Aggression and Violent Behavior 27 (2016) 30–41

Contents lists available at ScienceDirect

Aggression and Violent Behavior

Aggression replacement training (ART) for reducing antisocial behavior in adolescents and adults: A systematic review

Lars Brännström [a,*], Catrine Kaunitz [b], Anna-Karin Andershed [c], Sandra South [d], Geir Smedslund [e]

[a] Department of Social Work, Stockholm University, Sweden
[b] National Board of Institutional Care, Sweden
[c] School of Law, Psychology and Social Work, Örebro University, Sweden
[d] National Board of Health and Welfare, Sweden
[e] Knowledge Centre for the Health Services, Norway

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form 2 September 2015
Accepted 4 February 2016
Available online 12 February 2016

Keywords:
Intervention
Effectiveness
Aggression Replacement Training
ART
Delinquency

ABSTRACT

Objective: Aggression Replacement Training (ART) is a multimodal program aiming at replacing antisocial behaviors by actively teaching desirable behaviors. The program is frequently used and has been provided within a wide variety of settings, but its effectiveness in its own right has not been addressed in previous reviews. This systematic review examines the effect of ART on antisocial behavior in young people and adults.
Methods: Published and unpublished literature was searched to identify randomized and non-randomized studies comparing ART for adults and youth with usual care, other interventions, or no intervention. Primary outcomes included recidivism in antisocial behavior, while secondary outcomes were related to social skills, anger management and moral reasoning.
Findings: This review identified 16 studies with considerable clinical and methodological diversity. The methodological quality and the post-intervention follow-up of the studies were limited. Almost half of the studies were conducted by researchers who have vested interests in the intervention.
Conclusions: There is an insufficient evidence-base to substantiate the hypothesis that ART has a positive impact on recidivism, self-control, social skills or moral development in adolescents and adults. Further research is warranted by independent investigators exploring the effects of ART on clearly-defined target groups using high standard evaluation designs.

Contents

1. Introduction . . . 31
1.1. Aggression Replacement Training . . . 31
1.2. Previous research on the effectiveness of ART . . . 31
1.3. A systematic review of ART . . . 31
2. Methods . . . 32
2.1. Types of studies, participants and outcomes . . . 32
2.2. Search strategies, data collection and analysis . . . 32
3. Results . . . 32
3.1. Description of the included studies . . . 33
3.2. Risk of bias assessment . . . 33
3.3. Effects of the intervention . . . 37
3.3.1. Effects on recidivism . . . 37
3.3.2. Effects on anger control, social skills and moral reasoning . . . 37
4. Discussion . . . 39
4.1. Strengths and limitations . . . 40

* Corresponding author at: Department of Social Work, Stockholm University, SE-10691 Stockholm, Sweden. Tel.: +46 8 16 48 59.
E-mail address: lars.brannstrom@socarb.su.se (L. Brännström).

http://dx.doi.org/10.1016/j.avb.2016.02.006

4.2. Conclusions and recommendations for future research . . . 40
References . . . 40

1. Introduction

Antisocial behavior manifests itself in many different forms. Behaviors often referred to include aggressive and violent behavior, property violations or offenses, deceitful behavior, rule violations, and substance use/abuse (American Psychiatric Association, 2000). Much attention has been paid to the importance of self-control in regulating antisocial, delinquent and criminal behavior. This has included the evolution of several techniques and programs intended to improve self-control among children and adolescents (Piquero, Wesley, Jennings, & Farrington, 2010). Aggression Replacement Training (ART) is an example of such a program.

The ART-program is a multimodal program originally developed for aggressive delinquents in residential care in New York, USA, with the aim of replacing antisocial behaviors by actively teaching desirable behaviors. It is a structured program that combines the use of techniques from cognitive therapy (based on cognitive theories) and behavioral therapy (from learning theory). As ART has become a commonly used intervention for both youths and adults with antisocial behavior, it is increasingly important to systematically ascertain the program's effectiveness. Empirical studies have been conducted, and ART has been included as a cognitive behavioral intervention among others in previous systematic reviews (e.g.Landenberger & Lipsey, 2005; Lipsey, Chapman, & Landenberger, 2001; Lipsey, Landenberger, & Wilson, 2007). However, there has been no systematic review of ART, in its own right, even though such information would perhaps best advise decision makers and practitioners choosing interventions for their clients. Therefore, the purpose of this article is to provide a systematic review of studies of the effect of ART on antisocial behavior in young people and adults.

1.1. Aggression Replacement Training

The original manual states that ART is a 10-week, 30-hour intervention administered to groups of 8 to 12 youths three times a week (Goldstein, Glick, & Reiner, 1987). According to the developers, aggression has a behavioral component, an affective component, and a 'values' component. Thus, the main components of ART came to include skill streaming to teach pro-social behavior (behavioral component), anger control (affective component), and moral reasoning (cognitive component). This paper focuses on the original version of ART, in order to summarize research on the program as it was intended, and not on versions of the program that have been subject to "program drift" where "original standards have /…/ been compromised or diluted to suit the needs of the system using the program" (Glick & Gibbs, 2011, p. 2).

The behavioral component of ART consists of social skills training, a technique for teaching pro-social behavior to participants who have deficiency in or who lack these competencies, which is theoretically grounded in social learning theory and the work of Bandura (1973). The teaching of skills serves to replace the out-of-control destructive behaviors with constructive pro-social behavior. The anger control training component, building on the work of Novaco (1975) and Meichenbaum (1977), is designed to help reduce the frequency of anger arousal in those who are chronically aggressive, and to provide means of self-control when anger is aroused. While skill streaming is designed to teach what one should do in problematic situations, anger control training teaches what one should not do. Moral reasoning training is the third component of ART (Gibbs, Potter, & Goldstein, 1995) and has its foundation in Kohlberg's (1973) model of moral development. The primary purpose is to raise the individual's level of moral reasoning, thus enabling them to make more mature decisions in social situations. In ART, moral reasoning is promoted through group discussions of moral dilemmas, which are termed 'social decision-making meetings'.

Since ART was developed in the 1980s, the original program has been modified and applied for other populations, settings and outcomes.(Goldstein, Nensén, Daleflod, & Kalt, 2004), for example, adult violent offenders (Lipton, Thornton, McGuire, Porporino, & Hollin, 2000), EQUIP (Leeman, Gibbs, & Fuller, 1993), The Prepare Curriculum (Goldstein, 1999), The Peace Curriculum: expanded Aggression Replacement Training (Salmon, 2004), Family-ties (Calame & Parker, 2012), and Aggression Control Therapy (Hornsveld, Van Dam-Baggen, Leenaars, & Jonkers, 2004). The program has been provided across North America and Europe since the 1990s within a wide variety of social, educational, and correctional services, secure units, community services and prisons. People from many different professions have been educated as trainers, including teachers, counselors, youth-care workers, social workers, and correctional officers.

1.2. Previous research on the effectiveness of ART

Results from empirical studies have lent support for the effectiveness of the intervention (Goldstein, Glick, Irwin, Pask-McCartney, & Rubama, 1989; Goldstein et al., 1987). Also, as previously mentioned, ART-trials have been included in meta-analytic reviews of effects of a wider array of interventions with juvenile offenders (Landenberger & Lipsey, 2005; Lipsey et al., 2001, 2007) which have indicated positive results for the ART-program. ART has fallen under the broader category of cognitive-behavioral interventions, and results suggest that these types of interventions are among the more promising rehabilitative treatments for antisocial behavior in youth (Landenberger & Lipsey, 2005; Lipsey et al., 2001, 2007; Pearson, Lipton, Cleland, & Yee, 2002). The review authors have generally been positive as to ART's effectiveness, concluding, for example, that ART is an example of the kind of program "that policy makers should review for possible adoption" (Pearson et al., 2002, p. 493). None of these reviews report on the design and methodological quality separately for each included study, however.

Results of ART outcome studies have moreover been summarized in non-systematic reviews, several of which have suggested ART to be a promising empirically-based treatment for juvenile offenders (Howell, 1998; Springer, McNeece, & Arnold, 2003; Welsh & Hoshi., 2002). Other reviews also conclude that ART is effective (Cigno & Bourns, 1998; Sherman, Farrington, & MacKenzie, 2002). The United States Department of Justice claim in their Model Program Guide that ART is an effective program ("OJJDP Model Programs Guide, n.d.").

1.3. A systematic review of ART

ART seems to be a successful program in the sense that its use is wide-spread, well-known and frequently used for the purpose of turning the lives around for individuals with antisocial behavior. As such, it has claimed large investments in effort, time, and money from communities, professionals, families and participants. To our knowledge, however, no previous review has addressed the ART program specifically and none have reported on the study design or methodological quality separately for each included trial. Furthermore, none of the reviews to date has included studies of ART for adult participants. A systematic review of the ART intervention as a separate program in its own right is therefore needed and will be of interest to researchers, policymakers and practitioners.

This systematic review assesses the effect of ART on antisocial behavior in young people and adults in medium to high security correctional facilities, other residential facilities, schools/educational centers and other community-based services. We describe inclusion criteria and search strategies, and report characteristics, methodological quality and estimated treatment effects of each included study. A summary of the main findings and limitations of this review will be discussed alongside implications for practice and research.

2. Methods

2.1. Types of studies, participants and outcomes

We included studies that labeled the program under evaluation as "Aggression Replacement Training" (ART). Further, the study needed to state clearly that the evaluated program only included the main, original components of ART, (ii) include a reference to Goldstein, and (iii) include a statement that the core principles in the program were being adhered to. The justification for applying these restrictions are that the program developers did not want ART to be copyrighted, and instead encouraged the extension and modification of the program to other settings, populations and outcomes (Goldstein et al., 2004). However, the three original core components constituting ART were copyrighted by the co-developer Barry Glick in 2010, and later by the Swedish and Dutch partners in the form of a European Union trademark. Thus, in line with the copyright and in order to facilitate comparability of the included studies, this review does not include extended versions of ART.

We included studies of ART where individuals or groups (clusters) were randomly assigned to the different arms of the trial (that is, to ART or a control condition), as well as non-randomized studies with a control condition. Only ART programs containing all three original components were considered. Studies that compared ART with ART plus additional components/treatments were excluded. All follow-up durations reported in the primary studies were recorded. Both standardized and unstandardized measures were acceptable if documented psychometric properties were reported or available.

Participants were males and females (12 years and older) who had displayed a level of antisocial behavior sufficient for authorities or researchers to consider them eligible for the ART program (i.e., that they would benefit from participating in ART). Participants in residential- and correctional facilities (including prison, secure and open settings), as well as in community settings, were included. Voluntary, mandated and sentenced participants were also included.

Any assessment of recidivism in antisocial behavior (e.g., criminal behavior) measured in the studies was considered as a primary outcome. This information was collected either from registers, interviews with case workers, or through self-/parent-/teacher reports. Three secondary outcomes were selected, based on the mediators hypothesized by Goldstein and colleagues to be affected by the three ART components: (i) anger control (operationalized by Goldstein et al. (2004, p. 230) as self-control and a decrease in acting-out behaviors), (ii) social skills, and (iii) moral reasoning. The secondary outcomes were assessed through self-/parent-/teacher report questionnaires (e.g., Child Behavior Checklist [Achenbach, 1991]) or staff observations (e.g., Behavior Incident Report). Measures based on non-validated or non-standardized tests were only considered if information on psychometric properties were available. All included studies were examined for evidence of adverse events, measured as incidence of overall adverse events such as statistically significant negative effects of ART on the primary and secondary outcomes measured in the included studies.

The follow-up duration reported in the primary studies were recorded. The pre-post design assessments are considered a short-term follow-up, assessments from four to six months up to 12 months are considered a medium-term follow-up, and assessments one year post intervention are considered as having a long-term follow up.

2.2. Search strategies, data collection and analysis

Search strategies were developed by using various terms for aggression replacement, such as aggression control therapy, aggression prevention, positive peer culture, equipping youth to help one another, EQUIP program, Prepare curriculum, PEACE curriculum, Family ART, and the like (detailed search strategies are available upon request). To identify unpublished reports and on-going studies, ART-developers and independent investigators were contacted and reference lists of included studies and all obtained reviews were scanned for new leads. The following electronic reference databases, government databanks and professional websites were searched: ASSIA, The Cochrane Library (including CDSR, DARE, TRIALS, and HTA), The Campbell Library, Criminal Justice Abstracts, ERIC, PubMed, PsycINFO, Sociological Abstracts, and Social Services Abstracts. Additional searches were made using Google and Google Scholar. There were no restrictions regarding language or date of publication.

A minimum of two review authors independently screened titles and abstracts. Selection of primary studies was made according to criteria described above. Reasons for exclusion were documented for each study retrieved in full text and are available upon request. Guided by the checklist of items to consider in data collection and data extraction detailed in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2008), at least two independent coders extracted data. Attempts were made to contact the corresponding author of the study in question if the full text report contained insufficient information for a decision to be reached.

The methodological quality, here conceptualized as risk of bias in the included studies, was assessed independently by at least two reviewers on the basis of the revised CONSORT statement and the corresponding checklist for randomized controlled trials (Schulz, Altman, & Moher, 2010) as well as the Cochrane Handbook (Higgins & Green, 2008). Included studies were assessed based on whether or not they had adequately addressed sequence generation, allocation concealment, blinding of outcome assessors, incomplete outcome data, selective outcome reporting, and other sources of bias. Given the nature of the method, we assessed quality of blinding primarily by whether those who assessed and coded outcome measures were blind as to which arm of the trial the individual had been assigned to (i.e., ART intervention or control condition).

Since studies using non-randomized designs were included, methodological quality was assessed with an extension of the Cochrane risk of bias tool. Non-randomized studies require particular attention to selection bias and to confounding by examining various characteristics at baseline, as well as whether or not there were differences between arms of the control trial in these characteristics and the attempts made to control for any differences.

Internet searches were made for each of the included study investigators in attempt to assess the independence of the evaluator. To be regarded as independent, an evaluator cannot have vested interests in the intervention (e.g., financial or psychological as a developer or program proponent).

3. Results

Following an initial identification of 749 papers through the electronic databases and knowledge of existing papers, two of the review authors independently screened titles and abstracts for relevance according to the inclusion criteria. This process reduced the number of potentially relevant publications to 28. The full text screening identified 16 unique studies (reported in 18 publications) meeting the criteria for inclusion (Fig. 1). Most of the excluded studies (5 out of 12) did not meet the inclusion criteria of evaluating ART as a single program (e.g. ART was combined with other components/interventions). Other common reasons for exclusion was wrong population (participants were younger than 12 years) and/or no comparison group.

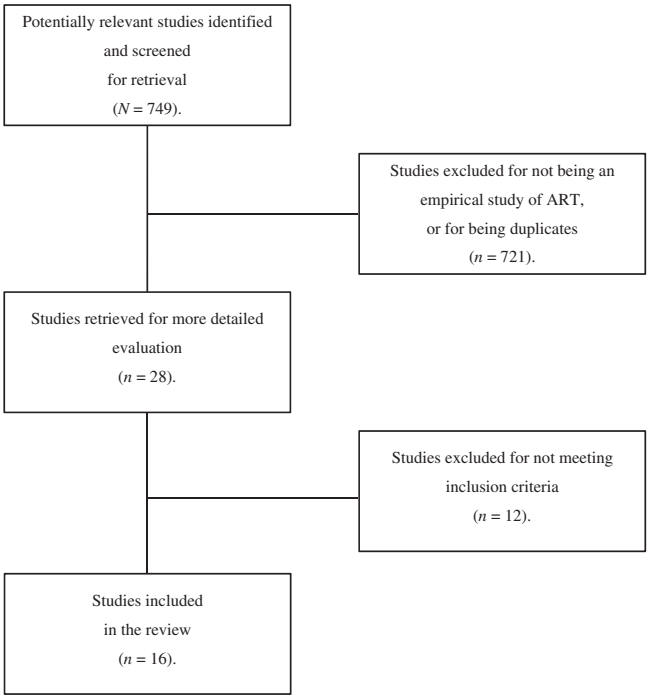

Fig. 1. Flow of literature through the review.

3.1. Description of the included studies

Table 1 summarizes the characteristics of the 16 included studies. The studies were divided into those that the investigators reported as including "youths" and those that included "adults", even though it was noted that there was an overlap in age for these two groups, and in some studies also limited information on the age of the participants. One study (Koposov, Gundersen, & Svartdal, 2014) reported the results for three age groups (6–9, 10–14 and 15 + years old). Since the inclusion criteria for this review stipulated participants 12 years and older we included only the oldest age group in the review. Moreover, since data from two studies were reported in multiple publications (Glick & Goldstein, 1987; Goldstein & Glick, 1994), we hereafter refer to these as Glick (1987) and Goldstein (1987) respectively. The study reported in Goldstein, Glick, Carthan, and Blancero (1994), Goldstein et al. (2004) is referred to as Goldstein (1994).

Published between 1987 and 2014, 12 studies focused primarily on youths and four on adults. Of the 16 studies, six were reported as controlled trials with random allocation of which four were cluster randomized trials. One of the included studies was reported as a quasi-randomized controlled trial and six were reported as non-randomized studies. In three studies it was not clear how individuals were allocated to different arms.

Sample size varied considerably, where two trials included over a thousand participants while the remaining 14 studies had sample sizes ranging from 18 to 151. Given the relatively small sample size of some of the studies, it is plausible that they did not have sufficient power to detect an effect of the intervention. Eight studies included only male participants, one study only females and seven studies included participants of both sexes. The trials were conducted in a wide variety of settings including community-based services, medium to high security correctional facilities, and schools/educational services. Most studies were conducted in the US.

Of the six studies that reported the primary outcome, three measured recidivism after program completion, two reported recidivism data both during the intervention and a set period after program completion, and the follow-up period varied between individuals in one study. The 12 studies that reported secondary outcomes did so over different time periods. Five reported data from standardized measurements administered weekly; three of these included in their pre- and post-treatment averages data from the period during which participants received ART. The time of measurement of other secondary outcomes was reported in only two studies, while unclear in eight of the 12 studies reporting other secondary outcome measures where general statements such as "pre-post basis" were used.

A history of aggressive, violent or criminal behavior was reported for participants in three studies (Barnoski, 2004; Erickson, 2013; Jones, 1991). Further, in two studies it was stated that a subset of the participants had such a history (Danielsson, Fors, & Freij, 2011; Hatcher et al., 2008). In eleven of the 16 included studies, it was unclear how many of the participants had a history of aggressive-, violent- and/or criminal behavior although the general statements made by the investigators suggest that at least some participants had such a history. A history of drug/alcohol abuse was reported in four studies (Barnoski, 2004; Coleman, Pfeiffer, & Oakland (1992); Danielsson et al., 2011; Erickson, 2013). The remaining twelve studies did not provide any information on the drug/alcohol abuse histories of the participants. One of the 16 included studies reported in detail whether participants also received other interventions during the trial period (Nodarse, 1998).

3.2. Risk of bias assessment

When properly implemented, RCTs typically prevent selection bias. Cluster-randomized and non-randomized trials present unique methodological challenges which must be addressed in order to effectively prevent selection bias, and therefore are often have a higher risk of bias than studies which randomized individuals to arms of the trial (Higgins & Green, 2008). Table 2 clearly shows that the included studies generally had either a high or unclear risk of bias on the majority of the domains examined.

The two RCTs, which had the potential to provide the most reliable results with respect to internal validity, had several serious other issues. Both RCTs had high or unclear selection bias and a high risk of other sources of bias (Coleman et al., 1992; Jones, 1991). One of the important other sources of bias was the potential for chance imbalance between the arms of the trial due to the small number of participants. Thus, the results of the RCTs included in this review must be interpreted with caution and attention must be paid to the potentially high risk of bias in these studies.

There were four studies that were reported as being cluster-RCTs, all of which included youths. Three of the cluster-RCTs had an unclear risk of selection bias (Glick, 1987, 1994; Zimmerman, 1987). The remaining cluster-RCT had a high risk of selection bias due to sequence generation as it is reported that clusters were indeed not truly randomized (Gundersen & Svartdal, 2006). All of the cluster-RCTs were judged as having a high risk of bias due to the evaluators being involved in the development of the program and/or having financial interests in the results. Importantly, all of the cluster-RCTs were judged to have a high risk of bias as a small number of clusters were randomized and because none of the studies undertook statistical control for the effects of clustering.

The examination of risk of bias due to confounding showed that only one study (Gundersen & Svartdal, 2006) reported a balance between arms of the trial for the potential confounder previous aggressive or violent behavior and also included this pre-measure as a covariate in the analyses exploring the effect of ART. It could therefore be argued that, although still at high risk of bias on multiple domains including selection bias, the results of Gundersen and Svartdal (2006) are potentially more reliable than those of the other three cluster-RCTs (Glick, 1987; Goldstein, 1994; Zimmerman, 1987).

Seven non-randomized controlled trials are included in this review, three including youths (Barnoski, 2004; Erickson, 2013; Koposov et al., 2014) and the remaining four including adults (Barto Lynch, 1995;

Table 1
Characteristics of included studies by study design and population.

Study ID, author (year)	Participants and setting	Interventions	Outcomes and measures	Follow-up	Results/authors' conclusion
Randomized controlled trials, youths					
1. Coleman et al. (1992)	*Participants*: adolescents with behavioral disorders ($N = 52$) in residential treatment. Information reported for $n = 39$. *Sex*: 10 females, 29 males. *Age*: $M = 15$ years 9 months (range: 13 years 2 months–18 years 11 months). *Setting*: the Devereux Foundation, Victoria, Texas. Residential treatment, USA.	*ART* ($n = 36$). *Control* ($n = 16$): Participated in regular seventh period classes, which were either academic or vocational in small group settings	*Anger control, social skills, moral reasoning.* Behavior Incident Report (BIR), Kendall–Wilcox Self-Control Scale (KWS), The Skill checklist, Socio-moral Reflection Measure (SRM)	Short term	No significant differences between experiment and control. No analyses or data reported.
2. Jones (1991)	*Participants*: high school students with a high degree of aggression ($n = 18$). *Sex*: 9 males, 9 females. *Age*: $M = 13$ years 9 months *Setting*: a Brisbane suburban high school, Australia.	*ART* ($n = 6$): Moral reasoning component only. ($n = 6$). *Control* ($n = 6$): no treatment condition.	*Anger control, moral reasoning.* BIR, KWS, SRM	Short term	Significant decrease in aggressive incidences in the ART condition (BIR), $F(1, 47) = 15.16$, $p = 0.0003$. One-way ANOVA of pre-test data showed that participants in the ART condition scored significantly higher in aggression, $F(2, 14) = 6.69$, $p = 0.009$. No difference between ART and control on KWS, $F(2, 14) = 2.48$, $p = 0.11$. All factors and interaction *n.s.*/not reported on the SRM.
Cluster-randomized controlled trials, youths					
3. Glick (1987)[a]	*Participants*: juvenile offenders ($N = 60$). *Sex*: male only. *Age*: $M = 15$ years, (range: 14–17 years). Setting: Annsville Youth Center, a New York State Division for Youth residential facility, USA.	Five units with 12 adolescents each, assigned to three conditions: *ART* ($n = 24$). *Control 1* ($n = 24$): participation in usual facility activities. *Control 2* ($n = 12$): no-treatment control group	*Anger control, moral reasoning.* BIR, KWS, SRM	Short term	ANCOVAs showed positive effect favoring ART on number of aggressive incidents ($F = 11.51$, $p < 0.01$), and intensity ($F = 9.34$, $p < 0.05$) (BIR), and KWS ($F = 10.67$, $p < 0.05$), No improvements on SRM ($F = 0.02$, *n.s.*).
4. Goldstein (1994)[b]	*Participants*: gang members ($N = 65$). *Sex*: mixed. No information on distribution. *Age*: not reported, "participating youths" (p. 94). Setting: two community-based agencies; Brooklyn, New York, USA.	*ART (6 gangs*, n = 38): 32 sessions, two sessions per week. *Control (6 gangs, n = 27)*: no-ART controls	*Anger control, social skills, recidivism.* Official data: rearrests, Anger Situation Inventory Scale (ASIS), The Skill checklist (SC)	Medium term for recidivism. Unclear for other outcomes.	Results favoring ART on re-arrest rates: RR = 0.25 (95% CI: 0.10–0.62). Stating results favoring ART on ASIS and SC but no data reported.
5. Gundersen and Svartdal (2006)	*Participants*: youths from local schools and institutions ($N = 65$). *Sex*: 48 males, 17 females *Age*: $M = 13$ years (range 11–17). *Setting*: schools and institutions in Norway.	*ART (8 groups, n = 47)*: school ($n = 26$), special school (n = 10) and institution ($n = 11$). *Control (3 groups, n = 18)*: school ($n = 5$), special school ($n = 10$) and institution ($n = 3$).	*Anger control, social skills, moral reasoning.* Social skills Rating System (SSRS), How I Think (HIT), Child and Adolescent Disruptive Behavior Inventory (CADBI), Child Behavior Checklist (CBCL)	Short term	*Parent ratings*: Improvement in ART condition compared to control condition (SSRS, ANOVA: $F(1, 44) = 5.33$, $p = 0.026$; CBCL, ANOVA: $F(1, 43) = 5.61$, $p = 0.02$). No significant interaction effect favoring ART (CADBI, ANOVA: $F(1, 38) = 3.27$, $p = 0.08$). *Teacher ratings*: No significant time or interaction effect on repeated measures (SSRS, CADBI, CBCL). Significant improvement in ART condition ($F(1, 51) = 11.07$, $p = 0.001$) but not in control condition: ($F(1, 51) = 1.7$, $p = 0.2$) assessed with SSRS. Reduction (improvement) in scores for both conditions, but contrast show only an effect in the ART condition $F(1, 14) = 8.11$, $p = 0.006$), assessed with CADBI. Contrast analysis showed improvement in scores for ART condition ($F(1, 47) = 13.49$, $p < 0.001$) but not for control ($F(1, 54) = 0.39$, $p = n.s$), assessed with CBCL. *Youth ratings*: No significant time or interaction effect on repeated measures (SSRS, HIT, CBCL). No significant change from pre-test to post test in ART condition ($F(1, 42) = 1.89$, $p = 0.176$) or in control condition ($F(1, 42) = 1.55$, $p = 0.22$), assessed with SSRS. Improvements in both conditions assessed with HIT (ART: $F(1, 54) = 9.58$, $p = 0.04$; control: $F(1, 54) = 7.35$, $p = 0.008$), and CBCL (ART, $F(1, 54) = 14.81$, $p < 0.001$; control, $F(1, 54) = 5.43$, $p = 0.002$).

6. Zimmerman (1987)	*Participants*: delinquent youth ($N = 36$). *Sex*: male only. *Age*: $M = 15.9$ years (range = 14–18). *Setting*: Annsville Youth Center, a New York state Division for Youth residential facility, USA.	Five units assigned to three conditions. *ART ($n = 21$)*. *Control 1 ($n = 10$)*: test instructions to enhance display of pro-socially skilled behaviors. *Control 2 ($n = 5$)*: no treatment.	*Anger control, moral reasoning*. BIR, KWS, SRM	Short term	The control condition had more severe acting-out behaviors on BIR, pre-test scores. ANOVA showed no significant effect of ART, $F(1, 34,) = 3.92$, *n. s.* ANCOVA with pre-test scores as covariate showed a significant positive effect of ART, $F(1, 33) = 11.504$, $p < 0.01$, on numbers of incidences, and $F(1, 33) = 9.342$, $p < 0.01$ (severity). Significant positive effect of ART on KWS. ANOVA: $F(1, 32) = 11.19$, $p = 0.025$. No effect of ART was seen on SRM, $F(1, 31) = 0.0$, *n.s.*
Non-randomized controlled trials					
Youths					
7. Barnoski (2004)	*Participants*: sentenced juvenile offenders ($N = 1229$) under supervision. *Sex*: 983 males, 246 females. *Age*: $M = 15.2$ years (range: 13–17). *Setting*: court probation staff or private contractors at 26 juvenile courts in Washington state, USA.	*ART ($n = 704$). Control ($n = 525$)*: No data collected on control condition.	*Recidivism (mis-demeanor, felony and violent felony)*. Official data: reconviction	Long term	Borderline significant difference in adjusted felony recidivism rate between ART condition (21%) and control condition (25%) ($p = 0.125$). No statistically significant differences in misdemeanor and felony recidivism (ART 46%, control 49%) and violent felony (ART 7%, control 6%) recidivism rates. Controlling for adherence to ART-program, results showed statistically significant differences in felony recidivism between those rated as competent or highly competent (19%) and the control condition (25%) ($p < 0.05$) (RR: 0.75, 95% CI: 0. 57–0.99). For misdemeanor and felony (control 49%; ART 45%) and violent felony (control 6%; ART 7%) recidivism there were no statistically significant reductions.
8. Erickson (2013)	*Participants*: sentenced female juvenile offenders ($N = 60$). *Sex*: female only. *Age*: $M = 16.8$ years *Setting* residential commitment program in Florida, USA.	*ART ($n = 30$)*: ART ® and treatment as usual at the facility (cognitive–behavioral, insight oriented, and supportive individual and group therapeutic interventions offered on a daily basis). *Control ($n = 30$)*: Treatment as usual at the facility.	*Aggressive behavior and rule-breaking behavior in and outside the classroom*. CBCL, Teacher Report Form, BIR, UCL PTSD Index for DSM-IV (Adolescent version).	Short term	Repeated measures 2 × 2 (time × group) ANOVA tests indicated no significant mean differences in rule breaking or aggressive behaviors pre-to posttest between ART and control group. Non-significant time × treatment condition interaction for aggressive behavior in the classroom $F(1, 58) = 0.78$, $p = 0.38$, for rule breaking behavior in the classroom $F(1, 58) = 0.21$, $p = 0.65$, and for aggressive behavior outside of the classroom, $F(1, 58) = 0.26$, $p = 0.61$.
9. Koposov et al., 2014	*Participants*: youths from social institutions and elementary schools ($N = 20$). *Sex*: male and female. *Age*: $M = 15.8$ (range 15–17 years). *Setting*: schools and social institutions in North-West Russia.	*ART ($n = 8$). Control ($n = 12$)*: no intervention	Social skills and problem behaviors. SSRS.	Short term	No significant effects of ART. No numbers reported.
Adults					
10. Barto Lynch (1995)	*Participants*: adult offenders ($N = 71$) with a history of violent and aggressive behavior. *Sex*: male only. *Age*: $M = 32.1$ years (range 21–50). *Setting*: medium security institution in La Grange, Kentucky, USA.	*ART ($n = 39$). Control ($n = 32$)*: similar group of incarcerated offenders who did not receive ART.	*Anger control, moral reasoning*. Aggression Scale (AS), Hostility Scale (HOS), SRM.	Short term	No significant differences between ART and control conditions in AS (no statistics reported). Significant decrease in SRM in ART condition, ANOVA: $F(1, 53) = 4.79$, $p = 0.03$.
11. Curulla (1991)	*Participants*: adult misdemeanor offenders with learning disabilities ($N = 67$). *Sex:* male only. Age: $M = 26$ years *Setting*: Learning Disability Association of Washington probation programme, Redmond, Washington, USA.	*ART ($n = 16$). Control 1 ($n = 18$)*: all components in ART except moral, combined with additional social skills training for a time period equivalent to that spent on moral discussions in ART. *Control 2 ($n = 33$)*: Clients who had been tested and found suitable for the program, but whom the court did not sentence to the program.	*Recidivism*. Official data: district court charges.	Medium term	ART: Recidivism rate = 1/16. Control 2: Recidivism rate = 8/33. RR = 0.26 (95% CI = 0.04–1.89).

(continued on next page)

Table 1 (*continued*)

Study ID, author (year)	Participants and setting	Interventions	Outcomes and measures	Follow-up	Results/authors' conclusion
12. Danielsson et al. (2011)	*Participants*: sentenced offenders (N = 3138). *Sex*: male only. *Age*: M = 24.5 years (range 17–56). *Setting*: Prison and community probation services, Sweden.	*ART* (n = 523). *Control (n = 2165)*: no ART participation, also no participation in a national prison or probation programme, the usual prison and probation services.	*Recidivism*. Official data: reconviction in any crime and violent crime.	Long term: varying follow-up times (M = 1.7 years, range = 0–4.7, SD = 1.4).	Increased risk of reconviction in any crime in the ART condition, compared to the control condition (HR: 1.25; 95% CI: 1.07–1.46). Subgroup analyses showed no significant difference in recidivism between those who completed the ART programme (59%) and those who did not (HR: 1.10, 95% CI: 0.91–1.34). Non-completers were more likely to be reconvicted (HR: 1.43, 95% CI: 1.17–1.76). Increased risk of reconviction for violent crime in the ART condition (HR: 1.31; 95% CI: 1. 07–1.60). No statistically significant difference in recidivism between those 59% who completed ART (HR: 1.05, 95% CI: 0.81–1.35). The sub-group analyses also showed that non-completers were more likely to be reconvicted (HR: 1.57, 95% CI: 1.21–2.03).
13. Hatcher et al. (2008)	*Participants*: Sentenced offenders (N = 106) in community rehabilitation. *Sex*: male only. *Age*: M = 27.42 years (range: 18–53). *Setting*: community probation. English and Welsh probation services, UK.	*ART (n = 53)*: Adult community-based version of ART. *Control (n = 53)*: male offenders convicted of a violent offense and received a community penalty, who had not participated in ART	*Recidivism*. Official data: reconviction.		ART: Reconviction rate 20/53. Control: Reconviction rate 27/53. RR = 0.74 (95% CI: 0.48–1.14). Program non-completers (n = 13) were more likely to be reconvicted than program completers (n = 15) and controls (n = 53). When removing non-completers and non-starters (n = 38) from the analysis, the reduction in reconviction was 15% in the ART condition (n = 15) as compared to the control condition.
Unclear design, youths					
14. Goldstein (1987)[c]	*Participants*: juvenile offenders (N = 51) in residential care *Sex*: male only *Age*: M = 18 years 8 months (range: 13–21). *Setting*: a New York State Division for youth maximum security facility, USA.	ART, Control 1 and Control 2. No information on numbers in each condition.	*Anger control, moral reasoning*. BIR, KWS, SRM	Short term	No significant difference in number of aggressive incidences (ANCOVA: F = 0.69, *n.s.*), but differences favoring ART on intensity (ANCOVA: F = 0.53, p < 0.001) (BIR). Results favoring ART on KWS: F = 0.28 (p < 0.01). Significant improvements in SRM in the ART condition versus control (ANCOVA: F = 14.73, p < 0.01).
15. Goldstein et al. (1989)	*Participants*: juvenile delinquents (N = 84) discharged from residential care. *Sex*: male only. *Age*: not reported. *Setting*: community, New York state Division for youth, USA.	*ART 1. (n = 13)*: youths and family got ART. *ART 2 (n = 39)*: youths only got ART. *Control (n = 32)*: no-ART control group youth.	*Anger* control, social skills, recidivism. Official data, agency case files, interviews with case workers: Rearrests, Anger Situation Inventory Scale, The Skill checklist	Short term	Results favoring ART on re-arrest rates: RR: 0.35, 95% CI: 0.15–0.81. Significant increase in interpersonal skills in the ART conditions (F (2, 83) = 7.64, p < 0.01) and mean anger arousal compared to control (F = 6.47, p = 0.001).
16. Nodarse (1998)	*Participants*: adolescents with emotional handicap (N = 50). *Sex*: 48 males, 2 females. *Age*: mean not reported (range: 12–14). *Setting*: Ruth Owens Kruse Educational Center, Florida, USA.	*ART (n = 25)*: ART + individual therapy (50 min a week) and project adventure training. *Control (n = 25)*: individual therapy (50 min a week) and project adventure training.	*Anger control, social skills*. Behavior Assessment System for Children (BASC), Teacher Rating Scale (TRS)	Short term	Significant results favoring the ART condition for anger control (ANOVA: F (1, 47) = 15.16, p = 0.0003), and social skills (F (1, 47) = 11.95, p = 0.004).

Note. ANOVA = analysis of variance. ANCOVA = analysis of covariance. HR = hazard ratio. RR = risk ratio. CI = confidence interval.

[a] Glick (1987) refers to study reported in Glick and Goldstein (1987) and Goldstein and Glick (1994).

[c] Goldstein (1987) refers to study reported in Glick and Goldstein (1987) and Goldstein and Glick (1994).

[b] Goldstein (1994) refers to study reported in Goldstein et al. (1994) and Goldstein et al. (2004).

Curulla, 1991; Danielsson et al., 2011; Hatcher et al., 2008). Two of these studies attempted to take the risk of bias due to confounding in non-randomized trials into account by matching pairs of individuals from the ART and control arms of the trial on the potential confounder history of criminal behavior (Danielsson et al., 2011; Hatcher et al., 2008). These studies were also judged as having low risk of detection bias (i.e., due to lack of blinding of outcome assessment) for the primary outcome recidivism, as the outcome was based on official data. These two studies could therefore be considered as having more robust results than the other non-randomized studies, although the risk of selection bias remains high due to the non-randomized design. The remaining non-randomized trials were considered at high risk of bias on multiple domains and the reliability of the results of these studies is therefore questionable. None of these studies attempted to account for the risk of bias due to confounding. The results of Barnoski (2004) should be interpreted with particular caution, as data collected during only a part of the study period was included in analyses. Barto Lynch (1995) and Koposov et al. (2014) was judged as having high risk of bias due to incomplete outcome data as only those that completed the program were included in analyses. Barto Lynch (1995) also had high risk of bias due to potentially serious statistical issues. Curulla (1991) was considered at high risk of bias as there was information to suggest that concurrent participation in other programs was frequent.

Three of the included studies (Goldstein, 1987; Goldstein et al., 1989; Nodarse, 1998) had an unclear design, and all included youths. As reflected in the inability to even judge the study design employed, these studies had poor reporting standards on the majority of the risk of bias domains. All of these studies had potentially high risk of bias on at least one category. We therefore consider the reliability of the results of these studies low or at best questionable.

Nearly half (7 out of 16) of the studies were conducted by researchers who can be considered as having vested interests in the program.

3.3. Effects of the intervention

Statistical meta-analysis is usually advisable when assessing the effectiveness of interventions, especially when included studies, as in our case, are underpowered (Higgins & Green, 2008). Also in this review, a meta-analytic approach was originally intended. However, for such an analysis to be meaningful, careful consideration has to be paid to the diversity of studies (Borenstein, Hedges, Higgins, & Rothstein, 2009). Moreover, meta-analyses of studies that are at high risk of bias may be particularly misleading. If bias is present in each (or some) of the individual studies, a meta-analysis will simply compound the errors, and produce a 'wrong' result that may be interpreted as having more credibility than it deserves (Deeks, Higgins, & Altman, 2008). Given the substantive clinical and methodological diversity depicted in Table 1, and the frequent occurrence of high risk of bias in multiple domains among the studies detailed in Table 2, we found it inappropriate to pool the estimated treatment effects into a precision weighted summary estimate. Moreover, due to the low number of studies reporting sufficiently adequate data, it was not possible to explore sources of heterogeneity using meta-regression. Consequently, we report the results narratively by outcome, population and design.

3.3.1. Effects on recidivism

Based on official justice-department records or other official records of recidivism, three studies reported the effects of ART on this primary outcome in youth (Barnoski, 2004; Goldstein, 1989, 1994) and three studies in adults (Curulla, 1991; Danielsson et al., 2011; Hatcher et al., 2008). All studies reported sufficient data to calculate and/or reported risk ratios (RR) and corresponding 95% confidence intervals (CI), although there was insufficient data reported in Goldstein (1994) to adjust for the design-effect of clustering.

3.3.1.1. Cluster-randomized trials. In Goldstein et al. (1994) there were significantly lower rearrest rates in the ART youth groups versus control: five of 38 (13%) ART participants and 14 of the 27 control group members (53%) were rearrested during the 8-month tracking period, see Table 1.

3.3.1.2. Non-randomized controlled trials. Barnoski (2004) reported no effect of ART for juvenile offenders compared to treatment-as-usual (RR: 0.95, 95% CI 0.85–1.07). When controlling for adherence, though, highly competently delivered ART reduced recidivism to a greater extent relative to the control group, see Table 1.

In Curulla (1991) there were no statistically significant differences between the treated group and a no-treatment control group, likely due to the low incidence of recorded charges: one participant was recorded in the ART-group, five in ART without moral component and eight in no intervention group, see Table 1.

Danielsson et al. (2011) estimate Cox proportional hazard regression models and report hazard ratios (HR) as effect sizes. Assuming that risk is constant over time, HRs may be interpreted as RRs (Singer & Willett, 2003). Controlling for age, previous convictions for crime of violence, drug offenses and mean time incarcerated. Adult ART participants had a 25% elevated risk of reconviction in any crime when compared to the control condition, see Table 1. Subgroup analyses showed that there was no statistically significant difference in recidivism between those who completed the ART program and those who were in the control condition. However, separate analyses of non-completers suggest that ART dropouts had a 43% elevated risk to be reconvicted. Similar elevated risks were also found for reconviction for violent crime, regardless whether ART participants were completers or non-completers, see Table 1.

Data reported in Hatcher et al. (2008) suggest a positive but non-significant effect of ART for adults on probation service, see Table 1. Comparisons of completers and non-completers in the two conditions showed no statistically significant differences between completers and control (RR 0.53, 95% CI: 0.18–1.54) or non-completers and control (RR: 1.63, 95% CI: 0.94–2.83).

In sum, Table 1 shows that there is considerable clinical and methodological diversity (including, but not limited to, varying follow-up times and lack of detail on control conditions) between the three studies. However, subgroup analyses indicated that non-completers had significantly higher reconviction rates than the control group (Curulla, 1991; Danielsson et al., 2011; Hatcher et al., 2008).

3.3.1.3. Studies with unclear design. In Goldstein et al. (1989) ART for young male offenders discharged from residential care appeared to result in a positive effect when compared to those receiving no intervention. The study, which had an unclear design, reported significantly lower re-arrest rates where 8 of the 52 (15%) ART and family ART participants were re-arrested, compared to 14 of the 32 control condition members (43%), see Table 1.

3.3.2. Effects on anger control, social skills and moral reasoning

Of the 16 included studies, 12 reported on at least one of the secondary outcomes. However, the reporting standards were generally poor. Due to data only being available for a small subset of studies, we found it to be inappropriate to attempt to display the effect sizes using a common metric. Consequently, we report the results of the analyses as they were reported by the evaluators.

Eleven studies assessed anger control, typically operationalized as acting-out behavior and self-control, in youths based on standardized measures or observations. Seven studies reported results of tests with documented psychometric properties or staff observations of social skills in youth (Coleman et al., 1992; Erickson, 2013; Goldstein et al., 1989; Goldstein, 1994; Gundersen & Svartdal, 2006; Koposov et al., 2014; Nodarse, 1998). Moral reasoning was explored in six trials including youths (Coleman et al., 1992; Jones, 1991; Glick, 1987;

Table 2
Risk of bias summary of included studies by design and population.

Study ID, author (year)	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias	Restricted participant selection	Balance between the arms	Matched on confounders	Adjusted for confounders in statistical analyses
Randomized controlled trials, youths										
1. Coleman et al. (1992)	Unclear	High	Unclear	High	Unclear	High	N/A	N/A	N/A	N/A
2. Jones (1991)	Unclear	Unclear	Unclear	Unclear	Unclear	High	N/A	N/A	N/A	N/A
Cluster-randomized controlled trials, youths										
•3. Glick (1987)[a]	Unclear	Unclear	Unclear	Unclear	Unclear	High	High	Unclear	High	High
•4. Goldstein et al. (1994)[b]	Unclear	Unclear	Low	Unclear	Unclear	High	High	Unclear	High	Unclear
•5. Gundersen and Svartdal (2006)	Low	High	Low	Unclear	Unclear	High	High	Low	High	Low
•6. Zimmerman (1987)	Unclear	Unclear	Unclear	High	Unclear	Low	Low	Low	Low	Low
Non-randomized controlled trials										
Youths										
7. Barnoski (2004)	High	High	Low	Unclear	High	Unclear	High	Unclear	High	Low
8. Erickson (2013)	High	High	High	High	Low	Low	Low	Low	High	High
•9. Koposov et al. (2014)	High	High	High	High	Low	High	Low	High	High	High
Adults										
10. Barto Lynch (1995)	High	High	Unclear	High	Unclear	High	High	Unclear	High	Unclear
11. Curulla (1991)	High	High	Low	Unclear	Unclear	High	High	Unclear	High	High
12. Danielsson et al. (2011)	High	N/A	Low	Low	Low	Unclear	High	Unclear	Low	Low
13. Hatcher et al. (2008)	High	N/A	Low	Low	Unclear	High	Low	Low	Low	High
Studies with unclear design, youths										
•14. Goldstein (1987)[c]	Unclear	Unclear	Unclear	Unclear	Unclear	High	High	Low	Low	Low
•15. Goldstein et al. (1989)	High	Unclear	Low	Unclear	Unclear	High	High	Low	High	Low
16. Nodarse (1998)	Unclear	Unclear	Unclear	Unclear	High	High	High	Unclear	High	Low

Note. • = non-independent investigator(s). N/A = not applicable. When judgments differ between outcomes/potential confounders, judgments for recidivism and the potential confounders anger control/aggressive behavior/criminal behavior are reported. Risk of bias-domains concerning confounding was only assessed for cluster-randomized and non-randomized controlled trials.

[a] Glick (1987) refers to study reported in Glick and Goldstein (1987) and Goldstein and Glick (1994).

[c] Goldstein (1987) refers to study reported in Glick and Goldstein (1987) and Goldstein and Glick (1994).

[b] Goldstein (1994) refers to study reported in Goldstein et al. (1994) and Goldstein et al. (2004).

Goldstein 1987; Gundersen & Svartdal, 2006; Zimmerman, 1987). All of these studies reported pre- and post-intervention measures, although the periods covered and lengths of the follow-up times differed considerably between studies. One non-randomized controlled trial presented results using standardized measures of the secondary outcomes anger control and moral reasoning, for adults (Barto Lynch, 1995).

3.3.2.1. Randomized controlled trials. Coleman et al. (1992) found no significant differences between the ART condition and the control condition in acting-out behavior self-control, social skills and moral reasoning, see Table 1. In Jones (1991), three conditions were compared: ART, the moral-reasoning component of ART only, and a treatment-as-usual (TAU) control. The investigators report that the ART condition resulted in a significant decrease in acting-out behavioral incidents, although a one-way ANOVA of pre-test data revealed a significantly higher level of behavioral incidents in the ART condition compared to the control condition. There was no statistically significant effect of ART on self-control. All factors and the interaction effect were reported to be not significant for moral reasoning.

3.3.2.2. Cluster-randomized controlled trials. Glick (1987) compared an ART condition a comparison condition that received brief instructions, and TAU control condition. This study reported a significant positive effect of ART on self-control, as well as on the number and the intensity of acting-out behavioral incidents. No improvements on the moral component were reported. Further, Goldstein (1994) reported a significant positive effect of ART on anger control and on social skills.

Gundersen and Svartdal (2006) compared ART with a TAU control condition. Two measures of acting-out behavior were administered: the Child and Adolescent Disruptive Behavior Inventory (CADBI) and the Child Behavior Checklist (CBCL). They also explored the effect of ART on social skills using the Social Skills Rating System (SSRS), and on moral reasoning (using How I Think, HIT). Interviews were conducted with parents, teachers, and youth. Results differed between instruments and respondent groups, see Table 1. For parent- and teacher ratings, there were generally improvements within the ART-condition, when the two conditions were analyzed separately. According to youth ratings, there were either no documented changes, or changes had occurred for both the ART-condition and the control condition, see Table 1.

Zimmerman (1987) compared ART to a brief-instructions comparison condition, and a TAU control condition, and reported a significant positive effect of ART on self-control. For acting-out behavior, an analysis of pre-test scores revealed that the control condition had more severe acting-out behavior. An ANOVA showed no significant effect of ART. However, an ANCOVA with pre-test scores as covariate revealed a significant positive effect of ART on both the number of incidences, and the severity of the reported incidences. No effect of ART on moral reasoning was reported.

3.3.2.3. Non-randomized controlled trials. Results in Erickson (2013) indicated no significant mean differences in rule-breaking or aggressive behaviors in female juvenile offenders in a residential commitment program, pre-to posttest between ART for and control (TAU), see Table 1. Nor did the degree of trauma which was used as a covariate have an impact on ART-intervention efficacy. Mean negative behaviors were reduced for all girls who participated in the study. Koposov et al. (2014) investigated the efficacy of ART on pro-social skills and decrease in problem behaviors. Based on youths self-reports, effects of ART were found among children 14-years old and younger, while no changes were observed among youths older than 15. There were no differences between ART and control, according to parent and teacher ratings. In both these studies, the evaluators explain the absence of effects with the low number of participants.

Barto Lynch (1995) measured incarcerated adults' anger control and found no statistically significant differences between the ART and control conditions in acting-out behavior, see Table 1. There was a significant difference in moral reasoning in the ART condition compared to the control condition, but the investigators report that "although the SRM results between the experimental and control conditions reveal a statistically significant quantitative difference, this finding may be negligible in practical terms" (Barto Lynch, 1995, p. 80).

3.3.2.4. Studies with unclear design. For Goldstein (1987), no information was available regarding the experimental design or the number of participants allocated to each condition. This study reported that ART had no effect on the number of acting-out behavioral incidences. In contrast, the investigators report significant positive effects of ART on the intensity of acting-out behavioral incidents, on self-control (KWR) and on moral reasoning (SRM), see Table 1.

Goldstein et al. (1989) compared ART with family-ART, and a TAU control. The investigators report that mean "anger arousal" was significantly lower in the combined ART and family-ART conditions when compared to the control condition, see Table 1. Moreover, the combined ART and family-ART conditions also had a significantly greater increase in social skills compared to the control condition. Finally, Nodarse (1998) reported a significant positive effect of ART on acting-out behavior and social skills for adolescents with an emotional handicap, see Table 1.

4. Discussion

This review assessed the effect of ART on antisocial behavior in young people and adults. Looking at each individual study, the results indicate positive effects of ART, both on recidivism and on the secondary outcomes anger control, social skills and moral reasoning. However, the majority of studies suffer from rather extensive flaws (e.g. selection bias, see Table 2), which greatly limit our ability to draw generalizable conclusions. The included studies showed substantive clinical and methodological diversity, the overall methodological quality of the studies was poor, and the post-intervention follow-up was generally limited. Overall, the reporting standard was low and obtaining data where it was missing or incomplete was not possible in all cases despite efforts to contact investigators. Moreover, almost half of the studies were conducted by researchers who could not be considered as independent.

The information was equally limited with regard to participants' socioeconomic backgrounds and structural factors such as poverty, unemployment and poor environments. Most important, the information on the characteristics of participants was inadequate and there was a general failure to report systematically on any mental health difficulties. For example, it is well known that aggression and violent behavior are highly comorbid with other problems such as attention deficit and hyperactivity (Cai, 2004; Waschbusch, 2002), which was not addressed or controlled for in any of the included studies. No distinctions between different types of offenders were made. To the best of our knowledge, there were no analyses of the possible implications of including participants with different psychiatric diagnoses. The applicability of the evidence is therefore limited. Furthermore, there was a clear possibility of contamination in a number of studies where both intervention and control participants were recruited from the same institution, center or school class.

Whilst biases arising from lack of blinding are challenging to overcome in trials of psychosocial interventions, other methodological weaknesses could have been addressed by the investigators, such as reporting the numbers of participants, reporting all outcome data, recording the level of adherence to program, and acknowledging the possibility of confounding effects. It is noteworthy that none of the cluster-RCTs took the potential effects of clustering into account in their analysis. A number of studies were moreover characterized by potentially serious statistical issues (e.g. choice of statistical methods),

and only two studies reported on the numbers drop-out from the program or performed intention-to-treat-analysis.

Finally, the reported effects of ART on primary and secondary outcomes varied, and the high risk of bias in the included studies means that any result should be interpreted with considerable caution. It is noteworthy that in the two studies that reported negative effects of ART on adult recidivism (Danielsson et al., 2011; Hatcher et al., 2008), such negative effects were apparent in non-completers only. Negative effects of programs for non-completers have also been reported in other studies (Dowden, Blanchette, & Serin 1999), and this must be considered as an implementation issue.

4.1. Strengths and limitations

Our comprehensive literature search included non-English electronic databases and non-English sources of gray literature. Although we consider it likely that we have obtained all of the existing evidence from controlled trials on the effects of ART, a potential bias in the review process is that our inclusion criteria were restricted to studies that labeled the program under evaluation as "Aggression Replacement Training", included a reference to Goldstein, and clearly stated that the three core components in the program were being adhered to. Hence, the requirement was that the evaluated program only included the main, original components of ART, in line with the reasoning of Glick and Gibbs (2011). This resulted in a limited number of studies being included, which clearly illustrates one of the difficulties in this field — "program drift". How can researchers, decision makers, and practitioners use conclusions of aggregated method evaluations to guide their work, when the methods under study are very different? Which components are the ones that create positive change? Are some components superfluous, or even harmful? Such problems can only be solved by meticulous inclusion criteria, and component analyses. The latter was not possible, in our case. As such, this review does not examine the effectiveness of extended/modified versions of ART. It remains possible that such studies would provide additional – and/or different – evidence as to the effectiveness of similar programs. A related limitation is that broader forms of evidence including issues related to feasibility, appropriateness, meaningfulness and how participants experience the intervention has been ignored. It seems reasonable to assume that a systematic review of qualitative studies of ART trials may show that other aspects of the program than effectiveness can be valued.

It has been shown that study biases operate systematically and typically inflate treatment effect estimates (Wortman, 1994). The approach used here, with a heavy focus on methodological quality, ensures that critical methodological issues have been examined systematically, which makes it less likely that problems or biases have been overlooked. On the other hand, it increases the level of detail and provides less room for generous interpretations of results.

4.2. Conclusions and recommendations for future research

A great number of the individual studies included in this systematic review indicate positive effects of ART. The first-hand conclusion after inspecting this research would be that ART is an effective program for changing the behavior of antisocial adolescents and adults. However, when looking into the quality of the research building the basis for such a conclusion, the picture becomes much more problematic. The results of the present review show that the primary studies of ART do not provide a sufficient base for substantiating the claim that the program is effective for reducing antisocial behavior in adolescents and adults. This is in stark contrast to the findings of previous reviews that have examined ART-focused trials and considered the quality of the evidence sufficient to draw conclusions about effectiveness (Cigno & Bourns, 1998; Howell, 1998; Landenberger & Lipsey, 2005; Lipsey et al., 2001, 2007; Pearson et al., 2002; Sherman et al., 2002; Springer et al., 2003; Welsh & Hoshi, 2002), even though there is a clear overlap between included studies in the present and some of the previous reviews (Landenberger & Lipsey, 2005; Lipsey et al., 2001, 2007). Naturally, this does not mean that we claim there is evidence for no effect. Rather, we cannot say whether ART helps, whether it has no effect, or even whether it is a harmful intervention.

Goldstein and colleagues have argued that aggression and other forms of antisocial behavior can be traced back to a general lack of pro-social behavior, a low level of anger control, and an immature, egocentric style of moral reasoning. Indeed, one of the theoretical bases for ART is that aggressive behavior and criminality in youths can be reduced through improved ability to manage anger and self-control. However, the extent to which aggressive behavior can be explained by a lack of self-control, especially in populations with comorbidity, is still under discussion among scholars (see e.g., McGuire, 2008; Polaschek & Reynolds, 2004). For example, Gottfredson & Hirschi (1990, see also Piquero et al., 2010) suggest, in contrast, that the possibility of improving individuals' self-control is restricted to late childhood, as self-control is determined early in life and is thereafter relatively stable.

Regardless of the theoretical support for the components of ART, the fact remains that the target group is not clearly defined in the ART-literature. The behavior expected to be targeted by ART is often generally described as aggressive or antisocial, although other behaviors have also been suggested to be affected by the intervention. Even when focusing on aggressive behavior, a more precise definition is desirable (see e.g., McGuire, 2008).

A common notion in prevention is that "one size does not fit all". In this spirit, an important task for further research is to explore the effects of ART on clearly-defined target groups. It will be important to provide detailed descriptions of the type of aggressive or antisocial behavior that the participants have previously expressed and which is expected to be affected by the program. It will also be important to take into consideration whether or not participants are lacking in social skills, ability to manage anger and moral reasoning, as these factors have been hypothesized as mediators. Finally, studies should measure, and include in analyses, characteristics such as attrition, adherence to the program, and background characteristics such as comorbidity and socioeconomic status. Such studies will thus not only provide insight into the effectiveness of ART but also into whether or not a relation exists between these hypothesized mediators and the targeted behaviors. This review has revealed that such studies are currently lacking.

One can question whether a systematic review of studies with methodological flaws should even be conducted. However, we consider this as an important piece of knowledge, both for researchers in designing future research, and for practitioners in deciding whether to use a particular intervention or not. It is not fair to claim that we know anything about a program's effectiveness, based on studies with great shortcomings. It may lead to wrongful conclusions, in either direction — results pointing toward a method being effective that is not, or results showing no effects of methods that actually are beneficial for the participants. Further research is therefore warranted by independent investigators exploring the effects of ART on clearly-defined target groups using high standard evaluation designs.

As our knowledge about the program's efficiency gradually accumulates, it will also be of importance that researchers report implementation costs such that a cost-effectiveness analysis can be performed. This may be particularly important for decision makers and practitioners working in the context of reducing aggression and other forms of antisocial behavior, because interventions in this field typically are expensive, and the negative externalities on family, friends and society of such behavior are large.

References[1]

Achenbach, T.M. (1991). *Manual for the child behavior checklist/4–18 and 1991 profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

[1] References marked with an asterisk are included in the review.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders (fourth edition)*. Washington DC: American Psychiatric Association.
Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice-Hall.
*Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile Offenders. Olympia*, Washington: The Evergreen State College, Washington State Institute for Public Policy.
*Barto Lynch, J. (1995). *The use of aggression replacement training with adult offenders. A program for violent and aggressive inmates*. Louisville, KY: Spalding university.
Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley.
Cai, X. (2004). Stability of externalizing problem behaviors with onset in early childhood: A meta-analytic review. Unpublished dissertation. Nashville: Vanderbilt University.
Calame, R., & Parker, K. (2012). *Family-TIES: A family-bas|ed Intervention to Complement Prepare, ART, and TIES Youth Groups*. Champaign, IL: Research Press Publishers.
Cigno, K., & Bourns, D. (1998). *Cognitive-behavioural social work in practice*. Aldershot: Ashgate/Arena.
*Coleman, M., Pfeiffer, S., & Oakland, T. (1992). Aggression replacement training with behaviorally disordered adolescents. *Behavioral Disorders*, *18*(1), 54–66.
*Curulla, V.L. (1991). *Aggression replacement training in the community for adult learning disabled offenders*. Washington, DC: University of Washington.
*Danielsson, M., Fors, A., & Freij, I. (2011). *Behandlingsprogrammet ART i kriminalvården. Utvärdering av återfall i brott för programdeltagare 2003–2006 [The treatment program ART in correctional treatment. An evaluation of recidivism for participants during 2003–2006]*. Stockholm: Kriminalvården [The Swedish Prison and Probation Service].
Deeks, J.J., Higgins, J.P.T., & Altman, D.G. (2008). Analysing data and undertaking meta-analyses. In J.P.T. Higgins, & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 243–296). John Wiley & Sons and The Cochrane Collaboration: Hoboken, NJ.
Dowden, C., Blanchette, K., & Serin, R. (1999). *Anger Management Programming for Federal Male Inmates: an Effective Intervention*. Research report R-82 Ottawa, ON: Correctional Service of Canada.
*Erickson, J.A. (2013). *The efficacy of aggression replacement training with female juvenile offenders in a residential commitment program*. Tampa, FL: University of South Florida.
Gibbs, J.C., Potter, G.B., & Goldstein, A.P. (1995). *The EQUIP program: Teaching youth to think and act responsibly through a peer-helping approach*. Champaign, IL: Research Press.
Glick, B., & Gibbs, J.C. (2011). *Aggression replacement training: A comprehensive intervention for aggressive youth* (3rd ed.). Champaign, IL: Research Press.
*Glick, B., & Goldstein, A.P. (1987). Aggression replacement training. *Journal of Counseling & Development*, *65*(7), 356–362.
Goldstein, A.P. (1999). *The prepare curriculum*. Champaign, IL: Research Press.
*Goldstein, A.P., & Glick, B. (1994). Aggression replacement training: curriculum and evaluation. *Simulation & Gaming*, *25*(1), 9–26.
Goldstein, A.P., Glick, B., & Reiner, S. (1987). *Aggression replacement training*. Champaign, IL: Research Press.
*Goldstein, A.P., Glick, B., Irwin, M.J., Pask-McCartney, C., & Rubama, I. (1989). *Reducing delinquency: intervention in the community*. Elmsford, NY,US: Pergamon Press.
*Goldstein, A.P., Glick, B., Carthan, M., & Blancero, D. (1994). The prosocial gang. implementing aggression replacement training. *Thousand oaks*. CA: Sage.
*Goldstein, A.P., Nensén, R., Daleflod, B., & Kalt, M. (2004). *New perspectives on aggression replacement training. Practice, Research and Application*. Chichester: John Wiley & Sons.
Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.
*Gundersen, K., & Svartdal, F. (2006). Aggression replacement training in Norway: outcome evaluation of 11 Norwegian student projects. *Scandinavian Journal of Educational Research*, *50*(1), 63–81.
*Hatcher, R.M., Palmer, E.J., McGuire, J., Hounsome, J.C., Bilby, C.A.L., & Hollin, C.R. (2008). Aggression replacement training with adult male offenders within community settings: a reconviction analysis. *Journal of Forensic Psychiatry & Psychology*, *19*(4), 517–532.
Higgins, J.P.T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. Hoboken, NJ: John Wiley & Sons and The Cochrane Collaboration.
Hornsveld, R., Van Dam-Baggen, R., Leenaars, E., & Jonkers, P. (2004). *Agressiehanteringstherapie voor forensischpsychiatrische patienten met gewelddelicten: ontwikkeling en praktijk* [aggression control therapy for forensic psychiatric patients with violent offences: development and application]. *Tijdschrift voor Psychotherapie*, *30*(1), 22–38.
Howell, J.C. (1998). Promising programs for youth gang violence prevention and intervention. In D. Loeber, & D. Farrington (Eds.), *Serious and violent juvenile offenders. risk factors and successful interventions* (pp. 284–312). Thousand Oaks, CA: Sage Publications.
*Jones, Y. (1991). Aggression replacement training in a high school setting. *Australian Journal of Guidance and Counselling*, *1*(1), 1–19.
Kohlberg, L. (Ed.). (1973). *Collected papers on moral development and moral education*. Cambridge. MA: Harvard University, Center for Moral Education.
*Koposov, R., Gundersen, K., & Svartdal, F. (2014). Efficacy of aggression replacement training among children from North-West Russia. *The International Journal of Emotional Education*, *6*(1), 14–24.
Landenberger, N.A., & Lipsey, M.W. (2005). The positive effects of cognitive behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, *1*(4), 451–476.
Leeman, L.W., Gibbs, J.C., & Fuller, D. (1993). Evaluation of a multi-component group treatment program for juvenile delinquent. *Aggressive Behavior*, *19*(4), 281–292.
Lipsey, M.W., Chapman, G.L., & Landenberger, N.A. (2001). Cognitive behavioral programs for offenders. *Annals of the American Academy of Political and Social Science*, *578*, 144–157.
Lipsey, M.W., Landenberger, N.A., & Wilson, S.J. (2007). *Effects of cognitive behavioral programs for criminal offenders*. Oslo: Campbell Collaboration.
Lipton, D., Thornton, D., McGuire, J., Porporino, F., & Hollin, C.R. (2000). Programme accreditation and correctional treatment. *Substance Use and Misuse*, *35*, 1705–1734.
McGuire, J. (2008). A review of effective interventions for reducing aggression and violence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1503), 2577–2597.
Meichenbaum, D.H. (1977). *Cognitive behaviour modification: An integrative approach*. New York: Plenum.
*Nodarse, M.V. (1998). *The effects of aggression replacement training on adolescents with an emotional handicap*. Miami, FL: Caribbean Center for Advanced Studies, Miami Institute of Psychology.
Novaco, R.W. (1975). *Anger control: The development and evaluation of an experimental treatment*. Lexington, MA: D.C. Heath.
OJJDP Model Programs Guide. Retrieved Jan 30 2013, from http://www.ojjdp.gov/mpg/Aggression%20Replacement%20Training%20%20174;%20(ART%20%20174;)-MPGProgramDetail-292.aspx
Pearson, F.S., Lipton, D.S., Cleland, C.M., & Yee, D.S. (2002). The effects of behavioural/cognitive-behavioral programs on recidivism. *Crime and Delinquency*, *48*, *476–496wa*.
Piquero, A., Wesley, G., Jennings, D., & Farrington, D. (2010). *Self-control interventions for children under age 10 for improving self-control and delinquency and problem behaviours*. Oslo: Campbell Collaboration.
Polaschek, D. L., & Reynolds, N. (2004). Assessment and treatment: Violent offenders. *The essential handbook of offender assessment and treatment*, 201–218.
Salmon, S. (2004). The peace curriculum: expanded aggression replacement training. In A.P. Goldstein, R. Nensén, B. Daleflod, & M. Kalt (Eds.), *New perspectives on aggression replacement training. Practice, Research and Application* (pp. 171–188). Chichester: John Wiley & Sons.
Schulz, K.F., Altman, D.G., & Moher, D. (2010). Consort 2010 statement: updated guidelines for reporting parallel group randomized trials. *BMC Medicine*, *8*, 18.
Sherman, L.W., Farrington, B.W., & MacKenzie, D. (2002). *Evidence-based crime prevention*. London: Routledge.
Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis. modeling change and event occurrence*. Oxford: Oxford University Press.
Springer, D.W., McNeece, C.A., & Arnold, E.M. (2003). *Substance abuse treatment for criminal and juvenile offenders: an evidence-based guide for practitioners*. Washington, DC: American Psychological Association.
Waschbusch, D.A. (2002). (2002) A meta-analytic examination of comorbidhyperactive impulse-attention problems and conduct problems. *Psychological Bulletin*, *128*, 118–150.
Welsh, B.C., & Hoshi, A. (2002). Communities and crime prevention. In L.W. Sherman, D.P. Farrington, B.C. Welsh, & D.C. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 97–197). London: Routledge.
Wortman, P.M. (1994). Judging research quality. In H. Cooper, & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97–109). Russel Sage Foundation: New York, NY.
*Zimmerman, D.L. (1987). *The efficacy of aggression replacement training with juvenile delinquents*. Syracuse, NY: Syracuse university.