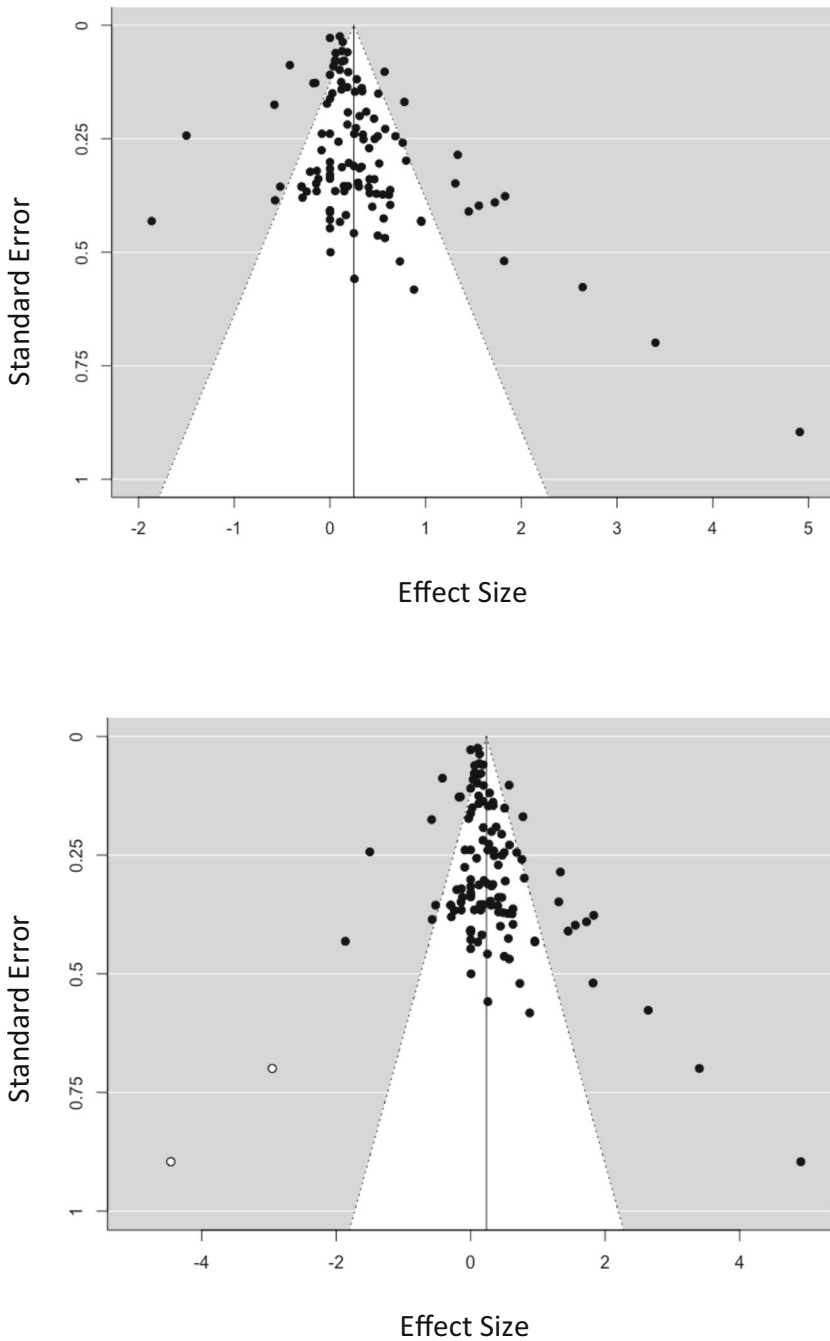
Post-intervention individual effect sizes ( $n_{es} = 294$ ) ranged between  $-1.89$  and  $4.91$  ( $M = 0.26$ ). Follow-up individual effect sizes ( $d_{es}$ ) ranged between  $-0.30$  and  $2.39$  ( $M = 0.18$ ) for the Follow-up I period, and between  $-1.25$  and  $2.23$  ( $M = 0.11$ ) for Follow-up period II. About one fifth (20.1%) of the post-intervention individual effect sizes were negative. At Follow-up I 18.3% and at Follow-up II 36.7% went in an (only partially significant) undesirable direction, i.e., the control group showed a better outcome than the treatment group. However, the majority of evaluations revealed a positive outcome (i.e., better results for the treated group). The mean weighted post-intervention effect size was  $d_+ = 0.11$  using the fixed effect model (FEM) and  $d_+ = 0.25$  using the random effect model (REM). The same pattern was observed for Follow-up I and Follow-up II effect sizes (see Table 1), indicating some impact of the aggregation method. However, as all mean weighted effect sizes showed significant heterogeneity (see Table 1) with a high amount of random variance (see  $I^2$  statistic), we restricted all further analyses to the most appropriate REM.

In general, the mean effect sizes for the different time periods could be interpreted as an improvement of 5–13% (at post-intervention) with an odds ratio of 1.57 (e.g., a 1.57 better chance for improvement in the intervention group). However, these values decreased for the follow-up (see Table 1). The mean weighted effect size for the Follow-up II period was no longer significantly different from zero. This pattern of results was not influenced by the type of control group. Comparisons with treated controls (placebo, minimal treatment, and treatment as usual) yielded only slightly lower effect sizes than comparisons with untreated controls (e.g.,  $d_+ = 0.23$ ,  $k = 37$  vs.  $d_+ = 0.26$ ,  $k = 82$  for the Post-intervention period).

## Sensitivity Analyses, Publication Bias, and Methodological Moderators

Some individual and comparison effect sizes were far outside the range of the effect size distribution and should be interpreted as outliers. This was especially apparent for five intervention–control comparisons that yielded untypically high negative ( $d = -1.86$ ; [46];  $d = -1.50$ , [47]) or high positive effect sizes ( $d = 2.64$ , [48];  $d = 3.40$  and  $4.91$  from two studies by [49]). All these effects contained problems because they had either very low standard deviations (in small samples), measures of low intensity of antisocial behavior (e.g., *speaking loud in classes* or *leaving seat in classes* as measure for disruptive behavior), or an unusual high increase of problem behavior within the control group. Therefore, we carried out sensitivity analyses with and without these outlier effects. The effect on the mean post-intervention effect size (REM) was small ( $d_+ = 0.23$ ,  $k = 114$ ) when all cited outliers were

omitted. It remained significant when the negative outliers were deleted ( $d_+ = 0.26$ ,  $k = 117$ ) and when only the positive outliers were omitted ( $d_+ = 0.21$ ,  $k = 116$ ). Although the



**Fig. 2** Funnel plot before (left) and after the trim and fill analysis (right)

sensitivity analyses did not reveal a strong impact, we will report results with and without outliers when they were significant in our moderator analyses (see below).

As the overall effect became smaller when the intervention–control comparisons were integrated by weighting for sample sizes, we tested our study pool also for publication bias. A publication bias may be assumed if a negative relation between sample sizes and effect sizes is observed. This suggests that studies with small sample sizes and non-significant results may have a lower chance to be published due to author, reviewer, or editor decisions. Indeed, in our study, Begg's rank correlation [50] showed a small but significant negative relation between sample size and post-intervention effects ( $\tau = -0.13$ ,  $p < 0.04$ ,  $k = 119$ ). A funnel plot (see Fig. 2) indicated a significant asymmetry of the effect size distribution (Egger test,  $t = 3.67$ ,  $df = 117$ ,  $p < 0.001$ ), and the trim and fill procedure [51] estimated two missing studies on the left side of the effect size distribution (i.e., two high negative effect sizes in Fig. 2). When we included these two effect sizes into the total post-intervention calculation, the mean effect size decreased slightly ( $d_+ = 0.23$ ,  $k = 121$ ). Finally, a comparison of effect sizes between studies published in journals or books and those reported in (sometimes unpublished) dissertations showed no significant difference, but a slightly higher effect for officially published reports ( $d_+ = 0.27$  vs.  $0.15$ ,  $k = 100$  vs.  $19$ ;  $Q_b = 0.94$ ,  $df = 1$ ,  $p = 0.33$ ). Hence, although our study pool may contain some publication bias, this was rather small.

Methodological moderators can have much impact on the findings of meta-analyses (e.g., [41]). As we only included RCTs, we controlled for variation in the basic evaluation design and could assume sound internal validity of the primary studies. However, even RCTs can suffer from serious threats to validity. One is the dropout of participants that may reduce internal validity of the randomization process. Indeed, we found a tendency in this direction. Effect sizes and reported dropout rates at post-intervention correlated at  $r = -0.15$  ( $k = 99$ ,  $p < 0.09$ , sample weighted correlation). When we carried out categorical (non-linear) analyses, this trend became more obvious. If the dropout rate was below 5%, we found the highest effect size ( $d_+ = 0.40 \pm 0.15$ ,  $k = 53$ ), if it was between 5 and 25%, we found a lower although still significant effect size ( $d_+ = 0.23 \pm 0.16$ ,  $k = 37$ ). Only a few comparisons ( $k = 9$ ) showed higher dropout rates (> 25%) with an effect-size of almost zero ( $d_+ = 0.00 \pm 0.32$ ). These differences were nearly significant ( $Q_b = 5.66$ ,  $df = 2$ ,  $p < 0.06$ ). Dropout seemed to have an impact on results, but not in the expected direction.

### Differential Effects for Different Antisocial Behavior Outcomes

Table 2 displays differential results for different outcomes of antisocial behavior and assessment methods. Outcomes were coded into four categories: aggressive behavior (e.g., ratings on physical and verbal aggression, aggressive interactions), delinquent behavior (e.g., damage of others' property, theft, school suspensions, school records, drug abuse), oppositional-disruptive behavior (e.g., scales on oppositional behavior, disruptive school behavior), and general antisocial measures (e.g., summary ratings on externalizing or antisocial behavior). In addition, outcomes were classified into four groups of assessment methods (rating scales, behavior observations, official records, and peer nominations). In all four categories of outcome behavior and the respective

**Table 2** Effectiveness according to different antisocial behavior outcome categories and assessment methods on the effect size and comparison level (post-intervention)

Outcome category/assessment method	Effect size level		Comparison level	
	$d_{es+}$	$n_{es}$	$d_+ \pm CI$	$k$
<b>Outcome category</b>				
Aggression	0.13**	139	0.23 ± 0.13 <sup>+</sup>	77
Delinquency	0.21*	33	0.27 ± 0.20	21
Oppositional/disruptive	0.22*	55	0.27 ± 0.17	35
General antisocial	0.21*	67	0.21 ± 0.10	43
<b>Assessment method <sup>a</sup></b>				
Ratings	0.15*	202	0.18 ± 0.08	93
Behavioral observation	0.26**	55	0.51 ± 0.29	32
Official records	0.21*	20	0.40 ± 0.37	12
Sociometric nominations	0.22*	12	0.11 ± 0.06	10

$d_{es}$ , mean weighted effect size on effect size level (random effect-size model);  $n_{es}$ , number of individual effect sizes;  $d_+ \pm CI$ , weighted mean effect size with 95% confidence interval (random effect-size model);  $k$ , number of comparisons

<sup>a</sup> Due to a low number of outcomes based on psychometric test scores these effect sizes are omitted

<sup>+</sup> Effect size showed significant heterogeneity.

\* $p < 0.05$

\*\* $p < 0.01$

\*\*\* $p < 0.001$

assessment methods, we found significant effects (see Table 2). There were also no significant differences between the effects in different types of operationalization or assessment at post-intervention. With regard to outcome assessment, rating scales were the most frequent method ( $n_{es} = 202$  of 294 individual effect sizes). However, the highest effect sizes yielded in behavior observations and official records ( $d_+ = 0.51$  and 0.40 on the comparison level). The rather high effect size for behavior observations was partially due to the outlier effect sizes because all positive outliers are based on behavior observations. However, even without these studies, the mean effect was still larger than for other assessment methods ( $d_+ = 0.36$ ,  $k = 28$ ).

### Differential Effects for Intervention and Sample Characteristics

Table 3 shows differential effects for selected moderators of the program content and samples. As most child skills programs used cognitive-behavioral concepts, there was not much variation in the basic type of program. Probably for this reason, no intervention characteristic had a significant impact on program outcomes. There were only some noteworthy results on a descriptive level. For example, programs that based on Adlerian and Humanistic concepts produced slightly higher effects than cognitive-behavioral programs at Post-intervention and at Follow-up period I ( $d_+ = 0.10$ ,  $k = 15$  vs.  $d_+ = 0.20$ ,  $k = 5$ ). Within the cognitive-behavioral category, programs with a mainly behavioral orientation (in contrast to social-cognitive orientation) yielded the highest

**Table 3** Overview of selected moderator analyses (post-intervention)

Characteristic	$d_+ \pm CI$	$k$	$Q_w$	$Q_b$ (df)
Intervention characteristics				
Intervention approach <sup>a</sup>				0.10 (1)
Cognitive-behavioral training	0.24 ± 0.09	102	152.58***	
Adlerian & humanistic	0.31 ± 0.23	16	10.13	
Intervention intensity <sup>b</sup>				3.72 (3)
Low	0.15 ± 0.17	35	29.36	
Low moderate	0.36 ± 0.16	40	74.54***	
High moderate	0.21 ± 0.16	31	37.60	
Intensive	0.26 ± 0.24	13	19.18	
Intervention format				3.35 (4)
Individual training	0.34 ± 0.33	10	8.16	
Group training	0.25 ± 0.09	101	134.36*	
Individual and group training	-0.14 ± 0.49	5	12.89*	
Self-help condition	0.52 ± 0.66	2	5.55*	
Individual supervision/care	0.18 ± 0.74	1		
Program administrator/facilitator <sup>c</sup>				1.23 (2)
Psychosocial practitioners	0.22 ± 0.11	67	108.27***	
Authors, university staff	0.32 ± 0.17	37	41.62	
Various	0.12 ± 0.46	4	0.53*	
Publication year (program recency)				5.23 (1)*
Up to 2000	0.33 ± 0.11	78	145.51***	
2001 to 2015	0.13 ± 0.12	41	19.02	
Sample characteristics				
Prevention type/risk level				21.81 (2)***
Universal	0.11 ± 0.12	36	49.22	
Selective	0.08 ± 0.15	29	19.41	
Indicated	0.49 ± 0.13	54	96.33***	
Mean age of participants (years)				14.09 (4)**
Up to 6	0.15 ± 0.18	22	19.46	
6.1–8.9	0.09 ± 0.18	24	27.53	
9–10	0.44 ± 0.18	29	26.19	
10.1–13.9	0.22 ± 0.16	34	35.22	
14–17.5	0.62 ± 0.32	10	50.71***	
Gender (proportion of male youngsters)				10.94 (3)*
Up to 40%	0.17 ± 0.40	4	0.60	
41–60%	0.13 ± 0.11	58	80.90*	
61–80%	0.41 ± 0.20	24	26.49	
81–100%	0.41 ± 0.19	33	59.85**	

$d_+ \pm CI$ , weighted mean effect size according to the random-effect-model with 95% confidence interval;  $k$ , number of comparisons;  $Q_w$ , heterogeneity within a category with  $k-1$  degrees of freedom;  $Q_b$ , heterogeneity between groups

\* $p < 0.05$

\*\* $p < 0.01$

\*\*\* $p < 0.001$

<sup>a</sup> One comparison could not be classified within these two categories

<sup>b</sup> Intensity was calculated as total hours by multiplying number and duration of sessions. If this was not possible (e.g., due to missing values) we used the number of sessions alone or the training period as additional information. Low = up to 8 h or up to 10 sessions/2 month period. Low-moderate = 9–15 h or 11–20 sessions/2- to 4-month period. High-moderate = 16–30 h or 21–40 sessions/4- to 8-month period. Intensive = more than 30 h or more than 40 sessions/8 months period

<sup>c</sup> Because some programs did not have an administrator or due to missing values, the number of comparisons do not add up to 119

effect size ( $d_+ = 0.36$ ,  $k = 29$ ) at Post-intervention. The intensity of the program had no significant influence, although programs with low intensity ( $d_+ = 0.15$ ) revealed a somewhat smaller (and non-significant) mean effect and programs with a low to moderate intensity a somewhat higher mean effect (see Table 3 for description of intensity). Interestingly, when we omitted one very positive outlier, the intensive programs even showed a low and non-significant mean effect size ( $d_+ = 0.15 \pm 0.17$ ).

A large number of programs had a group format ( $k = 101$ ). Individual training formats seemed to be a bit more effective, but the difference between formats was not significant. The negative mean effect size of programs with an individual and group format may have been due to the very low number of comparisons and one high negative outlier. Without the latter comparison, these programs had an average effect of  $d_+ = 0.25$  ( $k = 4$ ). There was also a tendency toward larger effects in comparisons where study authors, program developers, or university staff had been administrators or facilitators of the program. In contrast, programs that have been administered by practitioners (e.g., teachers) showed somewhat smaller effects. This comparison nearly reached significance when we omitted the above-mentioned outlier effect size ( $d_+ = 0.18$ ,  $k = 63$  vs.  $d_+ = 0.36$ ,  $k = 36$ ,  $Q_b = 5.30$ ,  $df = 2$ ,  $p < 0.07$ ). However, these differences disappeared at the Follow-up assessments ( $d_+ = 0.16$ ,  $k = 14$  vs.  $d_+ = 0.16$ ,  $k = 3$ , at Follow-up period I, and  $d_+ = 0.13$ ,  $k = 8$  vs.  $d_+ = 0.19$ ,  $k = 4$ , at Follow-up period II).

More recent evaluations yielded lower effect sizes. Evaluations of programs that have been published since our original meta-analysis (from 2001 onwards) showed a significantly lower post-intervention effect size of  $d_+ = 0.13$  ( $k = 41$ ) compared with older studies (see Table 3). The same pattern could be identified for Follow-up period I ( $d_+ = 0.20$ ,  $k = 12$  vs.  $d_+ = 0.07$ ,  $k = 8$ ,  $Q_b = 3.20$ ,  $df = 1$ ,  $p < 0.08$ ) but not for the Follow-up period II ( $d_+ = 0.06$ ,  $k = 8$  vs.  $d_+ = 0.15$ ,  $k = 6$ ,  $Q_b = 0.31$ ,  $df = 1$ ,  $p < 0.57$ ). These results were counterintuitive as one could expect that outcomes are getting better in more up-to-date programs. More detailed analyses showed that the more recent programs contained a universal approach. The proportion of universal programs was much lower in older studies (up to 2000, 16.7%) than in newer studies (from 2001 onward, 56.1%) and this was related to effect size (see below and Table 3). However, even within universal programs, older studies tended to produce larger effects ( $d_+ = 0.25$ ,  $k = 13$  vs.  $d_+ = 0.05$ ,  $k = 23$ ,  $Q_b = 2.07$ ,  $df = 1$ ,  $p < 0.15$ ) as well as within indicated samples ( $d_+ = 0.54$ ,  $k = 44$  vs.  $d_+ = 0.39$ ,  $k = 10$ ,  $Q_b = 0.65$ ,  $df = 1$ ,  $p < 0.42$ ). We also investigated another reason for this unexpected result. The control groups of older vs. more recent studies showed different developments. Whereas control children in older studies became somewhat worse from pre- to post-test ( $d_+ = -0.16 \pm 0.14$ ,  $k = 49$ ), this

was not the case in more recent studies ( $d_+ = 0.01 \pm 0.15$ ,  $k = 25$ ; lower number of comparisons due to missing data), although this difference was not significant ( $Q_b = 2.35$ ,  $df = 1$ ,  $p < 0.13$ ).

Turning to sample characteristics, child social skills training was significantly more effective in indicated prevention approaches, i.e., when the youngsters have already shown some behavioral problems (see Table 3). These groups benefited more from the programs than groups that participated in universal programs (i.e., for all students at a school) or selected groups (i.e., with some individual or family risk factors). For the latter groups, the effect sizes were not significantly different from zero. This could be partially due to the intensity of those programs. Almost half (48.3%) of the selective prevention programs were of low intensity (i.e., less than 8 h or 10 sessions, see Table 3) with a non-significant effect size ( $d_+ = 0.11 \pm 0.12$ ,  $k = 14$ ). This proportion was higher than for universal and indicated prevention programs. However, at Follow-up period I, programs for at-risk groups (selective prevention) yielded at least a small but significant effect size ( $d_+ = 0.20$ ,  $k = 9$ ) that disappeared at Follow-up period II ( $d_+ = 0.16$ ,  $k = 3$ ). Again, universal prevention yielded no significant follow-up effect ( $d_+ = 0.06$ ,  $k = 8$  for Follow-up period I and  $d_+ = 0.09$ ,  $k = 5$  for Follow-up period II). In contrast, indicated prevention had a significant follow-up effect at period I ( $d_+ = 0.81$ ,  $k = 3$ ), but not at period II ( $d_+ = 0.13$ ,  $k = 6$ ). As at post-intervention, these results let to a significant difference at Follow-up period I between different prevention strategies ( $Q_b = 17.81$ ,  $df = 2$ ,  $p < 0.001$ ).

In addition to different risk levels, age of participants was a significant predictor of program effectiveness. A meta-regression showed that the mean age of the sample explained a significant part of outcome heterogeneity ( $Q_R = 5.77$ ,  $df = 1$ ,  $p < 0.02$ ) with a significant beta coefficient ( $\beta = 0.19$ ,  $p = 0.02$ ) that indicated higher effectiveness in older groups. However, a slightly different picture with a nonlinear trend appeared when we compared five different age groups (see Table 3). For the two youngest groups, analyses revealed no significant effects. Larger effects were observed for the 9 to 10 years old youngsters (where social skills training was most often implemented) and for the oldest group from 14 years upward, resulting in a significant moderator effect (see Table 3). In contrast, this picture changed when we omitted the outlier effect sizes. First, the two youngest age groups now had low but significant effect sizes ( $d_+ = 0.19$ ,  $k = 21$  and  $d_+ = 0.14$ ,  $k = 23$ ). Second, the effect size for the oldest (adolescent) group decreased to a non-significant value of  $d_+ = 0.18$  ( $k = 7$ ). However, the 9 to 10 years old group have still the highest effect size and the moderator test was still significant ( $Q_b = 10.30$ ,  $df = 4$ ,  $p < 0.04$ ). In addition, for this age group, we found a high and significant effect size for indicated and even for universal prevention programs ( $d_+ = 0.59$ ,  $k = 13$  and  $d_+ = 0.52$ ,  $k = 7$ , respectively), whereas we did not find an interaction between age and prevention type in other age groups. All effects remained low for universal and selected prevention and higher for indicated prevention. Overall, the results on age as moderator were somewhat complicated, but we can summarize that there was a relatively low effect in younger groups, a medium effect size around the age of 9 to 10, and (dependent of how we weight outliers) a low to medium effect for adolescent groups.

In addition to risk and age, gender was a significant moderator of effect size. A meta-regression showed that the proportion of male participants explained a significant part of the heterogeneity ( $Q_R = 3.96$ ,  $df = 1$ ,  $p < 0.05$ ) with a significant beta coefficient ( $\beta = 0.16$ ,  $p < 0.05$ ) that indicated larger effects when more males were in the study



sample. We also grouped the gender proportion into categories and found higher effect sizes when more male youngsters participated (above 60%) as compared with studies with a more balanced gender ratio or more participating girls (see Table 3). These differences were not affected by outlier effect sizes. However, there was a clear confound between gender and risk group (or prevention type). Most comparisons evaluating indicated prevention with higher effect sizes had predominantly male participants (e.g., more than 60%;  $d_+ = 0.44$ ,  $k = 41$  out of 54). In contrast, the majority of comparisons evaluating universal prevention programs with lower and non-significant effect sizes had balanced gender proportions ( $d_+ = 0.05$ ,  $k = 31$  out of 39).

Finally, we put these three significant moderators into a meta-regression to determine the most relevant characteristic (universal and selective groups vs. indicated groups as dummy variable, mean age and proportion of male participants as continuous variables). The results were quite clear. The total model was significant ( $Q_R = 23.14$ ,  $df = 3$ ,  $k = 119$ ,  $p < 0.001$ ), but the only variable that still showed a significant impact was the type of prevention. Indicated prevention programs revealed higher effect sizes ( $\beta = 0.33$ ,  $p < 0.001$ ), whereas age and proportion of male gender were non-significant ( $\beta = 0.09$  and  $-0.03$ , respectively). This pattern was even more pronounced if we omitted the outlier effect sizes ( $Q_R = 17.60$ ,  $df = 3$ ,  $k = 114$ ,  $p < 0.001$ ;  $\beta = 0.36$ ,  $p < 0.001$  for the prevention type and  $\beta = -0.04$  and  $-0.01$ , for age and male gender proportion).

## Discussion

### General Findings

Our meta-analysis of child social skills training integrates 130 RCTs on this type of developmental prevention of antisocial behavior. This substantial body of sound research shows a significant mean positive effect on aggressive, delinquent, and related child behavior problems. Although rather small ( $d = 0.25$ ), the overall effect is robust despite a potentially slight influence of publication bias. Nonetheless, most studies had only short follow-up periods, and the long-term effect after more than 1 year is not significant. Looking beyond the details of our analysis that we discuss below, this is one particularly important message of our study: It shows that systematic reviews do not just provide replicated knowledge on specific interventions; they also map deficits in the research landscape. Although we found more studies with a follow-up period of at least 1 year than in our previous meta-analysis, very long follow-ups are still very rare. This corresponds to a review of various types of developmental crime prevention that revealed only about a handful of studies with follow-ups of 10 years or more [9]. Some of these studies contained combined child- and family-oriented approaches, but none for targeted child social skills training alone. This shows that the favorable state of many RCTs on child social skills training is not sufficient, but calls for more long-term outcome evaluations. For example, in a follow-up of about 10 years that compared early child skills training and parent training, Lösel, Stemmler, and Bender [52] found at least some significant effects of the former. However, the lack of long-term effects in this analysis may point to the need for booster sessions or even a repeated application of child social skills training matched with the different social developmental tasks at different ages or developmental periods [53].

The overall rather small effect of child social skills training in the present review is in the same range as that in our previous meta-analysis about 15 years ago ( $d=0.25$  for antisocial behavior outcomes; [13]). At first glance, this may be surprising, because research and practice have invested some effort in improving prevention programs. However, this finding is not unique. For example, the effects in meta-analyses on offender treatment have also not increased over time (e.g., [54, 55]). This may be due in part to more rigorous evaluation designs leading to smaller effects in some fields of intervention [56]. However, all the primary studies in our meta-analysis used an RCT, making this explanation implausible—as indicated by our results on dropout rates (higher effects in studies with lower dropout). We assume two other reasons for the lack of increased effectiveness over time: First, our updated review contained a larger proportion of universal prevention programs that, on average, had lower effects than indicated approaches (see Table 3). Second, in recent decades, many countries have become more aware of the need for developmental prevention, and correspondingly good routine services in child and family care are a reason for lower effects [57]. Our results on different control conditions revealing no differences between “treated” and “nontreated” control groups also confirm this view. As the second author of this article described at the Twentieth Anniversary Session of the Academy of Experimental Criminology [58], conditions in the control groups or “treatment as usual” may be highly relevant when evaluating effects of specific programs.

Although overall effect sizes are mainly low, they are not without practical relevance. Even small effects in developmental prevention can lead to a desirable benefit–cost ratio for society [2]. Many evaluations of child skills programs followed a universal prevention approach. Therefore, the respective small effects are plausible, because a majority of the children in these programs had no risk of antisocial development. Although universal prevention for whole schools or neighborhoods serves large subgroups that may not be in need of an intervention and can be only relatively short for financial reasons, this approach also has its advantages: It avoids potential stigmatization [59], and, as public health research has shown, it may reduce the overall population risk level [60]. Even in universal early prevention programs, children with some behavior problems seem to benefit more than other participants (e.g., [61]).

### **Moderator Effects**

The most robust finding on moderators in our study is the larger effect of indicated prevention programs. In the meta-regression, this reduced the weight of other moderators, although the latter should not be seen as irrelevant for practice. In regression analyses, it is typical for the strongest predictor to suppress the weight of other correlated variables that may have only a slightly smaller bivariate weight. According to our findings on indicated versus universal prevention, policy and practice should prioritize those children and families who already exhibit some problem behavior and therefore face an enhanced risk for later aggressive and delinquent development. This is also plausible for financial reasons. In cases of enhanced risk, child-oriented interventions should be combined with multimodal family, school, and neighborhood approaches (e.g., [62–64]). Compared with mere child skills training, these comprehensive “packages” of programs are more promising, but they are also more expensive, more difficult to implement, and more challenging for evaluation. For example, family-oriented prevention programs showed similar effect

sizes to child-oriented programs (e.g., [2, 4, 8]). Therefore, one can expect that combinations of such programs may be more effective than single programs by themselves. However, this is not always the case (e.g., [52]), and in the case of nonsignificant effects in complex “packages,” we need differentiated data to disentangle potential reasons for failure (or for stronger effects). There is widespread consensus among prevention experts that programs should not be too isolated, but embedded in broader and structural policies that address those children and families who are most in need [53, 65]. Within this context, child skills training can be a first step toward prevention and the assessment of further risks and needs.

The stronger effect of indicated prevention programs is in accordance with our previous meta-analysis [12, 13] and with reviews on other types of prevention and offender treatment programs [18, 38]. At first glance, the finding of larger effects for children who already show some antisocial behavior problems and who are beyond preschool age would seem to be counterintuitive, because one may expect that it would be easier to intervene when the risk is lower and during earlier childhood. However, our finding of a smaller effect of universal programs and at preschool age is to be expected, because, in these cases, many children are included who do not have a clear risk of future antisocial behavior. As developmental and life course research has shown, there is only a small group of relatively persistently antisocial youngsters [24, 66]. Many “difficult” children at preschool age recover relatively quickly due to “natural” protective factors [67]. There is also a statistical “floor effect” when risk is low that makes further risk reduction difficult. In contrast, indicated programs are more focused and thus may lead to somewhat better effects. Of course, this finding does not contradict the use of very early prevention programs. For example, some of the most effective family-oriented programs intervened in early childhood and even prenatally [68]. Our finding suggests only that *child-oriented* programs can be effective at later ages as well (i.e., age-appropriate preventive interventions are never too early and never too late; [69]).

Although most programs are based on cognitive-behavioral concepts of social problem solving, self-control, perspective taking, thinking about alternatives to antisocial behavioral schemes, skills practicing, and other constructs that are relevant for antisocial development, there is much heterogeneity in the respective outcomes (see Table 3). This heterogeneity might be due to characteristics beyond the program content (see below). There is no significant difference in effect sizes between cognitive-behavioral approaches and programs based on psychodynamic or humanistic concepts. At first glance, this may seem somewhat surprising, because cognitive-behavioral approaches are widely seen as most appropriate [3, 38]. However, in recent years, the differences between theoretical concepts seem to have become less important, and modern interventions are more eclectic and problem-oriented. As Beelmann and Eisner [70] have shown for parent training programs, differences in effect size between different theoretical concepts disappear after controlling for the methodological characteristics of evaluations. In addition, it should be borne in mind that beyond the content of the program, the client–trainer relationship and the implementation quality also play an important role (e.g., [71, 72]).

Other moderator effects of our study suggest that programs implemented by program authors or university staff have a somewhat larger effect. This may be due in part to the above-mentioned slight influence of a publication bias. However, more generally, model or demonstration projects implemented by researchers often show larger effects than

evaluations of either routine practice or evaluations by independent researchers (e.g., [5, 10]). Better supervision of the implementation and staff engagement in model projects may explain this finding [11].

Another significant moderator is the proportion of male youngsters in the samples. Despite some heterogeneity in the respective findings, the overall tendency is for larger effects in studies treating mainly males. This is in accordance with the above-mentioned relevance of the aggression and delinquency baseline. Typically, boys have a higher risk of antisocial behavior in childhood and youth, whereas girls exhibit more internalizing problems such as anxious, depressive, and eating symptoms in later adolescence. However, the gender difference in our findings may also be due in part to contents of programs being more suitable for the larger group of males. Although basic risk factors of antisocial development are similar for males and females, there are also some differences [73, 74].

The data source of the outcome assessment is another moderator of effect size in our meta-analysis. Behavior observations and official records of aggression, delinquency, and other forms of antisocial behavior show larger effects than ratings by parents, teachers, and others. The only low correlations between ratings of child behavior problems by different informants are well known [75]. Cross-sectional correlations of ratings by different informants are even lower than ratings by the same informants in longitudinal studies [76]. These findings indicate that informants may have developed a stereotype of a child and her or his behavior that remains relatively stable over time and is therefore less sensitive when it comes to measuring change. Accordingly, one of our studies on developmental prevention shows only desirable effects when different teachers rate child behavior before and after the intervention [77].

Most programs were delivered in a group format and this has a significant mean effect. However, as a tendency, individual formats seem to score somewhat better, potentially suggesting more individual adaptations of program content and delivery. Adaptation and individualization are currently much discussed topics in prevention research (e.g., [78, 79]). One conclusion of these discussions is that individual and cultural adaptations of program content could have some benefits as long as they do not violate a program's theoretical foundation and principles. For example, it makes sense to adapt exercises for social problem solving to those situations that are highly relevant for the individual child or adolescent, but such changes should not modify the principles of social problem-solving.

Implementation quality is highly important for effectiveness in delinquency prevention programs [72] and in prevention research in general [80], but unfortunately many of the studies in our review reported no details on this. Perhaps, as mentioned above, differences in the therapeutic alliance also play a role in the outcome variance, but most studies did not provide details on this issue. In addition, program intensity is not a significant moderator, although very short programs have a nonsignificant effect on size. We investigated this issue in more detail and checked whether the overall duration of the program or the number of sessions per month had a specific effect, but this was not the case. More information on program delivery is necessary to enable a sound interpretation. Preliminarily, we conclude that very short training programs are probably unsuccessful, but beyond this, the intensity of a child skills training does not seem to be a core factor for success. This may also be due to problems of early dropout especially for very long or intensive programs.

## Limitations

Our review contains encouraging results on the effectiveness of child social skills training on various forms of antisocial development. However, overall effects are small. This is a common finding in developmental prevention for different types of programs (e.g., [2, 3]) and needs to be interpreted within a realistic framework of what single interventions at only one time in development can achieve in prevention.

Although small effects have practical and financial relevance, this review shows that there is a clear lack of long-term evaluations of child social skills training. Insofar, our principally sound basis of integrating many RCTs has its limits. Evaluations suggest that there is some promise of success, but we do not know whether programs affect long-term antisocial development. Therefore, our partially positive message should not be interpreted as proof that child skills' training is an effective measure for preventing aggression and delinquency *in the long term*. As even long-term correlational predictions of antisocial development have limited validity, our findings on prevention are plausible within the broader context of developmental and life course criminology.

When feasible, child skills training should be combined with family-oriented and other interventions (although these are more difficult to implement in routine practice). Moreover, child skills training should not be limited to one time in development but refreshed and intensified at later ages. With regard to combined multimodal programs, it is often difficult to disentangle the most relevant components for success or failure. Our meta-analysis is unable to give guidance on this important issue, because it focused on child skills training alone.

Although our moderator analyses provide various significant and theoretically meaningful results, we need to emphasize that these data do not differentiate between short- and long-term effects. To achieve a sufficiently large number of primary studies for differential analyses, we had to include both immediate and follow-up outcomes. It is a general problem of meta-analyses that differential effects can be based on only a small number of studies containing the respective subgroups [41].

Our meta-analysis is also limited insofar as it analyzed studies only in the English and German languages. However, our pilot search procedures did not reveal eligible reports in other languages. This may have been due to the high threshold of an RCT design. Perhaps the focus on RCTs may be not only a strength but also a limit of our study. Unbiased RCTs provide high internal validity, but they are used less frequently in evaluations of routine practice. Some general research in criminology suggests that RCTs have smaller effects than (sound) quasi-experiments [56]. However, this has yet to be proven for person-oriented measures [10, 37]. We suggest investigating this issue in more detail and comparing RCTs and sound, practice-oriented quasi-experiments on child social skills training.

## Conclusions

Our meta-analysis on child social skills training as a measure of preventing aggression, delinquency, and other antisocial development in children and youth reveals favorable mean effects. However, most of these effects are rather small, and findings depend on various moderators. This probably indicates that programs may be strengthened by

applying more differentiated concepts to specific target groups [81]. Our study shows that indicated prevention for children who already have some behavior problems is most promising. Therefore, indicated approaches should be implemented when feasible, but more easily implemented universal programs could be used as a first step. Although the substantial number of RCTs is a clear strength in this field, the lack of long-term follow-ups is a weakness. In comparison with the many correlational long-term studies in developmental and life course criminology, there is a clear imbalance when it comes to long-term intervention studies. This is relevant not only for practical aims of prevention but also because sound (quasi-) experimental studies are important for cross-validating correlational findings. This shows the need to link these two strands of developmental and life course criminology more closely together [82, 83].

**Acknowledgements** Open Access funding provided by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Farrington, D. P., & Welsh, B. C. (2007). *Saving children from a life of crime*. Oxford: Oxford University Press.
2. Farrington, D. P., Gaffney, H., Lösel, F., & Tfofi, M. (2017). Systematic reviews of the effectiveness of developmental prevention programs in reducing delinquency, aggression, and bullying. *Aggression and Violent Behavior, 33*, 91–106. <https://doi.org/10.1016/j.avb.2016.11.003>.
3. Lösel, F. (2012). Entwicklungsbezogene prävention von gewalt und kKriminalität: ansätze und wirkungen [developmental prevention of violence and crime: approaches and effects]. *Forensische Psychiatrie, Psychologie und Kriminologie, 6*, 71–84. <https://doi.org/10.1007/s11757-012-0159-2>.
4. Lösel, F., & Bender, D. (2017). Parenting and family-oriented programs for the prevention of child behavior problems: what the evidence tells us. *Journal of Family Research, 11*(SI), 217–239.
5. Beelmann, A., & Raabe, T. (2009). The effects of preventing antisocial behavior and crime in childhood and adolescence. Results and implications of research reviews and meta-analyses. *European Journal of Developmental Science, 3*, 260–281. <https://doi.org/10.3233/DEV-2009-3305>.
6. O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people. Progress and possibilities*. Washington: The National Press Academy.
7. Sandler, I., Wolchik, S. A., Cruden, G., Mahrer, N. E., Ahn, S., Brincks, A., & Brown, C. H. (2014). Overview of meta-analyses of the prevention of mental health, substance use, and conduct problems. *Annual Review of Clinical Psychology, 10*, 243–273. <https://doi.org/10.1146/annurev-clinpsy-050212-185524>.
8. Weiss, M., Schmucker, M., & Lösel, F. (2015). Meta-analyse zur wirkung familienbezogener präventionsmaßnahmen in Deutschland [meta-analysis of the effects of family-based prevention programs in Germany]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 44*, 27–44. <https://doi.org/10.1026/1616-3443/a000298>.
9. Farrington, D. P., & Welsh, B. C. (2013). Randomized experiments in criminology: what has been learned from long-term follow-ups? In B. C. Welsh, A. A. Braga, & G. J. N. Bruinsma (Eds.), *Experimental criminology: prospects for advancing science and public policy* (pp. 111–140). New York: Cambridge University Press.



10. Lösel, F. (2018). Evidence comes by replication, but needs differentiation: the reproducibility issue in science and its relevance for criminology. *Journal of Experimental Criminology*, *14*, 257–278. <https://doi.org/10.1007/s11292-017-9297-z>.
11. Mihalic, S. F., & Elliott, D. S. (2015). Evidence-based programs registry: blueprints for healthy youth development. *Evaluation and Program Planning*, *48*, 124–131. <https://doi.org/10.1016/j.evalprogplan.2014.08.004>.
12. Beelmann, A., & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. *Psicothema*, *18*, 602–609.
13. Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: a systematic review of randomized evaluation. *Annals of the American Academy of Political and Social Science*, *587*, 84–109. <https://doi.org/10.1177/0002716202250793>.
14. Dirks, M. A., Treat, T. A., & Weersing, V. A. (2007). Integrating theoretical, measurement, and intervention models of youth social competence. *Clinical Psychology Review*, *27*, 327–347.
15. Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gulotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. New York: Guilford.
16. Greenberg, M. T., & Kusché, C. A. (2006). Building social and emotional competence: The PATHS curriculum. In S. R. Jimerson & M. J. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 395–412). Mahwah: Lawrence Erlbaum.
17. Webster-Stratton, C. (2005). The incredible years: A training series for the prevention and treatment of conduct problems in young children. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders. Empirically based strategies for clinical practice* (2nd ed., pp. 507–556). Washington: American Psychological Association.
18. Lösel, F., & Bender, D. (2012). Child social skills training in the prevention of antisocial development and crime. Crime prevention and public policy. In B. C. Welsh & D. P. Farrington (Eds.), *The Oxford handbook of crime prevention* (pp. 102–129). New York: Oxford University Press.
19. Caldarella, P., & Merrell, K. W. (1997). Common dimensions of social skills of children and adolescents: A taxonomy of positive behaviors. *School Psychology Review*, *26*, 264–279.
20. Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health*, *63*(Suppl. 1), i37–i57.
21. Rose-Krasnor, L. (1997). The nature of social competence: a theoretical review. *Social Development*, *6*, 111–135. <https://doi.org/10.1111/1467-9507.00029>.
22. Jennings, W. G., & Reingle, J. M. (2012). On the number and shape of developmental/life-course violence, aggression, and delinquency trajectories: a state-of-the-art review. *Journal of Criminal Justice*, *40*, 472–489.
23. Laub, J. H., & Sampson, R. J. (2003). *Shared beginnings, divergent lives: delinquent boys to age 70*. Cambridge: Cambridge University Press.
24. Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review*, *100*, 674–701. <https://doi.org/10.1037/0033-295X.100.4.674>.
25. Garrido Genovés, V., Anyela Morales, L., & Sánchez-Meca, J. (2006). What works for serious juvenile offenders? A systematic review. *Psicothema*, *18*, 611–619.
26. Bender, D., & Lösel, F. (2011). Bullying at school as predictor of delinquency, violence and other antisocial behaviour in adulthood. *Criminal Behaviour and Mental Health*, *22*, 99–106. <https://doi.org/10.1002/cbm.799>.
27. Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent and serious delinquency in adolescence and early adulthood: a synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious & violent juvenile offenders. Risk factors and successful interventions* (pp. 86–105). Thousand Oaks: Sage.
28. Robins, L. N., & Price, R. K. (1991). Adult disorders predicted by childhood conduct problems: results from the NIMH epidemiologic catchment area project. *Psychiatry: Interpersonal and Biological Processes*, *54*, 116–132.
29. Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology*, *39*, 349–371. <https://doi.org/10.1037//0012-1649.39.2.349>.
30. Lösel, F., & Beelmann, A. (2005). Social problem solving programs for preventing antisocial behavior in children and youth. In M. McMurran & J. McGuire (Eds.), *Social problem solving and offenders: evidence, evaluation and evolution* (pp. 127–143). New York: Wiley.
31. Beelmann, A., Pfingsten, U., & Lösel, F. (1994). The effects of training social competence in children: a meta-analysis of recent evaluation studies. *Journal of Clinical Child Psychology*, *23*, 260–271.



32. Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>.
33. Ang, R. P., & Hughes, J. N. (2002). Differential benefits of skills training with antisocial youth based on group composition: a meta-analytic investigation. *School Psychology Review, 31*, 164–185.
34. Farrington, D. P., & Hawkins, J. D. (2019). The need for long-term follow-ups of delinquency prevention experiments. *JAMA Network Open, 2*(3), e190780. <https://doi.org/10.1001/jamanetworkopen.2019.0780>.
35. Koehler, J. A., Lösel, F., Humphreys, D. K., & Akoensi, T. D. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology, 9*, 19–43. <https://doi.org/10.1007/s11292-012-9159-7>.
36. Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: a synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: risk factors and successful interventions* (pp. 313–345). Thousand Oaks: Sage.
37. Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: a review of systematic reviews. *Annual Review of Law and Social Science, 3*, 297–320. <https://doi.org/10.1146/annurev.lawsocsci.3.081806.112833>.
38. Wilson, D. P. (2016). Correctional programs. In D. Weisburd, D. P. Farrington, & C. Gill (Eds.), *What works in crime prevention and rehabilitation: lessons from systematic reviews* (pp. 193–217). New York: Springer.
39. Piquero, A. R., Jennings, W. G., Farrington, D. P., Diamond, B., & Reingle Gonzales, J. M. (2016). A meta-analysis update on the effectiveness of early self-control improvement programs to improve self-control and reduce delinquency. *Journal of Experimental Criminology, 12*, 249–264. <https://doi.org/10.1007/s11292-016-9257-z>.
40. Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. *American Journal of Preventive Medicine, 33*, 130–143. <https://doi.org/10.1016/j.amepre.2007.04.011>.
41. Schmucker, M., & Lösel, F. (2012). Meta-analysis as a method of systematic reviews. In D. Gadd, S. Karstedt, & S. F. Messner (Eds.), *The sage handbook of criminological research methods* (pp. 425–443). Thousand Oaks: Sage.
42. Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage Publications.
43. Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364–386. <https://doi.org/10.1177/1094428106291059>.
44. Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
45. Wilson, D. B. (2010). *SPSS macros for performing meta-analytic analyses*. Retrieved from <http://mason.gmu.edu/~dwilsonb/ma.html>.
46. Mize, J., & Ladd, G. W. (1990). A cognitive-social learning approach to social skills training with low-status preschool children. *Developmental Psychology, 26*, 388–397. <https://doi.org/10.1037/0012-1649.26.3.388>.
47. Prinz, R. J., Blechman, E. A., & Dumas, J. E. (1994). An evaluation of peer coping-skills training for childhood aggression. *Journal of Clinical Child Psychology, 23*, 193–203. [https://doi.org/10.1207/s15374424jccp2302\\_8](https://doi.org/10.1207/s15374424jccp2302_8).
48. Block, J. (1978). Effects of a rational emotive mental health program on poorly achieving disruptive high school students. *Journal of Counseling Psychology, 25*, 61–65. <https://doi.org/10.1037/0022-0167.25.1.61>.
49. Etscheid, S. (1991). Reducing aggressive behavior and improving self-control: a cognitive behavioral training program for behaviorally disordered adolescents. *Behavioral Disorders, 16*, 107–115.
50. Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101. <https://doi.org/10.2307/2533446>.
51. Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology, 4*, 61–81. <https://doi.org/10.1007/s11292-007-9046-9>.
52. Lösel, F., Stemmler, M., & Bender, D. (2013). Long-term evaluation of a bimodal universal prevention program: effects on antisocial development from kindergarten to adolescence. *Journal of Experimental Criminology, 9*, 429–449. <https://doi.org/10.1007/s11292-013-9192-1>.
53. Malti, T., Noam, G. G., Beelmann, A., & Sommer, S. (2016). Toward dynamic adaptation of psychological interventions for child and adolescent development and mental health. *Journal of Clinical Child and Adolescent Psychology, 45*, 827–836. <https://doi.org/10.1080/15374416.2016.1239539>.
54. Schmucker, M., & Lösel, F. (2017). Sexual offender treatment for reducing recidivism among convicted sex offenders: a systematic review and meta-analysis. *Campbell Systematic Reviews, 13*, 1–75. <https://doi.org/10.4073/csr.2017.8>.

55. Tong, L. S. J., & Farrington, D. P. (2006). How effective is the reasoning and rehabilitation programme in reducing offending? A meta-analysis of evaluations in four countries. *Psychology, Crime and Law*, *12*, 3–24.
56. Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, *578*, 50–70.
57. Sundell, K., Hansson, K., Löfholm, C. A., Olsson, T., Gustle, L. H., & Kadesjö, C. (2008). The transportability of multisystemic therapy to Sweden: ahort-term results from a randomized trial of conduct-disordered youth. *Journal of Family Psychology*, *22*, 550–560. <https://doi.org/10.1037/a0012790>.
58. Farrington, D. P., Lösel, F., Braga, A. A., Mazerolle, L., Raine, A., Sherman, L. W., & Weisburd, D. (2019). Experimental criminology: Looking back and forward on the 20th anniversary of the academy of experimental criminology. *Journal of Experimental Criminology*. Published online December 4. <https://doi.org/10.1007/s11292-019-09384-z>.
59. Offord, D. R. (2000). Selection of levels of prevention. *Addictive Behaviors*, *25*, 833–842. [https://doi.org/10.1016/S0306-4603\(00\)00132-5](https://doi.org/10.1016/S0306-4603(00)00132-5).
60. Coid, J. W. (2003). Formulating strategies for the primary prevention of adult antisocial behaviour. “High risk” or “population” strategies? In D. P. Farrington & J. W. Coid (Eds.), *Early prevention of adult antisocial behavior* (pp. 32–78). Cambridge: Cambridge University Press.
61. Lösel, F., Beelmann, A., Jaurisch, S., & Stemmler, M. (2006). Prävention von problemen des sozialverhaltens im vorschulalter. Evaluation des Eltern- und Kindertrainings EFFEKT [prevention of problem behavior in preschool age: Evaluation of the parent and child training EFFEKT]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *35*, 127–139. <https://doi.org/10.1026/1616-3443.35.2.127>.
62. Conduct Problems Prevention Research Group. (2010). Fast track intervention effects on youth arrests and delinquency. *Journal of Experimental Criminology*, *6*, 131–157. <https://doi.org/10.1007/s11292-010-9091-7>.
63. Hawkins, J. D., Brown, E. C., Oesterle, S., Arthur, M. W., Abbott, R. D., & Catalano, R. F. (2008). Early effects of communities that care on targeted risks and initiation of delinquent behavior and substance use. *Journal of Adolescent Health*, *43*, 15–22. <https://doi.org/10.1016/j.jadohealth.2008.01.022>.
64. Homel, R., Kate Freiberg, K., & Sara Branch, S. (2015). CREATE-ing capacity to take developmental crime prevention to scale: A community-based approach within a national framework. *Australian and New Zealand Journal of Criminology*, *48*, 367–385. <https://doi.org/10.1177/0004865815589826>.
65. Homel, R., & McGee, T. R. (2012). Community approaches to crime and violence prevention: Building prevention capacity. In R. Loeber & B. Welsh (Eds.), *The future of criminology* (pp. 172–177). New York: Oxford University Press.
66. Stemmler, M., & Lösel, F. (2015). Developmental pathways of externalizing behavior from preschool age to adolescence: an application of general growth mixture modeling. In M. Stemmler, A. von Eye, & W. Wiedermann (Eds.), *Dependent data in social sciences research: Forms, issues, and methods of analysis* (pp. 91–106). New York: Springer.
67. Lösel, F., & Bender, D. (2003). Resilience and protective factors. In D. P. Farrington & J. W. Coid (Eds.), *Early prevention of adult antisocial behaviour* (pp. 130–204). Cambridge: Cambridge University Press.
68. Olds, D. L., Kitzman, H. J., Cole, R. E., Hanks, C. A., Arcoletto, K. J., Anson, E. A., Luckey, D. W., Knudtson, M. D., Henderson Jr., C. R., Bondy, J., & Stevenson, A. J. (2010). Enduring effects of prenatal and infancy home visiting by nurses on maternal life course and government spending: Follow-up of a randomized trial among children at age 12 years. *Archives of Pediatrics and Adolescent Medicine*, *164*, 419–424. <https://doi.org/10.1001/archpediatrics.2010.49>.
69. Lösel, F. (2007). It’s never too early and never too late: Towards an integrated science of developmental intervention in criminology. *Criminologist*, *35*(2), 1–8.
70. Beelmann, A., & Eisner, M. (2018). Parent training programs for preventing and treating antisocial behavior in childhood and adolescence: a meta-analysis. Presentation at the 74th annual meeting of the American Society of Criminology, November 14–17, Atlanta, GA.
71. Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work*, *19*, 531–540. <https://doi.org/10.1177/1049731509335549>.
72. Lipsey, M. W. (2018). Effective use of the large body of research on the effectiveness of programs for juvenile offenders and the failure of the model programs approach. *Criminology & Public Policy*, *17*, 1–10. <https://doi.org/10.1111/1745-9133.12345>.
73. Lösel, F., & Stemmler, M. (2012). Continuity and patterns of externalizing and internalizing behavior problems in girls: a variable- and person-oriented study from preschool to youth age. *Psychological Test and Assessment Modeling*, *54*, 307–319.

74. Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behavior: conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. New York: Academic Press. <https://doi.org/10.1017/CBO9780511490057>.
75. Achenbach, T. M. (2006). As others see us: clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science*, *15*, 94–98. <https://doi.org/10.1111/j.0963-7214.2006.00414.x>.
76. Lösel, F. (2002). Risk/need assessment and prevention of antisocial development in young people: basic issues from a perspective of cautionary optimism. In R. Corrado, R. Roesch, S. D. Hart, & J. Gierowski (Eds.), *Multiproblem violent youth: a foundation for comparative research needs, interventions and outcomes* (pp. 35–57). Amsterdam: IOS/NATO Book Series.
77. Hacker, S., Lösel, F., Stemmler, M., Jaurisch, S., Runkel, D., & Beelmann, A. (2007). Training im problemlösen (TIP): implementation und evaluation eines sozial-kognitiven kompetenztrainings für kinder [training in problem solving: implementation and evaluation of a social-cognitive competence training for children]. *Heilpädagogische Forschung*, *23*, 11–21.
78. Beelmann, A., Malti, T., Noam, G., & Sommer, S. (2018). Innovation and integrity: Desiderata and future directions for prevention and intervention science. *Prevention Science*, *19*, 358–365. <https://doi.org/10.1007/s11121-018-0869-6>.
79. Sundell, K., Beelmann, A., Hasson, H., & von Thiele Schwarz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child and Adolescent Psychology*, *45*, 784–796. <https://doi.org/10.1080/15374416.2015.1020540>.
80. Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350.
81. Beelmann, A. (2011). The scientific foundation of prevention. The status quo and future challenges of developmental crime prevention. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behavior and crime. Contributions of developmental and evaluation research to prevention and intervention* (pp. 137–164). Cambridge: Hogrefe Publishing.
82. Farrington, D. P., Ohlin, L. E., & Wilson, J. Q. (1986). *Understanding and controlling crime: toward a new research strategy*. New York: Springer.
83. Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, *43*(2S1), 8–23. <https://doi.org/10.1016/j.amepre.2012.04.029>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Andreas Beelmann<sup>1</sup> · Friedrich Lösel<sup>2,3</sup>

✉ Andreas Beelmann  
andreas.beelmann@uni-jena.de

Friedrich Lösel  
fal23@cam.ac.uk; friedrich.loesel@fau.de

<sup>1</sup> Department of Research Synthesis, Intervention and Evaluation, Institute of Psychology, Friedrich-Schiller-University, Jena, Germany

<sup>2</sup> University of Cambridge, Cambridge, UK

<sup>3</sup> University of Erlangen-Nuremberg, Erlangen, Germany