# Treatment, prediction, and assessment of childhood aggression

Treatment, prediction, and assessment of childhood aggression

Anne Hendriks

Anne Hendriks

VRIJE UNIVERSITEIT

# TREATMENT, PREDICTION, AND ASSESSMENT OF CHILDHOOD AGGRESSION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op maandag 18 november 2019 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Anne Myriem Hendriks

geboren te Amsterdam

promotoren: prof.dr. C. Finkenauer
prof.dr. M. Bartels

# CONTENTS

Chapter 1

# Introduction

Childhood aggression consists of a broad spectrum of behaviors including overt and physical behaviors such as fighting, stealing, or disobedience, and covert behaviors such as gossiping, social exclusion, or becoming friends with other children as revenge (Achenbach & Rescorla, 2001; Björkqvist, Lagerspetz, & Kaukiainen, 1992; Goodman, 2001; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003). This dissertation focuses on overt and physical childhood aggression. Childhood aggression increases the likelihood for children to display other behavioral and emotional problems (Bartels et al., 2018), not only in childhood but also later in life (Althoff, Verhulst, Rettew, Hudziak, & Van Der Ende, 2010; Copeland, Wolke, Shanahan, & Costello, 2015). Moreover, childhood aggression is burdensome for parents (Meltzer, Ford, Goodman, & Vostanis, 2011; Roberts, McCrory, Joffe, de Lima, & Viding, 2017) and produces high financial costs for society due to higher conviction rates and use of health and social welfare services (Rivenbark et al., 2018; Romeo, Knapp, & Scott, 2006).

To gain insight into the etiology of individual differences in childhood aggression and co-occurring behavioral and emotional problems, the ACTION (Aggression in Children: Unravelling gene-environment interplay to inform Treatment and InterventiON strategies; http://www.action-euproject.eu/) consortium was founded in 2014. ACTION brings together multiple large cohort studies including childhood prospective twin, population-based, and clinical cohorts in genetically informative populations. The focus of ACTION is to inform on the etiology of differences in aggression between children by unravelling its genetic architecture using univariate, multivariate, and longitudinal genetic and epigenetic modelling in twin and genetic and epigenetic association studies (Bartels et al., 2018). A strong focus of ACTION includes biomarker and metabolomics research (Boomsma, 2015). As part of the large-scale international ACTION collaboration, this dissertation project focuses on treatment, prediction, and assessment of childhood aggression.

Despite the large amount of attention paid to treatments, their effectiveness for childhood aggression is generally low (Weisz et al., 2017). Research that reveals for whom and under which circumstances (i.e., moderators) treatments are more effective is needed to optimize treatment effectiveness. Several arguments underline why it is important to continue research on treatments for childhood aggression. For instance, the high genetic stability of childhood aggression demonstrates that a wait-and-see policy will not work because, without treatment, children with heightened levels of aggression likely remain aggressive (Lubke, Mcartor, Boomsma, & Bartels, 2017; Porsch et al., 2016; Wesseldijk et al., 2018). In addition, diagnosis at a later age predicts worse outcomes later in life, such as a higher probability of having a criminal record and a lower income (Campbell, Lundstrom, Larsson, Lichtenstein, & Lubke, 2018). Because the limited success of treatments so far, and the stability and worse off outcomes of childhood aggression throughout the lifespan, it is necessary to better grasp the moderators for treatment effectiveness.



Chapter 2 provides an overview of the current knowledge on the effectiveness of treatments for childhood aggression and its moderators. In Chapter 2, I present a literature synthesis of 72 meta-analyses and systematic reviews on the effectiveness of psychosocial treatments for childhood aggression. This study provides an indication of the overall treatment effectiveness and moderators that might explain why some children respond better to treatment than others and under which circumstances.

For early prevention, however, it would also be useful to better predict aggressive behavior problems to identify children more likely to develop aggressive behavior problems. Both genetic and environmental factors are important contributors to individual differences in childhood aggression (Burt, 2009; Ligthart, Bartels, Hoekstra, Hudziak, & Boomsma, 2005; Rhee & Waldman, 2002). Therefore, a comprehensive understanding of the etiology of childhood aggression requires taking both genetic and environmental factors into account, as well as examining the possible interplay between them. Therefore, in this dissertation, I consider aggression within a socio-ecological framework (Sameroff, 2010), that acknowledges the importance of biological and psychological systems at the individual level together with factors at distal levels including parents and the family, the community, and finally the geopolitical level. Predictors of childhood aggression may associate with higher or lower levels, but their effects may also be more complex. It is possible that children from a certain background are much more heterogeneous, resulting in a larger variance in certain groups compared to others. For example, children growing up in a poor neighborhood in which there are lower levels of social control more be more heterogenic that children growing up in a more well-off neighborhood (Lynam et al., 2000; Shanahan & Hofer, 2005; South, Hamdi, & Krueger, 2015). In addition, it is possible that the importance of genetic and environmental factors for the expression of a trait varies as a result of different backgrounds, such as levels of socioeconomic status (SES; Tucker-Drob & Bates, 2015; Tuvblad, Grann, & Lichtenstein, 2006). It has for example been shown that heritability was lower, the influence of the shared environment was higher, and the influence of the nonshared environment was lower on aggressive behavior in adolescents from lower SES families compared to adolescents from higher SES families (Tuvblad et al., 2006). Nevertheless, it is not clear yet how and to what extent the etiology of childhood aggression is affected by the interplay between genetic and environmental factors.

Comprehension of the etiology of childhood aggression requires acknowledging the many different levels of influence and different ways in which they may influence childhood aggression. For these reasons, the studies in Chapter 3 and Chapter 4 include predictors from the country level to the individual level. In addition, Chapter 5 examines a possible gene-environment interaction of socioeconomic strata and the genetic architecture of childhood aggression.

First, focusing on the country level, Chapter 3 examines the association between adolescent aggressive behavior and national-level policies for child and adolescent mental health in 30 European countries. Data are from the 2013/2014 Health Behaviour in School Aged Children (HBSC) study, which includes 172,829 eleven- to fifteen-year-olds from 30 European countries (Currie et al., 2014, 2012). To assess whether the association between child and adolescent mental health policies (Erskine et al., 2017; Eurostat, 2016; OECD Social Policy Division, 2016; Signorini et al., 2017) and adolescent mental health varies across indicators of adolescent mental health, the study included adolescent life satisfaction and psychosomatic symptoms, in addition to adolescent aggressive behavior.

Second, after a focus on country-level predictors, Chapter 4 focuses on proximal predictors of childhood aggression. To this end, I analyze data from 62,227 children from the Child and Adolescent Twin Study in Sweden (Anckarsäter et al., 2011) and the Netherlands Twin Register (Van Beijsterveldt et al., 2013). The outcome is a psychometrically harmonized physical/overt aggression score for 9-year-old children (Luningham et al., submitted). Predictor variables include demographics, prenatal characteristics, physical development, family environment, parenting, parental education level, life events, and behavioral symptoms. Simultaneous assessment of these predictors provides insight in the relative importance of each predictor variable in relation to other predictors. The large sample allows for sophisticated analysis steps in independent parts of data. These steps include 1) exploratory data analysis and tuning meta-parameters for the data mining, 2) fitting increasingly complex data mining models to assess which predictors have which type of effects (i.e., linear, non-linear, interaction), 3) assessment of model performance and importance of predictor variables, and 4) a confirmatory prediction model of childhood aggression that integrates the results of the data mining analyses. As such, the analyses allow us to explore the type of effects of predictor variables and examine their effects simultaneously to obtain a robust prediction of childhood aggression.

Third, to explore complexity in the etiology of childhood aggression due to gene-environment interaction, Chapter 5 examines a possible moderator of the contribution of genetic and environmental variables to individual differences in childhood aggressive behavior. Chapter 5 investigates the moderating effect of socioeconomic status (SES) on the genetic architecture of childhood aggression in large samples of 7-year-old twins from the Netherlands Twin Register (N = 24,112; Van Beijsterveldt et al., 2013) and the Twins Early Development Study (N = 19,644; Haworth, Davis, & Plomin, 2013) from the United Kingdom.

The advantages of consortia such as ACTION are that they make efficient use of a wealth of existing data, combine experts from different backgrounds, and increase generalizability through examination of different samples. Also, combining samples allows us to answer new questions compared to single-cohort studies. Nevertheless, it induces heterogeneity, for instance, because data are collected with different research purposes and consequently cohorts vary in their instruments to assess aggression. In ACTION, several cohorts participate, such as the Child and Adolescent Twin Study in Sweden (CATSS; Anckarsäter et al., 2011), the Netherlands Twin Register (NTR; Van Beijsterveldt et al., 2013), and the Twin Early Development Study (TEDS; Haworth, Davis, & Plomin, 2013). These cohorts use various instruments to assess childhood aggression, among which are the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, Ivanova, & Rescorla, 2017; Achenbach & Rescorla, 2001) and the Strengths and Difficulties Questionnaire (Goodman, 2001; Goodman & Scott, 1999). It is unclear to what extent this heterogeneity affects comparability of results across cohorts, and how to best deal with heterogeneity due to different measures for childhood aggression.

Therefore, Chapter 6 examines the agreement between different measures of aggressive behavior in Chapter 6. The sample consists of 1,254 twin pairs from the Netherlands Twin Register for whom mother- and father-reports are available on aggressive behavior as assessed with the Autism - tics, attention-deficit hyperactivity disorder, and other comorbidities (A-TAC; Hansson et al., 2005), the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). This study assesses agreement between these measures with regard to convergence of their item content, concordance at the recommended clinical cutoff, correlation among the scores of the different scales, and the extent to which they measure the same underlying genetic mechanisms.

Finally, Chapter 7 provides a discussion of the results from this dissertation with the aim to improve the development and implementation of treatment and prevention programs for childhood aggressive behavior. Moreover, Chapter 7 translates the findings of this dissertation into its implications for our understanding of childhood aggression, its risk factors, and assessment. To this end, Chapter 7 provides a summary, a discussion of the implications, and a general conclusion of this dissertation.

Chapter 2.

# Childhood Aggression: A Synthesis of Reviews and Meta-Analyses to Reveal Patterns and Opportunities for Prevention and Intervention Strategies.

Published as: Hendriks, A. M., Bartels, M., Colins, O. F., & Finkenauer, C. (2018). Childhood aggression: A synthesis of reviews and meta-analyses to reveal patterns and opportunities for prevention and intervention strategies. *Neuroscience and Biobehavioral Reviews, 91*. https://doi.org/10.1016/j.neubiorev.2018.03.021

## ABSTRACT

This study provides a synthesis of meta-analyses and systematic reviews on non-pharmacological treatments for childhood aggression. Treatments referred to universal prevention, selective prevention, indicated prevention, or intervention (Mrazek & Haggerty, 1994). Seventy-two meta-analyses and systematic reviews met the inclusion criteria. We describe their characteristics, effect sizes across types of treatments, and the effects of various moderators. For universal and selective prevention, effects were mostly absent or small; for indicated prevention and interventions, effects were mostly small or medium. Only two moderators had a positive effect on treatment effectiveness, namely pre-test levels of aggression and parental involvement. These results identified similarities between indicated prevention and intervention treatments, on the one hand, and universal prevention and selective prevention, on the other. Our findings suggest that research distinguishing between targets of treatments (i.e., factors associated with childhood aggression vs. present aggressive behaviors) would be promising. Moreover, to further increase effectiveness of treatments for childhood aggression, individual differences warrant scientific attention.


### Highlights:

- The synthesis included 72 meta-analyses and systematic reviews
- Effect sizes for treatments for childhood aggression were mostly small
- Promising distinction between treating aggression vs. treating associated factors
- Treatment might benefit from a stronger emphasis on individual differences


**Keywords:**

Childhood aggression, prevention, intervention, meta-analysis, systematic review.




Childhood aggression and its social impairment inflict a tremendous personal and financial burden on affected children, their relatives, peers, and society as a whole (e.g., Dretzke et al., 2005; Fergusson, Horwood, & Ridder, 2005; Foster & Jones, 2005; Hunter, 2003; Knapp, Scott, & Davies, 1999; Scott, Knapp, Henderson, & Maughan, 2001). The prevalence of clinical aggression in children ranges from 2-16% (e.g., American Psychiatric Association, 1994; Merikangas, Nakamura, & Kessler, 2009; Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). Early onset childhood aggression continues into adolescence and adulthood for a substantial number of children (e.g., American Psychiatric Association, 1994; Huesmann, Dubow, & Boxer, 2009). Although treatments for childhood aggression are the most commonly studied amongst childhood disorders, their mean effect sizes are lower than those found for, for example, for childhood anxiety ($d$ = 0.46 vs. $d$ = 0.61; Weisz et al., 2017). Thus, insights in the treatment of aggression are essential.

Childhood aggression is a broad and complex construct. Problematic levels of aggression have their onset at different ages, with different underlying processes, and problems associated with aggression can express themselves in myriad forms (e.g., Barnes, Smith, & Miller, 2014; Bolhuis et al., 2017; Frick, 2001; Frick & Dickens, 2006; Tremblay, 2000). This diversity is reflected in various conditions in which aggression is the primary problem that are studied in the literature (e.g., conduct disorder, oppositional defiant disorder, externalizing behavior problems, antisocial behavior, disruptive behavior problems. In addition, the heterogeneity of childhood aggression is reflected in the many proposed subsets and dimensions of aggressive behaviors, for example, overt versus covert aggression (Crick, Casas, & Mosher, 2007), destructive versus nondestructive aggression (Frick et al., 1993), direct versus indirect aggression (Card, Stucky, Sawalani, & Little, 2008), and reactive versus proactive aggression (Raine et al., 2006). Yet, the only consensus in studies examining childhood aggression is that childhood aggression is common, that it may predict various psychosocial problems later on, and that it should be treated at early stages of development (e.g., Baker, 2009; Coie et al., 1993; Comer, Chow, Chan, Cooper-Vince, & Wilson, 2013; Connor et al., 2006; Frick & Dickens, 2006; Johnson et al., 2014).

Since 2000, the number of prevention and intervention strategies for childhood aggression has increased tremendously, an increase which is accompanied by a similar increase in scientific papers (Chorpita et al., 2011). Research shows, however, that prevention strategies and interventions for childhood aggression are more effective for some children than for others (Frick, 2001). The vast amount of information and the boundary conditions (i.e., moderators) of treatment effectiveness make it increasingly difficult to translate research results to practice and translate scientific findings to help those who suffer from childhood aggression, including children, parents, and teachers. Meta-analyses and reviews

have been published with the goal to structure and synthesize the abundance of findings and studies. Nevertheless, these studies offer little integration and mostly fail to consider prevention and intervention components simultaneously to identify effective components in the treatment of childhood aggression. Thus, to the authors' knowledge, no comprehensive systematic review and synthesis of the existing reviews and meta-analyses on treatments for childhood aggression exists. The present study seeks to fill this gap.

To distinguish between different types of prevention and intervention strategies for childhood aggression, we adopt the categorization presented by Mrazek and Haggerty (1994), consisting of universal prevention, selective prevention, indicated prevention, and intervention. *Universal prevention* aims at a population without any specified risk-factors for developing childhood aggression. *Selective prevention* aims at subgroups who have an elevated risk of developing childhood aggression (e.g., due to socioeconomic status, single-parent status), but who have not yet displayed behaviors associated with childhood aggression. *Indicated prevention* aims at subgroups who have an elevated risk to develop childhood aggression, and are identified as showing behaviors associated with childhood aggression but do not meet diagnostic criteria. Finally, *interventions* aim to treat diagnosed childhood aggression.

Although the literature typically differentiates between prevention and intervention research, we will focus on patterns between prevention and intervention of childhood aggression, given that they often include similar and overlapping components and clinical change strategies (Hoagwood, 2002; Sawyer, Borduin, & Dopp, 2015). As an example, indicated prevention and interventions mainly seem to differ in whether targeted children score above or below a certain diagnostic threshold of childhood aggression related disorders (Grove, Evans, Pastor, & Mack, 2008; Mrazek & Haggerty, 1994). Nevertheless, some authors suggest such a differentiation could be considered an arbitrary or artificial distinction (Boyle et al., 1996; Hoagwood, 2002; Sawyer et al., 2015). Therefore, we will refer to prevention and intervention as treatments in the following.

In this synthesis, we will follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement guidelines to identify, screen, and describe the reviews (Moher, Liberati, Tetzlaff, & Altman, 2009). It includes all non-pharmacological types of prevention and intervention identified above: Universal prevention, selective prevention, indicated prevention, and intervention. First, we provide a systematic review on the meta-analyses and systematic reviews on treatment effectiveness for childhood aggression. Second, we investigate the effectiveness of the types of treatments. Third, the present study reviews the influence of moderators – participant, treatment, and methodological variables – on the effectiveness of the treatment of childhood aggression. In the discussion, we will elaborate on patterns that occurred within the results and on the implications of those patterns for research and clinical practice.





# METHOD

## Literature Search

To identify the reviews and meta-analyses, we conducted a systematic literature search for systematic reviews and meta-analyses published in English between January 2000 and October 2017 in accordance with the PRISMA protocol (Moher et al., 2009). Table 1 provides an overview of the search terms and databases. In addition, we searched through reference lists of the identified articles for articles that did not appear in the electronic literature search.

## Inclusion and Exclusion Criteria

Articles were included in the present study if they: (1) were a meta-analysis and/or a systematic review studying treatment effectiveness on childhood aggression, (2) focused mainly on children aged 6 to 12, (3) were published in a peer-reviewed journal, and (4) were published in English. Childhood aggression in this study comprised of aggressive behavior, externalizing behavior, disruptive behavior problems, conduct disorder, oppositional behavior, oppositional defiant disorder problems, and antisocial behavior. Articles were eligible for inclusion if they mentioned effectiveness of a non-pharmacological treatment on childhood aggression in the title or abstract.

Because the focus of the present study was on childhood aggression in general populations, we excluded articles that examined aggression as comorbid symptom of another disorder (e.g., autism), traumatic life events, and developmental disabilities. For the same reason, we excluded articles examining the effect of treatment on specific variants and expressions of aggression, such as (cyber) bullying, delinquency, gang membership, truancy, recidivism, and violence. In addition, we excluded reviews or meta-analyses of single-subject/case studies.

**Table 1.** Search strategy: Databases and search terms. Keywords of different groups were combined with 'AND'.

| **Databases** | | | |
|---|---|---|---|
| **ERIC** | **PsycINFO** | **Pubmed** | **Review initiatives** |
| **Method** | **Method** | **Method** | Campbell Collaboration |
| Meta-analysis | Meta-analysis | Meta-analysis | Centre for Reviews and |
| Review | Review | Review | Dissemination |
| Systematic | Systematic | | Cochrane Collaboration |
| **Sample** | **Sample** | **Sample** | |
| Child | Child | Child | |
| **Outcome measure** | **Outcome measure** | **Outcome measure** | |
| Aggression | Aggression | Aggression | |
| Externalizing | Externalizing | Externalizing | |
| Externalising | Externalising | Externalising | |
| Oppositional | Oppositional | Oppositional | |
| Conduct disorder | Conduct disorder | Conduct disorder | |
| **Treatment** | **Treatment** | | |
| Intervention | Intervention | | |
| Prevention | Prevention | | |

## Data Extraction

We developed a coding sheet containing 41 variables, including age of participants, year of publication, language of the included articles, the number of included studies, moderators, and the results of the reviews and meta-analyses to extract information from the included reviews and meta-analyses. We also coded discrepancies between the study's definition of the treatment and our classification. To take the quality of each included systematic review and meta-analysis into account, we coded whether the study provided a description of the search terms and databases; whether it specified criteria for studies, participants, treatments, and measurement instruments; whether it explicitly described the process of inclusion and exclusion of the studies; whether it took study quality of the included studies into account; and whether it discussed the possibility of publication bias.

The first author extracted the data. To control for reliability, a trained graduate student coded a randomly drawn sample of 50% of the included articles. Questions and differences in coding were resolved through discussion until both coders reached full agreement. For the quantitative variables (i.e., number of included articles, effect sizes, lower and upper bound of included years), Cronbach's alphas for rater agreement based on 50% of the studies ranged between .99 and 1.00.

## Synthesis Strategy

We first described the literature search and discussed the characteristics of the included systematic reviews and meta-analyses. These characteristics consisted of variables related to sample size, range of years included, and study quality.



Second, for each treatment type (i.e., universal prevention, selective prevention, indicated prevention, and intervention), we extracted the effect sizes for comparison and discussion. We categorized all available effect sizes into no effect, small, medium, and large. For standardized mean differences (i.e., *Cohen's d, Hedges' g*), we considered effect sizes ranging from 0.2 to 0.49 to be small effects, effect sizes ranging from 0.5 to 0.79 to be medium effects, and effect sizes from 0.8 to be large effects (Lipsey & Wilson, 2000). Moreover, we included effect sizes below 0.2 that were significant in the category of small effects. For studies using an effect size measure that was less common (i.e., standard deviation reduction; Epstein, Fonnesbeck, Potter, Rizzone, & McPheeters, 2015), we adopted the size as reported by the authors. For unstandardized test statistics (weighted mean difference; Michelson, Davenport, Dretzke, Barlow, & Day, 2013), we reported the values without interpreting the size of the effect. When studies reported both weighted and unweighted effect sizes, we used the weighted effect size to avoid overestimation of effect sizes.

Third, we investigated the results for the moderators identified during the data extraction. These moderators included participant characteristic (e.g., child age, child gender, pre-treatment level of aggression, socioeconomic status), intervention characteristic (e.g., implementation, treatment, and session-related factors), and methodological characteristic (e.g., informant and research quality).

# RESULTS

## Literature Search

The literature search yielded 8,818 articles. Figure 1 displays the selection process. After removal of duplicates, the titles and abstracts of the identified papers were screened to determine their eligibility. Based on the initial screening of the abstract, we selected 111 papers for full-text screening; 72 articles fulfilled the criteria and were included. Because some systematic reviews also included effect sizes, for reasons of clarity, from here on we adopted the term study for each article, both systematic reviews and meta-analyses.

## Study Characteristics

The studies included articles published between 1950 and 2016. The amount of included articles in the studies ranged between 3 and 254. Ten percent of the studies (seven studies) included a maximum of ten articles, 26% (19 studies) included between 11 and 20 articles, 35% (25 studies) included between 21 and 50 articles, 19% (14 studies) included between 51 and 100 articles, 4% (three studies) included between 101 and 200 articles, 3% (two studies) included more than 200 studies. For 3% (two studies), it was uncertain how many articles related to childhood

aggression were included, because they only reported the total number of included articles (Chorpita et al., 2002, 2011). Seventy-two percent (52 studies) reported which databases and search terms were used, 25% (18 studies) reported only the databases, and 3% (two studies) reported neither. Sixty-four percent (46 studies) included only published articles, 36% (26 studies) also included book chapters and dissertations. Thirty-one percent (22 studies) evaluated publication bias. Forty-seven percent (34 studies) assessed the quality of the included articles, either by assessing methodological rigor, or with criteria including: Cochrane criteria, Critical Appraisal Skills Program, Jadad Scale, JAMA criteria, Methods Guide for Effectiveness and Comparative Effectiveness Reviews, Outcome Research Coding Protocol, PRISMA guidelines, Quality Index, Quality of Reporting Meta-analyses, and Task Force criteria.

The different type of treatment programs that were examined in the studies were: psychosocial treatments, cognitive behavioral treatments, parent training programs, school-based treatments, and other types, such as solution-focused brief therapy, (multi)systemic therapy, family therapy, media-based treatments, after-school programs, child-centered play therapy, and martial arts. Table 2 presents the frequencies of the different types of treatment programs across universal prevention, selective prevention, indicated prevention, and intervention. The most commonly studied moderators associated with participant characteristics were child age, child gender, pre-test levels of aggression, and socioeconomic status. The most commonly studied moderators associated with treatment characteristics were implementation, treatment, and session-related factors (i.e., intensity, frequency, and duration). The most commonly studied moderators associated with methodological characteristics were the informant and research quality. Table 3 presents moderator frequency across universal prevention, selective prevention, indicated prevention, and intervention.

## Effectiveness of Treatments for Childhood Aggression

We first examined the effectiveness of the four types of treatments. The effect sizes, type of treatments, and the outcome measures are displayed in Table 4, the percentages of the effect sizes are displayed in Table 5.



**Figure 1.** Flow chart of the literature search

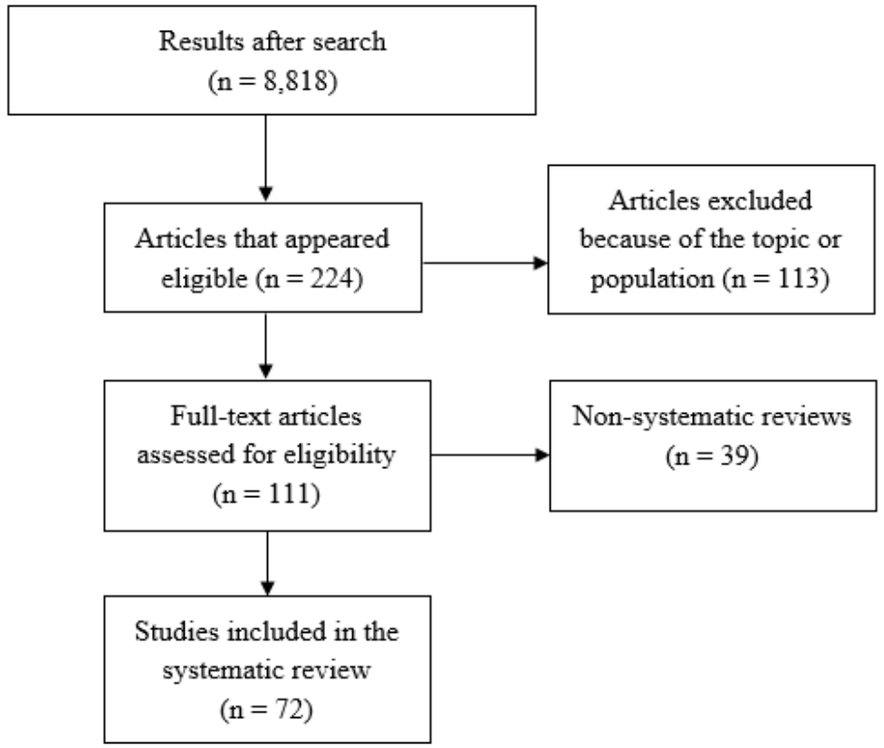


**Table 2.** Number of treatment programs for childhood aggression across types of treatments.

| Intervention components | Universal prevention | Nr. | Selective prevention | Nr. | Indicated prevention | Nr. | Intervention | Nr. |
|---|---|---|---|---|---|---|---|---|
| **Psychosocial treatments** | Franklin et al., 2017; Grove et al., 2008 | 2 | Chorpita et al., 2011; Connor et al., 2006; Farmer et al., 2002; Franklin et al., 2017; Greenberg, Domitrovich, & Bumbarger, 2001; Grove et al., 2008; Sawyer et al., 2015 | 7 | Chorpita et al., 2011; Connor et al., 2006; Epstein et al., 2015; Eyberg, Nelson, & Boggs, 2008; Franklin et al., 2017; Greenberg et al., 2001; Grove et al., 2008; Rosato et al., 2012; Sawyer et al., 2015; Weisz et al., 2013 | 10 | Bakker et al., 2017; Bradley & Mandell, 2005; Chorpita et al., 2002, 2011; Comer et al., 2013; Connor et al., 2006; Epstein et al., 2015; Farmer et al., 2002; Fossum et al., 2016, 2008; Franklin et al., 2017; Greenberg et al., 2001; Lee, Horvath, & Hunsley, 2013; Rosato et al., 2012; Sawyer et al., 2015; Weisz et al., 2013, 2017 | 17 |
| **Cognitive-behavioral** | Bennett & Gibbons, 2000; Grove et al., 2008; Smedler, Hjern, Wiklund, Anttila, & Pettersson, 2015 | 3 | Bennett & Gibbons, 2000; Chorpita et al., 2011; Grove et al., 2008; Smedler et al., 2015 | 4 | Bennett & Gibbons, 2000; Chorpita et al., 2011; Grove et al., 2008; McCart et al., 2006; Smedler et al., 2015; Smeets et al., 2015; Sukhodolsky et al., 2004; Weisz et al., 2013 | 8 | Battagliese et al., 2015; Bennett & Gibbons, 2000; Chorpita et al., 2002, 2011; Fossum et al., 2016; Furlong et al., 2012; McCart et al., 2006; Smeets et al., 2015; Sukhodolsky et al., 2004; Weisz et al., 2017, 2013 | 11 |
| **Parent training** | Kaminski, Valle, Filene, & Boyle, 2008; Leijten et al., 2013; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Shelleby & Shaw, 2014; Thomas, Abell, Webb, Avdagic, & Zimmer-Gembeck, 2017 | 6 | Chorpita et al., 2011; De Graaf et al., 2008; Farmer et al., 2002; Kaminski et al., 2008; Leijten et al., 2013; Lundahl et al., 2006; Menting et al., 2013; Michelson et al., 2013; Nowak & Heinrichs, 2008; Shelleby & Shaw, 2014; Thomas et al., 2017; Tully & Hunt, 2016 | 12 | Barlow & Stewart-Brown, 2000; Briggs, Cox, Sharkey, Briggs, & Black, 2015; Buchanan-Pascall et al., 2017; Chorpita et al., 2011; De Graaf et al., 2008; Dretzke et al., 2005; Farmer et al., 2002; Gavita & Joyce, 2008; Kaminski et al., 2008; Leijten et al., 2013; Lundahl et al., 2006; McCart et al., 2006; Menting et al., 2013; Michelson et al., 2013; Nowak & Heinrichs, 2008; Reyno & McGrath, 2006; Shelleby & Shaw, 2014; Thomas et al., 2017; Thomas & Zimmer-Gembeck, 2007; Tully & Hunt, 2016 | 20 | Barlow & Stewart-Brown, 2000; Briggs et al., 2015; Buchanan-Pascall et al., 2017; Chorpita et al., 2002, 2011; De Graaf et al., 2008; Dretzke et al., 2005, 2009; Furlong et al., 2012; Gavita & Joyce, 2008; Kaminski et al., 2008; Lee et al., 2013; Leijten et al., 2013; Lundahl et al., 2006; Maughan et al., 2005; McCart et al., 2006; Menting et al., 2013; Michelson et al., 2013; Nowak & Heinrichs, 2008; Reyno & McGrath, 2006; Shelleby & Shaw, 2014; Tarver et al., 2014; Thomas et al., 2017; Thomas & Zimmer-Gembeck, 2007; Tully & Hunt, 2016 | 25 |

**Table 2.** Continued

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **School-based treatment** | Barnes et al., 2014; Durlak & Weissberg, 2007; Durlak et al., 2011; Farahmand, Grant, Polo, & Duffy, 2011; Franklin et al., 2017; Gansle, 2005; Hahn et al., 2007; Oliver et al., 2011; Park-Higgerson et al., 2008; Ray, Armstrong, Balkin, & Jayne, 2015; Wilson, Gottfredson, & Najaka, 2001; Wilson & Lipsey, 2006, 2007; Wilson et al., 2003 | 14 | Barnes et al., 2014; Durlak & Weissberg, 2007; Farahmand et al., 2011; Franklin et al., 2017; Gansle, 2005; Hahn et al., 2007; Ray et al., 2015; Wilson et al., 2001; Wilson & Lipsey, 2007; Wilson et al., 2003 | 10 | Durlak & Weissberg, 2007; Farahmand et al., 2011; Franklin et al., 2017; Gansle, 2005; Ray et al., 2015; Stoltz et al., 2012; Wilson et al., 2001; Wilson & Lipsey, 2007; Wilson et al., 2003 | 9 | Durlak & Weissberg, 2007; Farahmand et al., 2011; Franklin et al., 2017; Ray et al., 2015; Wilson et al., 2001; Wilson & Lipsey, 2007; Wilson et al., 2003 | 7 |
| **Other** | Candelaria, Fedewa, & Ahn, 2012; Hale et al., 2014; Harwood, Lavidor, & Rassovsky, 2017; Montgomery & Maunders, 2015; Ray et al., 2015 | 5 | Candelaria et al., 2012; Chorpita et al., 2011; Farmer et al., 2002; Hale et al., 2014; Harwood et al., 2017; Kremer et al., 2014; Montgomery & Maunders, 2015; Ray et al., 2015; Von Sydow, Retzlaff, Beher, Haun, & Schweitzer, 2013 | 9 | Bunge et al., 2016; Candelaria et al., 2012; Chorpita et al., 2011; Harwood et al., 2017; Montgomery & Maunders, 2015; Ray et al., 2015; Von Sydow et al., 2013 | 7 | Bond, Woods, Humphrey, Symes, & Green, 2013; Bunge et al., 2016; Candelaria et al., 2012; Chorpita et al., 2002, 2011; Farmer et al., 2002; Fossum et al., 2016; Harwood et al., 2017; Montgomery, Bjornstad, & Dennis, 2006; Montgomery & Maunders, 2015; Ray et al., 2015; Tarver et al., 2014; Von Sydow et al., 2013 | 13 |

**Table 3.** Numbers of the studied moderators for treatment effectiveness for childhood aggression.

| Intervention components | Universal prevention | Nr. | Selective prevention | Nr. | Indicated prevention | Nr. | Intervention | Nr. |
|---|---|---|---|---|---|---|---|---|
| Child age | Barnes et al., 2014; Franklin et al., 2017; Grove et al., 2008; Hahn et al., 2007; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Park-Higgerson et al., 2008; Wilson & Lipsey, 2006 | 8 | Barnes et al., 2014; Franklin et al., 2017; Grove et al., 2008; Hahn et al., 2007; Kremer et al., 2014; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 8 | Franklin et al., 2017; Grove et al., 2008; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Smeets et al., 2015; Stoltz et al., 2012; Sukhodolsky et al., 2004 | 8 | Bakker et al., 2017; Comer et al., 2013; Erford et al., 2014; Fossum et al., 2016, 2008; Franklin et al., 2017; Lundahl et al., 2006; Maughan et al., 2005; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Smeets et al., 2015; Stoltz et al., 2012; Sukhodolsky et al., 2004 | 13 |
| Child gender | Barnes et al., 2014; Franklin et al., 2017; Grove et al., 2008; Nowak & Heinrichs, 2008; Wilson & Lipsey, 2006 | 5 | Barnes et al., 2014; De Graaf et al., 2008; Franklin et al., 2017; Grove et al., 2008; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 6 | De Graaf et al., 2008; Franklin et al., 2017; Grove et al., 2008; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Smeets et al., 2015 | 6 | Bakker et al., 2017; Comer et al., 2013; De Graaf et al., 2008; Erford et al., 2014; Fossum et al., 2008; Franklin et al., 2017; Maughan et al., 2005; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Smeets et al., 2015 | 10 |
| Pre-test levels of aggression | Bennett & Gibbons, 2000; Leijten et al., 2013; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Wilson et al., 2003 | 5 | Bennett & Gibbons, 2000; De Graaf et al., 2008; Leijten et al., 2013; Lundahl et al., 2006; Menting et al., 2013; Nowak & Heinrichs, 2008; Wilson et al., 2003 | 7 | Bennett & Gibbons, 2000; De Graaf et al., 2008; Leijten et al., 2013; Lundahl et al., 2006; Menting et al., 2013; Nowak & Heinrichs, 2008; Stoltz et al., 2012; Sukhodolsky et al., 2004; Wilson et al., 2003 | 9 | Bennett & Gibbons, 2000; De Graaf et al., 2008; Leijten et al., 2013; Lundahl et al., 2006; Menting et al., 2013; Nowak & Heinrichs, 2008; Stoltz et al., 2012; Sukhodolsky et al., 2004; Wilson et al., 2003 | 9 |
| SES | Leijten et al., 2013; Lundahl et al., 2006; Wilson & Lipsey, 2006 | 3 | Leijten et al., 2013; Lundahl et al., 2006 | 2 | Leijten et al., 2013; Lundahl et al., 2006 | 2 | Leijten et al., 2013; Lundahl et al., 2006 | 2 |
| Implementation | Barnes et al., 2014; Durlak et al., 2011; Franklin et al., 2017; Lundahl et al., 2006; Park-Higgerson et al., 2008; Wilson & Lipsey, 2006; Wilson et al., 2003 | 7 | Barnes et al., 2014; Franklin et al., 2017; Lundahl et al., 2006; Sawyer et al., 2015; Wilson et al., 2003 | 5 | Franklin et al., 2017; Lundahl et al., 2006; Sawyer et al., 2015; Smeets et al., 2015; Wilson et al., 2003 | 5 | Bakker et al., 2017; Erford et al., 2014; Franklin et al., 2017; Lundahl et al., 2006; Maughan et al., 2005; Sawyer et al., 2015; Smeets et al., 2015; Wilson et al., 2003 | 8 |

**Table 3.** Continued

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Treatment | Dymnicki, Weissberg, & Henry, 2011; Grove et al., 2008; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Park-Higgerson et al., 2008 | 5 | Farmer et al., 2002; Grove et al., 2008; Kremer et al., 2014; Lundahl et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 6 | Epstein et al., 2015; Farmer et al., 2002; Grove et al., 2008; Lundahl et al., 2006; McCart et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Stoltz et al., 2012 | 8 | Bakker et al., 2017; Battagliese et al., 2015; Epstein et al., 2015; Fossum et al., 2008; Lundahl et al., 2006; McCart et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Stoltz et al., 2012 | 9 |
| Intensity, frequency, duration | Gansle, 2005; Wilson & Lipsey, 2006; Wilson et al., 2003 | 3 | Kremer et al., 2014; Sawyer et al., 2015; Wilson et al., 2003 | 3 | Buchanan-Pascall et al., 2017; Gansle, 2005; McCart et al., 2006; Sukhodolsky et al., 2004; Wilson et al., 2003 | 5 | Bakker et al., 2017; Battagliese et al., 2015; Buchanan-Pascall et al., 2017; Erford et al., 2014; Fossum et al., 2016, 2008; Gansle, 2005; Maughan et al., 2005; McCart et al., 2006; Sawyer et al., 2015; Sukhodolsky et al., 2004; Wilson et al., 2003 | 12 |
| Informant | Bennett & Gibbons, 2000; Wilson & Lipsey, 2006 | 2 | Bennett & Gibbons, 2000; Menting et al., 2013; Sawyer et al., 2015 | 3 | Bennett & Gibbons, 2000; Menting et al., 2013; Sawyer et al., 2015 | 3 | Battagliese et al., 2015; Bennett & Gibbons, 2000; Dretzke et al., 2009; Fossum et al., 2008; Maughan et al., 2005; Menting et al., 2013; Sawyer et al., 2015; Tarver et al., 2014; Weisz et al., 2017 | 9 |
| Research quality | Barnes et al., 2014; Bennett & Gibbons, 2000; Nowak & Heinrichs, 2008; Wilson & Lipsey, 2006 | 4 | Barnes et al., 2014; Bennett & Gibbons, 2000; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 4 | Barnes et al., 2014; Bennett & Gibbons, 2000; Buchanan-Pascall et al., 2017; McCart et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 6 | Barnes et al., 2014; Bennett & Gibbons, 2000; Buchanan-Pascall et al., 2017; Erford et al., 2014; Fossum et al., 2008; Maughan et al., 2005; McCart et al., 2006; Nowak & Heinrichs, 2008; Sawyer et al., 2015 | 9 |

**Table 4.** Effect sizes for different treatments of childhood aggression. Effect sizes of multiple types of treatments calculated simultaneously were indicated with an asterisk.

| Article | Universal prevention | Selective prevention | Indicated prevention | Intervention | Type of program | Outcome measure |
|---|---|---|---|---|---|---|
| Bakker et al., 2017 | | | | *d* = 0.01 – 0.37 | Psychological treatments | Conduct disorder or conduct problems |
| Barlow & Stewart-Brown, 2000 | | | x | x | Group parent education programs | Externalizing behavior |
| Barnes et al., 2014 | *d* = -0.14* | *d* = -0.14* | | | School-based cognitive-behavioral intervention | Aggression |
| Battagliese et al., 2015 | | | | *d* = -0.52 | School-based cognitive behavioral interventions | Externalizing behavior (including ADHD) |
| Bennett & Gibbons, 2000 | *d* = 0.23* | *d* = 0.23* | *d* = 0.23* | *d* = 0.23* | Cognitive behavior therapy | Antisocial behavior |
| Bond et al., 2013 | | | | x | Solution-focused brief therapy | Externalizing behavior |
| Bradley & Mandell, 2005 | | | | *SMD* = 0.25 – 1.06 | Treatments for ODD | ODD |
| Briggs et al., 2015 | | | x | x | Single-parent group interventions | Child compliance |
| Buchanan-Pascall et al., 2017 | | | *g* = 0.38 | *g* = 0.38 | Parent group interventions | Externalizing problems |
| Bunge et al., 2016 | | | x | x | Cognitive behavioral intervention technologies | CD, ODD, ADHD |
| Candelaria, Fedewa, & Ahn, 2012 | *d* = 0.34 | *d* = 0.34 | *d* = 0.34 | *d* = 0.34 | Anger management programs | Aggression |
| Chorpita et al., 2002 | | | | x | Empirically based treatments | CD and ODD |
| Chorpita et al., 2011 | | x | x | x | Empirically based treatments | Disruptive behavior |
| Comer et al., 2013 | | | | *g* = 0.71 – 0.90 | Psychosocial treatments | Disruptive behavior |
| Connor et al., 2006 | | x | x | x | Prevention programs and psychosocial treatments | Aggression, conduct problems, antisocial behaviors, and violence. |
| De Graaf et al., 2008 | | *d* = 0.88* | *d* = 0.88* | *d* = 0.88* | Triple P level 4 | Disruptive behavior problems |
| Dretzke et al., 2005 | | | *SMD* = -0.35 – -0.73* | *SMD* = -0.35 – -0.73* | Parent training | CD |

**Table 4.** Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Dretzke et al., 2009 | | | | *SMD* = 0.40 – 0.67 | Parent training | CD |
| Durlak et al., 2011 | *g* = 0.17 – 0.26 | | | | School-based social and emotional learning | Conduct problems |
| Durlak & Weissberg, 2007 | *SMD* = 0.18 | *SMD* = 0.18 | *SMD* = 0.18 | *SMD* = 0.18 | After school programs | Noncompliance, aggression, delinquent acts, disciplinary referrals, rebelliousness, and other types of conduct problems |
| Dymnicki et al., 2011 | *g* = 0.11 | | | | Elementary school based programs | Overt aggression |
| Epstein et al., 2015 | | | 1.2 SD decrease | 1.2 SD decrease | Psychosocial interventions | Disruptive behaviors |
| Erford et al., 2014 | | | | *d+* = 0.36 – 0.68 | Counseling or psychotherapy | ODD |
| Eyberg et al., 2008 | | | x | x | Psychosocial treatments | Disruptive behavior |
| Farahmand et al., 2011 | *g* = 0.02* | *g* = 0.02* | *g* = 0.02* | *g* = 0.02* | School-based mental health and behavioral programs | Conduct problems |
| Farmer et al., 2002 | | x | x | | Treatment approaches with demonstrated evidence | Disruptive behavior |
| Fossum et al., 2008 | | | | *d* = 0.62 | Psychotherapy | Disruptive, aggressive, and oppositional behaviors |
| Fossum et al., 2016 | | | | *d* = 0.64 | Psychological interventions | Conduct problems |
| Franklin et al., 2017 | *d* = 0.02 | *d* = 0.02 | *d* = 0.02 | *d* = 0.02 | Teacher-delivered psychosocial interventions | Externalizing behaviors |
| Furlong et al., 2012 | | | | *SMD* = -0.53 | Behavioral and cognitive-behavioral group-based parenting programs | Conduct problems |
| Gansle, 2005 | *d* = 0.54* | | *d* = 0.54* | *d* = 0.54* | School-based interventions | Anger and externalizing behavior |
| Gavita & Joyce, 2008 | | | *d* = 0.75* | *d* = 0.75* | Group based cognitively enhanced parent training | Disruptive or externalizing behavior |

**Table 4.** Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Greenberg et al., 2001 | x | x | x | | Prevention programs | Externalizing behavior |
| Grove et al., 2008 | *d* = 0.17* | *d* = 0.17* | *d* = 0.17* | | Prevention programs | Behavior referrals, aggression, delinquency, conduct problems, arrests or court contacts, oppositional symptoms |
| Hahn et al., 2007 | x | x | | | Universal school-based programs | Violent and aggressive behavior |
| Hale et al., 2014 | x | x | | | Family, community, or Web-based prevention | Aggression |
| Harwood et al., 2017 | *d* = 0.38 – 0.65 | *d* = 0.38 – 0.65 | *d* = 0.38 – 0.65 | *d* = 0.38 – 0.65 | Martial arts | Externalizing behaviors |
| Kaminski et al., 2008 | *SMD* = 0.25* | *SMD* = 0.25* | *SMD* = 0.25* | *SMD* = 0.25* | Parent training | Child behavior problems |
| Kremer et al., 2014 | | *g* = 0.11 | | | After-school programs | Externalizing behavior |
| Lee et al., 2013 | | | | x | Psychological treatments | Disruptive behavior problems |
| Leijten et al., 2013 | x | x | x | x | Parent training | Disruptive behavior problems |
| Lösel & Beelmann, 2003 | *d* = 0.26* | *d* = 0.26* | *d* = 0.26* | | Child skills training | Antisocial behavior |
| Lundahl et al., 2006 | *d* = 0.42* | *d* = 0.42* | *d* = 0.42* | *d* = 0.42* | Parent training | Externalizing behavior |
| Maughan et al., 2005 | | | | *d* = 0.30 | Behavioral parent training | Externalizing behavior |
| McCart et al., 2006 | | | *d* = 0.40* | *d* = 0.40* | Behavioral parent training and cognitive-behavioral therapy for youth | Antisocial behavior |
| Menting et al., 2013 | | *d* = 0.27* | *d* = 0.27* | *d* = 0.27* | Incredible Years parent training | Disruptive behavior |
| Michelson et al., 2013 | | ECBI-I -20.90 (p < .001), ECBI-P -6.03 (p < .001), CBCL Ext -3.66 (p = .006), and SDQ CD -.059 (p < .001) | ECBI-I -20.90 (p < .001), ECBI-P -6.03 (p < .001), CBCL Ext -3.66 (p = .006), and SDQ CD -.059 (p < .001) | ECBI-I -20.90 (p < .001), ECBI-P -6.03 (p < .001), CBCL Ext -3.66 (p = .006), and SDQ CD -.059 (p < .001) | Parent Management Training | Disruptive behavior |
| Montgomery & Maunders, 2015 | x | x | x | x | Creative bibliotherapy | Externalizing behavior |

**Table 4.** Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Montgomery et al., 2006 | | | | x | Media-based cognitive-behavioral treatments | Externalizing behavior problems |
| Nowak & Heinrichs, 2008 | *g* = 0.35* | *g* = 0.35* | *g* = 0.35* | *g* = 0.35* | Triple P | Child behavior problems |
| Oliver et al., 2011 | *SMD* = 0.18 – 0.20 | | | | Teachers' classroom management | Problem behaviors |
| Park-Higgerson et al., 2008 | *SMD* = -0.09 | | | | School-based prevention programs | Externalizing, aggressive, or violent behavior |
| Ray et al., 2015 | *d* = 0.34* | *d* = 0.34* | *d* = 0.34* | *d* = 0.34* | Child-centered play therapy | Externalizing behaviors |
| Reyno & McGrath, 2006 | | | x | x | Parent training | Externalizing behavior problems |
| Rosato et al., 2012 | | | x | x | Psychosocial interventions | Overt aggression |
| Sawyer et al., 2015 | | *d* = 0.25* | *d* = 0.25* | *d* = 0.41 | Psychosocial interventions | Antisocial behavior |
| Shelleby & Shaw, 2014 | x | x | x | x | Parenting interventions | Conduct problems |
| Smedler et al., 2015 | x | x | x | | Prevention programs | Externalizing behavior |
| Smeets et al., 2015 | | | *d* = 0.50* | *d* = 0.50* | Cognitive behavior therapy | Aggressive behavior |
| Stoltz et al., 2012 | | | *d* = 0.30* | *d* = 0.30* | Individual interventions with or without additional components | Externalizing behavior |
| Sukhodolsky et al., 2004 | | | *d* = 0.67* | *d* = 0.67* | Cognitive behavior therapy | Anger or aggression |
| Tarver et al., 2014 | | | | *SMD* = 1.01 | Self-directed parenting interventions | Externalizing behavior |
| Thomas et al., 2017 | *SMD* = .87 | *SMD* = .87 | *SMD* = .87 | *SMD* = .87 | Parent-child interaction therapy | Externalizing behavior problems |
| Thomas & Zimmer-Gembeck, 2007 | | | *d* = -2.16 - -0.31* | *d* = -2.16 - -0.31* | PCIT and Triple P | Aggression, extreme tantrums, and opposition |
| Tully & Hunt, 2015 | | x | x | x | Brief parenting interventions | Externalizing behavior |
| Von Sydow et al., 2013 | | | x | x | Systemic therapy | Externalizing disorders |

**Table 4.** Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Weisz et al., 2013 | | | *d* = 0.31* | *d* = 0.31* | Evidence-based psychotherapies | Externalizing behaviors |
| Weisz et al., 2017 | | | | *d* = 0.46 | Psychological therapy | Conduct problems |
| Wilson et al., 2001 | *d* = 0.17* | *d* = 0.17* | *d* = 0.17* | *d* = 0.17* | School-based prevention programs | Aggressive behavior |
| Wilson et al., 2003 | *d* = 0.16 – 0.32* | *d* = 0.16 – 0.32* | *d* = 0.16 – 0.32* | *d* = 0.16 – 0.32* | School-based demonstration and routine practice programs | Aggressive behavior |
| Wilson & Lipsey, 2006 | *SMD* = 0.21 | | | | School-based social information processing interventions | Aggressive behavior |
| Wilson & Lipsey, 2007 | *d* = 0.21 | *d* = 0.29* | *d* = 0.29* | | School-based prevention programs | Aggressive behavior |

**Table 5.** Frequencies and percentages of effect sizes of different types of treatments. For standardized mean differences, we considered effect sizes from 0.2 and effect sizes below 0.2 that were significant to be small effects, from 0.5 to be medium effects, and from 0.8 to be large effects. Other effects include: no effect to small, small to medium, small to large, and medium to large effects. If a systematic review or meta-analysis reported effect sizes for multiple types of treatments, we included them all.

| | **Universal prevention** | **Selective prevention** | **Indicated prevention** | **Intervention** | **Total** |
|---|---|---|---|---|---|
| No effect | 4 (17%) | 4 (19%) | 2 (7%) | 2 (5%) | 12 (11%) |
| Small effect | 16 (70%) | 14 (67%) | 18 (60%) | 17 (44%) | 65 (58%) |
| Medium effect | 1 (4%) | 0 (0%) | 5 (17%) | 9 (23%) | 15 (13%) |
| Large effect | 1 (4%) | 2 (10%) | 2 (7%) | 3 (8%) | 8 (7%) |
| Other | 1 (4%) | 1 (5%) | 3 (10%) | 8 (21%) | 13 (12%) |
| Total | 23 (20%) | 21 (19%) | 30 (27%) | 39 (35%) | 113 (100%) |

**Universal prevention.** Twenty-three studies (32% of total) reported effect sizes for the effectiveness of universal prevention programs. Seventeen percent of these studies found no effect. Seventy percent of these studies found a small effect. Four percent of these studies found a medium effect. Four percent of these studies found a large effect size. Four percent of these studies found a small to medium effect.

**Selective prevention.** Twenty-one studies (29% of total) reported effect sizes for selective prevention. Nineteen percent of these studies found no effect. Sixty-seven percent of these studies found a small effect. None of these studies found a medium effect. Ten percent of these studies found a large effect. Five percent of the studies found a small to medium effect.

**Indicated prevention.** Thirty studies (42% of total) reported effect sizes for indicated prevention. Seven percent of these studies found no effect. Sixty percent of these studies found a small effect. Seventeen percent of these studies found a medium effect. Seven percent of these studies found a large effect. Six percent of these studies found effects ranging between small and medium. Three percent of these studies found effects ranging between small and large.

**Intervention.** Thirty-nine studies (54% of total) reported effect sizes for intervention. Five percent of these studies found no effect. Forty-four percent of these studies found a small effect. Twenty-three percent of these studies found a medium effect. Eight percent of these studies found a large effect. Three percent of these studies found effects ranging between no effect and a small effect. Eleven percent of these studies found effects ranging between small and medium. Five percent of these studies found effects ranging between small and large. Three percent of these studies found effects ranging between medium and large.

**Summary.** Overall, the majority of reported effect sizes (61%) were on indicated prevention and interventions. The most prevalent category of effects for all types of treatments was a small effect (65%). For universal and selective prevention effects were mostly absent or small, whereas for indicated prevention and intervention effects were mostly small or medium.

## Moderating Variables

We investigated the results of the included studies for commonly investigated moderators. These moderators included participant characteristics, intervention characteristics, and methodological characteristics.

**Participant characteristics. *Child age.*** Nineteen of the studies (26% of total) took age into account as a moderator of treatment effectiveness for childhood aggression. Sixteen percent of these studies found larger treatment effectiveness for younger children (Fossum, Handegård, Adolfsen, Vis, & Wynn, 2016; Nowak & Heinrichs, 2008; Stoltz, Londen, Dekovic, Castro, & Prinzie, 2012). Eleven percent found larger treatment effectiveness for older children (Comer et al., 2013; Park-

Higgerson, Perumean-Chaney, Bartolucci, Grimley, & Singh, 2008). Five percent found no effect of age between groups, but did find stronger effects for younger children when looking at within-group effect sizes (Fossum, Handegård, Martinussen, & Mørch, 2008). Five percent found that treatments were more effective for younger (3-5 years old) and older children (9-11 years old) but less effective in between for children aged 6-8 (Maughan, Christiansen, Jenson, Olympia, & Clark, 2005). Finally, five percent found that treatments were less effective for children in elementary and middle school compared to kindergarten and high school (Hahn et al., 2007). Fifty-eight percent of these studies found that child age did not have a significant moderating effect (Bakker, Greven, Buitelaar, & Glennon, 2017; Barnes et al., 2014; Erford, Paul, Oncken, Kress, & Erford, 2014; Franklin et al., 2017; Grove et al., 2008; Kremer, Maynard, Polanin, Vaughn, & Sarteschi, 2014; Lundahl, Risser, & Lovejoy, 2006; Sawyer et al., 2015; Smeets et al., 2015; Sukhodolsky, Kassinove, & Gorman, 2004; Wilson & Lipsey, 2006).

***Gender.*** Thirteen studies (18% of total) included child gender as a moderator for treatment effectiveness in reducing childhood aggression. Eight percent found that treatment effectiveness was larger for boys (Comer et al., 2013), while the remaining eight percent found that treatment effectiveness was larger for girls (De Graaf, Speetjens, Smit, De Wolff, & Tavecchio, 2008). Eighty-five percent of these studies found no significant moderating effect (Bakker et al., 2017; Barnes et al., 2014; Erford et al., 2014; Fossum et al., 2008; Franklin et al., 2017; Grove et al., 2008; Maughan et al., 2005; Nowak & Heinrichs, 2008; Sawyer et al., 2015; Smeets et al., 2015; Wilson & Lipsey, 2006).

***Pre-treatment level of aggression.*** Nine studies (13% of total) included children's levels of aggression prior to treatment as a moderator. Sixty-seven percent of these studies found a positive association between pre-treatment levels of aggression and treatment effectiveness for childhood aggression (De Graaf et al., 2008; Leijten, Raaijmakers, De Castro, & Matthys, 2013; Lundahl et al., 2006; Menting, Orobio de Castro, & Matthys, 2013; Sukhodolsky et al., 2004; Wilson, Lipsey, & Derzon, 2003). Thirty-three percent found that this factor did not moderate treatment effectiveness (Bennett & Gibbons, 2000; Nowak & Heinrichs, 2008; Stoltz et al., 2012).

***Socioeconomic status.*** Three studies (4% of total) included socioeconomic status (SES) as a moderator. The first of these studies found that treatments were more effective for families with a higher SES (Lundahl et al., 2006). In contrast, the second study found that treatments were more effective for low SES compared to higher/mixed SES (Wilson & Lipsey, 2006). Finally, the third study found that SES interacted with pre-treatment levels of aggression, suggesting that disadvantaged samples improved less due to treatment when they had lower levels of aggression at pre-test (Leijten et al., 2013).



**Treatment characteristics.** ***Implementation.*** Seven studies (10% of total) examined whether a treatment was implemented to groups or individuals. Forty-three percent of these studies found that treatments for childhood aggression were more effective when implemented individually (Lundahl et al., 2006; Maughan et al., 2005; Nowak & Heinrichs, 2008). Fifty-seven percent of did not find that including group vs. individual implementation moderated treatment effectiveness (Bakker et al., 2017; Erford et al., 2014; Franklin et al., 2017; Smeets et al., 2015).

Seven studies (10% of total) included the person who implemented the treatment. Fourteen percent of these studies found larger effects for specialist-implemented programs compared to teacher-implemented programs (Park-Higgerson et al., 2008). Fourteen percent found that treatments implemented by researchers had larger effects compared to treatments implemented by professionals and paraprofessionals (Sawyer et al., 2015). Fourteen percent found that treatments implemented by teachers had a larger effect than interventions implemented by researchers (Wilson et al., 2003). Forty-three percent found that whether the treatment was implemented by a professional did not moderate treatment effectiveness (Barnes et al., 2014; Maughan et al., 2005; Wilson & Lipsey, 2006). Fourteen percent did not find a difference between implementation by teachers or non-school personnel (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

***Treatment.*** Five studies (7% of total) examined whether the global type of treatment moderated effectiveness. Twenty percent of these studies found a positive effect for selective prevention compared to universal prevention (Park-Higgerson et al., 2008) and 20 percent found a positive effect for universal prevention compared to selective prevention (Barnes et al., 2014). Forty percent found stronger effects for intervention compared to prevention (Nowak & Heinrichs, 2008; Sawyer et al., 2015). Twenty percent found no moderating effect of prevention type (i.e., universal vs. selective vs. indicated prevention; Grove et al., 2008).

Five studies (7% of total) included the specific type of treatment component as a moderator (e.g., cognitive behavioral therapy, parent training). Twenty percent of these studies found larger effects for behavioral therapy than for family therapy (Fossum et al., 2008), while in contrast 20 percent found larger effects for behavioral parent training than for cognitive behavioral therapy (McCart, Priester, Davies, & Azen, 2006). Sixty percent found no effect (Kremer et al., 2014; Sawyer et al., 2015; Stoltz et al., 2012).

Five studies (7% of total) examined the moderating effect of parental involvement. Twenty percent of these studies found that treatments with a parent component were more effective, either alone or combined with other components (Epstein et al., 2015). Forty percent found that cognitive-behavioral treatments were more effective when they were delivered to both parents and children (Battagliese et al.,

2015; Farmer, Compton, Bums, & Robertson, 2002). Forty percent found no difference between treatments aimed at parents, children, or multiple systems (Bakker et al., 2017; Lundahl et al., 2006).

***Session-related factors.*** Fourteen studies (19% of total) focused on treatment intensity, including number of sessions, session duration, and treatment intensity, yielding 19 moderator effects. Five percent of these studies found that number of sessions per week in one study did not have an effect (Battagliese et al., 2015) and 26 percent found that session duration had no effect (Bakker et al., 2017; Buchanan-Pascall, Gray, Gordon, & Melvin, 2017; Erford et al., 2014; Sawyer et al., 2015; Wilson & Lipsey, 2006). In contrast, 11 percent found larger effects for longer durations of treatment (Gansle, 2005; Wilson & Lipsey, 2006) and five percent found larger effects for higher treatment intensity (Wilson et al., 2003). Finally, five percent found a negative moderating effect of number of sessions, indicating smaller effects for more sessions (Maughan et al., 2005). Forty-seven percent found that number of sessions did not significantly moderate treatment effectiveness (Bakker et al., 2017; Battagliese et al., 2015; Erford et al., 2014; Fossum et al., 2016, 2008; Kremer et al., 2014; McCart et al., 2006; Sawyer et al., 2015; Sukhodolsky et al., 2004)

**Methodological characteristics.** ***Informant.*** Ten studies (14% of total) included the informant of childhood aggression as a moderator. Thirty percent of these studies found larger effects for parent-reports compared to independent observations (Dretzke et al., 2009; Maughan et al., 2005; Tarver, Daley, Lockwood, & Sayal, 2014). Ten percent found larger effects for parent-reports compared to teacher-reports (Battagliese et al., 2015). Ten percent found larger effects for parent-reports compared to teacher- and self-reports (Weisz et al., 2017). Ten percent yielded larger effect for observations by researchers compared to parent- or teacher-report (Menting et al., 2013). Forty percent found no effect (Bennett & Gibbons, 2000; Fossum et al., 2016; Sawyer et al., 2015; Wilson & Lipsey, 2006).

***Research quality.*** There were ten studies (14% of total) that included research quality as a moderator, yielding 14 effect sizes. Twenty-one percent of these studies found a negative effect of an overarching measure of research quality (e.g., a score based on sample size, random assignment, low attrition rates, inclusion of one normed/blinded outcome measure, presence of an attention placebo control group, and whether posttest data was reported for all pre-test measures; (Bennett & Gibbons, 2000; McCart et al., 2006; Nowak & Heinrichs, 2008). Seven percent found greater treatment effectiveness for studies with a low risk of bias compared to studies with a high or unknown risk (Buchanan-Pascall et al., 2017). Twenty-one percent found that whether a sample was assigned randomly did not moderate treatment effectiveness (Barnes et al., 2014; Sawyer et al., 2015; Wilson & Lipsey, 2006). Seven percent found that random assignment had a positive effect on treatment effectiveness (Nowak & Heinrichs, 2008). Seven percent found that random assignment had a negative effect on treatment effectiveness (Maughan et al., 2005). Seven percent found that whether research included an assessment of reliability had a negative moderating effect on treatment effectiveness (Maughan et al., 2005). Seven percent found that the presence of diagnostic information positively moderated treatment effectiveness (Fossum et al., 2008). Seven percent found no effect of whether the program was studied by the developer (Wilson & Lipsey, 2006). Seven percent found no effect of blind assessment (Erford et al., 2014). Seven percent found no effect of sample size (Erford et al., 2014).

**Summary.** To sum up, the effects of moderating variables on the effectiveness of treatments for childhood aggression were mixed. In the majority of studies including age as a moderator (58%), there was no moderating effect. For studies including child gender, 85% of the studies found no moderating effect. For studies including pre-test levels of aggression, 67% of the studies found a positive moderating effect, indicating larger treatment effectiveness for children with higher pre-test levels of aggression. The moderating effects of SES were mixed. Of the studies comparing implementation to groups or individual, 57% of studies found no effect of implementation to individuals compared to implementation to groups. Of studies investigating the moderating effect of the person implementing the treatment, 57% found no moderating effect. In the studies comparing the moderating effects of different treatment programs, 60% found no effect. The moderating effect of type of treatment was mixed. Of studies investigating the moderating effect of parental involvement, 60% found positive moderation of parent involvement. Of the studies examining the moderating effect of session-related factors or treatment intensity, 78% of the moderator effects were not significant. The moderating effect of the informant was mixed. The moderating effect of research quality was mixed.

## DISCUSSION

This study provided a synthesis of systematic reviews and meta-analyses to obtain a comprehensive overview of the existing literature on the effectiveness of treatments for childhood aggression. The included studies were heterogeneous in the types of treatments and moderators, and in levels of study quality. The most prevalent effect size for treatments for childhood aggression was small. Two moderators had an effect in the majority of studies in which they were included. First, a positive moderation of pre-test levels of aggression on treatment effectiveness indicated that treatments were more effective for children with higher pre-test levels of aggression. Second, parental involvement had a positive moderating effect on treatment effectiveness, indicating that treatments were more effective when parents were involved. For the other moderators, effects were absent or mixed. Additionally, two overarching patterns emerged. In the following, we will discuss these patterns and describe their theoretical and clinical implications.

## Effect Sizes Vary as a Function of Treatment Targets

The literature differentiates between prevention and intervention (Grove et al., 2008; Sawyer et al., 2015). Prevention pertains to universal prevention (i.e., for children without any specified risk-factors for developing childhood aggression), selective prevention (i.e., for children with an elevated risk for developing childhood aggression), and indicated prevention (i.e., for children with an elevated risk for developing childhood aggression identified as showing behaviors associated with childhood aggression). Interventions pertain to treating children with diagnosed aggression (Mrazek & Haggerty, 1994).

Our results suggest that rather than clustering indicated prevention with prevention strategies, it shares more features with intervention. First, the effect sizes for universal prevention and selective prevention were almost all absent or small, whereas effects for indicated prevention and intervention were mostly small or medium. Second, studies assessing treatment effectiveness of indicated prevention and intervention focused on similar treatment programs, namely psychosocial treatment programs, cognitive-behavioral treatment programs, and parent training programs. Likewise, studies assessing treatment effectiveness of universal prevention or selective prevention examined similar types of programs, namely mainly school-based programs.

These patterns reflect an important difference between the two clusters of treatments. While universal and selective prevention target risk factors of childhood aggression (Durlak et al., 2011; Oliver, Wehby, & Reschly, 2011; Park-Higgerson et al., 2008; Wilson & Lipsey, 2006), indicative prevention and intervention target the (sub-clinical) symptoms of childhood aggression itself. Most risk factors associated with aggression, such as a lack of cognitive, social, and behavioral skills, are nonspecific and influence multiple dimensions of mental disorders and psychosocial problems, rather than being predictive of a single outcome, such as childhood aggression (Bradley & Corwyn, 2002; Lahey, Krueger, Rathouz, Waldman, & Zald, 2017; Mcmahon, Grant, Compas, Thurm, & Ey, 2003). For most children, such risk factors do not lead to childhood aggression. Consequently, it is more challenging for universal and selective prevention programs to be effective than for indicated prevention and intervention programs. Therefore, treatment effectiveness may be less determined by the type of treatment program than by the treatment targets (i.e., risk factors vs. (sub)clinical symptoms of childhood aggression).

A focus on treatment targets may also have implications for research and treatment practices. Treatments are often studied separately for children with diagnosed disorders (e.g., Sawyer et al., 2015), leaving out children without a diagnosis or with sub-clinical symptom levels. Nevertheless, our synthesis suggests that indicated prevention effectiveness is comparable to interventions, suggesting that children with sub-clinical aggression may benefit from treatment. Furthermore, children displaying aggression are likely to profit more from earlier treatment (Baker, 2009; Coie et al., 1993; Comer et al., 2013; Connor et al., 2006; Frick & Dickens, 2006; Johnson et al., 2014). In addition to preventing the development of full-blown childhood aggression, indicated prevention may attenuate the development of other disorders (e.g., anxiety disorders, substance use disorders; Shankman et al., 2009) and sub-clinical disorders (e.g., sub-clinical anxiety, sub-clinical substance use disorders; Lewinsohn, Shankman, Gau, & Klein, 2004). These findings underline the possible gains of clustering subclinical and diagnosed intervention programs when examining treatment effectiveness.



## Role of Moderators in Treatment Effectiveness

A majority of the studies that included pre-test levels of aggression found that higher levels were associated with higher treatment effectiveness for childhood aggression. One explanation for this effect could be that there is more room for improvement for individuals with higher levels of aggression. It is also possible that higher levels of aggression allow clinicians to assign indicated prevention or interventions targeting aggression rather than nonspecific risk factors, thereby increasing effectiveness of the treatment (Mrazek & Haggerty, 1994). Finally, some children may be more susceptible to treatment than others (Belsky & Pluess, 2009). If high levels of aggression indicate that children are more susceptible to environmental influences conducive to the development of childhood aggression, this may also indicate that they are more susceptible to benefitting from a treatment. To examine this suggestion, longitudinal, genetically informed designs would be particularly promising.

A majority of studies that included the moderating effect of parental involvement found that it had a positive effect on treatment effectiveness for childhood aggression. Consistent with this finding, research suggests that treatments focusing only on parents, parental psychopathology, and parenting strategies already may have a positive effect on child behavior (Hudziak & Bartels, 2008; Hudziak & Ivanova, 2016; Weissman et al., 2006). Childhood aggression is strongly influenced by both genetic factors and the environment (e.g., Burt, 2009; Fedko et al., 2016; Hudziak et al., 2003; Porsch et al., 2016; Van Beijsterveldt, Bartels, Hudziak, & Boomsma, 2003; Wesseldijk et al., 2016). Given the genetic influence on aggression, it is not unlikely that parents of children with (symptoms of) aggression show aggression-related symptoms themselves (Frick et al., 1992). Given the environmental influence on aggression, parents may amplify their children's (risk to develop) aggression through negative or ineffective parenting strategies (Belsky, Hsieh, & Crnic, 1998; Berg-Nielsen, Vikan, & Dahl, 2002). Therefore, an opportunity for future research may be to focus more on parental influences as possible moderators of treatment

effectiveness. Factors such as parental dysfunction, parental psychopathology, and family stress are associated with a higher risk to develop childhood aggression (Frick et al., 1992; Goodman et al., 2011; Loeber & Hay, 1997).

Finally, the majority of the commonly included moderators (e.g., age, gender, SES, treatment characteristics, methodological characteristics) were not consistently associated with treatment effectiveness. Overall, treatments for childhood aggression yielded small effects, and only two of the commonly included moderators explained why some children responded better to treatment than others. Recognizing childhood aggression as multidimensional disorder – both in development (Nock, Kazdin, Hiripi, & Kessler, 2006; Tremblay, 2000) and expression (Bolhuis et al., 2017; Tremblay, 2010) - may be more auspicious than the current often applied 'one size fits all approach'. Given this multidimensionality, more customized approaches for treatment of childhood aggression seem promising. The present study included diagnostic classifications of childhood aggression that are neither simple nor specific. Individuals with the same diagnosis can have remarkably distinct symptoms and/or combinations of symptoms. New approaches that examine the heterogeneity in aggressive behavior by including, for example, biological and physiological information and change of behavior over time (e.g., Fanti, 2016), hold promise for identifying predictors and correlates of specific types of aggression and subsequently develop and apply more targeted treatments.

The heterogeneity of childhood aggression in the present study underlines the need for a clearer taxonomy for childhood aggression. It was beyond the scope of the present study to examine whether the heterogeneity in population influenced treatment effectiveness. Childhood aggression and related disorders often rely on identifying combinations of subsets of symptoms, or criteria, to define diagnoses. To illustrate, Bolhuis and colleagues (2017) discerned multiple dimensions from the Child Behavior Checklist Aggression scale and Rule Breaking scale including physical aggression, irritability, oppositional or disobedient behavior, and rule breaking. Burt (2013) demonstrated that aggressive and non-aggressive rule-breaking dimensions of antisocial behavior show both similarities and differences. These findings highlight that the utility of different diagnoses and thresholds of symptoms for the evaluation of treatment effects is limited.

In addition to classifying childhood aggression with a more concise and clear taxonomy, biological information may contribute to more customized treatment approaches. Increasingly, researchers unravel the interplay between genes and the environment to inform treatment practices and identify novel treatment targets (Boomsma, 2015; Burt, 2013).



## Limitations and Future Recommendations

Synthesis studies play an important role in cumulative science by combining and integrating information across multiple studies and, in our case, a time period of more than 60 years. Despite its contributions, there were also some limitations. One limitation concerns a weakness of each systematic review and meta-analysis, namely that the results reflect the quality of the included studies. Second, there is some overlap in the articles included by the studies (e.g., 27 of the articles in Hahn et al. (2007) were also included in Wilson and Lipsey (2007)), and it is not unlikely that studies with larger effect sizes were included more often. This may have implications for the reported treatment effectiveness and moderator effects. Nevertheless, the considerable number of systematic reviews and meta-analyses included strengthens our confidence in the robustness of our findings.

## Conclusion

The present study provided a comprehensive synthesis of the literature on treatment effectiveness for childhood aggression. We identified patterns in the literature on treatment effectiveness and identified opportunities for future research. Overall, treatments for childhood aggression yielded small effects. Our results suggest that there is merit in clustering treatment programs based on treatment targets (i.e., risk factors vs. (sub)clinical symptoms of childhood aggression). More systematic research examining the moderating role of risk factors associated with parental factors, individual development, and expression would be promising to further our understanding of treatment effectiveness. Such work has the potential to inform the tailoring of treatments for individual children to augment existing strategies for prevention and intervention for childhood aggression.

Chapter 3

# National child and adolescent health policies as indicators of adolescent mental health: A multilevel analysis of 30 European countries.

Accepted as: Hendriks, A. M., Bartels, M., Stevens, G. W. J. M., Walsh, S. D., Torsheim, T., Elgar, F. J., & Finkenauer, C. (in press). National child and adolescent health policies as indicators of adolescent mental health: A multilevel analysis of 30 European countries. *Journal of Early Adolescence.*

## ABSTRACT

There is little evidence on the association between child and adolescent mental health (CAMH) policies and adolescent mental health. This study examined this association using data on indicators of adolescent mental health— aggressive behavior, life satisfaction, and psychosomatic symptoms—in 172,829 eleven- to fifteen-year-olds from 30 European countries in the 2013/14 Health Behaviour in School-aged Children (HBSC) study. Individual records were linked to national-level policies for CAMH, controlling for national-level adult violence, adult well-being, and income inequality. Multi-level analyses revealed lower adolescent aggressive behavior in countries with more CAMH policies, even after controlling for other national-level indicators. Adolescent life satisfaction and psychosomatic symptoms were not associated with CAMH policies. Results may inform policy recommendations regarding investments in adolescent mental health.



**Keywords:**
Adolescent mental health, national policies, HBSC






Adolescent mental health is important for daily academic, social, and family functioning, and for developmental trajectories of mental health through adulthood (Althoff, Verhulst, Rettew, Hudziak, & Van Der Ende, 2010; Kessler et al., 2007). Positive adolescent mental health is related to lower risks of delinquency and crime, physical illness, and mental health problems throughout life (Kleinert, 2007; Patton et al., 2018). Adverse outcomes from adolescent mental health problems not only impose a burden on individuals and their immediate environment but also on society (Caspi et al., 2016; Fergusson & Woodward, 2002; Kessler et al., 2007; Romeo, Knapp, & Scott, 2006; World Health Organization, 2005, 2015). Therefore, adolescence is a critical period for researchers, policy makers, and practitioners to invest in to ensure and maintain positive mental health (Patton et al., 2014).

Recognizing the important role of adolescent mental health and the lack of policies to support it (Shatkin & Belfer, 2004), the World Health Organization called for more emphasis on policies for child and adolescent mental health (CAMH; World Health Organization, 2005). Since then, several large projects have gathered information on the current state of policies for CAMH in Europe and provided recommendations to strengthen their implementation (e.g., Carral Bielsa, Braddick, Jané-Llopis, Jenkins, & Puras, 2010; Coppens et al., 2015). Nevertheless, to our knowledge, the extent to which these policies relate to adolescent mental health remains unresearched. The aim of the present study is to explore the relation between CAMH policies and different indicators of adolescent mental health across 30 European countries.

In this study, we focus on three indicators of adolescent mental health that are reliably associated with other mental health indicators and contextual risk factors, namely aggressive behavior, life satisfaction, and psychosomatic symptoms. To illustrate, childhood aggressive behavior is robustly related to other externalizing and internalizing symptoms (Bartels et al., 2018; Granic, 2014; King & Waschbusch, 2010), and psychopathology later in life (Althoff et al., 2010). Life satisfaction is generally considered a hallmark of superior mental and physical health and resilience throughout the life course (Cohn, Fredrickson, Brown, Mikels, & Conway, 2009; Lyubomirsky, King, & Diener, 2005; Proctor, Linley, & Maltby, 2009), and psychosomatic symptoms are associated with mental health problems such as depression and anxiety, and with stressors such as bullying and school stress (Berntsson, Köhler, & Gustafsson, 2001; Piko, 2007).

## National CAMH Policies and Adolescent Mental Health

According to the Social Ecological Model (SEM), adolescent mental health development occurs across different social levels (Mcleroy, Bibeau, Steckler, & Glanz, 1988; UNICEF, 2016). National policies for CAMH take place at the most external and distal social level (i.e., policy/enabling environment), however they may affect adolescents directly at the individual level or through intermediate levels

specified in the SEM, including the organizational, community, and interpersonal level. We identified four types of national-level policies relevant to CAMH: monitoring adolescent mental health, structural facilities for adolescent mental health, investment in family benefits, and investment in education.

First, monitoring of the prevalence of adolescent mental health problems may be an indication of the priority governments give to this topic (World Health Organization, 2005). Robust evidence on the incidence and prevalence of mental health problems informs governments about the mental health needs of adolescents in a country and potential gaps for prevention and intervention strategies, which may have implications that affect adolescent mental health at the organizational level (Jenkins, 2003; Wittchen & Jacobi, 2005; Wittchen et al., 2011). Second, a country's number of structural facilities and resources (i.e., psychiatrists, psychologists, community-based mental health services) that deliver medical and psychosocial interventions reflects the capacity to respond to the needs of adolescents with mental health problems, which may express itself at a community level, and thus improve adolescent mental health (Tylee, Haller, Graham, Churchill, & Sanci, 2007). Third, policies aimed at vulnerable families through investment in family benefits (i.e., child payments and allowances, parental leave benefits, child care support) are important to decrease the risk of mental health problems due to poverty and low socioeconomic status and to improve adolescent mental health through implications at the interpersonal (e.g., stress-alleviation in the family environment) and individual level (Currie et al., 2012; Viner et al., 2012). Fourth, policies that aim to ensure universal access to education may improve adolescent mental health through an increase of both quality of (i.e., interpersonal level) and access to (i.e., individual level) education. Consistent evidence links higher educational attainment with better health, well-being, and mental health (Link et al., 2008; Ottova et al., 2012; Sawyer et al., 2012; Viner et al., 2012). Taken together, we expected a positive association between the above four types of CAMH policy assessment at the national level with adolescent life satisfaction and a negative association with adolescent aggressive behaviors and psychosomatic symptoms.

### Other National-Level Indicators Related to Adolescent Mental Health

There may be other social determinants (e.g., cultural and social norms) that influence adolescent mental health that are associated with CAMH policy influences, or could be misinterpreted as such (Viner et al., 2012), possibly leading to overestimating the effect of CAMH policies. For example, levels of adult violence and well-being may reflect the cultural norms of society that adolescents grow up with and as a consequence may influence their mental health. Adolescence is a sensitive period for social learning (e.g., Sawyer et al., 2012). Therefore, cultural norms regarding the (in)



appropriateness of certain behaviors or feelings at the country level may affect the development and reporting of adolescent mental health. Also, national-level income inequality may highlight the (absence or presence of) economic differences, social competition, and strength of social trust (Pickett & Wilkinson, 2007, 2010). Previous research showed that levels of national-level income inequality were associated with adolescent mental health indicators such as psychosomatic symptoms (Elgar et al., 2015) and life satisfaction (Levin et al., 2011). Therefore, we control for adult violence, adult well-being, and income inequality when examining the association between policies for CAMH and indicators of adolescent mental health.



### The Present Study

The Health Behaviour of School-Aged Children (HBSC) study is a collaborative cross-national study which aims to gain insights in adolescents' health and well-being, and their social context (Currie et al., 2014, 2012). Combining HBSC data with country-level indicators of national CAMH policies and national indicators reflecting social and cultural norms provides a powerful tool to examine factors that explain differences in adolescent mental health between countries (Currie & Alemán-Díaz, 2015). Using HBSC data, we assessed adolescent aggressive behavior, life satisfaction, and psychosomatic symptoms as indicators of adolescent mental health in 30 countries and examined their association with national indicators of CAMH policies, controlling for adult violence, adult well-being, and income inequality.

## METHOD

Data consisted of survey data on 11-, 13-, and 15-year-olds who participated in the HBSC Study in 2013/2014, from the following countries: Austria, Belgium (i.e., Flanders and Wallonia), Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom (i.e., England, Scotland, and Wales). Table 1 depicts the sample sizes of each age group in each country. The HBSC study gathers data every four years from adolescents aged 11, 13 or 15 years through school-based self-report surveys. It started in 1983/1984 with five countries, while it included 44 countries in 2013/2014. All countries use standard methodology (e.g., Currie & Alemán-Díaz, 2015; World Health Organization, 2016).

### Individual-Level Measurements

**Aggressive behavior.** Aggressive behavior was assessed with two items. One item evaluated the frequency of physical fights, "During the past 12 months, how many times were you in a physical fight?", using a five-point scale "*I have not been in a*

**Table 1.** Sample descriptives per country

| | | Individual variables | | | National-level variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **AGG** | **LS** | **PS** | **Monitoring** | **Facilities** | **Family benefits** | **Education** | **Mean policy** | **Adult VIO** | **Adult WB** | **Gini** |
| Austria | 3458 | 1.67 | 7.95 | 1.80 | -0.86 | -0.51 | 0.15 | 0.29 | -0.23 | 0.73 | 35.6 | 30.48 |
| Belgium | 10285 | 1.60 | 7.33 | 2.03 | -0.37 | -0.01 | 0.91 | 0.85 | 0.34 | 1.84 | 18.2 | 27.59 |
| Bulgaria | 4796 | 1.68 | 7.85 | 2.07 | -0.86 | -1.31 | -1.08 | -1.20 | -1.11 | 1.50 | 16.6 | 36.01 |
| Croatia | 5741 | 1.51 | 7.91 | 1.95 | -0.86 | -0.90 | -1.60 | -0.91 | -1.07 | 1.08 | 11.5 | 32.51 |
| Czech Republic | 5082 | 1.53 | 7.20 | 2.04 | -0.86 | -0.88 | 0.71 | -0.68 | -0.43 | 0.87 | 16.9 | 26.13 |
| Denmark | 3891 | 1.47 | 7.63 | 1.97 | -0.37 | -0.20 | 1.26 | 2.50 | 0.80 | 0.73 | 37.0 | 29.08 |
| Estonia | 4057 | 1.53 | 7.79 | 1.99 | -0.37 | 0.40 | -0.36 | -0.19 | -0.13 | 3.94 | 16.2 | 33.15 |
| Finland | 5925 | 1.34 | 7.69 | 2.17 | 2.07 | 2.94 | 0.78 | 1.01 | 1.70 | 1.66 | 26.2 | 27.12 |
| France | 5691 | 1.59 | 7.33 | 2.23 | 1.58 | -0.09 | 1.25 | 0.20 | 0.73 | 1.18 | 22.8 | 33.35 |
| Germany | 5961 | 1.53 | 7.38 | 1.92 | 1.09 | 0.15 | 0.59 | -0.33 | 0.37 | 0.77 | 28.0 | 30.13 |
| Greece | 4141 | 1.54 | 7.72 | 1.95 | -0.37 | 0.35 | -1.27 | -1.00 | -0.57 | 1.28 | 10.3 | 36.68 |
| Hungary | 3935 | 1.58 | 7.54 | 2.13 | -0.37 | -0.82 | 1.16 | -0.53 | -0.14 | 1.39 | 16.5 | 30.55 |
| Iceland | 10602 | 1.31 | 7.73 | 2.16 | -0.37 | - | 1.22 | 1.46 | 0.77 | 0.31 | - | 26.94 |
| Ireland | 4098 | 1.36 | 7.59 | 2.00 | 0.11 | -0.59 | 0.98 | 0.55 | 0.26 | 1.81 | 26.3 | 32.52 |
| Italy | 4072 | 1.42 | 7.36 | 2.30 | 1.09 | 0.58 | -0.53 | -0.85 | 0.07 | 0.84 | 15.3 | 35.16 |
| Latvia | 5557 | 1.87 | 7.35 | 2.07 | -0.37 | 0.26 | -1.05 | -0.34 | -0.38 | 3.41 | 14.3 | 35.48 |
| Lithuania | 5730 | 1.81 | 7.85 | 1.96 | -0.86 | 0.00 | -0.36 | -0.19 | -0.35 | 5.79 | 11.7 | 35.15 |
| Luxembourg | 3318 | 1.59 | 7.44 | 2.19 | -0.86 | 1.62 | 1.20 | -1.71 | 0.06 | 0.19 | 30.2 | 34.79 |
| Malta | 2265 | 1.49 | 7.61 | 2.17 | -0.86 | -0.51 | -1.30 | 1.91 | -0.19 | 1.42 | 24.4 | - |
| Netherlands | 4301 | 1.43 | 7.75 | 1.94 | 1.58 | 0.04 | -0.70 | 0.38 | 0.33 | 0.74 | 31.9 | 27.99 |

**Table 1.** Continued

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Norway | 3422 | 1.38 | 7.89 | 1.88 | 1.58 | - | 0.70 | 0.93 | 1.07 | 0.91 | 32.0 | 25.90 |
| Poland | 4545 | 1.59 | 7.38 | 2.11 | -0.86 | -0.52 | -0.91 | -0.36 | -0.66 | 0.78 | 20.1 | 32.55 |
| Portugal | 4989 | 1.43 | 7.49 | 1.70 | -0.37 | -0.88 | -1.10 | -0.11 | -0.62 | 1.37 | 18.3 | 36.04 |
| Romania | 3980 | 1.75 | 8.12 | 2.07 | -0.86 | -0.77 | -0.89 | -1.77 | -1.07 | 1.68 | 20.8 | 27.45 |
| Slovakia | 6099 | 1.72 | 7.40 | 2.07 | -0.86 | -0.90 | -0.44 | -1.02 | -0.80 | 1.44 | 21.4 | 26.12 |
| Slovenia | 4997 | 1.56 | 7.73 | 1.91 | -0.86 | 0.10 | -0.54 | 0.20 | -0.28 | 0.58 | 14.7 | 25.59 |
| Spain | 11136 | 1.38 | 7.85 | 1.85 | 1.09 | -0.06 | -1.07 | -0.45 | -0.12 | 0.65 | 20.6 | 35.89 |
| Sweden | 7700 | 1.30 | 7.36 | 2.19 | 0.11 | 2.58 | 1.23 | 1.05 | 1.24 | 0.91 | 29.1 | 27.32 |
| Switzerland | 6634 | 1.51 | 7.73 | 2.11 | 0.11 | - | -0.47 | -0.10 | -0.15 | 0.71 | 39.4 | 31.64 |
| UK | 16421 | 1.43 | 7.51 | 1.97 | 2.07 | -0.06 | 1.56 | 0.42 | 1.00 | 0.92 | 23.5 | 32.57 |

*Note.* AGG, aggressive behavior; LS, life satisfaction; PS, psychosomatic symptoms; Monitoring, monitoring of adolescent mental health; Facilities, adolescent mental health facilities; Family benefits, investment in family benefits; Education, investment in education; Mean policy, the total policy score; Adult VI, adult violence; Adult well-being, adult well-being; Gini, income inequality.

*physical fight*," "*1 time*," "*2 times*," "*3 times*," and "*4 times or more*" (Centers for Disease Control and Prevention et al., 2006). This item was developed as part of the Youth Risk Behavior Survey Questionnaire (Brener, Collins, Kann, Warren, & Williams, 1995) and validated in adolescents both as a continuous (Pickett, 2005) and dichotomous (Pickett et al., 2013; Walsh et al., 2013) variable. The second item assessed the frequency of bullying others, "How many times have you bullied others at school in the previous months?" with the following response options "*I haven't*" "*once or twice*," "*2 or 3 times a month*," "*about once a week*," and "*several times a week*" (Olweus, 1992). Based on item content and the correlation between the items ($r$ = .28), the two items were combined into a mean score of aggressive behavior; higher scores indicated more aggressive behavior.

**Life satisfaction.** Life satisfaction was measured with the Cantril ladder of life satisfaction (Cantril, 1965), adapted for use with adolescent samples by omitting the part asking to imagine the best and worst possible life (Levin & Currie, 2014). The item asks adolescents where on a ladder they would rate their satisfaction with life. Possible answers ranged from "0 = *worst possible life"* at the bottom of the ladder to "10 = *best possible life*" at the top of the ladder. We included the individual scores as a continuous variable.

**Psychosomatic symptoms.** Psychosomatic symptoms were assessed with the following questions from the HBSC symptom checklist: "In the last 6 months: how often have you had the following? 1) headache, 2) stomach-ache, 3) back ache, 4) feeling low, 5) irritability or bad temper, 6) feeling nervous, 7) difficulties in getting to sleep, and 8) feeling dizzy". Possible responses were: "*about every day*", "*more than once a week*", "*about every week*", "*about every month*", "*rarely or never*". The measure has demonstrated reliability and validity for research on adolescents (Cronbach's α in the present study was .83; Ravens-Sieberer et al., 2008). Following previous research (Ravens-Sieberer et al., 2009), we computed the mean score. Participants with missing information on more than two items were not included in the analyses. A higher score indicated more psychosomatic symptoms.

**Demographic variables.** We included age, gender, and socioeconomic status as demographic variables because of their associations with adolescent mental health (e.g., Archer, 2004; Cavallo et al., 2006; Sweeting & West, 2003). Age was included as a categorical variable, with the categories 11, 13, and 15 years. Socioeconomic status was measured with the Family Affluence Scale (Currie et al., 2008; Torsheim et al., 2016) comprising of six items: "Does your family own a car, van or truck?" (*No*=0, *Yes*=1, *Yes, two or more*=2); "During the past 12 months, how many times did you travel away on holiday with your family?" (*Not at all*=0, *Once*=1, *Twice or more*=2); "How many computers does your family own?" (*None*=0, *One*=1, *Two or more*=2); "Do you have your own bedroom for yourself?" (*No*=0, *Yes*=1); "Does your family own a dishwasher?" (*No*=0, *Yes*=1, *Yes, two or more*=2); "How many bathrooms are there in your house?" (*None*=0, *One*=1, *Two or more*=2). Following previous HBSC studies (Elgar, Gariépy, Torsheim, & Currie, 2017; Levin et al., 2011; Torsheim et al., 2004; Walsh, Bruckauf, & Gaspar, 2016), we used the ridit transformation of the sum scores to render relative FAS scores comparable across countries. In this transformation, FAS scores within each country were assigned a distribution between 0 and 1, with a country mean of 0.5 (Torsheim et al., 2004). A higher score was indicative of a higher socioeconomic status.

## National-Level Indicators for Policies

Descriptives of all policy indicators per country as transformed for thz analyses are described in Table 1. The original values of the policy indicators are presented in Supplementary Table 1. For analyses on the separate policies, we used unstandardized values; to create a total policy score we standardized the values. First, we derived the extent to which countries monitor adolescent mental health through collection of epidemiological data using information from the Global Burden of Disease Study (GBD, Institute for Health Metrics and Evaluation, 2018). Countries could report on six types of children's psychopathologies, including conduct disorder, attention-deficit hyperactivity disorder, autism spectrum disorders, eating disorders, depression, and anxiety. Countries received a score (i.e., 0 - 6) based on the number of children's psychopathology types countries reported to have epidemiological data on, in either the 2010 or 2013 GBD study (Erskine et al., 2017; supplemental material). We standardized the variable to have a mean of zero and a standard deviation of one. Monitoring scores before standardization ranged between 0 (Bulgaria, Croatia, Czech Republic, Lithuania, Luxembourg, Malta, Poland, Romania, Slovakia, Slovenia) and 6 (Finland and the United Kingdom).

Second, as indicators of adolescent mental health facilities in a country, we used the number of child and adolescent mental health services (CAMHS), the number of psychiatrists, and the number of psychologists per 100.000 young people (i.e., younger than 18 or the age of majority in a country). CAMHS were defined as specialist, community-based, multidisciplinary mental health services, delivering medical and psychosocial interventions. We derived this information from the article by Signorini and colleagues (2017), which describes the current status of CAMHS in the European Union based on a questionnaire completed by child psychiatrists and representatives of national child psychiatry associations. We standardized these variables to have a mean of zero and a standard deviation of one and combined them into a single factor score (Cronbach's α = .63). For Iceland, Norway, and Switzerland we did not find a source to complement this variable, therefore we coded them as missing. The three variables had the following ranges per 100.000 young people: number of CAMHS, 0.50 (Bulgaria) to 12.90 (Finland); number of CAMH psychiatrists, 1.90 (Bulgaria) to 36.00 (Finland); number of psychologists, 1.70 (Bulgaria) to 104.20 (Sweden).

Third, investment in family benefits to support (vulnerable) families was measured by the percentage of a country's Gross Domestic Product (GDP) spent on family benefits. We retrieved this information from the Organisation for Economic Co-operation and Development (OECD) family database describing country percentages from 2012 or 2013 (http://www.oecd.org/els/family/database.htm; OECD Social Policy Division, 2016). For countries for which the OECD did not report this information (indicated with an asterisk in Supplementary Table 1), we retrieved information from the World Social Protection Report (correlation between estimates from the OECD and the World Social Protection Report for countries with data available from both was .89; ILO, 2014). Family benefits referred to payments solely for families and children, namely child payments and allowances, parental leave benefits, and childcare support. The percentage of the GDP was a combination of cash transfers to families, payments to services to support families with children, and financial support for families provided through tax benefits. We standardized this variable to have a mean of zero and a standard deviation of one. Before transformation, scores ranged between 0.96% (Croatia) and 3.95% (the United Kingdom).

Fourth, investment in education was measured by national-level percentages of GDP spent on education, derived from Eurostat, the statistical office of the European Union (Eurostat, 2016). The variable represented the percentage of the national GDP spent on education. This was through funding by the public sector given to educational institutions or students and their families, or through public subsidies to private firms or non-profit organizations to support educational activities. We standardized the values to have a mean of zero and a standard deviation of one. Before transformation the values ranged between 3.07% (Romania) and 8.75% (Denmark).

**Total policy score**. Due to the transformation, the separate policy variables had a similar scale, allowing for combining them into a single total policy score. We calculated a total policy score for each country; Cronbach's α was .65. Because there were some missing values, we used the mean of each country instead of a sum score.

## Other National-Level Indicators Related to Adolescent Mental Health

**Adult violence.** We used countries' homicide rates as a proxy for adult violence. We obtained this information from Eurostat, which reported the number of intentional homicides per 100.000 inhabitants reported in 2013 based on police records (Eurostat, 2017). The number of intentional homicide rates was lowest in Iceland (0.31) and highest in Lithuania (5.79).



**Adult well-being.** We derived data on national-level adult well-being from the Gallup-Healthways Global Well-Being Index based on interviews administered in 2014. This index comprises the world's largest dataset on well-being. The interviews investigated five dimensions of well-being (i.e., purpose, social, financial, community, and physical). The reported numbers refer to the percentage of the population thriving on at least three dimensions (Gallup Healthways Well-Being Index, 2014). Adult well-being was lowest in Croatia (11.5) and highest in Switzerland (39.4).

**Income inequality.** To measure income inequality, we used the Gini index (The World Bank Group, 2017). In the present study, we report the most recent estimates; these were obtained by the World Bank between 2011 and 2013. These estimates derived from the World Bank were percentages ranging between 0 and 100; 0 indicated perfect equality, 100 indicated perfect inequality. Some articles report Gini estimates ranging from 0-1; the scale, however, did not have consequences for the results. Slovenia had the lowest income inequality (25.59); Greece had the highest (36.68).

## Analyses

First, we explored the data by performing descriptive analyses and obtaining correlations between all variables. Second, we performed multi-level analyses separately for aggressive behavior, life satisfaction, and psychosomatic symptoms to examine the relationship between national-level indicators and indicators of adolescent mental health. To adjust for oversampling of subpopulations, we applied survey weights in the multi-level analyses. We fitted two-level random intercept models with adolescents (Level 1) nested within countries (Level 2). To examine the contribution of all variables, we performed blockwise model fitting. We fitted a model containing only the dependent variable (adolescent aggressive behaviors, also separately for fighting and bullying, reported in Supplementary Table 2 and 3, life satisfaction, or psychosomatic symptoms; Model 1), subsequently adding the demographic variables (Model 2), the total policy score (Model 3.1) or the separate policy variables (Model 3.2). Next, we added adult violence, adult well-being, and income inequality both with the total policy score (Model 4.1), and the separate policy variables (Model 4.2). To correct for the large number of tested variables, and control the familywise error rate, the significance threshold was $p < .01$. We Z-standardized the adolescent aggressive behavior and psychosomatic symptoms for interpretability of the parameter estimates. For life satisfaction, we retained the scale because of the interpretation of a single unit increase as a higher step on the ladder.

# RESULTS

Table 1 displays the country means for aggressive behavior, life satisfaction, and psychosomatic symptoms, and the values for the national-level indicators. Mean adolescent aggressive behavior was lowest for Sweden (1.30) and highest for Latvia (1.87). Adolescent life satisfaction was lowest in the Czech Republic (7.20) and highest in Romania (8.12). Adolescents in Portugal scored lowest on psychosomatic symptoms (1.70) and adolescents in Italy scored highest (2.30). The country scoring the lowest on policy was Bulgaria (-1.11), the country scoring the highest was Finland (1.70). Figure 1 graphically displays the country means of CAMH policies, and adolescent aggressive behaviors, life satisfaction, and psychosomatic symptoms.

Table 2 shows the disaggregate pooled correlations between all indicators of adolescent mental health and national-level variables. As expected, a positive association was found between aggressive behavior and psychosomatic symptoms. Aggressive behavior and psychosomatic symptoms were both negatively associated with life satisfaction. The correlations between the indicators of adolescent mental health and the national-level indicators (i.e., total policy, monitoring, structural facilities, investment in family benefits, investment in education, adult violence, adult well-being, income inequality) were small but almost all significant. As expected, adolescent aggressive behavior was weakly negatively associated with all national indicators except for positive associations with adult violence and income inequality. Adolescent life satisfaction, unexpectedly, was negatively associated with all policy variables and positively associated with adult violence and income inequality; there was no association with adult well-being. Adolescent psychosomatic symptoms had, similarly, and unexpectedly, weak positive associations with most national-level indicators, except for a negative association with adult violence and income inequality, and no association with monitoring. The associations between the national-level indicators were stronger than the associations between the (individual) adolescent mental health indicators and the national-level indicators, and in the expected direction. For instance, adult well-being was positively associated with all policy variables, but negatively associated with adult violence and income inequality. Adult violence, on the other hand, had negative associations with all policy variables and adult well-being, and a positive association with income inequality.



**Figure 1.** Country means for total policy, adolescent aggressive behaviors (AGG), life satisfaction (LS), and psychosomatic symptoms (PS). For a more interpretable scale, we standardized values of AGG, LS, and PS.

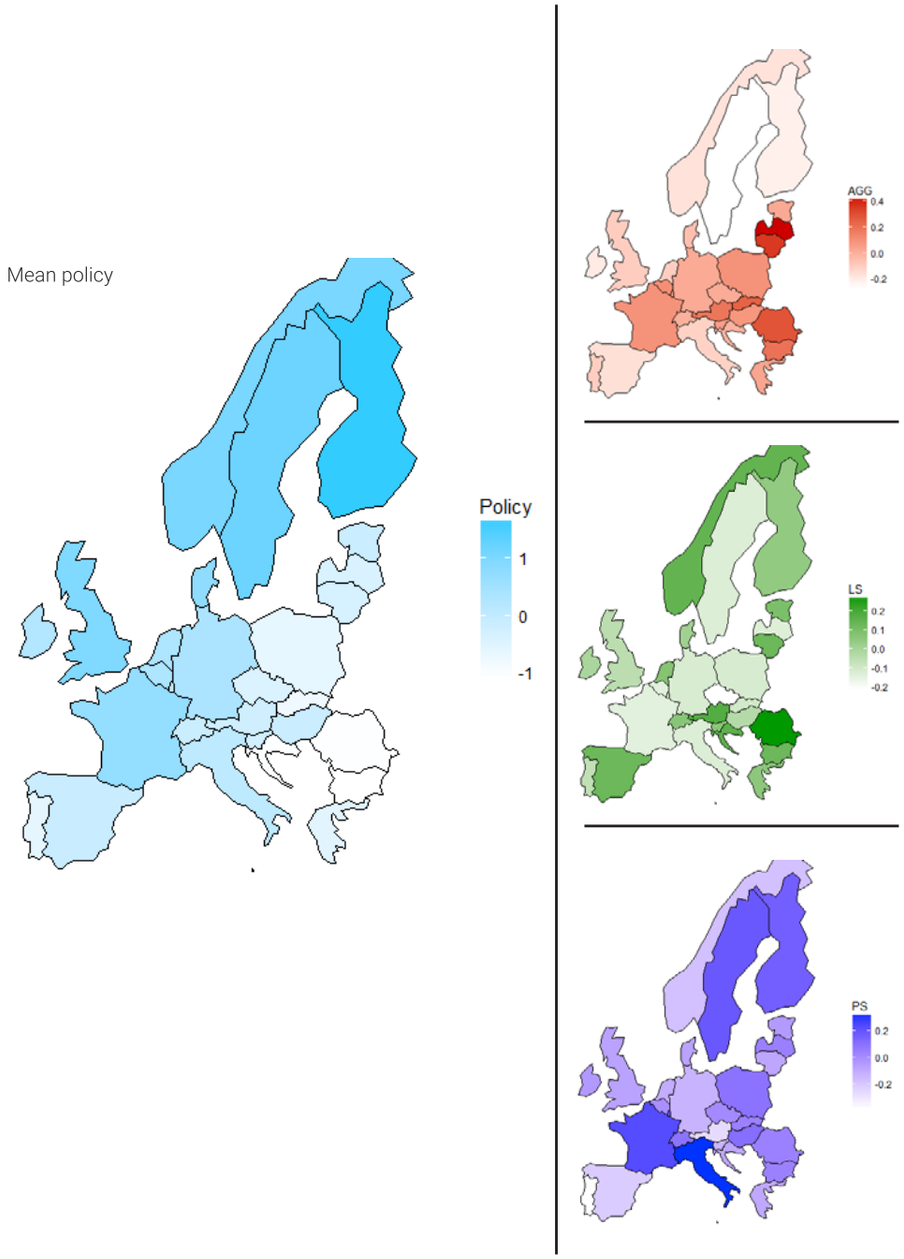


**Table 2.** Pooled correlations between the individual variables and the national-level variables

| | AGG | LS | PS | Total policy | Monitoring | Facilities | Family benefits | Education | Adult VIO | Adult WB | Gini |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AGG | 1.00 | | | | | | | | | | |
| LS | -.10 | 1.00 | | | | | | | | | |
| PS | .16 | -.41 | 1.00 | | | | | | | | |
| Total policy | -.12 | -.03 | .04 | 1.00 | | | | | | | |
| Monitoring | -.09 | -.02 | -.01 | .66 | 1.00 | | | | | | |
| Facilities | -.08 | -.02 | .06 | .70 | .30 | 1.00 | | | | | |
| Family benefits | -.08 | -.04 | .04 | .80 | .46 | .36 | 1.00 | | | | |
| Education | -.09 | -.01 | .01 | .74 | .20 | .37 | .53 | 1.00 | | | |
| Adult VIO | .11 | .01 | -.01 | -.22 | -.26 | -.01 | -.24 | -.13 | 1.00 | | |
| Adult WB | -.07 | .00 | .01 | .48 | .29 | .33 | .43 | .45 | -.46 | 1.00 | |
| Gini | .04 | .02 | -.03 | -.34 | .03 | -.21 | -.39 | -.45 | .27 | -.32 | 1.00 |

*Note.* AGG, aggressive behavior; LS, life satisfaction; PS, psychosomatic symptoms; Total policy, the total policy score; Monitoring, monitoring of adolescent mental health; Facilities, adolescent mental health facilities; Family benefits, investment in family benefits; Education, investment in education; Adult VIO, adult violence; Adult well-being, adult well-being; Gini, income inequality. All correlations except the following had a *p*-value < .01: Adolescent life satisfaction with adult well-being, and adolescent psychosomatic symptoms with monitoring.

## Aggressive Behavior

Table 3 displays the results of the linear multilevel regression analyses of aggressive behavior. According to the null model (Model 1), there was significant but small country-level variance in aggressive behavior. Dividing the country-level variance by the total variance (0.031/(0.031+0.967) = 0.031) yielded an intraclass correlation of .031, which revealed that 3.1% of total variance in aggressive behavior was due to between-country variation. The model that included the demographic variables (Model 2) indicated that aggressive behavior was lowest at age 15, was higher for boys, and was not related to socioeconomic status. Adding total policy in Model 3.1 showed a significant negative association with aggressive behavior, indicating that there was more aggressive behavior in countries with fewer policies. Model 3.2 revealed that none of the separate policies was associated with aggressive behavior. Controlling for the other national-level indicators (i.e., adult violence, adult well-being, income inequality; Model 4.1) yielded a positive association between adolescent aggressive behavior and adult violence, indicating that adolescent aggressive behavior was higher in countries with higher levels of adult violence. Additionally, the significant association between policy and adolescent aggressive behaviors remained after controlling for the other national-level indicators. In line with Model 3.2, Model 4.2 showed no effect of separate policies on adolescent aggressive behavior. Across the models, country-level variance decreased from 0.031 to 0.011, indicating that the variables explained 65% of the country-level variance in adolescent aggressive behavior.

## Life Satisfaction

Table 4 displays the results of the linear multilevel regression analyses for adolescent life satisfaction. The null model (Model 1) showed significant but small country-level variance in life satisfaction. Of the total variance in life satisfaction, 1.4% (0.051/(0.051+3.628) = 0.014) was at the country-level. Model 2 indicated that life satisfaction decreased with age, was higher for boys, and was higher for adolescents with a higher socioeconomic status. Adding policy to the model (Model 3.1) revealed no significant association with life satisfaction. Similarly, the model including the separate policies (Model 3.2) showed no effect on life satisfaction. When taking the other national-level indicators into account (Model 4.1 and Model 4.2), there was no association between policy (i.e., total policy and separate policies) and adolescent life satisfaction either. Moreover, no associations between life satisfaction and the other country-level variables (adult violence, adult well-being, and income inequality) were found. Country-level variance decreased from 0.051 to 0.036 across models, indicating that the variables explained (0.051 – 0.036) / 0.051 = 29% of the between-country variance in adolescent life satisfaction.

**Table 3.** Results of linear multilevel regression analyses of aggressive behavior.

| **Fixed effects** | **Model 1 AIC = 474984** | | **Model 2 AIC = 422321** | | **Model 3 AIC = 422308** | | **Model 3 sub AIC = 375993** | | **Model 4 AIC = 394627** | | **Model 4 sub AIC = 370533** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | 0.019 | 0.032 | -0.297** | 0.032 | -0.295** | 0.026 | -0.282** | 0.026 | -0.357 | 0.232 | -0.342 | 0.230 |
| Age | | | | | | | | | | | | |
| 11 | | | 0.092** | 0.006 | 0.092** | 0.006 | 0.085** | 0.006 | 0.088** | 0.006 | 0.086** | 0.006 |
| 13 | | | 0.094** | 0.006 | 0.094** | 0.006 | 0.096** | 0.006 | 0.093** | 0.006 | 0.096** | 0.006 |
| 15 (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| Gender | | | | | | | | | | | | |
| Boy | | | 0.499** | 0.005 | 0.499** | 0.005 | 0.508** | 0.005 | 0.510** | 0.005 | 0.508** | 0.005 |
| Girl (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| SES | | | 0.007 | 0.008 | 0.007 | 0.008 | 0.003 | 0.009 | 0.010 | 0.009 | 0.004 | 0.009 |
| Policy | | | | | -0.153** | 0.036 | | | -0.160** | 0.036 | | |
| Monitoring | | | | | | | -0.075 | 0.029 | | | -0.052 | 0.025 |
| Facilities | | | | | | | -0.035 | 0.028 | | | -0.046 | 0.024 |
| GDP family benefits | | | | | | | 0.016 | 0.030 | | | 0.007 | 0.029 |
| GDP education | | | | | | | -0.056 | 0.027 | | | -0.080 | 0.030 |
| Adult violence | | | | | | | | | 0.076* | 0.021 | 0.080* | 0.022 |
| Adult well-being | | | | | | | | | 0.005 | 0.004 | 0.006 | 0.004 |
| Income inequality | | | | | | | | | -0.005 | 0.006 | -0.006 | 0.007 |
| Variance components | | | | | | | | | | | | |
| Level 1 (child) variance | 0.967** | 0.003 | 0.881** | 0.003 | 0.881** | 0.003 | 0.911** | 0.003 | 0.903** | 0.003 | 0.913** | 0.004 |
| Level 2 (country) variance | 0.031** | 0.008 | 0.030** | 0.008 | 0.019** | 0.005 | 0.017** | 0.005 | 0.013** | 0.003 | 0.011** | 0.003 |

*Note. N individual* = 172829; *N countries* = 30; * $p < .01$; ** $p < .001$

**Table 4.** Results of linear multilevel regression analyses of life satisfaction

| **Fixed effects** | **Model 1 AIC = 477309** | | **Model 2 AIC = 424795** | | **Model 3.1 AIC = 424796** | | **Model 3.2 AIC = 373648** | | **Model 4.1 AIC = 391760** | | **Model 4.2 AIC = 367775** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | 0.003 | 0.029 | 0.440** | 0.030 | 0.439** | 0.029 | 0.448** | 0.029 | 0.543 | 0.307 | 0.744* | 0.259 |
| Age | | | | | | | | | | | | |
| 11 | | | -0.419** | 0.006 | -0.419** | 0.006 | -0.435** | 0.006 | -0.421** | 0.006 | -0.433** | 0.006 |
| 13 | | | -0.176** | 0.006 | -0.176** | 0.006 | -0.183 | 0.006 | -0.174** | 0.006 | -0.180** | 0.006 |
| 15 (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| Gender | | | | | | | | | | | | |
| Boy | | | -0.412** | 0.005 | -0.412** | 0.005 | -0.418** | 0.005 | -0.415** | 0.005 | -0.419** | 0.005 |
| Girl (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| SES | | | -0.083** | 0.008 | -0.083** | 0.008 | -0.086** | 0.009 | -0.076** | 0.009 | -0.086** | 0.009 |
| Policy | | | | | 0.052 | 0.040 | | | 0.078 | 0.048 | | |
| Monitoring | | | | | | | -0.005 | 0.032 | | | 0.003 | 0.028 |
| Facilities | | | | | | | 0.060 | 0.032 | | | 0.069 | 0.026 |
| GDP family benefits | | | | | | | 0.030 | 0.034 | | | 0.084 | 0.033 |
| GDP education | | | | | | | -0.026 | 0.031 | | | -0.062 | 0.033 |
| Adult violence | | | | | | | | | -0.012 | 0.028 | -0.007 | 0.024 |
| Adult well-being | | | | | | | | | -0.005 | 0.005 | -0.008 | 0.005 |
| Income inequality | | | | | | | | | 0.001 | 0.009 | -0.004 | 0.007 |
| Variance components | | | | | | | | | | | | |
| Level 1 (child) variance | 0.979** | 0.003 | 0.902** | 0.003 | 0.902** | 0.003 | 0.904** | 0.003 | 0.893** | 0.003 | 0.902** | 0.003 |
| Level 2 (country) variance | 0.024** | 0.006 | 0.025** | 0.007 | 0.024** | 0.006 | 0.021** | 0.006 | 0.022** | 0.006 | 0.014** | 0.004 |

*Note. N individual* = 172829; *N countries* = 30; * $p < .01$; ** $p < .001$

**Table 5**. Results of linear multilevel regression analyses of psychosomatic symptoms.

| **Fixed effects** | **Model 1 AIC = 477309** | | **Model 2 AIC = 424795** | | **Model 3.1 AIC = 424796** | | **Model 3.2 AIC = 373648** | | **Model 4.1 AIC = 391760** | | **Model 4.2 AIC = 367775** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | | | 0.440** | 0.030 | 0.439** | 0.029 | 0.448** | 0.029 | 0.543 | 0.307 | 0.744* | 0.259 |
| Age | | | | | | | | | | | | |
| 11 | | | -0.419** | 0.006 | -0.419** | 0.006 | -0.435** | 0.006 | -0.421** | 0.006 | -0.433** | 0.006 |
| 13 | | | -0.176** | 0.006 | -0.176** | 0.006 | -0.183 | 0.006 | -0.174** | 0.006 | -0.180** | 0.006 |
| 15 (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| Gender | | | | | | | | | | | | |
| Boy | | | -0.412** | 0.005 | -0.412** | 0.005 | -0.418** | 0.005 | -0.415** | 0.005 | -0.419** | 0.005 |
| Girl (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| SES | | | -0.083** | 0.008 | -0.083** | 0.008 | -0.086** | 0.009 | -0.076** | 0.009 | -0.086** | 0.009 |
| Policy | | | | | 0.052 | 0.040 | | | 0.078 | 0.048 | | |
| Monitoring | | | | | | | -0.005 | 0.032 | | | 0.003 | 0.028 |
| Facilities | | | | | | | 0.060 | 0.032 | | | 0.069 | 0.026 |
| GDP family benefits | | | | | | | 0.030 | 0.034 | | | 0.084 | 0.033 |
| GDP education | | | | | | | -0.026 | 0.031 | | | -0.062 | 0.033 |
| Adult violence | | | | | | | | | -0.012 | 0.028 | -0.007 | 0.024 |
| Adult well-being | | | | | | | | | -0.005 | 0.005 | -0.008 | 0.005 |
| Income inequality | | | | | | | | | 0.001 | 0.009 | -0.004 | 0.007 |
| Variance components | | | | | | | | | | | | |
| Level 1 (child) variance | 0.979** | 0.003 | 0.902** | 0.003 | 0.902** | 0.003 | 0.904** | 0.003 | 0.893** | 0.003 | 0.902** | 0.003 |
| Level 2 (country) variance | 0.024** | 0.006 | 0.025** | 0.007 | 0.024** | 0.006 | 0.021** | 0.006 | 0.022** | 0.006 | 0.014** | 0.004 |

*Note.* *N individual* = 172829; *N countries* = 30; * $p < .05$; ** $p < .01$; *** $p < .001$

## Psychosomatic Symptoms

Table 5 displays the results of the linear multilevel regression analyses for adolescent psychosomatic symptoms. The null model (Model 1) revealed significant but small country-level variance in psychosomatic symptoms. Of the total variance in psychosomatic symptoms, 2.4% (0.024/(0.024+0.977) = 0.024) was at the country-level. Psychosomatic symptoms increased with age, were lower for boys, and were lower for adolescents with a higher socioeconomic status (Model 2). Including total policy (Model 3.1) yielded no significant association with psychosomatic symptoms. Similarly, when examining the separate policies (Model 3.2), there were no significant associations with psychosomatic symptoms. Controlling for the other national-level variables (which were not significantly related to psychosomatic symptoms, Model 4.1), the association of total policy with adolescent psychosomatic symptoms remained nonsignificant and the same was true for the separate policies (Model 4.2). There was almost no decrease in country-level variance between the models (0.024 to 0.022).

# DISCUSSION

The present study aimed to examine the association between policies for CAMH and indicators of adolescent mental health (i.e., aggressive behavior, life satisfaction, and somatic complaints), controlling for indicators of cultural and social norms (i.e., adult violence, adult well-being, income inequality). We found a relation between combined policies for CAMH and adolescent aggressive behavior, whereby more policies were associated with fewer reported adolescent aggressive behaviors. Policies for CAMH were not associated with adolescent life satisfaction and psychosomatic symptoms. Our findings for aggressive behavior are in line with earlier research that reports a more robust decrease in externalizing behavior problems as a result of improving SES than for internalizing problems (Reiss, 2013). Some of the included CAMH policies comprised of investments to decrease the effects of SES (i.e., investments in family benefits, investments in education) on CAMH, thus this may explain why CAMH policies were associated with adolescent aggressive behaviors, but not with life satisfaction and psychosomatic symptoms. Also, we know more about prevention and treatment of mental health problems than about the promotion of positive mental health (Barry, 2009; Bartels, Cacioppo, Van Beijsterveldt, & Boomsma, 2013; Weisz et al., 2017; Welsh et al., 2015), which may explain that we found an association with CAMH policy for adolescent aggressive behaviors but not for life satisfaction. Finally, high levels of aggressive behaviors may be a more salient societal issue than low levels of life satisfaction and/or high levels of psychosomatic symptoms. Aggressive behaviors are among the most prevalent disorders (NICE, 2013), often co-occur with other disorders or adversity

(Bartels et al., 2018; Sousa, Correia, Ramos, Fraga, & Barros, 2010; Walsh et al., 2013), and are costly for society (Rivenbark et al., 2018; Scott, Knapp, Henderson, & Maughan, 2001). These characteristics may make them a more likely target for policies compared to life satisfaction or psychosomatic complaints.

Although the multi-level analyses failed to find country-level variance for policies for CAMH in adolescent life satisfaction and psychosomatic symptoms, we have reasons to suspect their importance. In contrast to the adolescent mental health indicators, the correlations between adult violence and well-being with the policy variables were strong and in the expected direction (i.e., more policies, less violence and higher levels of well-being, respectively). The adult correlations may indicate that adults have been exposed longer to the positive effect of policies or that effects of policies are more easily accessible to adults (e.g., family benefits) than to adolescents (Gulliver, Griffiths, & Christensen, 2010; Rocha, Graeff-Martins, Kieling, & Rohde, 2015; Tylee et al., 2007). In addition, policies such as investment in family benefits may alleviate stress among low-SES adults in a family more strongly and directly than among low-SES adolescents (e.g., Evans, 2004), and consequently relate more strongly to adult mental health than adolescent mental health. Nevertheless, although not detected in the present study, through alleviation of stress of the adults in the family, these policies still may have a positive effect on adolescent mental health. Future research on the mechanisms underlying the differential relations between policies and adult and adolescent mental health indicators may enhance our understanding of how national CAMH policies may benefit adolescent mental health.

## Strengths and Limitations

The present study included large, nationally representative samples from the HBSC study, allowing us to include individual-level data from 30 countries combined with national-level indicators from renowned statistical institutes (Currie & Alemán-Díaz, 2015). Also, we included both income inequality and adult violence and well-being to account for some confounding differences between countries.

Despite these strengths, a number of limitations should be noted. First, the data were cross-sectional, preventing causal inferences. Examining adolescent mental health in the years after implementing a policy for CAMH may provide important insights in the direction of effects, mechanisms, and changes over time (e.g., some policies may become increasingly effective). Furthermore, longitudinal research may illuminate whether and why the association between CAMH policies and adolescent mental health varies across different indicators of adolescent mental health. It would be interesting to investigate this prospectively profiting from the increase of big data and recording of national-level information (e.g., Fuller, Buote, & Stanley, 2017), but also retrospectively with existing data on adolescent mental health and national-level data from statistical institutes across the last decades.

Second, by combining national-level information with individual-level data, our analyses regarding policies for CAMH only allowed for national-level comparisons. Adolescent mental health may, however, be (more) susceptible to influences of more proximal level contexts such as regions, schools, peers, and the family (Mcleroy et al., 1988; UNICEF, 2016). To add to complexity, these levels may be affected differently by national CAMH policies. To illustrate, adolescents from more disadvantageous backgrounds may benefit more from more structural facilities such as schools or mental health care facilities than adolescents from more advantageous backgrounds (e.g., Hatch et al., 2012; Phelan, Link, Diez-Roux, Kawachi, & Levin, 2004). In addition, implementation of CAMH policies may vary within countries, and therefore national-level information may not yet provide a complete impression of the association between CAMH policies and adolescent mental health. Research on these mechanisms and interactions of the effects of policies for CAMH on adolescent mental health including more proximal levels of influence would therefore be particularly promising.

## Conclusion

Combining national-level information from renowned institutes with the individual-level data from the HBSC study allowed us to examine whether policies for CAMH are associated with adolescent mental health. We found less adolescent aggressive behaviors in countries with more policies for CAMH. There was no association between policies for CAMH and adolescent life satisfaction and psychosomatic symptoms. To examine whether and how policies may affect adolescent life satisfaction and psychosomatic symptoms, more research is needed. Nevertheless, our findings provide a good starting point for further research on the implications of policies for CAMH for adolescent mental health.

**Supplementary Table 1**. Untransformed values of the separate policies.

| Countries | GBD coverage | Number of CAMHS per 100000 young people | Number of CAMH psychiatrists per 100000 young people | Number of CAMH psychologists per 100000 young people | OECD % of GDP spent on family benefits in 2013, 2012, or 2011 | % of GDP spent on education in 2011/2010 |
|---|---|---|---|---|---|---|
| Austria | 0 | 0.7 | 6.0 | - | 2.61 | 5.8 |
| Belgium | 1 | 2.4 | 11.1 | - | 3.33 | 6.55 |
| Bulgaria | 0 | 0.5 | 1.9 | 1.7 | 1.45* | 3.82 |
| Croatia | 0 | 1.3 | 6.3 | 3.1 | 0.96* | 4.21 |
| Czech Republic | 0 | 0.8 | 6.8 | 4.4 | 3.14 | 4.51 |
| Denmark | 1 | 1.1 | 10.3 | 22.4 | 3.66 | 8.75 |
| Estonia | 1 | 2.1 | 16.8 | 25.2 | 2.13 | 5.16 |
| Finland | 6 | 12.9 | 36.0 | 36.9 | 3.21 | 6.76 |
| France | 5 | 3.0 | 9.1 | - | 3.65 | 5.68 |
| Germany | 4 | 4.1 | 8.0 | 32.9 | 3.03 | 4.98 |
| Greece | 1 | 2.4 | 16.3 | - | 1.27* | 4.09 |
| Hungary | 1 | 3.1 | 3.4 | 8.4 | 3.57 | 4.71 |
| Iceland | 1 | - | - | - | 3.63 | 7.36 |
| Ireland | 2 | 5.2 | 5.2 | 5.1 | 3.40 | 6.15 |
| Italy | 4 | 2.1 | 20.0 | - | 1.97 | 4.29 |
| Latvia | 1 | 5.3 | 11.2 | - | 1.48* | 4.96 |
| Lithuania | 0 | 0.9 | 14.0 | 21.0 | 2.13* | 5.17 |
| Luxembourg | 0 | 1.9 | 21.4 | 65.3 | 3.61 | 3.15 |
| Malta | 0 | 3.0 | 3.0 | - | 1.24* | 7.96 |
| Netherlands | 5 | 3.2 | 10.7 | - | 1.81 | 5.93 |
| Norway | 5 | - | - | - | 3.13 | 6.66 |
| Poland | 0 | 2.5 | 3.5 | - | 1.61 | 4.94 |
| Portugal | 1 | 1.8 | 5.4 | 4.7 | 1.43 | 5.27 |
| Romania | 0 | - | 3.1 | - | 1.63* | 3.07 |
| Slovakia | 0 | 3.6 | 3.6 | 2.7 | 2.06 | 4.06 |
| Slovenia | 0 | 9.7 | 6.0 | 15.4 | 1.96 | 5.68 |
| Spain | 4 | 2.4 | - | - | 1.46 | 4.82 |
| Sweden | 2 | 1.0 | 23.4 | 104.2 | 3.64 | 6.82 |
| Switzerland | 2 | - | - | - | 2.03 | 5.28 |
| UK | 6 | 7.0 | 4.5 | - | 3.95 | 5.98 |

*Note.* Values with an asterisk were derived from the World Social Protection Report.

**Supplementary Table 2.** *Results of linear multilevel regression analyses of fighting separately*

| Fixed effects | Model 1 AIC = 465348 | | Model 2 AIC = 420961 | | Model 3 AIC = 420953 | | Model 3 sub AIC = 373715 | | Model 4 AIC = 391276 | | Model 4 sub AIC = 367636 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | | | -0.357** | 0.019 | -0.356** | 0.017 | -0.342** | 0.015 | -0.016 | 0.158 | -0.018 | 0.129 |
| Age | | | | | | | | | | | | |
| 11 | | | 0.181** | 0.006 | 0.181** | 0.006 | 0.173** | 0.006 | 0.177** | 0.006 | 0.174** | 0.006 |
| 13 | | | 0.119** | 0.006 | 0.119** | 0.006 | 0.115** | 0.006 | 0.112** | 0.006 | 0.114** | 0.006 |
| 15 (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| Gender | | | | | | | | | | | | |
| Boy | | | 0.523** | 0.005 | 0.523** | 0.005 | 0.530** | 0.005 | 0.529** | 0.005 | 0.531** | 0.005 |
| Girl (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| SES | | | 0.003 | 0.008 | 0.003 | 0.008 | 0.003 | 0.009 | 0.006 | 0.009 | 0.003 | 0.009 |
| Policy | | | | | -0.074* | 0.022 | | | -0.072* | 0.024 | | |
| Monitoring | | | | | | | -0.028 | 0.015 | | | -0.022 | 0.014 |
| Facilities | | | | | | | -0.059* | 0.015 | | | -0.057** | 0.013 |
| GDP family benefits | | | | | | | 0.031 | 0.016 | | | 0.040 | 0.016 |
| GDP education | | | | | | | -0.015 | 0.015 | | | -0.031 | 0.017 |
| Adult violence | | | | | | | | | -0.000 | 0.014 | 0.003 | 0.012 |
| Adult well-being | | | | | | | | | -0.002 | 0.002 | -0.003 | 0.002 |
| Income inequality | | | | | | | | | -0.010 | 0.004 | -0.009 | 0.004 |
| Variance components | | | | | | | | | | | | |
| Level 1 (child) variance | 0.989** | 0.003 | 0.895** | 0.003 | 0.895** | 0.003 | 0.918** | 0.004 | 0.905** | 0.003 | 0.916** | 0.004 |
| Level 2 (country) variance | 0.011** | 0.003 | 0.010** | 0.003 | 0.007** | 0.002 | 0.005** | 0.001 | 0.006** | 0.002 | 0.003** | 0.001 |

*Note. N individual* = 172829; *N countries* = 30; * $p < .01$; ** $p < .001$

**Supplementary Table 3.** *Results of linear multilevel regression analyses of bullying separately*

| **Fixed effects** | **Model 1 AIC = 461611** | | **Model 2 AIC = 420063** | | **Model 3 AIC = 420056** | | **Model 3 sub AIC = 375808** | | **Model 4 AIC = 395786** | | **Model 4 sub AIC = 371086** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | | | -0.084 | 0.049 | -0.082 | 0.042 | -0.077 | 0.043 | -0.707 | 0.336 | -0.631 | 0.337 |
| Age | | | | | | | | | | | | |
| 11 | | | -0.080** | 0.006 | -0.080** | 0.006 | -0.081** | 0.007 | -0.080** | 0.006 | -0.080** | 0.007 |
| 13 | | | 0.016* | 0.006 | 0.016* | 0.006 | 0.027** | 0.006 | 0.021** | 0.006 | 0.027** | 0.006 |
| 15 (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| Gender | | | | | | | | | | | | |
| Boy | | | 0.244** | 0.005 | 0.244** | 0.005 | 0.251** | 0.005 | 0.256** | 0.005 | 0.251** | 0.005 |
| Girl (ref) | | | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| SES | | | 0.006 | 0.008 | 0.006 | 0.008 | 0.000 | 0.009 | 0.009 | 0.009 | 0.001 | 0.009 |
| Policy | | | | | -0.194* | 0.059 | | | -0.213** | 0.052 | | |
| Monitoring | | | | | | | -0.108 | 0.048 | | | -0.071 | 0.036 |
| Facilities | | | | | | | 0.015 | 0.048 | | | -0.011 | 0.034 |
| GDP family benefits | | | | | | | -0.013 | 0.051 | | | -0.045 | 0.043 |
| GDP education | | | | | | | -0.088 | 0.046 | | | -0.111 | 0.043 |
| Adult violence | | | | | | | | | 0.150** | 0.031 | 0.154** | 0.032 |
| Adult well-being | | | | | | | | | 0.012 | 0.005 | 0.015 | 0.006 |
| Income inequality | | | | | | | | | 0.005 | 0.009 | 0.001 | 0.010 |
| Variance components | | | | | | | | | | | | |
| Level 1 (child) variance | 0.929** | 0.003 | 0.898** | 0.003 | 0.898** | 0.003 | 0.0942** | 0.004 | 0.942** | 0.004 | 0.949** | 0.004 |
| Level 2 (country) variance | 0.070** | 0.018 | 0.070** | 0.018 | 0.052** | 0.013 | 0.047** | 0.013 | 0.026** | 0.007 | 0.024** | 0.007 |

*Note. N individual* = 172829; *N countries* = 30; * $p < .01$; ** $p < .001$

Chapter 4

# Predicting Childhood Aggression: Mining Large Datasets Followed by Confirmatory Modeling.

Submitted as: Hendriks, A. M., Luningham, J., Hong, M., Jaccobucci, R., Lundström, S., Larsson, H., … Lubke, G. (2019). Predicting childhood aggression: Mining large data followed by confirmatory models.

## ABSTRACT

Background: The aim was to predict childhood aggression, by analyzing data from two large European cohorts (N = 62,227), with a novel methodological approach.

Data: Participants came from the Child and Adolescent Twins Study in Sweden and the Netherlands Twin Register. The outcome measure was physical overt aggression as assessed around age 9, psychometrically harmonized across multiple European cohorts. The large set of predictor variables consisted of demographics, prenatal characteristics, physical development, family environment, parenting, parental education level, life events, and behavioral symptoms.

Method: To avoid capitalization of chance, data were partitioned in four parts for different analysis steps. These included 1) exploratory data analysis and tuning meta-parameters for data mining, 2) fitting increasingly complex data mining models to assess which predictors had which types of effects, 3) assessment of model performance and importance of the predictor variables, and 4) fitting a confirmatory prediction model of aggression that integrated results of the data mining analyses.

Results: The prediction model confirmed linear main effects of predictor variables and included interactions of predictors with sex and cohort. Associations between the main predictors (non-physical aggression, ADHD, conduct disorder, maternal smoking during pregnancy, parenting style, and life events) and childhood aggression were in line with previous research, yet weaker, likely because we considered more predictors simultaneously.

Conclusion: Fitting all predictors simultaneously provided clear insight in the importance of predictors relative to each other. Externalizing, non-aggressive behaviors had the strongest effects, and may act as salient targets for early detection and prevention of childhood aggression.



**Keywords:**
Childhood aggression, data mining, confirmatory prediction model




## GENERAL SCIENTIFIC SUMMARY

Salient behaviors associated with ODD, CD, and ADHD such as arguing, being easily distracted, and hyperactivity appear to best predict childhood aggression, above prenatal, physical, and environmental predictors (e.g., maternal smoking during pregnancy, parental education level). This is consistent with previously reported high comorbidity of childhood aggression with other behavioral and emotional problems. These behaviors, which could be noticed by people in the environment, may facilitate early detection and prevention of (behavior problems related to) childhood aggression.

Although childhood aggression has been the focus of a large body of research, insight in the mechanisms of predictors on the development of childhood aggression and its psychosocial factors is still limited (Jaffee, Strait, & Odgers, 2012). Childhood aggression receives much attention due to its high prevalence (NICE, 2013), comorbidity with other disorders (Bartels et al., 2018), adverse outcomes for the individual (Copeland, Wolke, Shanahan, & Costello, 2015), negative consequences for parents (Meltzer, Ford, Goodman, & Vostanis, 2011; Roberts, McCrory, Joffe, De Lima, & Viding, 2018), and high costs for society (Rivenbark et al., 2018; Romeo, Knapp, & Scott, 2006). Although there is a small decrease in the prevalence over the past years (Erskine et al., 2014; Pickett et al., 2013), current treatment effects are still generally small (Hendriks, Bartels, Colins, & Finkenauer, 2018; Weisz et al., 2017). Given the burden on the individual and her/his surroundings as well as society at large, further research is needed to increase knowledge about the precursors of aggression. The ACTION (Aggression in Children: Unraveling gene-environment interplay to inform Treatment and interventION strategies) consortium aims to combine multidisciplinary information from multiple research groups to enhance knowledge on childhood aggression (Bartels et al., 2018; Boomsma, 2015). The present study utilizes the combined sample size and wealth of information in the ACTION consortium data to find a set of robust predictors for childhood aggression, to assess their respective importance using advanced analysis, and to describe their relation to childhood aggression.

The cohorts within the ACTION consortium contain a wide range of variables, which allowed us to include a heterogeneous set of variables to predict childhood aggression. Variable categories comprise of demographics, prenatal characteristics, physical development, the family environment, parenting, parental education level, life events, and behavioral symptoms as reported by mothers. For example, pre- and perinatal characteristics include variables such as birth weight or maternal smoking during pregnancy. Behavioral symptoms as reported by the mother may refer to behaviors associated with forms of childhood psychopathology (e.g., attention-deficit/hyperactivity disorder, anxiety).

While multiple studies have examined some of these predictors of childhood aggression, most focused on individual predictor variables, for example, sex of the child (Archer, 2004; Card, Stucky, Sawalani, & Little, 2008) or socioeconomic status (Piotrowska, Stride, Croft, & Rowe, 2015). Assessing predictors separately may however result in an incorrect estimation of predictor effects because predictors are correlated. Examining predictor variables simultaneously provides unbiased estimates as well as information concerning the importance of predictor variables relative to others (e.g., Sabina & Banyard, 2015). The present study combines two data-rich cohorts in order to investigate which of the different available predictor variables are associated with childhood aggression through main and/or interaction effects. The combined sample size in the present study, however, permits not only investigating large numbers of predictors and assessing potential nonlinear and interaction effects using data mining, but also to follow up with confirmatory analyses in a hold-out set of the data.

Data mining techniques are ideal for handling large data with many variables because there is no need to a priori specify what type of effect (linear/non-linear, main effect/interaction) a given predictor has on the outcome (Miller, Lubke, McArtor, & Bergeman, 2016). This is an important advantage since it is usually impossible to estimate a confirmatory model that includes main and interaction effects of all potential predictors. In this study we used an approach termed "Deductive Data Mining" (DDM; Hong et al., submitted) to inform which effects and variables need to be included in a final confirmatory model to predict childhood aggression. In DDM, increasingly complex data mining models are fitted to the data that differ with respect to the type of permitted effect a predictor can have on the outcome. For instance, the lasso (Tibshirani, 1996) only fits linear main effects whereas tree methods permit nonlinear and interaction effects. By comparing the performance of the different data-mining models one can deduce which predictors and which types of effects lead to the best model performance. Subsequently, the found effects are included in a confirmatory model that is fitted to a hold-out set of the data (i.e., a part of the data that has not previously been used for modeling; e.g., Faraway, 2016) in order to estimate effect sizes and perform statistical significance testing.

In addition to executing a larger scale search for potential predictors of childhood aggression, our study also combined two large European cohorts. Combining data sets from multi-county cohorts increases generalizability, but also poses methodological challenges because predictors and outcomes are often assessed by different instruments. Different item wording can introduce bias in parameter estimates when fitting models to the combined data, which in turn complicates interpretation (Curran & Hussong, 2009; Curran et al., 2008). Within the ACTION consortium, a physical aggression phenotype was harmonized using psychometric modeling (Luningham et al., submitted). This harmonized phenotype served as the outcome in the present study. Regarding the predictor variables, psychometric harmonization was not feasible due to the lack of a phenotypic reference set (e.g., a subsample with overlapping data on all questionnaires). Therefore, we followed the general practice of aligning items based on inspection of item wording followed by thorough exploratory analyses to detect potential cohort differences.

In conclusion, the goal of the present study was to find a set of robust predictors of childhood aggression and investigate how they relate to physical/overt childhood aggression using data from two large ACTION consortium cohorts. The present study provides a significant methodological innovation as this is a first large data mining study followed by a confirmatory analysis applied to substantive data from multiple cohorts, thus permitting recommendations for future collaborative projects. The paper is organized as follows. After describing the data resources, we provide an overview of the analysis strategy followed by a detailed explanation of each analysis step. This part includes the rationale of DDM and information concerning the specific data mining models as well as the measures to evaluate their performance. The results are presented accordingly. We finish by discussing the clinical conclusions and the limitations of our study.



# METHOD

## Data

The ACTION consortium comprises multiple large general population data sets from different cohorts, most of which assessed childhood aggression differently. To facilitate multi-cohort analyses within the ACTION consortium, a harmonized aggression score was created of childhood aggression scores through psychometric modeling (Luningham et al., submitted). The harmonization of the outcome variable was carried out using data from the Child and Adolescent Twins Study in Sweden (CATSS, N = 27,894; Anckarsäter et al., 2011), FinnTwin12 (FT12, N = 4,884; Kaprio, 2013), the Netherlands Twin Register (NTR, N = 34,333; Van Beijsterveldt et al., 2013), The Swedish Twin study of CHild and Adolescent Development (TCHAD, N = 2,181; Lichtenstein, Tuvblad, Larsson, & Carlström, 2007), and the Twins Early Development Study (TEDS, N = 17,267; Haworth, Davis, & Plomin, 2013). The overlap of data available as predictors for aggression across all cohorts was small, with CATSS and NTR forming the exception. CATSS and NTR are the two largest ACTION cohorts (total N = 62,227), and had 27 comparable items tapping into the domains of child and parent characteristics as well as mother-rated ADHD indicators. All analyses were carried out in these two cohorts.

**CATSS.** CATSS was launched in 2004 in order to longitudinally follow development of Swedish twins born in Sweden since 1992 during childhood and adolescence. First data collection was through a telephone interview with the parents of 9/12 year-old twins, followed by questionnaires at ages 9, 12, 15, and 18 (Anckarsäter et al., 2011). The sample for which the harmonized aggression score was available consisted of 27,894 CATSS participants.

**NTR.** NTR is a nationwide population-based register founded in 1987 in the Netherlands to investigate individual differences in mental and physical health. Data collection starts with a questionnaire shortly after birth of the twins, which is followed by age-specific questionnaires at age 2, 3, 5, 7, 9/10, 14, 16, and 18 (Van Beijsterveldt et al., 2013). The sample for which the harmonized aggression score was available consisted of 34,333 NTR participants.

## Variables

**The outcome: A harmonized factor score of overt/physical aggression.** The five ACTION cohorts employed different questionnaires to assess childhood aggression, namely the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 2001) and the Autism - tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC; Hansson, Svanstrom Rojvall, Rastam, Gillberg, & Anckarsater, 2005). Prior to the present study, the aggression phenotype was harmonized across five major participating cohorts of ACTION. Items pertaining to overt/physical aggression were assessed in each of the different questionnaires and combined to model a harmonized aggression score (items listed in Supplementary Table 1). In order to facilitate psychometric modeling additional data were collected on a subsample of 9-year old twins in the NTR (N = 2,316; 2,263 twins with mother report and 1,548 with father report). Mothers and fathers rated their children on all questionnaire items used in ACTION cohorts, resulting in a so-called “reference set”. The reference set permitted modeling item level data and the extraction of a factor score of childhood aggression while controlling for sex, rater, and cohort differences. This factor score represented a generalizable measure of overt or physical aggression (Luningham et al., submitted)

**The predictors.** Questionnaire items were selected with similar content across the two cohorts. We obtained variables in the following categories: demographics, prenatal characteristics, physical development, family environment, parenting, parental education level, life events, and behavioral symptoms as reported by the mother. Demographics comprised of sex, age of the child at the assessment of aggression, and cohort (i.e., CATSS, NTR). Pre- and perinatal characteristics were birth weight, gestational age, maternal smoking during pregnancy, and maternal alcohol use during pregnancy and were assessed at age 9/12 for CATSS and at first contact for the NTR. For CATSS, data collection began at 9/12, so most of the predictor variables were collected at this age. For NTR, data were collected from shortly after birth up to 18 years, but because the sample for age 7 years was largest, we selected this cohort for most variables. Physical development variables consisted of height, weight, asthma, eczema, and medication use. The category family environment included whether a child had siblings (not present in CATSS data), age of mother at birth, age of father at birth, and whether both parents lived in the same household. Parenting assessed parental monitoring. Parental education level comprised of maternal and paternal education level. Life events referred to the proportion of life events that children experienced; both CATSS and the NTR included a list of life events, however containing a different number of items and with different content. To harmonize this variable, we calculated the proportion of life events to which the response was “yes” out of the total number of life events. Behavioral symptoms as reported by the mother (assessed at age 9/12 in CATSS and age 7 in NTR) consisted of motor skills, arguing, lying, bragging, feeling no guilt, short attention, daydreaming, easily distracted, not finishing things, hyperactivity, feeling superior, being bullied, impulsivity, fear of situations, anxiety, and nightmares. Supplementary Table 2 presents information on all variables, whether they were available across both cohorts, and exact item wording for the variables with more than just minor differences in wording.

Because children in the NTR were around age 7 at the included assessment of height and weight and children in CATSS were either age 9 or 12, we corrected height based on an average difference of 11.30 cm between age 7 and 9, and an average difference of 17.60 cm between age 9 and 12 (Bonthuis et al., 2012). Similarly, for weight we corrected for a difference of 5.25 kg between age 7 and 9 (World Health Organization, 2007). Because of growth spurts after age 10, however, the WHO did not report average weight beyond this age, so we corrected for the mean difference in the CATSS data between age 9 and 12 (11.40 kg), which was in line with growth numbers reported for the United Kingdom (Royal College of Pediatrics and Child Health, 2012). Unrealistic values for height (i.e., below 105 cm, above 145 cm) and weight (i.e., below 15 kg, above 45 kg) were considered as missing data (World Health Organization, 2007).

The items related to the quality of parenting differed for CATSS and NTR. NTR started to include these variables in 2010, resulting in a high missingness rate (88%) in the current sample. We therefore tested the effects in cohort specific analyses because of their theoretical importance (e.g., Racz & McMahon, 2011; Stattin & Kerr, 2000). Similar as for parenting, the life events items varied across cohorts, and were thus included in cohort specific analyses (e.g., Grant, Compas, Thurm, McMahon, & Gipson, 2004; Guerra, Huesmann, Tolan, Van Acker, & Eron, 1995).

The behavioral symptoms in the CATSS data were from the A-TAC, with response options 0 = *"No"*, 0.5 = *"Yes, to some extent"*, and 1 = *"Yes"*. The behavioral symptoms in the NTR were from the CBCL, with response options 0 = *"Not true"*, 1 = *"Somewhat or sometimes true"*, and 2 = *"Very true or often true"*. Although the items were measured on 3-point scales in both cohorts, response options were dichotomized due to a differential use of the zero and mid point response options in the two cohorts, with CATSS having a higher frequency of zero than the NTR whereas the pattern was reverse for the mid category. Dichotomization was done by collapsing the zero and midpoint response options in both cohorts, which resulted in very similar item distributions across questionnaires.

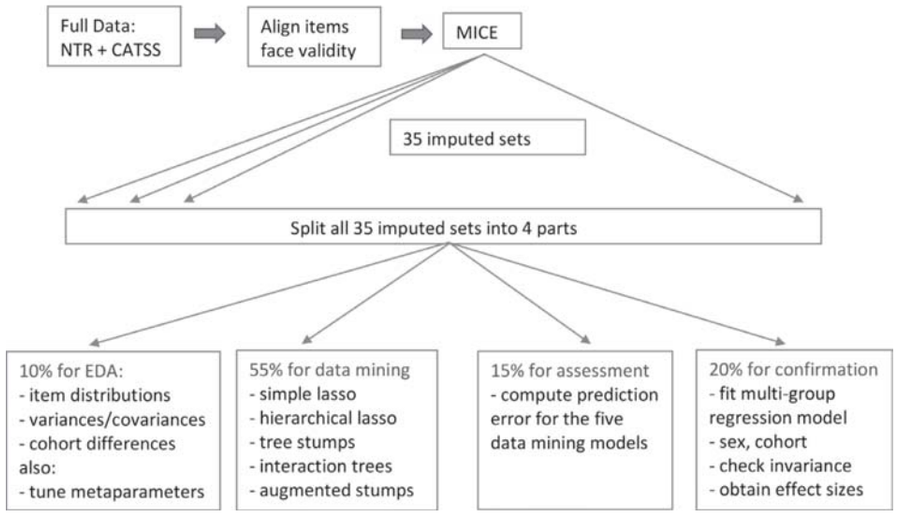
**Table 1.** Workflow of the analyses

| Part | Percentage | Analysis step |
|---|---|---|
| | 100% | Imputation of missing data<br>- Method was multiple imputation by chained equations (MICE)<br>- It creates multiple imputed data sets over which model outcomes may be averaged to obtain more robust results and insight in the variability of outcomes<br>- followed by splitting the data into 4 parts |
| 1 | 10% | Exploratory data analysis<br>- Structural stratification in the predictor variables. E.g., cohort differences, sex differences<br>- Check predictor variables for near-zero variances<br>- Examine whether there are high correlations (> .70) among predictor variables<br>- Tune meta-parameters for the data mining analyses |
| 2 | 55% | Data mining<br>- Simple lasso for linear main effects<br>- Hierarchical lasso for linear main and interaction effects<br>- Boosted tree stumps for linear and non-linear main effects<br>- Boosted interaction trees for linear and non-linear main and interaction effects |
| 3 | 15% | Model performance assessment<br>- Use prediction error and proportion of variance explained as a measure of model performance for the different data mining methods<br>- Variable importance measures to assess the relative importance of each predictor in predicting childhood aggression<br>- Look for convergence across models |
| 4 | 20% | Confirmatory prediction model<br>- Fit a model based on the outcomes of the data mining analyses to predict childhood aggression to obtain effect sizes of the relations between the predictor variables and childhood aggression |

## Analysis Strategy

The analysis of the ACTION data is outlined here and described in detail in the next paragraphs. The workflow is shown in Table 1 and graphically displayed in Figure 1. Prior to the main analyses we carried out multiple imputation by chained equations (MICE to impute missing data, resulting in 35 imputed sets. Next, each of the imputed sets was split into four mutually exclusive parts. This is necessary to avoid capitalization of chance and inflation of Type I error that occurs in sequential analyses without data splitting (Hastie, Tibshirani, & Friedman, 2009; Lubke & Campbell, 2016). Steps 1-4 of the main analyses were carried out in the four different parts of the data. Step 1 involved Exploratory Data Analysis checks and selection of meta-parameters for the data mining models (Part 1 of the data comprised 10% of the sample). The 2nd analysis step implemented the deductive data mining (DDM) procedure using Part 2 (55% of the data). The performance of the different models was assessed in Part 3 of the data (step 3, 15% of the data). The 4th and final step consisted of integrating the results from steps 2 and 3 in a traditional confirmatory model prediction aggression fitted to Part 4 of the data (20%).

**Figure 1.** Process flow




### Step 1: Imputation.

Multiple imputation is a well-known and powerful approach for dealing with complex missing data. In multiple imputation, missing values are filled in by creating a prediction model for the missingness and drawing values from the predictive distribution repeatedly to create multiple complete datasets (Rubin, 1987; Van Buuren & Groothuis-Oudshoorn, 2011). The analysis of interest is carried out in each of the complete datasets and subsequently pooled (Rubin, 1987). The imputation process is more challenging when missing data are present in multiple variables within a dataset. Multivariate missing data can be handled by MICE, in which an iterative process is set up establishing a predictive distribution for each variable with missingness conditional on all other variables (Van Buuren & Groothuis-Oudshoorn, 2011). The procedure is as follows. For the first variable with missing values,

a predictive distribution is established by predicting the variable from all other variables in the dataset for cases in which the variable was not missing. Missing values are filled in by draws from this predictive distribution. The next variable with missing values is then predicted and filled in conditional on all other variables in the data. One iteration is complete after cycling through all variables with missing values. The process continues for multiple iterations to create an imputed dataset, and this is repeated to build multiple imputed datasets.

MICE is advantageous because of its flexibility and its utility with large, complex data. In MICE, the researcher can utilize different models for each variable to establish the predictive distribution; for example, a linear regression can be specified for a continuous variable such as age and a multinomial regression used for a categorical survey item. Additionally, creating multiple imputations reflects uncertainty in the prediction of missing data. Predictive mean matching (PMM) is a common imputation procedure that introduces missing data uncertainty into the imputation. PMM utilizes a linear model to predict a variable with missingness from the other variables, but it includes two steps to add randomness to the procedure. In each iteration, the regression coefficients are first drawn from a multivariate normal distribution based on their estimates and covariance matrix. Then, a predicted value is found for cases in which the variable is missing. However, the imputed value is not the predicted value from the linear model, but is instead a random draw of observed values that are nearest to the predicted value. The number of candidate “neighbors” is typically around 5, 10, or 15. These two steps prevent bias due to overconfidence of the predicted values for missing data.

For the current study, nearly all variables of interest contained some missing data, with the exception of the harmonized aggression score. Multiple imputation was used to obtain the largest number of complete cases possible for the variable selection step, as some of the chosen data mining algorithms rely on complete cases. However, there is evidence that imputing variables with extreme levels of missingness leads to bias and lower power, so we only imputed variables with reasonable amounts of missing data (Kontopantelis, White, Sperrin, & Buchan, 2017). Variables were not included in the imputation procedure if a) the variable had more than 50% missing values in the data combined across NTR and CATSS, or b) the variable was 100% missing within either NTR or CATSS. In all, 26 covariates were imputed prior to data-splitting and subsequent analyses. Each covariate was imputed from all other covariates that met the missing proportion threshold for imputation plus the harmonized aggression outcome, consistent with recommended practices (Sterne et al., 2009; Van Buuren & Groothuis-Oudshoorn, 2011). Imputation was implemented using the R package mice (Van Buuren & Groothuis-Oudshoorn, 2011).



## Step 2: Partitioning of the data.

The motivation to split the data and carry out the different steps of the analysis in separate parts of the data is to prevent capitalization on chance, which occurs if sequential steps of an analysis are carried out in the same data. For instance, assessing the performance of a model in the same data that were used to fit the model would result in overestimating the performance. This is because a model adapts not only to the structure in the data but also to some extent to the idiosyncrasies of the sample. An evaluation of a model in new data shows how well the model can predict the actual structure.

We chose to partition the data into parts of different sizes to optimally leverage the total sample. The partitioning was carried out in the same way in all 35 imputed sets, and was based on random selection at the subject level. A different option would have been to select at the family level, which would have the advantage of creating independent partitions. However, currently the data mining methods used in this study do not offer to account for family relatedness. Since the largest part of the data is used for mining, we preferred to minimize the relatedness within partition by selecting at the subject level. A partition of 10% (N = 6,222) was sufficient to carry out exploratory data checks, and to tune the data mining algorithms. The main part of the analysis focused on fitting different types of data mining approaches. Due to fact that the low signal to noise ratio in behavioral data requires large samples (e.g., Ivanova et al., 2007), we choose to allocate 55% of the data (N= 34,225) for this step. Prediction error served to compare the different data mining models and was calculated in 15% (N = 9,334) of the data. Here we also computed variable importance measures and selected important predictors for the last step. The remaining 20% (N = 12,446) were used to fit a confirmatory model and assess statistical significance and effect sizes of the selected predictors.

## Step 3: Exploratory Data Analysis and Meta-parameters.

Predictor variables were selected based on item content. Since the items stemmed from different questionnaires, the item wording was not identical. Therefore, exploratory analyses were necessary to check whether there was structural stratification across cohorts. In addition, we assessed whether variables had extremely low endorsement rates leading to near-zero variances, and whether there were high correlations between predictor variables (i.e., above .7). We then used the R package caret (Kuhn, 2018) to tune all data mining procedures to obtain optimal meta-parameters. The drawback of using a partition of the data that is smaller compared to the one used the main analysis is the potential dependence of meta-parameters on sample size. Therefore, a small subset of the meta-parameters (but not the entire grid) was checked again during the main analysis.

## Step 4: Data Mining

In this analysis we used an approach termed "Deductive Data Mining" (DDM; Hong et al., submitted). Data mining methods in the regression setting are generally geared towards predictor selection and/or obtaining an optimal prediction performance (i.e., lowest prediction error in future data). Individual methods do not provide guidance regarding the type of effect of the predictors (e.g., linear/non-linear main and/or interaction effects). Deductive Data Mining (DDM) introduces the concept of model comparison that is well-known in confirmatory analyses such as Structural Equation Modeling into the area of data mining. The rationale of DDM is to compare the performance of data mining models that differ with respect to the type of effect that is included in the model. For instance, a simple lasso fits only linear main effects but no non-linear effects or interactions, whereas boosted trees adapt to linear as well as non-linear effects, and can include linear and non-linear interaction effects. By comparing the prediction error of a simple lasso and boosted trees one can assess whether the inclusion of non-linear and/or interaction effects is necessary to improve prediction performance. Table 2 provides an overview of the characteristics of the approaches used in this analysis, and the type of information they can provide in terms of effects. All methods afford the means to rank the predictors according to their importance of predicting aggression, and therefore permit variable selection.

In order of increasing complexity of effects, the different data mining approaches were (1) simple lasso (linear main effects), (2) hierarchical lasso (linear main and linear interaction effects), (3) boosted tree stumps (i.e., trees with a single split, linear and non-linear main effects), and (4) boosted interaction trees (linear and non-linear main and interaction effects). In the next section these four models are explained in more detail.

**Table 2.** Model descriptions for deductive data mining approach

| Models | Linear marginal effects | Non-linear marginal effects | Linear interaction effects | Non-linear interactions |
|---|---|---|---|---|
| 1) Lasso | All | N/A | N/A | N/A |
| 2) Hierarchical lasso | All | N/A | All | N/A |
| 3) Boosted stump tree (tree depth = 1) | All | All | N/A | N/A |
| 4) Boosted interaction tree (tree depth = 5) | All | All | All | All |
| 5) Confirmatory regression model with new data | Specified | Specified | Specified | Specified |



### *Regularization methods: Lasso and hierarchical lasso*

The least absolute shrinkage and selection operator (*lasso;* Tibshirani, 1996, 2011) as well as the hierarchical lasso (Bien, Taylor, & Tibshirani, 2013; Choi, Li, & Zhu, 2010; Haris, Witten, & Simon, 2016) are multiple regression models where regression coefficients are regularized. Regularization involves placing constraints (penalties) on the coefficients in a regression model such that the sum of square coefficients cannot exceed a penalty value, thus shrinking the coefficients from their OLS values. The aim of regularization is to improve the stability and the generalizability of the model. The lasso can shrink the coefficients all the way to 0, thereby performing variable selection.

Whereas the simple lasso only fits linear main effects, the *hierarchical lasso* includes interactions between variables and quadratic terms for each variable that is identified as a non-zero main effect. This produces a computationally efficient method to identify interactions and quadratic effects in a regression framework. Both lasso and hierarchical lasso only model linear effects. By comparing their performance to that of boosted stumps and interaction trees, respectively, one can deduct whether non-linear effects are necessary.

The amount of regularization in the lasso and hierarchical lasso is a meta-parameter that in our study was optimized in Part 1 of the data. The R package glmnet (Friedman et al., 2018; R Core Team, 2018) was used for the lasso to compare the fit of 100 penalty values using the lowest value of MSE to select the meta-parameter. A number of packages in R exist to implement the hierarchical lasso. We used the hierNet package (Bien & Tibshirani, 2014) to compare 20 penalty values, again with MSE as the selection criterion.

### *Boosted tree stumps and interaction trees*

Trees are built in a recursive fashion (Breiman, 1984; Friedman, 2001). Similar to step-wise regression, the first step is to select the predictor that has the strongest association with the outcome. Rather than estimating a regression coefficient, a cut point on the predictor is obtained that optimally partitions the sample into two groups that are more homogeneous with respect to the outcome. For instance, suppose age is the strongest predictor of income. The algorithm searches for the optimal cut point (e.g., 25 years) that results in two groups (older/younger than 25) that are jointly the most homogenous with respect to income. The algorithm is then repeated in both partitions (called daughter nodes). The recursive partitioning of the nodes results in a tree structure. Single decision trees are popular because they are easy to interpret and visualize. However, the structure of a single tree structure depends heavily on the sample data. A new sample can result in a different choice of the splitting variables, and therefore in a very different tree structure.

Boosted trees combine a large number of trees to improve the prediction quality of a single decision tree (Friedman, 2001). The individual trees in such an ensemble can be specified to feature only a single split, resulting in so-called stumps. Since there are no conditional splits, tree stumps only capture main effects. Since many tree stumps are combined that may feature the same predictor but different cut points, stumps adapt to linear and non-linear main effects. Interaction effects are captured by permitting subsequent conditional splits. By comparing boosted stumps to trees with more splits (called interaction trees in this paper), one can deduct whether interactions are contributing to the prediction performance.

Trees can form the basis to rank all predictors according to their importance in predicting the outcome, but these methods do not lend themselves to deduct which of the individual variables are involved in the interactions. In order to evaluate specific second order interactions, we computed pairwise product terms of the predictors in each of the 35 imputed sets. These product terms were then included in augmented data sets to which we fitted tree stumps. The resulting variable importance measures provide the sought-after indication of which second order interactions are associated with the outcome. If the augmented stump model does not underperform the full-fledged boosted interaction model then one can deduct that only second order interactions are required.

Boosted trees require specification of three meta-parameters. The shrinkage parameter controls the contribution of each tree to the model, and therefore controls the speed with which the model adapts to the data. The shrinkage parameter is interrelated with the second meta-parameter, the number of trees to include in the model, with slower adaptation requiring more trees to be added to the model. The third parameter is the number of splits (also called tree depth), which controls the maximum order of interactions that can be modeled. In the current study we use the R package caret (Kuhn, 2018) to select these meta-parameters in Part 1 of the data and the package gbm (Greenwell, Boehmke, Cunningham, & GBM Developers, 2019, version number 2.1.5) to fit boosted trees to the aggression data.

### *Data mining analyses*

We fitted the lasso, hierarchical lasso, boosted stump, and boosted interaction trees to part 2 of each of the 35 imputed data sets. Several potentially interesting predictors were only available in the NTR sample, and were investigated in additional analyses. These variables were (1) maternal alcohol use during pregnancy, (2) bragging, (3) feeling no guilt, (4) feeling superior, and (5) impulsivity. Items relating to parenting quality and the proportion of serious life events were only available in a small number of subjects in the NTR and were measured rather differently in the Dutch and Swedish samples. Therefore these items were only investigated in CATSS.



### Step 5: Prediction Error and Variable Importance Measures

The performance of the four data mining models was assessed in part 3 of the data. We computed mean squared error (MSE) to quantify the prediction error of each model in new data. In addition, we obtained variable importance measures (VIMs; Friedman, 2001), which provide some guidance regarding which variables explain the most variance in childhood aggression.

### Step 6: Confirmatory Model Predicting Overt/Physical Aggression in Children

Based on the comparison of the performance of the different data mining models it is possible to deduct which type of effects need to be included in a confirmatory model for an optimal prediction of childhood aggression. In the last part of the analysis we fitted a multi-group multiple regression model to part 4 of the data. The model included the effects of variables that appeared to be associated with childhood aggression model in deductive data mining.

## RESULTS

### Imputation

The imputation procedure was carried out with 30 iterations per dataset and 35 imputations in total. Thirty iterations were chosen to ensure proper convergence of the imputation for each dataset, though fewer iterations are typically required (e.g., 10-20; Van Buuren & Groothuis-Oudshoorn, 2011). Thirty-five datasets were imputed because the general recommendation is to create a number of datasets roughly equal to the percentage of missing data (Bodner, 2008). The average marginal missingness for the 26 imputed variables was 26.9%, so 35 imputations with 30 iterations each were expected to obtain a sufficient representation of the missing data. For all imputations, variability was introduced into the regression coefficients as described above. For continuous data, imputed values were drawn based on the 10 nearest neighbors. For binary outcomes, imputed values were found by drawing from a Bernoulli distribution using the predicted response probability.

All variables in the imputation were found to demonstrate good consistency and levels of convergence. The behavioral symptom variables generally reached a stable predicted value and standard deviation after around 20 iterations in each imputation, and the other variables reached convergence after about five iterations (convergence plots are presented in Supplementary Figure 1). The variables

generally demonstrated a small amount of between-imputation variability, especially the binary variables. This is likely due to the preponderance of zero's found in complete data, leading to a high likelihood of the imputed value also being zero.

Each imputed dataset was partitioned into the four analysis partitions. The imputed data were then checked by calculating descriptive statistics in the exploratory partition. Table 3 presents means, between- and within-imputation variability, and whether the variable was imputed based on the criteria of less than 50% missing values in combined data or 100% missing within a single cohort. Additionally, the proportion of missingness for each variable is presented in the combined data and within cohort.

## Results Part 1 Data Analyses: Exploratory Data Analysis and Meta-Parameters

Low variances in the combined NTR/CATSS sample were detected for the following dichotomous items: motor skills (0.0262), lying (0.0142), stealing (0.0083), daydreaming (0.0524), being bullied (0.0214), fear of situations (0.0443), anxious (0.099), and nightmares (0.0157). Given the sample sizes in parts 2-4 of the analyses, these variances were deemed unproblematic as they translate to a sufficient number of endorsements in all parts of the data. Low variances in the NTR and CATSS cohort specific samples were very similar to the combined sample. A check for multicollinearity revealed that none of the correlations between predictors exceeded 0.75. Correlations with cohort that exceeded 0.1 were height (r=0.289, Dutch children taller), living with both parents (r=0.11, Dutch children more likely), eczema (r=.254, Swedish more likely), as well as parenting and life events. The latter two variables were expected to correlate with cohort since they were measured differently across cohort. These two variables were only included in CATSS specific analyses since the missingness in the NTR was high due to only including these questions in later surveys (see Table 3).

We also used Part 1 of the data to select the optimal set of meta-parameters of the different data mining methods such as the maximum number of trees necessary, number of splits for a given tree, shrinkage rate, and number of minimum observations in a terminal node in the tree methods. The motivation to train the model on a smaller portion of the data is to reduce the computational burden. The optimal tuning parameters are presented in Table 4.



**Table 3.** Imputation coefficients.

| Variable | Imputed? | Mean | Total SE | W/in SE | B/w SE | Total Missing | NTR Missing | CATSS Missing |
|---|---|---|---|---|---|---|---|---|
| Aggression | No | 0.2194 | 0.5780 | 0.5780 | 0.0058 | 0 | 0 | 0 |
| Birth weight | Yes | -0.0184 | 0.9784 | 0.9783 | 0.0116 | 0.0435 | 0.0556 | 0.0286 |
| Gest. Age | Yes | 36.5213 | 2.6520 | 2.6518 | 0.0305 | 0.106 | 0.0482 | 0.1772 |
| Maternal Smoking | Yes | 0.2113 | 0.4083 | 0.4083 | 0.0057 | 0.2785 | 0.0463 | 0.5642 |
| Maternal Alc. | No | 0.1801 | 0.3843 | 0.3842 | 0.0069 | 0.4746 | 0.0477 | 1 |
| Height | Yes | 126.4288 | 6.5617 | 6.5614 | 0.0665 | 0.263 | 0.4217 | 0.0677 |
| Weight | Yes | 25.4686 | 5.2574 | 5.2570 | 0.0629 | 0.269 | 0.41 | 0.0955 |
| Asthma | Yes | 0.1154 | 0.3195 | 0.3195 | 0.0033 | 0.1669 | 0.2869 | 0.0192 |
| Eczema | Yes | 0.1292 | 0.3355 | 0.3354 | 0.0043 | 0.1661 | 0.2871 | 0.0172 |
| Medication | Yes | 0.1246 | 0.3303 | 0.3303 | 0.0036 | 0.438 | 0.7692 | 0.0304 |
| Siblings | No | 0.9033 | 0.8999 | 0.8998 | 0.0167 | 0.5811 | 0.2408 | 1 |
| Both parents | Yes | 0.8555 | 0.3516 | 0.3516 | 0.0038 | 0.3419 | 0.6063 | 0.0164 |
| Mother Age | Yes | 31.0550 | 4.1657 | 4.1653 | 0.0591 | 0.2013 | 0.0173 | 0.4278 |
| Father Age | Yes | 33.4881 | 5.0243 | 5.0239 | 0.0624 | 0.2191 | 0.0341 | 0.4468 |
| Parenting score | No | 3.6493 | 0.3061 | 0.3060 | 0.0056 | 0.4934 | 0.8816 | 0.0157 |
| Mother Edu. | Yes | 3.1880 | 0.7700 | 0.7699 | 0.0092 | 0.3667 | 0.2853 | 0.4668 |
| Father Edu. | Yes | 3.1356 | 0.8311 | 0.8310 | 0.0091 | 0.4151 | 0.3344 | 0.5143 |
| Prop. LE | No | 0.1119 | 0.1358 | 0.1358 | 0.0020 | 0.6155 | 0.756 | 0.4425 |
| Motor skills | Yes | 0.0269 | 0.1618 | 0.1618 | 0.0022 | 0.1869 | 0.2794 | 0.073 |
| Argues | Yes | 0.0662 | 0.2487 | 0.2487 | 0.0031 | 0.179 | 0.279 | 0.0559 |
| Lying | Yes | 0.0144 | 0.1191 | 0.1191 | 0.0015 | 0.1819 | 0.2798 | 0.0614 |
| Stealing | Yes | 0.0083 | 0.0909 | 0.0909 | 0.0011 | 0.1897 | 0.2776 | 0.0815 |
| Brags | No | 0.0585 | 0.2348 | 0.2348 | 0.0037 | 0.6033 | 0.281 | 1 |
| Feels no guilt | No | 0.9394 | 0.2386 | 0.2385 | 0.0044 | 0.6022 | 0.279 | 1 |
| Short attention | Yes | 0.0832 | 0.2761 | 0.2761 | 0.0035 | 0.2355 | 0.2879 | 0.1709 |
| Daydream | Yes | 0.0554 | 0.2289 | 0.2288 | 0.0026 | 0.2304 | 0.2776 | 0.1724 |
| Distracted | Yes | 0.0910 | 0.2876 | 0.2876 | 0.0035 | 0.4994 | 0.7675 | 0.1694 |
| Doesn't finish things | Yes | 0.0571 | 0.2321 | 0.2320 | 0.0032 | 0.4761 | 0.7677 | 0.1172 |
| Hyperactive | Yes | 0.0773 | 0.2670 | 0.2670 | 0.0040 | 0.2091 | 0.2774 | 0.1251 |
| Superior | No | 0.8962 | 0.3050 | 0.3050 | 0.0057 | 0.6021 | 0.2788 | 1 |
| Bullied | Yes | 0.0219 | 0.1462 | 0.1462 | 0.0015 | 0.18 | 0.2789 | 0.0583 |
| Impulsive | No | 0.0617 | 0.2406 | 0.2406 | 0.0040 | 0.603 | 0.2804 | 1 |
| Fearful | Yes | 0.0464 | 0.2104 | 0.2104 | 0.0027 | 0.4191 | 0.2862 | 0.5826 |
| Anxious | Yes | 0.0100 | 0.0997 | 0.0997 | 0.0013 | 0.4454 | 0.2785 | 0.6508 |
| Nightmare | Yes | 0.0159 | 0.1251 | 0.1251 | 0.0014 | 0.2199 | 0.2783 | 0.148 |

*Note.* Summary stats across imputed datasets for combined NTR and CATSS data in the exploratory partitions. Variables were excluded from imputation if the total missingness exceeded 50%, or if the variable was 100% missing in one of the two cohorts. Variability is expressed in the Total Standard Error (SE), which is comprised of SE for the estimate within each imputed dataset and variability between imputed datasets. Between SE is present due to missing data uncertainty. Gest. Age denotes gestational age; Alc. denotes alcohol use during pregnancy; Mother Edu. denotes maternal educational qualifications; Father Edu. denotes paternal educational qualifications; Prop. LE denotes the proportion of life events for which the response was yes; LE denotes Life Events. Note that small between-imputation variability is present for non-imputed variables because the partitions were randomized within imputed dataset.

**Table 4.** Optimal set of tuning parameters.

| | Tree depth = 1 | Tree depth = 5 |
|---|---|---|
| Number of trees | 18000 | 10000 |
| Shrinkage | 0.001 | 0.001 |
| Minimum observations | 100 | 100 |

**Table 5.** Average prediction error across imputed data sets for each data mining model.

| Model | Average $R^2$ | Average MSE |
|---|---|---|
| Hier_Lasso | 0.219 (0.010) | 0.260 (0.003) |
| Lasso | 0.265 (0.009) | 0.244 (0.003) |
| Stump | 0.265 (0.009) | 0.244 (0.003) |
| Interaction Tree | 0.280 (0.009) | 0.239 (0.003) |
| Augmented Stump | 0.279 (0.009) | 0.239 (0.003) |

*Note.* Standard Deviations in parenthesis are for across imputed data sets.

## Results Part 2 and Part 3 Data Analyses: Effect Type Selection and Predictor Selection

### Effect Type Selection

Table 5 presents $R^2$ and MSE for the five models averaged across imputed data sets that were calculated in part 3 of the data. MSE measures the squared divergence of the model predicted outcome and the observed outcome aggregated over all subjects whereas $R^2$ expresses MSE as a proportion of the variance of the outcome. When computed in new data, prediction error quantifies how well a new observation would be predicted by the model while $R^2$ quantifies how well a model "explains" the individual differences in the outcome, in this case aggression.

The $R^2$ of the boosted interaction model was consistently higher than those of the other models; the MSE was consistently lower (see Table 5). These results provide evidence that there were at least some interactions and potentially non-linear effects. We then fit a boosted stump model to data that were augmented with pre-calculated product terms of the predictors. This model served to evaluate second order interaction effects, and is listed as augmented stump model in Table 5. The augmented stump model had an MSE and $R^2$ similar to the interaction tree model, permitting the conclusion that higher than second-order interaction effects did not contribute substantially to the prediction of physical/overt aggression. The augmented stump model was therefore used to select specific predictors and interaction effects to be included in the confirmatory prediction model fitted to Part 4 of the data.



### Predictor Selection

We used Variable Importance Measures (VIMs) described by Friedman (2001) and integrated in the R package gbm to select predictors (Greenwell et al., 2019). The VIMs of a given model are scaled to sum up to 100, thus resulting in a percentage scale (i.e., percentage of the contribution of an individual predictor to the total prediction of a model). Since the objective of this study was to cast the net wide and investigate all possibly interesting predictors of aggression we used a cutoff of 0.5 to select predictors for the confirmatory last part of the analyses. This criterion resulted in selecting the following 12 main effects for the prediction model: argues (34.935), distracted (17.050), hyperactive (11.905), lying (4.023), stealing (1.927), daydreaming (1.549), being bullied (1.209), age mother (1.032), maternal smoking during pregnancy (0.957), sex of the child (0.836), education father (0.636), short attention (0.572), and living with both parents (0.519). In addition, the following 8 product terms had VIMs larger than 0.5: cohort x sex (3.874), cohort by daydreaming (3.330), argues x sex (2.653), argues x cohort (2.589), argues x distracted (1.208), stealing x cohort (1.153), hyperactive x sex (0.838), argues x age (0.557). As can be seen, with the exception of argues x distracted and argues x age the interaction terms involved either cohort or sex. These six interactions can be modeled using multi-group modeling while permitting group-specific regression coefficients. The two interactions involving arguing were investigated as product terms in the confirmatory model.

## Results Part 4 Data Analysis: Multi-Group Regression Model

The effect type and predictor selection carried out in parts 2 and 3 of the data permitted a substantial reduction of possible effects from the initial 27 potentially interacting predictors to a total of 13 main and two interaction effects, and additionally six potentially group-specific effects.

We fit two multi-group regression models to the 35 imputed sets of Part 4 of the data, with sex and cohort as group defining variables. The base model featured the harmonized aggression outcome predicted by the 12 variables listed in the previous section and the two interaction terms (argues by distracted and argues by age), all with invariant regression coefficients across groups. Age was included as a main effect due to its participation in the interaction argues by age. This constrained model was compared to a model in which the variable argues was permitted to have sex and cohort specific effects, and the variable hyperactive to have sex specific effects. Group-specific regression coefficients capture the interactions with the grouping variable. The comparison of these two models evaluates the necessity of permitting interactions of arguing, daydreaming, stealing, and hyperactive with sex and/or cohort, respectively.

Comparing the two models in a likelihood ratio test aggregated over 35 imputed sets rejected the constrained model (chisq = 62.60, df = 6, p-value = 0). All other fit indices (AIC, BIC, RMSEA, and SRMR) were also consistently supporting the model with group-specific effects.

Table 6 shows the percentage out of 35 imputed sets each of the regression coefficients was significant. The table also shows the standardized regression coefficients using the group-specific variance of aggression for standardization. Note that standardized regression coefficients of dichotomous variables need to be interpreted with caution.

As can be seen in Table 6, the regression coefficients are in line with the ranking of VIMs in part 3 of the data, and reveal mild group differences. In general, arguing had the largest positive association with physical/overt aggression, followed by the items distracted and hyperactive. The variable daydreaming was significant in the CATSS sample in all 35 imputed sets but only in 31.4% in the NTR data, and had larger coefficients in CATSS. This difference is most likely due to item wording differences in the two cohorts (i.e., in CATSS but not in NTR the item wording included not listening when spoken to). Maternal smoking during pregnancy contributed to higher aggression, whereas living with both parents, higher educational level of the father, and higher age of the mother resulted in lower aggression. These effects were significant in almost all imputed sets, and can therefore be considered as robust. Short attention and the interactions of argues with distracted and with age were not significant in most imputed sets. Effect sizes can be found in Tables 6, 7 and 8, and are compared in detail to previous research in the discussion section.

**Table 6.** Part 4 analyses: Regression coefficients joint model/unequal coefficients.

| | NTR male | NTR female | % of sets significant | CATSS male | CATSS female | % of sets significant |
|---|---|---|---|---|---|---|
| Maternal smoking | 0.034 | 0.042 | 0.94 | 0.039 | 0.045 | 0.94 |
| Both parents | -0.026 | -0.031 | 0.91 | -0.035 | -0.040 | 0.91 |
| Age mother | -0.029 | -0.036 | 0.97 | -0.034 | -0.039 | 0.97 |
| Education mother | -0.010 | -0.013 | 0.43 | -0.014 | -0.016 | 0.43 |
| Education father | -0.022 | -0.027 | 0.97 | -0.033 | -0.039 | 0.97 |
| Child argues | 0.378 | 0.385 | 1.00 | 0.266 | 0.299 | 1.00 |
| Child lies | 0.079 | 0.072 | 1.00 | 0.066 | 0.069 | 1.00 |
| Child steals | 0.053 | 0.049 | 1.00 | 0.077 | 0.058 | 0.94 |
| Child has short attention | 0.007 | 0.006 | 0.09 | 0.007 | 0.006 | 0.09 |
| **Child daydreams** | **0.021** | **0.020** | **0.31** | **0.086** | **0.069** | **1.00** |
| Child is distracted | 0.138 | 0.118 | 1.00 | 0.157 | 0.136 | 1.00 |
| Child is hyperactive | 0.133 | 0.107 | 1.00 | 0.114 | 0.082 | 1.00 |
| Child is being bullied | 0.038 | 0.028 | 1.00 | 0.056 | 0.054 | 1.00 |
| Arguing by distracted | -0.007 | -0.006 | 0.03 | -0.005 | -0.004 | 0.03 |



**Table 7.** Part 4 analyses: Regression coefficients NTR-specific model.

| | NTR male | NTR female | % of sets significant |
|---|---|---|---|
| Age of the child | 0.040 | 0.050 | 0.97 |
| Maternal alcohol | 0.009 | 0.011 | 0.09 |
| Child brags | 0.127 | 0.067 | 1.00 |
| Child feels no guilt | -0.035 | -0.044 | 0.74 |
| Child feels superior | -0.070 | -0.082 | 1.00 |
| Child is impulsive | 0.099 | 0.089 | 1.00 |
| Arguing by age | -0.228 | -0.255 | 0.71 |

**Table 8.** Part 4 analyses: Regression coefficients CATSS-specific model.

| | CATSS male | CATSS female | % of sets significant |
|---|---|---|---|
| Proportion life events | 0.071 | 0.075 | 1.000 |
| Parenting | -0.048 | -0.054 | 0.914 |

The NTR-specific analyses revealed effects of bragging, impulsivity, and feelings of superiority (significant in all imputed sets, see Table 7). A small effect of age was significant in 97% of the imputed sets. Lack of guilt and the interaction arguing x age were significant in 74% and 71% of the imputed sets. Maternal alcohol consumption was only reaching statistical significance in 9% of the sets, thus not supporting a robust effect.

The CATSS specific analyses showed that the prevalence of serious life events was associated with an increased level of physical/overt aggression, whereas better parenting quality was associated with reduced aggression. These effects were significant in 100% and 91.4% of the data sets, respectively, and can therefore be considered as robust effects.

## DISCUSSION

The present study modeled the relationship between childhood aggression and a wide range of predictor variables, using a novel methodological approach, in 62,227 children from two different cohorts. The large sample allowed for splitting the data in independent parts for exploration, variable selection, assessment of model performance, and fitting an interpretable confirmatory model. Employment of different data mining techniques provided the opportunity to investigate a large number of predictors simultaneously without the need to a priori specify which types of effects (i.e., linear or nonlinear main effects, linear or nonlinear interaction effects) were present in the data.

The most important variables were non-physical aggression (arguing), and two ADHD indicators (being easily distracted, and hyperactive), which were significant in 100% of the imputed sets. We report results concerning regression coefficients as "% of imputed sets" in which an effect was significant because this provides a measure of robustness of an effect whereas significance itself is less informative in large samples. Other variables that were significant predictors of higher childhood aggression in over 90% of the imputed sets were maternal smoking during pregnancy, the child not living with both parents, lower age of mother at birth, lower educational qualification of the father, lying, stealing, and being bullied. In addition, daydreaming was a significant predictor in the Swedish data. The cohort specific analyses revealed for the Dutch children that in more than 90% of the imputed sets, aggression was significantly predicted by older age, bragging, feeling superior, and impulsivity. For the Swedish sample, analyses revealed that a higher proportion of life events and lower levels of parenting (i.e., monitoring) were significantly associated with childhood aggression in more than 90% of the imputed sets. Variables with above-threshold VIMs that did not consistently predict childhood aggression across imputed data sets included lower educational qualification of the mother (significant in 43% of imputed data sets), short attention (significant in 9% of imputed data sets), an interaction between arguing and being easily distracted (significant in 3% of imputed data sets), and specifically for the Dutch children maternal alcohol use during pregnancy (significant in 9% of imputed data sets), feeling no guilt (significant in 74% of imputed data sets), and an interaction between arguing and age (i.e., the association between age and aggression varies between children who do and do not argue and vice versa; significant in 71% of imputed data sets). Variables not selected based on the zero or close to zero VIMs included birth weight, gestational age, height, weight, asthma, eczema, medication use, having siblings, age father at birth, motor skills, not finishing things, fear of situations, anxiety, and nightmares.

With regards to the demographic variables, sex and cohort interacted with some of the variables (e.g., argues x sex, argues x cohort, hyperactive x sex), which led to multi-group models. Age only appeared to have an effect for the Dutch children, with higher aggression for older children. The group differences for boys and girls, and for the Netherlands and Sweden provide evidence for etiological differences between these groups.

Of the prenatal characteristics, maternal smoking during pregnancy significantly predicted childhood aggression in 94% of the imputed sets; maternal alcohol use during pregnancy was only significant in 9% of the imputed sets (only measured in Dutch children), and birth weight and gestational age were not selected based on their VIMs. Possibly, the influence of birth weight and gestational age on childhood aggression are attenuated by the environment in which children grow up (LaPrairie, Schechter, Robinson, & Brennan, 2011). The effect of .009 to .011 of maternal alcohol use during pregnancy was similar to the correlation of .008 found in a sample from the United Kingdom (Malanchini et al., 2018). Maternal smoking during pregnancy had an effect on childhood aggression ranging between .034 and .045, which was slightly smaller than the correlation of .085 found in a recent meta-analysis, with partly overlapping samples (Malanchini et al., 2018). Differences in effect sizes between previous research and the present study could be explained by the fact that we investigated all predictors simultaneously, which avoids overestimation of correlated effects.

None of the physical development variables were selected based on their importance (i.e., height, weight, asthma, and medication use). A meta-analysis on the association between asthma and externalizing behavior revealed an association of .29 (Pinquart & Shen, 2011). In addition, previous research found associations between height and weight during early childhood and later aggressive behaviors (i.e., $d$ = 0.25 – 0.30), but when controlling for other factors such as socioeconomic status, the associations disappeared (Raine, Reynolds, & Venables, 1998). Overall, it appears that the association between physical development and childhood aggression might be overruled when taking measures of the family environment and behavioral symptoms into account.

From the family environment variables (i.e., siblings, whether both parents live in the same household, age mother at birth, and age father at birth), only whether both parents lived in the same household (i.e., effects from -.040 to -.026), and age of mother at birth (i.e., effects from -.039 to -.029) were included in the confirmatory model. That having siblings was not a strong predictor for childhood aggression might be explained by the fact that all the children in the samples were twins, which may attenuate the impact of having other siblings. A possible explanation for higher aggression when parents do not live in the same household could be an increase in parental stress due to single parenthood, such as a lowered income (Briggs, Cox, Sharkey, Briggs, & Black, 2016). The positive effect of higher maternal age was in line with previous research (Tearne et al., 2015). It could be due to older mothers having better socioeconomic circumstances (Bornstein, Putnick, Suwalsky, & Gini, 2006), higher satisfaction with parenting, and more time spent with children (Ragozin, Basham, Crnic, Greenberg, & Robinson, 1982).

The parenting variable pertained to parental monitoring, an established predictor for childhood aggression (Racz & McMahon, 2011). In the present study, the regression coefficients (only measured in Swedish children) were -.048 and -.054, indicating higher aggression for children whose parents monitor them less. Previous research found an association of .31 between poor supervision and oppositional defiant disorder (ODD) and .39 between poor supervision and conduct disorder (CD; Burke, Pardini, & Loeber, 2008), which was stronger than the regression coefficients in the present study. Both disorders co-occur with aggressive behaviors, together

with symptoms that were important predictors in the present study, namely arguing for ODD, and stealing and lying for CD. Possibly, the regression coefficients for parenting were diminished due to inclusion of these variables in the confirmatory model.

Both maternal and paternal education were included in the confirmatory model. Paternal education was significant in 97% of the imputed sets, maternal education only in 43%. Both indicated lower levels of aggression for children when parents are more highly educated. The effects for paternal education ranged from -.039 to -.022, for maternal education they ranged from -.016 to -.010; these estimates were close to the correlation found of -.099 between aggression and socioeconomic status (which can be assessed through parental education level; Winkleby, Jatulis, Frank, & Fortmann, 1992) in a meta-analysis (Piotrowska et al., 2015)

A higher proportion of life events predicted higher aggression with regression coefficients of .071 and .075. Previous research found associations of .16 and .28 between life events and aggression (Guerra et al., 1995; Mcknight, Huebner, & Suldo, 2002). The discrepancy between the coefficients in the present study and previous research could be caused by heightened behavioral symptoms as a result of exposure to life events and thus absorbing the effects.

The importance of the mother-reported behavioral symptoms is in line with the high comorbidity between childhood aggression and other behavior problems (Bartels et al., 2018). The most important predictors were symptoms related to attention-deficit hyperactivity disorder (ADHD; i.e., hyperactivity, being easily distracted, and impulsivity), oppositional defiant disorder (ODD; i.e., arguing), and conduct disorder (CD; i.e., lying, stealing). The importance of arguing (i.e., regression coefficients between .266 and .385), lying (i.e., regression coefficients between .066 and .079), and stealing (i.e., regression coefficients between .049 and .077) confirms the overlap between childhood aggression and ODD and CD (American Pyschiatric Association, 1994), although the symptoms reflect distinguishable constructs in 9-year-old children (Lubke, McArtor, Boomsma, & Bartels, 2018). In addition, ADHD often co-occurs with aggressive behavior (e.g., Harvey, Breaux, & Lugo-Candelas, 2016; Rhee, Willcutt, Hartman, Pennington, & DeFries, 2008), which was supported by the regression coefficients in this study for being easily distracted (i.e., .118 - .157), hyperactivity (i.e., .082 - .133), impulsivity (i.e., .089 - .099), and daydreaming (i.e., 0.20 – 0.21 in the Dutch sample and 0.69 – 0.86 in the Swedish sample). This could partially be explained by a shared genetic liability; research found high genetic correlations (e.g., .46 - .74) between ADHD behaviors and forms of childhood aggression (Dick, Viken, Kaprio, Pulkkinen, & Rose, 2005; Kuja-Halkola, Lichtenstein, D'Onofrio, & Larsson, 2015). Previous research finding no association between short attention and aggression, was confirmed by the fact that, in the present study, short attention was only significant in 9% of the imputed sets (i.e., regression coefficients of .006-.007; e.g., Nagin & Tremblay, 2001).



Feeling no guilt, bragging, and feeling superior are all related to psychopathic traits including callous/unemotional (CU) traits and narcissism (Salekin, 2017). In the present study, these were all significant predictors for childhood aggression (i.e., respectively, regression coefficients of -.035, .127, and -.070 for boys and -.044, .067, and -.082 for girls). Taking the direction of item wording into account, this indicates higher aggression when children feel no guilt, brag, and feel superior, in line with previous research (Kerig & Stellwagen, 2010; Svensson et al., 2018). Being bullied had an effect between .028 and .056, with higher aggression for children being bullied. These effects were slightly smaller than a correlation of .14 found in a meta-analysis (Reijntjes et al., 2011). Motor skills were not selected based on VIMs, which was in line with previous research on behavior problems and motor skills mainly finding effects for ADHD, and thus not aggression (Emck, Bosscher, Beek, & Doreleijers, 2009). Internalizing symptoms (i.e., fear of situations, anxiety, nightmares) were not selected based on their VIMs. This could be explained by the lower comorbidity between aggression and internalizing disorders compared to the comorbidity between aggression and externalizing disorders (Bartels et al., 2018), indicating that internalizing symptoms are less important in predicting childhood aggression compared to externalizing symptoms.

Taking many variables into account simultaneously could explain that most of the regression coefficients in the present study were smaller than reported in previous research. While the findings are correlational, and should thus not be interpreted as causal relations, they do provide direction for variables that are valuable to examine through longitudinal research. The prediction effect of externalizing behavior symptoms remained taking all other selected variables into account. Practically, this implies that, although environmental variables may be important for the development of childhood aggression, paying attention to behaviors of the child such as arguing, being easily distracted, and hyperactivity will yield a better prediction of childhood aggression.

Heterotypic continuity causes childhood aggression to express differently across the life span (e.g., Hannigan, Walaker, Waszczuk, McAdams, & Eley, 2017; Lubke, McArtor, Boomsma, & Bartels, 2018), which hampers effective diagnosis and treatment referral. The high co-occurrence of childhood aggression and other symptoms suggests that having elevated levels for one disorder will also imply elevated levels for other disorders (Bartels et al., 2018). Screening for behavioral symptoms that underlie multiple mental disorders may target children that will likely develop childhood aggression but could also develop another disorder, such as ADHD. This may aid early detection of a liability to develop psychopathology and implementation of treatment before children develop a full-blown disorder. Moreover, treating one disorder could also positively affect levels for other disorders. For example, research found improved levels of aggression as a result of treatment

for ADHD (Chan, Fogler, & Hammerness, 2016). Our results, together with previous research, suggest merit in monitoring children's behavioral symptoms, because they might predict later aggression or any other disorder that often co-occurs with childhood aggression.

Although we included a wide range of predictor variables, some known risk factors for childhood aggression were not collected by the included cohorts or not available for the present study, imposing a limitation on the comprehensiveness of the prediction model. Examples are exposure to domestic violence (Evans, Davies, & DiLillo, 2008) and parental psychopathology (Connell & Goodman, 2002; Goodman et al., 2011). Nevertheless, these unavailable variables apply to more extreme cases, whereas the predictors in the present study would apply to the general population. Moreover, it may be easier to observe the variables included in the present study than to obtain information on sensitive topics. The predictors that were most important in the present study comprise salient behaviors that could be noticed by, for instance, parents or teachers, making the findings feasible to apply in common practices.

Throughout the study, we applied rigorous and novel methodological approaches. First of all, partitioning the data into four independent parts provided us with the possibility to fit different models with different types of predictor effects (i.e., linear, nonlinear, interaction) without having to pre-specify them. Moreover, partitioning the data allowed for testing the models in independent sets of data, thereby preventing the risk of overfitting. Second, analyses were able to detect measurement non-invariance variance related to differences between cohorts. For example, the behavioral item on daydreaming did not come up in the EDA as different between cohorts, but the data mining analyses followed by confirmatory modeling revealed that there was in fact measurement non-invariance. Therefore, we are confident that with these rigorous methods and the data available to us, we have obtained the most robust prediction model for childhood aggression.

Childhood aggression is a very heterogeneous disorder (Bolhuis et al., 2017). This may explain that research so far is inconclusive on the etiology of childhood aggression and that treatment effectiveness for childhood aggression is still limited (Hendriks et al., 2018; Weisz et al., 2017). Therefore, it is important to clearly define the type of childhood aggression under scrutiny (Hofvander, Ossowski, Lundström, & Anckarsäter, 2009). In the present study, we accounted for this by specifically examining physical and overt aggression. For future research, it would be interesting to study whether our findings also apply to other types of childhood aggression.

Finally, childhood aggression is a developmental disorder with a strong genetic component (e.g., 32% - 83%; Hudziak et al., 2003; Porsch et al., 2016; Rhee & Waldman, 2002). Moreover, many of the predictor variables included in the present study are partly explained by genetic factors such as parenting and being bullied in secondary school (Veldkamp et al., submitted; Vinkhuyzen, Van Der Sluis, De Geus, Boomsma, & Posthuma, 2010), and thus possibly even overlap in genetic liability, which may have led to biased estimates. One way to take this into account would be to correct for genetic information, for instance by including a polygenic risk score (e.g., Wray, Goddard, & Visscher, 2007) for childhood aggression or strongly related variables in the prediction model. The currently available GWAS for childhood aggression (Pappa et al., 2016) and antisocial behavior (Tielbeek et al., 2017) were not sufficiently powered to use for a polygenic risk score. Nevertheless, this will likely become possible in the near future, which then will provide opportunities to obtain a clearer understanding of the relationship between childhood aggression and predictor variables.

In conclusion, the presented research is the first large-scale study that included a large number of potential predictors for childhood aggression. The large number of variables allows one to assess the presence of all possible main and interaction effects simultaneously. Effects were detected using deductive data mining, and were tested using a confirmatory model fitted to a holdout partition of the data. Investigating multiple predictors simultaneously results in more unbiased effect sizes compared to one-at-a-time analyses, and form a more reliable basis for future research into the prediction of childhood aggression. The most important predictors were salient behaviors such as arguing, being easily distracted, and hyperactivity. Recommendations for future research include testing the found relations in longitudinal data to establish direction of causality and adding genetic information to control for genetic overlap between variables in the prediction model. Altogether, the present study applied rigorous methods on a wide range of predictor variables and yielded a set of variables that may facilitate early detection and prevention of childhood aggression.

**Supplementary Table 1.** Overt/Physical Aggression Items in ACTION

| Item Code | Item |
|---|---|
| A-TAC63 | Has there ever been a time when he/she would be angry to the extent that he/she cannot be reached? |
| A-TAC65* | Does he/she often tease others by deliberately doing things that are perceived as provocative? |
| A-TAC70 | Has he/she ever been deliberately been physical cruel to anybody? |
| A-TAC71 | Does he/she often get into fights? |
| CBCL016 | Cruelty, bullying or meanness to others |
| CBCL020 | Destroys his/her own things |
| CBCL021 | Destroys things belonging to his/her family or others |
| CBCL023 | Disobedient at school |
| CBCL037 | Gets in many fights |
| CBCL057 | Physically attacks people |
| CBCL094 | Teases a lot |
| CBCL095 | Temper tantrums or hot temper |
| SDQ05 | Often has temper tantrums or hot tempers |
| SDQ07 | Generally obedient, usually does what adults request |
| SDQ12 | Often fights with other children or bullies them |
| MPNI13 | Teases other kids or attacks them for no reason at all |
| MPNI21 | Hurts other kids when angry, e.g. by hitting, kicking, or throwing things at them |
| MPNI25 | Bullies smaller and weaker kids |
| MPNI27 | Calls people names when angry at them |
| MPNI33 | Is disobedient at school/home |

**Supplementary Table 2.** Variables per category and their description

| Variable category | Variable coding | CATSS | NTR |
|---|---|---|---|
| Outcome variable | Harmonized aggression factor score | x | x |
| Demographics | Sex | x | x |
| | Age | x | x |
| | Cohort | x | x |
| Prenatal characteristics | Birth weight, standardized separately per cohort | x | x |
| | Gestational age, in weeks | x | x |
| | Maternal smoking during pregnancy; yes/no | x | x |
| | Maternal alcohol use during pregnancy; yes/no | - | x |
| Physical development | Height, in cm, parent report. This variable was corrected for age differences. | x | x |
| | Weight, in kilos, parent report. This variable was corrected for age differences. | x | x |
| | Asthma, parent report; yes /no | Have or had … Asthma | Asthma, chronic bronchitis or CARA …, |
| | Eczema, parent report ; yes/no | Have or had … Eczema | Serious skin disease or eczema |
| | Medicine use, parent report; yes/no | Does s/he use prescribed medicine? | Does the child currently use prescription medication? |
| Family environment | Siblings, other than the co-twin, parent report; yes/no | - | x |
| | Whether both parents live in the same household, parent report; yes/no | x | x |
| | Age mother at birth | x | x |
| | Age father at birth | x | x |

**Supplementary Table 2.** Continued

| | | | |
|---|---|---|---|
| Parenting | Parenting, reported by parents. Mean score of >66% of the items, ranging from 1-4. A higher score indicates more parental monitoring | Do you know what your child does during his/her free time?<br>Do you usually know what kind of homework your child has?<br>Does your child keep a lot of secrets from you about what s/he does during her/his free time? RECODED<br>Does your child need to have your permission to stay out late on weekday evening?<br>Do you always require that your child tells you where s/he is at night, who s/he is with and what they do together?<br>How often do you initiate a conversation with your child about things that happened during normal day at school?<br>Do you usually ask your child to talk about things that happened during her/his free time (whom s/he met when s/he was out in the city, free time activities, etc.)?<br>Does your child usually want to tell how school was when s/he gets home (how s/he did on different exams, her/his relationship with teachers, etc.)?<br>In original data coded 1, 2, 3, 4, 5, for data aligning 3 and 4 combined. | I always know where and with whom my child is if he - she is not at home<br>I know the interest of the child and what kind of hobbies he -she likes<br>I know the daily activities of the child<br>I take into account to the view of the child |
| Parental education level | Maternal education level, 1-4: 1 = elementary school; 2 = high school, but not finished; 3 = finished high school; 4 = university. | x | x |
| | Paternal education level, 1-4: 1 = elementary school; 2 = high school, but not finished; 3 = finished high school; 4 = university. | x | x |

**Supplementary Table 2.** Continued

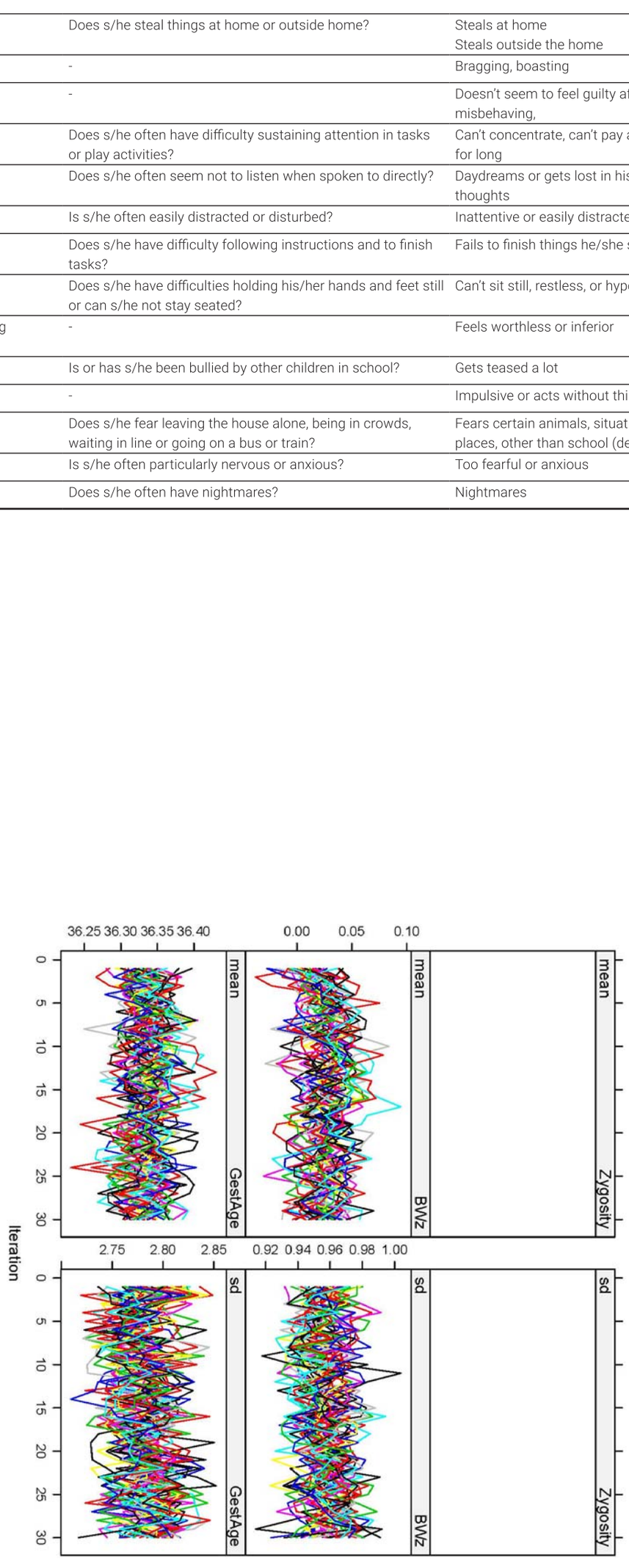
| | | | |
|---|---|---|---|
| Life events | Proportion of life events, calculated for all observations with more than 66% of the items filled in. Parent report for CATSS, self-report for NTR. | Parent-reported.<br>Has the child ever been in a serious car accident where he/she got modest to severe physical injuries or needed medical care?<br>Has the child ever been in any other serious accident where he/she got injured or hospitalized?<br>Has the child ever been emotionally abused or neglected? For example, being frequently shamed, embarrassed, ignored, or repeatedly told that he/she were “no good”<br>Has the child ever been physically neglected? For example, not fed, not properly clothed, or left to take care of him/herself<br>Has the child ever been physically abused – for example, hit, choked, burned, or beaten or severely punished by someone he/she knew well?<br>Was the child ever touched or made to touch someone else in a sexual way, because he/she felt forced in some way or threatened by harm to him/herself or someone else ?<br>Has the child ever had sex because he/she felt forced in some way or threatened by harm to him/herself or someone else? With sex, we mean orally, anally, and/or genitally ‘.<br>Has the child ever observed physical violence between family members? For example, hitting, kicking or punching<br>Has the child ever witnessed a threatening or violent criminal incident? (not film, internet or TV) e.g. someone was attacked, seriously hit, robbed or stabbed<br>Has the child ever been a direct witness (not film, Internet or TV) to any other serious incident that you have not mentioned?<br>Is there any other cruel or terrifying event the child has been exposed to? | Self-reported.<br>You moved to another neighborhood or city<br>A good friend moved house<br>You changed schools (not from elementary to high school)<br>You were seriously ill or had a serious accident<br>Someone close to you was or is seriously ill<br>Someone close to you died<br>Your parents have serious conflicts - fights<br>Your mother or father left home or your parents got divorced<br>Your mother’s or father’s new partner came to live with you<br>Your brother or sister left home<br>Your mother or father became unemployed<br>Your mother or father started working again after a long period at home<br>A little brother or sister was born or adopted |
| Behavioral symptoms mother report | Motor skills; 0-1 | Does s/he have problems coordinating movements smoothly? | Poorly coordinated or clumsy |
| | Arguing; 0-1 | Does s/he often argue with adults? | Argues a lot |
| | Lying; 0-1 | Does s/he often lie or cheat? | Lying or cheating |

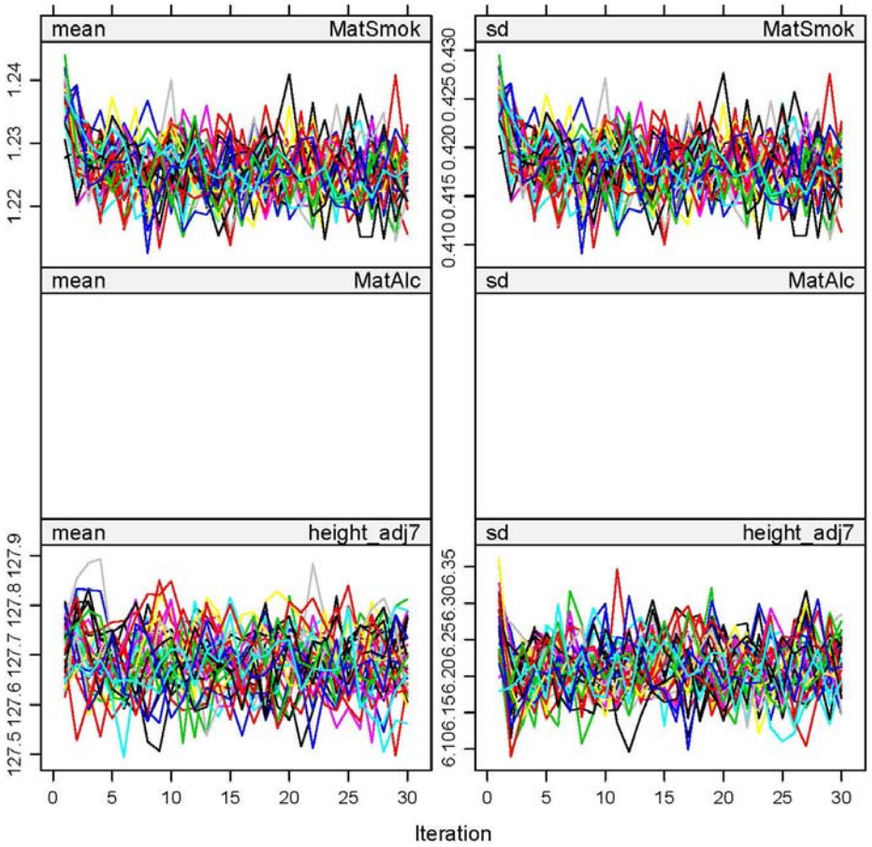
**Supplementary Table 2.** Continued

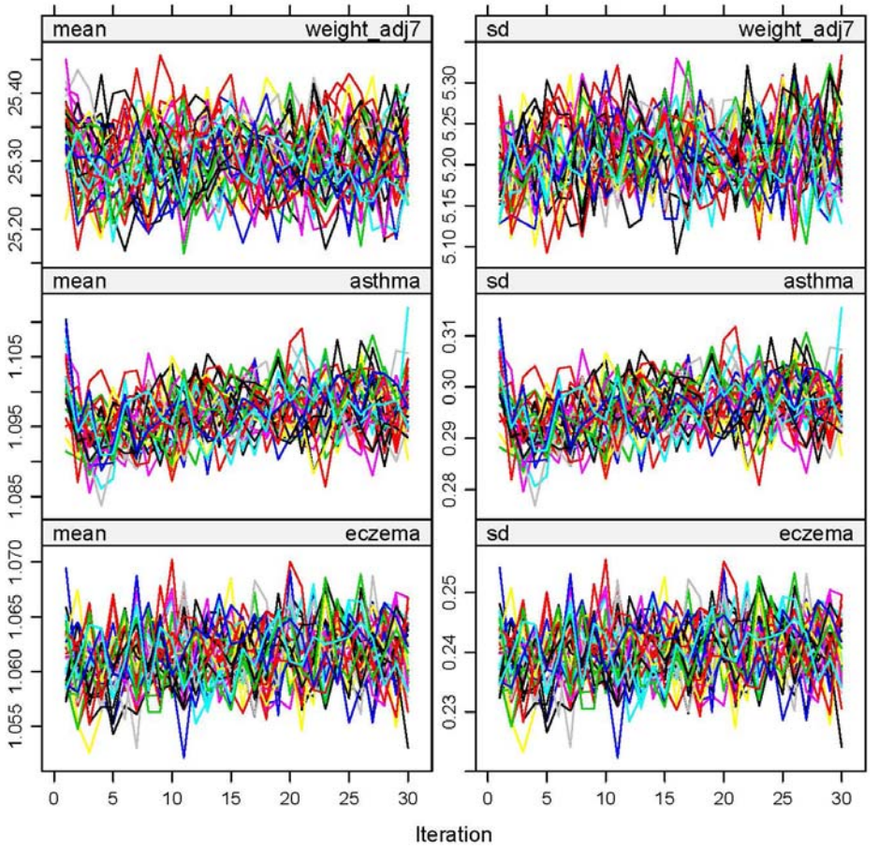
| | | | |
|---|---|---|---|
| | Stealing; 0-1 | Does s/he steal things at home or outside home? | Steals at home<br>Steals outside the home |
| | Bragging; 0-1 | - | Bragging, boasting |
| | No guilt; 0-1, 0 indicates less guilt | - | Doesn't seem to feel guilty after misbehaving, |
| | Short attention; 0-1 | Does s/he often have difficulty sustaining attention in tasks or play activities? | Can't concentrate, can't pay attention for long |
| | Daydreaming; 0-1 | Does s/he often seem not to listen when spoken to directly? | Daydreams or gets lost in his/her thoughts |
| | Easily distracted; 0-1 | Is s/he often easily distracted or disturbed? | Inattentive or easily distracted |
| | Not finishing things; 0-1 | Does s/he have difficulty following instructions and to finish tasks? | Fails to finish things he/she starts |
| | Hyperactivity; 0-1 | Does s/he have difficulties holding his/her hands and feet still or can s/he not stay seated? | Can't sit still, restless, or hyperactive |
| | Feeling superior; 0-1, 0 indicates feeling superior | - | Feels worthless or inferior |
| | Being bullied; 0-1 | Is or has s/he been bullied by other children in school? | Gets teased a lot |
| | Impulsivity; 0-1 | - | Impulsive or acts without thinking |
| | Fear of situations; 0-1 | Does s/he fear leaving the house alone, being in crowds, waiting in line or going on a bus or train? | Fears certain animals, situations, or places, other than school (describe): |
| | Anxiety; 0-1 | Is s/he often particularly nervous or anxious? | Too fearful or anxious |
| | Nightmares; 0-1 | Does s/he often have nightmares? | Nightmares |

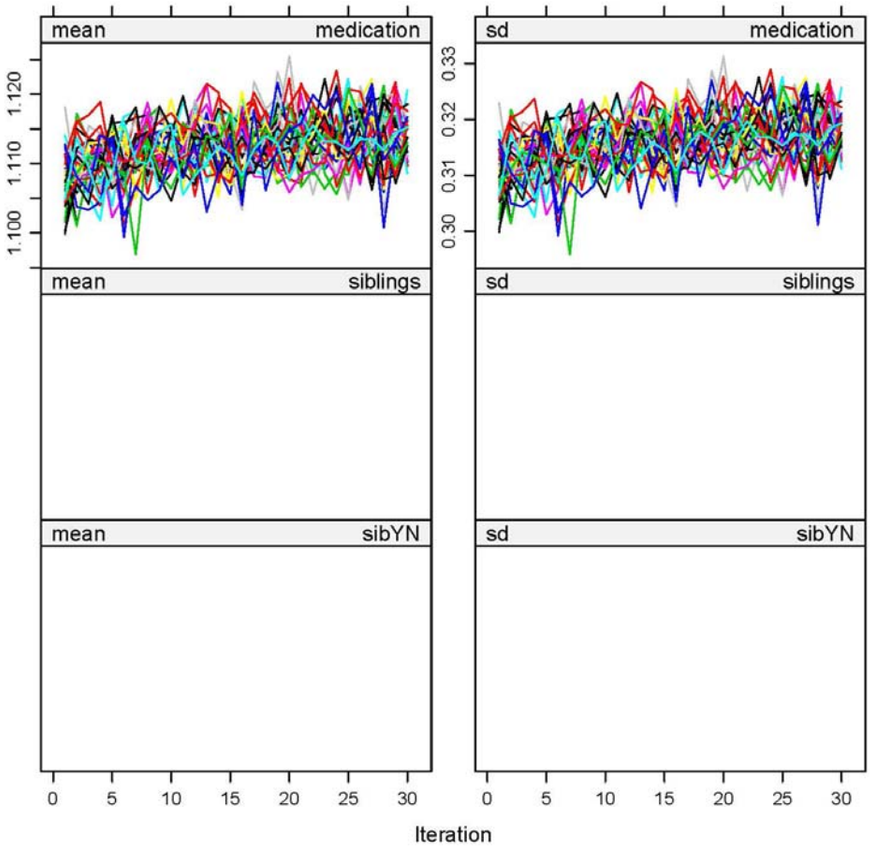


**Supplementary Figure 1.** Imputation convergence plots

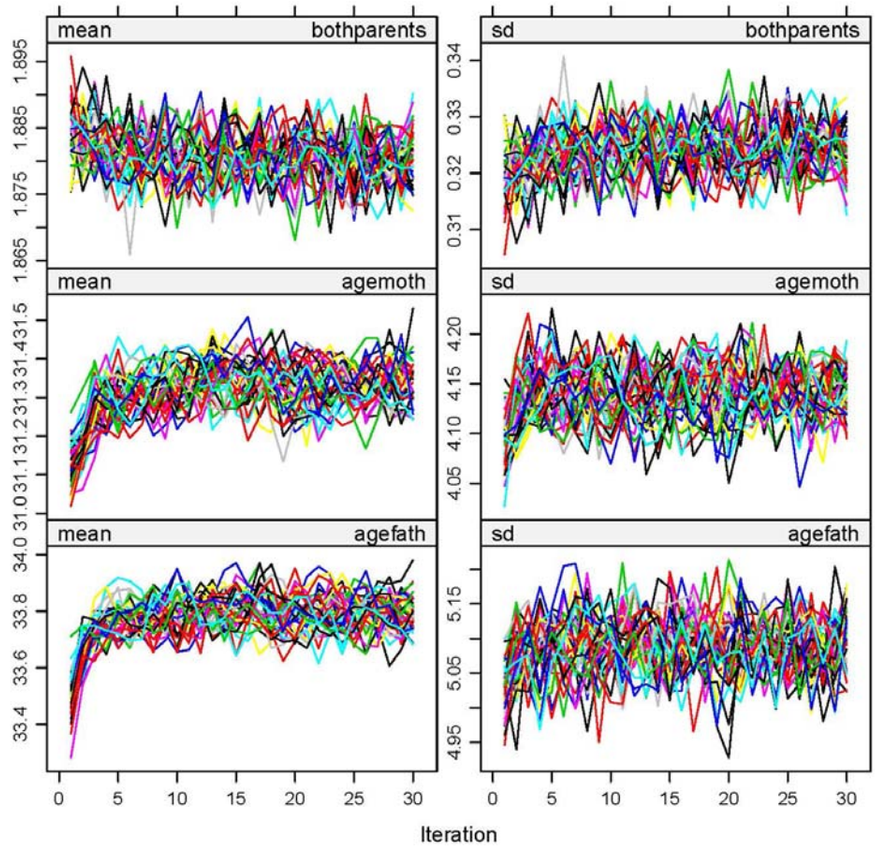


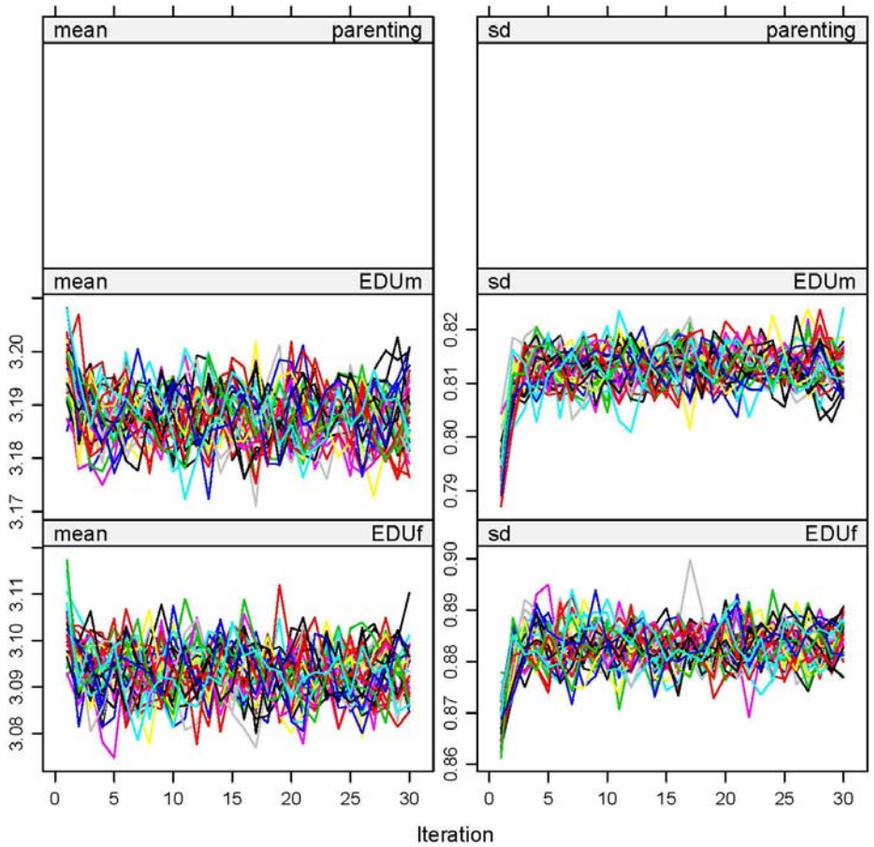

mean
MatSmok
sd
MatSmok
mean
MatAlc
sd
MatAlc
mean
height_adj7
sd
height_adj7
1.22
1.23
1.24
0.410
0.415
0.420
0.425
0.430
127.5
127.6
127.7
127.8
127.9
6.10
6.15
6.20
6.25
6.30
6.35
0
5
10
15
20
25
30
Iteration

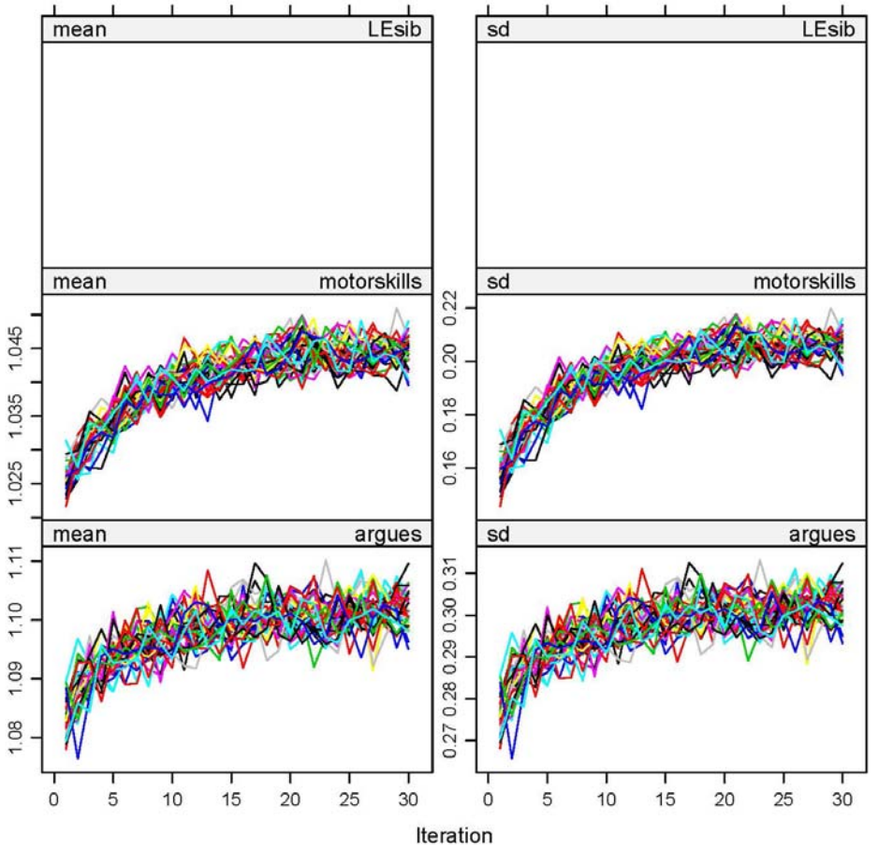

mean
weight_adj7
sd
weight_adj7
mean
asthma
sd
asthma
mean
eczema
sd
eczema
25.20
25.30
25.40
5.10
5.15
5.20
5.25
5.30
1.055
1.060
1.065
1.070
1.085
1.095
1.105
0.23
0.24
0.25
0.28
0.29
0.30
0.31
0
5
10
15
20
25
30
Iteration

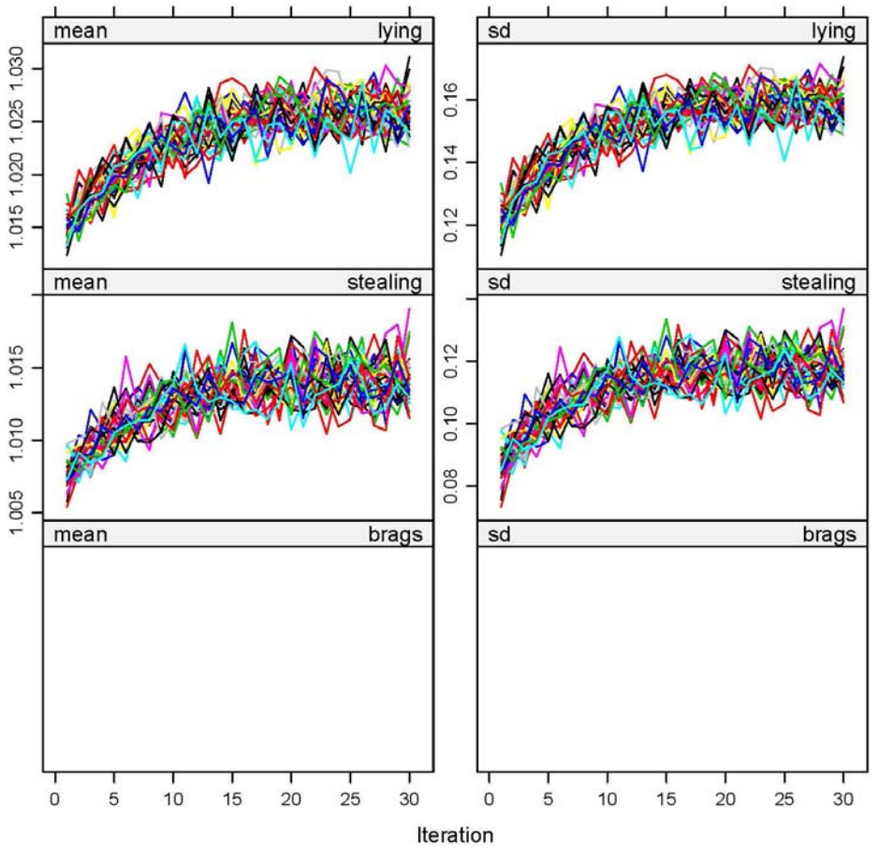

mean
medication
sd
medication
mean
siblings
sd
siblings
mean
sibYN
sd
sibYN
1.100
1.110
1.120
0.30
0.31
0.32
0.33
0
5
10
15
20
25
30
Iteration

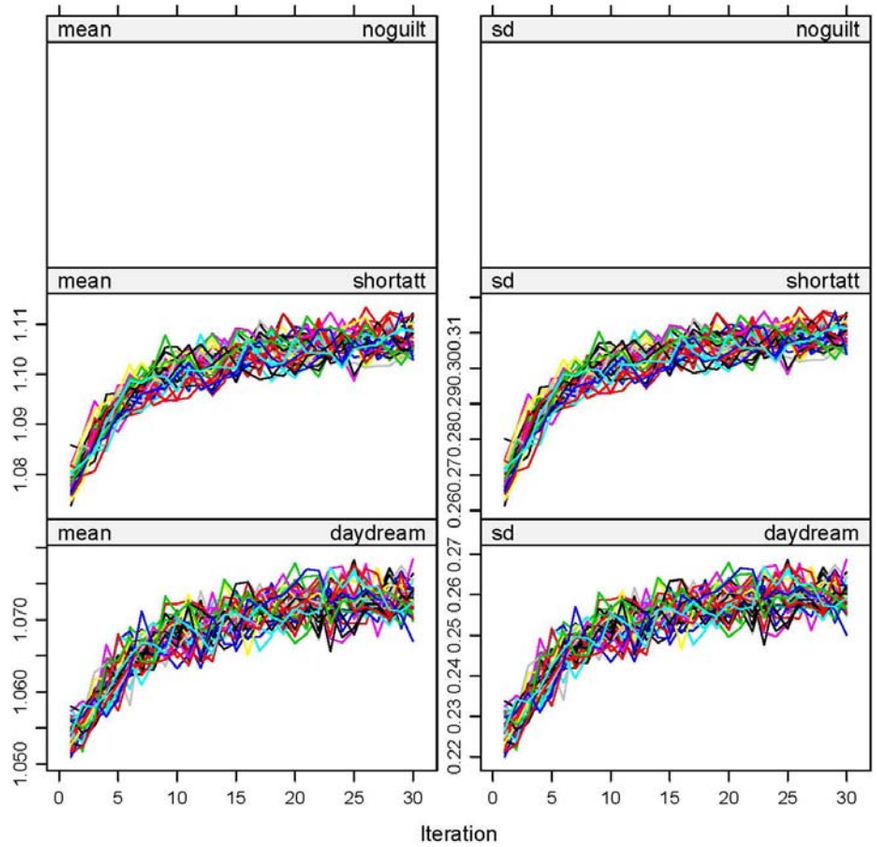


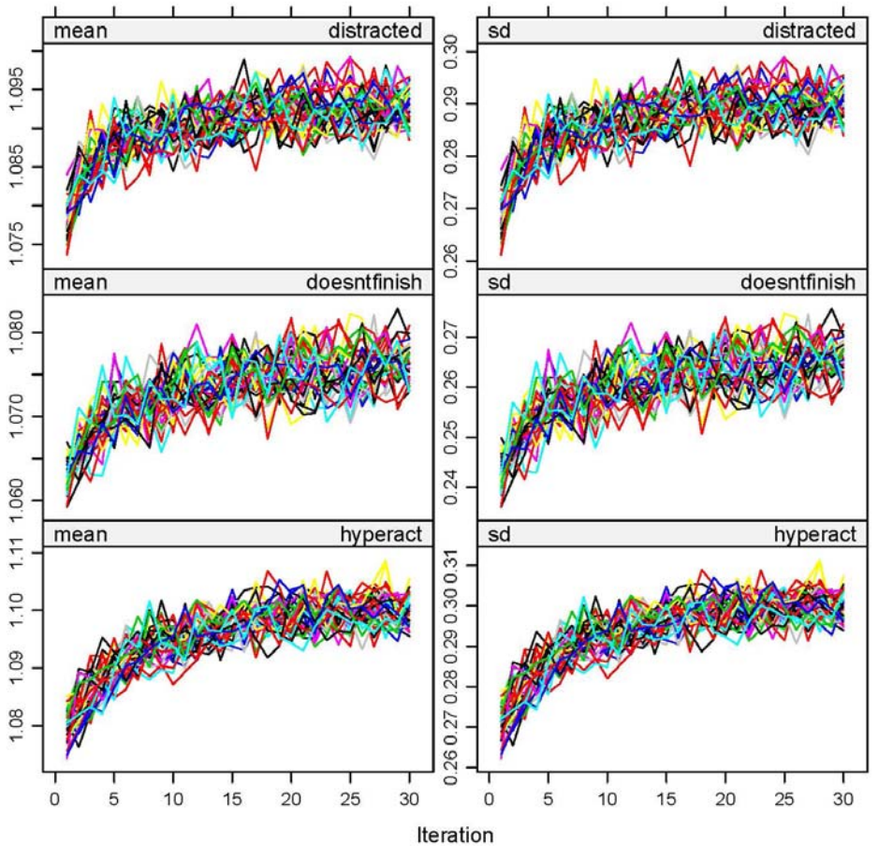

mean
bothparents
sd
bothparents
mean
agemoth
sd
agemoth
mean
agefath
sd
agefath
1.865
1.875
1.885
1.895
0.31
0.32
0.33
0.34
31.0
31.1
31.2
31.3
31.4
31.5
4.05
4.10
4.15
4.20
33.4
33.6
33.8
34.0
4.95
5.05
5.15
0
5
10
15
20
25
30
Iteration

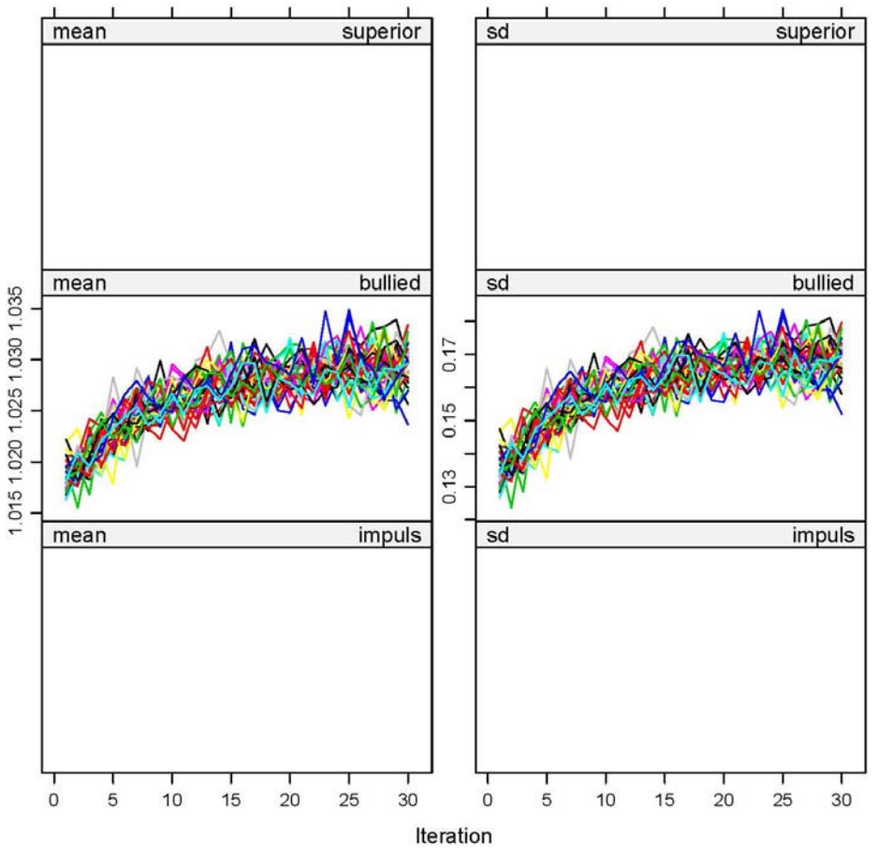

mean
parenting
sd
parenting
mean
EDUm
sd
EDUm
mean
EDUf
sd
EDUf
3.20
3.19
3.18
3.17
3.11
3.10
3.09
3.08
0.82
0.81
0.80
0.79
0.90
0.89
0.88
0.87
0.86
0
5
10
15
20
25
30
Iteration

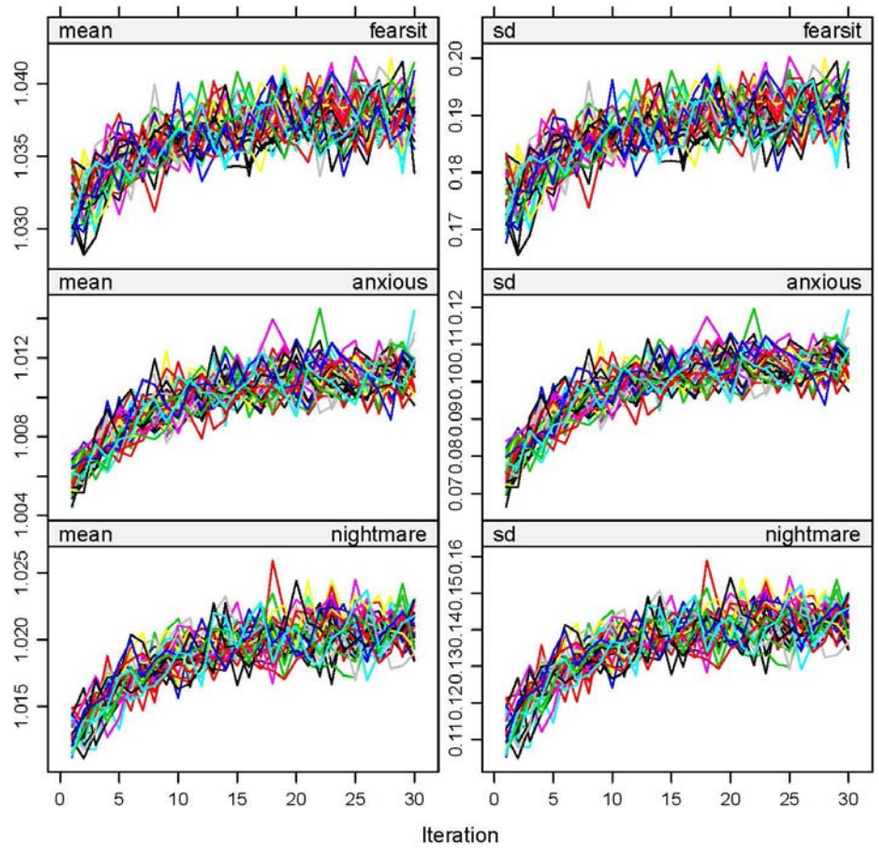

mean
LEsib
sd
LEsib
mean
motorskills
sd
motorskills
mean
argues
sd
argues
1.045
1.035
1.025
1.11
1.10
1.09
1.08
0.22
0.20
0.18
0.16
0.31
0.30
0.29
0.28
0.27
0
5
10
15
20
25
30
Iteration


mean
lying
sd
lying
mean
stealing
sd
stealing
mean
brags
sd
brags
1.005 1.010 1.015
1.015 1.020 1.025 1.030
0.08 0.10 0.12
0.12 0.14 0.16
0 5 10 15 20 25 30
Iteration


mean
noguilt
sd
noguilt
mean
shortatt
sd
shortatt
mean
daydream
sd
daydream
1.08 1.09 1.10 1.11
1.050 1.060 1.070
0.22 0.23 0.24 0.25 0.26 0.27
0.260.270.280.290.300.31
0 5 10 15 20 25 30
Iteration


mean
distracted
sd
distracted
mean
doesntfinish
sd
doesntfinish
mean
hyperact
sd
hyperact
1.075
1.085
1.095
1.060
1.070
1.080
1.08
1.09
1.10
1.11
0.26
0.27
0.28
0.29
0.30
0.24
0.25
0.26
0.27
0.26
0.27
0.28
0.29
0.30
0.31
0
5
10
15
20
25
30
Iteration


mean
superior
sd
superior
mean
bullied
sd
bullied
mean
impuls
sd
impuls
1.015
1.020
1.025
1.030
1.035
0.13
0.15
0.17
0
5
10
15
20
25
30
Iteration


mean
fearsit
sd
fearsit
mean
anxious
sd
anxious
mean
nightmare
sd
nightmare
1.040
1.035
1.030
0.20
0.19
0.18
0.17
1.012
1.008
1.004
0.12
0.11
0.10
0.09
0.08
0.07
1.025
1.020
1.015
0.16
0.15
0.14
0.13
0.12
0.11
0
5
10
15
20
25
30
Iteration

Chapter 5

# Comparing the Genetic Architecture of Childhood Behavioral Problems Across Socioeconomic Strata in the Netherlands and the United Kingdom.

Accepted as: Hendriks, A. M., Finkenauer, C., Nivard, M. G., Van Beijsterveldt, C. E. M., Boomsma, D. I., Plomin, R., & Bartels, M. (2019). Comparing the genetic architecture of childhood aggression across socioeconomic strata in the Netherlands and the United Kingdom. *European Child & Adolescent Psychiatry.*

## ABSTRACT

Socioeconomic status (SES) affects the development of childhood behavioral problems. It has been frequently observed that children from low SES background tend to show more behavioral problems. There also is some evidence that SES has a moderating effect on the causes of individual differences in childhood behavioral problems, with lower heritability estimates and a stronger contribution of environmental factors in low SES groups. The aim of the present study was to examine whether the genetic architecture of childhood behavioral problems suggests the presence of protective and/or harmful effects across socioeconomic strata, in two countries with different levels of socioeconomic disparity; the Netherlands and the United Kingdom. We analyzed data from 7-year-old twins from the Netherlands Twin Register (*N* = 24,112 twins) and the Twins Early Development Study (*N* = 19,644 twins). The results revealed a nonlinear moderation effect of SES on the contribution of genetic and environmental factors to individual differences in childhood behavioral problems. The heritability was higher, the contribution of the shared environment was lower, and the contribution of the nonshared environment was higher, for children from high SES families, compared to children from low or medium SES families. The pattern was similar for Dutch and UK families. We discuss the importance of these findings for prevention and intervention goals.



**Keywords:**
childhood behavioral problems, socioeconomic status, Netherlands, UK




Childhood behavioral problems comprise of aggressive and non-aggressive behaviors such as fighting, lying, stealing, or disobedience (Achenbach & Rescorla, 2001; Goodman, 2001). Childhood behavioral problems co-occur with many other symptoms of childhood psychopathology such as hyperactivity, inattention, and anxiety (Bartels et al., 2018). Moreover, they are considered a potential marker for psychopathology in later life (Althoff, Verhulst, Rettew, Hudziak, & Van Der Ende, 2010; Copeland et al., 2013). Childhood behavioral problems are, in addition, associated with substantial costs for families and society (Meltzer, Ford, Goodman, & Vostanis, 2011; Romeo, Knapp, & Scott, 2006). Given its profound effects for healthy development, it is important to gain more insight into risk factors for childhood behavioral problems.

One factor that is consistently found to be associated with childhood behavioral problems is socioeconomic status (SES), with children from low SES families displaying more behavioral problems than children from high SES families (Letourneau, Duffett-Leger, Levac, Watson, & Young-Morris, 2013; Piotrowska, Stride, Croft, & Rowe, 2015). SES refers to indicators of people's standing in the stratification system and is usually measured by education, occupation, employment, income, and/or wealth. SES determines a families' access to social, material, or cultural resources, for example as a result of the parents' educational qualifications (Phelan, Link, & Tehranifar, 2010; Winkleby, Jatulis, Frank, & Fortmann, 1992).

Despite evidence for an overall association between childhood behavioral problems and SES, an underexplored question in the field is whether differences in childhood behavioral problems emerge due to different underlying genetic and environmental processes for children from low SES families compared to children from high SES families. A powerful approach to gain more insight in the etiology of childhood behavioral problems is through twin studies. Twin studies have the ability to discern the contribution of genetic factors, shared environmental factors, and nonshared environmental factors to differences in childhood behavioral problems (Boomsma, Busjahn, & Peltonen, 2002). So far, twin studies have demonstrated that genetic differences between children explain about 52 to 78% of the differences in parent-reported behavioral problems in children. Shared environmental factors account for about 1% to 36% of the differences in childhood behavioral problems (Burt, 2009; Fedko et al., 2016; Hudziak et al., 2003; Ligthart, Bartels, Hoekstra, Hudziak, & Boomsma, 2005; Porsch et al., 2016; Van Beijsterveldt, Bartels, Hudziak, & Boomsma, 2003; Van der Valk, Van den Oord, Verhulst, & Boomsma, 2003). In addition, twin studies have shown that the contribution of genetic and environmental factors to individual differences in childhood behavioral problems varies across SES strata, indicating a moderator effect of SES on the genetic architecture of childhood behavioral problems. For example, research in children from the Netherlands (Middeldorp et al., 2014) and research on adolescents from Sweden (Tuvblad,

Grann, & Lichtenstein, 2006) report lower heritability, higher influence of the shared environment, and lower influence of the nonshared environment on behavioral problems in children from lower SES families compared to children from higher SES families. These findings suggest that the influence of shared environmental factors is amplified in disadvantageous environments, such as low SES, or reduced by advantageous environments.

The literature typically assumes that the effect of SES on both means and variances is linear, such that children with lower SES show more behavioral problems than children with higher SES, or that the heritability is lower in low SES groups than in high SES groups respectively. Alternatively, there may be separate protective and harmful effects, which can be examined using a trichotomization approach (Brumley & Jaffee, 2016). This approach compares the mid-range of a variable's distribution, in our case SES, with the upper and lower ends. An effect limited to low SES would be distinguishable if, for example, only children with low SES were to show more behavioral problems than the children in the mid-range and high SES. Put differently, children with mid-range and high SES would not differ in behavioral problems, but would show less behavioral problems than those with low SES, indicating that low SES is a risk factor for childhood behavioral problems. Furthermore, trichotomization of variance components could distinguish shifts in heritability (or a common environment) due to protective effects from shifts in heritability due to harmful effects. Gaining knowledge about the SES conditions that are promotive versus risky for the development of childhood behavioral problems is critical to our understanding of factors associated with childhood behavioral problems and can be used to tailor interventions to fit the needs of children from different social strata.

In order to get a better hold of the possible effects of SES on levels and variation of behavioral problems in middle childhood we, in the current study, compare two countries with different levels of socioeconomic disparity; the Netherlands and the United Kingdom (The World Bank Group, 2017). While, the contribution of genetic and environmental factors to individual differences in childhood behavioral problems appears to be similar in the Netherlands and the United Kingdom (Porsch et al., 2016), comparability of effects of SES on the variance decomposition is unknown. Because of the larger difference between low and high SES in the United Kingdom compared to the Netherlands (The World Bank Group, 2017), we expect larger differences in the etiology of childhood behavioral problems (e.g., the contribution of genetic and environmental factors to individual differences in childhood behavioral problems) across SES strata in the United Kingdom than in the Netherlands.

The present study aims to (1) investigate linearity of the moderating effect of SES on the genetic architecture of childhood behavioral problems and (2) investigate whether the moderating effect of SES differs between the Netherlands and the United Kingdom. To this end, we analyzed twin data from two large longitudinal prospective twin cohorts in the two countries. To allow for nonlinear effects of SES, we categorized SES into three strata.



# METHOD

## Participants and Measures

**Data from the Netherlands.** The Netherlands twin register (NTR) is a nation-wide population-based register founded in 1987 in the Netherlands (Van Beijsterveldt et al., 2013). For the present study, we included mother-reported data for 7-year-old twins (*N* = 24,826) born between 1986 and 2006. We excluded twin pairs in which one or both twins had a disease or disability that interfered with daily functioning (*N* = 714). The final sample consisted of 12,056 twin pairs (*N* = 24,112 twins, *M age* = 7.45 years, *SD* = 0.40, 49.7% males). Socioeconomic status (SES) was based on parental level of education. Based on the highest educational qualification of either the mother or the father assessed at age seven, we categorized children's SES as low, medium, or high.

Behavioral problems were assessed using maternal ratings on the Aggressive Behavior syndrome subscale of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001). This scale consisted of 18 items that assessed aggressive and non-aggressive behaviors such as "Disobedient at home", "Gets in many fights", and "Sulks a lot". Mothers were asked to report on their child's behavior in the past six months. Response categories were: 0 = "*Not true*", 1 = "*Sometimes or somewhat true*", and 2 = "*Very true or often true*". If more than three items were missing, participants were not included; otherwise the mean score was imputed for missing items.

**Data from the United Kingdom.** The Twins early development study (TEDS) is a twin register that longitudinally follows the development of twins born between 1994 and 1996 in England and Wales (from here on referred to as the United Kingdom; Haworth, Davis, & Plomin, 2013). For the present study, we included parent-reported data for 7-year-old twins (*N* = 20,685). We excluded 515 twin pairs in which one or both twins had a disease or disability that interfered with daily functioning. The final sample consisted of 9,822 twin pairs (*N* = 19,644 twins, *M age* = 7.07, *SD* = 0,25, 48,7% males). Like in the NTR, SES was based on parental level of education for the TEDS sample. Based on the highest educational qualification of either the mother or the father, assessed at first contact, we categorized children's SES as low, medium, or high. Although the educational system differs between the Netherlands and the United Kingdom, we established comparable categories as displayed in Table 1.

**Table 1.** SES categories

| SES | The Netherlands | The United Kingdom |
|---|---|---|
| Low | Elementary school | No qualifications |
| | Few years of more extensive primary education (mulo) | CSE grade 2-5 or 0-level/GCSE grade |
| | Graduated mulo or general secondary education (mavo) | D-G |
| | Few years of lower technical education (lts) | |
| | Few years higher general secondary education (havo)/ pre-university education (vwo) | |
| Medium | Graduated havo/vwo | CSE grade 1 or 0-level/GCSE grade |
| | Few years intermediate vocational education (mbo) | A-C |
| | Graduated mbo | A-level or S-level |
| | Few years of higher vocational education (hbo) or university | Higher National Certificates |
| High | Graduated hbo | Higher National Diplomas |
| | Graduated university | Undergraduate |
| | Postgraduate | Postgraduate qualification |

Behavioral problems were assessed using parental (97.3% maternal ratings) ratings on the Conduct Problem subscale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). This scale consisted of five items that assessed aggressive and non-aggressive behaviors such as "Often fights with other children or bullies them", "Generally obedient, usually does what parents request", and "Often has temper tantrums or hot tempers". Parents were asked to report on their child's behavior. Response categories were: 0 = "*Not true*", 1 = "*Somewhat true*", and 2 = "*Certainly true*". If more than two items were missing, participants were not included; otherwise the mean score was imputed for missing items.

## Statistical Analysis

To gain insight in the distribution of childhood behavioral problems across sex, SES, and countries, we obtained descriptive statistics using R. Next, we performed twin analyses in R (version 3.4.3) with the package OpenMx (version 2.8.3) with the NPSOL optimizer (Neale et al., 2016).

## Twin Analyses

With twin models, by comparing resemblance on a trait between monozygotic (MZ) and dizygotic (DZ) twins, it is possible to disentangle to which extent individual differences in a trait can be explained by genetic variance (A), variance due to the shared environment (C), and variance due to the nonshared environment (E; Boomsma et al., 2002). We extended the model by including two moderators to test whether the contribution of genetic and environmental variance to individual differences in childhood behavioral problems interacted with these moderators (i.e., SES strata, country).



Because the distribution of childhood behavioral problems was highly skewed, we categorized the variable by applying two thresholds, partitioning the sample in the 33% lowest scoring, the middle 33%, and 33% highest scoring children on childhood behavioral problems. This method has the advantage of optimizing parameter estimates (Derks, Dolan, & Boomsma, 2004). We fitted the thresholds for the Netherlands and the United Kingdom separately, before entering them into the model. To simultaneously compare SES strata and countries, we fitted a 30 group model containing all groups (e.g., MZ male, DZ male, MZ female, DZ female, DZ opposite sex * SES low, SES medium, SES high * the Netherlands and the United Kingdom). Categorizing SES into low, medium, and high allowed us to test for both protective and negative moderating effects of (high/low) SES. Because studies so far did not find evidence for qualitative or quantitative sex differences for childhood behavioral problems (Porsch et al., 2016; Vink et al., 2012), we constrained the correlations and A, C, and E components to be equal for boys and girls and opposite-sex twin correlations to be equal to DZ correlations. To account for the frequently observed mean differences in behavioral problems, thresholds were allowed to vary across sex.

## Model Fitting

We tested moderator effects by stepwise testing whether constraining parameters to be equal across moderator categories significantly deteriorated goodness of fit (i.e., $p < .01$). If a constraint did not significantly deteriorate model fit, we proceeded with applying this constraint in the later models. Based on the best fitting model, we estimated the influence of genetic and environmental factors on childhood behavioral problems across SES strata and countries.

First, we fitted a saturated model with thresholds freely estimated across sex, SES strata, and countries, and with correlations freely estimated across SES strata and countries. Next, we fitted the following models to test for the moderating effects of SES and country: thresholds constrained to be equal across SES strata; correlations constrained to be equal across SES strata; thresholds constrained to be equal across country; and correlations constrained to be equal across country.

Based on the findings from the saturated model, we specified the ACE model with the same 30 groups, constraining thresholds and A, C, and E in line with the results from previous models to test the moderating effect of SES and country on the contribution of genetic and environmental factors to childhood behavioral problems. We first compared the ACE model to the best saturated model. Next, we tested the moderator effects by constraining A, C, and E across moderator categories.

For interpretational purposes, we performed additional analyses; we fitted the best ACE model but then with the thresholds constrained and the means and variances freely estimated. This model allowed us to examine the absolute variance of childhood behavioral problems across SES strata and countries. Based on this model, we reported the absolute values of A, C, and E.

## RESULTS

The descriptives in Table 2 reveal that means were higher for boys than girls, behavioral problems decreased as SES increased, and the variance decreased as SES increased. The patterns were similar between the Netherlands and the United Kingdom. The twin correlations as displayed in Table 3 suggest that for all three SES strata genetic factors played a role, because the MZ correlations were higher than the DZ correlations. Because the DZ correlations were larger than half of the MZ correlations, we suspected shared environmental effects. MZ and DZ twin correlations were highest for low SES, were slightly lower for medium SES, and lowest for high SES, suggesting an increased contribution of the nonshared environment for higher SES; the difference between MZ and DZ correlations was constant across SES strata. These patterns occurred both in the Netherlands and in the United Kingdom.

**Table 2.** Descriptive statistics

| | The Netherlands | | | | | | |
|---|---|---|---|---|---|---|---|
| **SES strata** | **N pairs** | **Mean boys** | **Variance boys** | **Mean girls** | **Variance girls** | **Total mean** | **Total variance** |
| SES low | 2,109 | 6.64 | 31.08 | 5.29 | 23.16 | 5.96 | 27.55 |
| SES medium | 5,143 | 6.07 | 28.20 | 4.68 | 20.52 | 5.37 | 24.77 |
| SES high | 4,730 | 5.04 | 24.57 | 3.75 | 14.68 | 4.40 | 20.05 |
| Total | 12,046 | 5.77 | 27.74 | 4.43 | 19.08 | 5.09 | 23.83 |
| | **The United Kingdom** | | | | | | |
| **SES strata** | **N pairs** | **Mean boys** | **Variance boys** | **Mean girls** | **Variance girls** | **Total mean** | **Total variance** |
| SES low | 1,372 | 2.59 | 3.91 | 2.06 | 3.13 | 2.27 | 3.49 |
| SES medium | 5,059 | 1.96 | 3.13 | 1.60 | 2.33 | 1.74 | 2.66 |
| SES high | 3,265 | 1.61 | 2.40 | 1.24 | 1.82 | 1.40 | 2.08 |
| Total | 9,822 | 1.88 | 2.98 | 1.51 | 2.29 | 1.69 | 2.66 |



**Table 3**. Twin correlations

| | **The Netherlands** | | **The United Kingdom** | |
|---|---|---|---|---|
| **SES strata** | **MZ** | **DZ** | **MZ** | **DZ** |
| SES low | .87 [.84, .89] | .55 [.49, .60] | .83 [.76, .88] | .53 [.43, .62] |
| SES medium | .85 [.83, .87] | .53 [.49, .57] | .79 [.75, .82] | .49 [.43, .54] |
| SES high | .76 [.73, .79] | .41 [.36, .46] | .70 [.64, .76] | .41 [.34, .48] |

The results of the model fitting are displayed in the upper half of Table 4. We fitted a saturated model allowing for sex differences on the thresholds, but with correlations constrained to be equal for boys and girls (model 0). To test for the moderating effect of SES on childhood behavioral problems, we first constrained thresholds to be equal across SES strata (model 1). Applying this constraint significantly deteriorated model fit, indicating that the prevalence of behavioral problems varied over SES strata, with higher liability for children of low SES families to be in the group scoring highest on behavioral problems in both countries. Next, we constrained twin correlations to be equal across SES strata (model 2). This constraint significantly decreased model fit, indicating that there were SES effects on the correlations. MZ and DZ correlations were highest for low SES, slightly lower for medium SES, and lowest for high SES, with the difference between MZ and DZ remaining constant, suggesting that the contribution of genetic factors increased, shared environment decreased, and the nonshared environment increased for higher SES. To test the moderating effect of country on childhood behavioral problems, we constrained thresholds to be equal across countries (model 3). This constraint significantly deteriorated model fit, indicating that the thresholds were significantly different between countries. This difference indicated that the liability of children to fall within the lowest, middle or highest scoring group of behavioral problems was different between the Netherlands and the United Kingdom. The first threshold was lower for the Netherlands, the second threshold was only lower in the Netherlands for boys from medium or high SES, the other second thresholds were higher in the Netherlands. This indicated a higher liability to belong to the middle scoring group of behavioral problems, but a lower liability to belong to the highest scoring group for boys and for girls from low SES for children in the Netherlands compared to children in the United Kingdom. Next, we constrained correlations to be equal across countries (model 4). This constraint did not decrease model fit, indicating that there were no effects of country on the correlations. Table 5 displays the parameter estimates of the final saturated model (model 4).

**Table 4.** Statistics of the fitted models.

| Saturated model | Model number | Model | Estimated parameters | -2 LL | df | AIC | Compared to model | Δ -2 LL | Δ df | p |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Saturated model with thresholds equal for MZ and DZ | 36 | 61852.35 | 30693 | 466.35 | | | | |
| | 0 | twins, correlations equal for boys and girls. | 20 | 62132.26 | 30709 | 714.26 | 0 | 279.91 | 16 | 3.63e-50 |
| | 1 | Thresholds equal across SES. | 28 | 61915.42 | 30701 | 513.42 | 0 | 63.07 | 8 | 1.16e-10 |
| | 2 | Correlations equal across SES. | 24 | 62155.55 | 30705 | 745.55 | 0 | 303.19 | 12 | 1.00e-57 |
| | 3 | Thresholds equal across countries | **30** | **61865.90** | 30699 | **467.90** | **0** | **13.54** | **6** | **.04** |
| | **4** | **Correlations equal across countries** | | | | | | | | |
| **ACE model** | **Model number** | **Model** | **Estimated parameters** | **-2 LL** | **df** | **AIC** | **Compared to model** | **Δ -2 LL** | **Δ df** | **p** |
| | **5** | **ACE model based on the results of the final saturated model** | 33 | 62175.85 | 30696 | 783.85 | | | | |
| | 6 | A, C, and E equal across low, medium, and high SES | 27 | 62234.37 | 30702 | 830.37 | 5 | 58.52 | 6 | 9.01e-11 |
| | 6.1 | A, C, and E equal across low and medium SES | 30 | 62179.76 | 30699 | 781.76 | 5 | 3.91 | 3 | 0.27 |
| | 6.2 | A, C, and E equal across low and high SES | 30 | 62218.81 | 30699 | 820.81 | 5 | 42.96 | 3 | 2.51e-09 |
| | 6.3 | A, C, and E equal across medium and high SES | 30 | 62215.38 | 30699 | 817.38 | 5 | 39.53 | 3 | 1.34e-08 |



**Table 5.** Parameter estimates of the best fitting saturated model (model 3) with 95% confidence intervals.

| | | The Netherlands | | The United Kingdom | |
|---|---|---|---|---|---|
| | | Correlation | | Correlation | |
| | | MZ | DZ | MZ | DZ |
| SES low | | .86 [.83, .88] | .54 [.49, .59] | .86 [.83, .88] | .54 [.49, .59] |
| SES medium | | .83 [.81, .85] | .51 [.48, .54] | .83 [.81, .85] | .51 [.48, .54] |
| SES high | | .74 [.71, .77] | .41 [.37, .45] | .74 [.71, .77] | .41 [.37, .45] |
| | | **Mean** | **Variance** | **Mean** | **Variance** |
| SES low | boys | -0.86 [-0.93, -0.79] | 0.09 [0.03, 0.16] | -0.67 [-0.78, -0.57] | 0.01 [-0.09, 0.11] |
| | girls | -0.64 [-0.71, -0.57] | 0.35 [0.28, 0.41] | -0.45 [-0.55, -0.35] | 0.21 [0.11, 0.30] |
| SES medium | boys | -0.75 [-0.80, -0.71] | 0.19 [0.15, 0.24] | -0.49 [-0.54, -0.43] | 0.27 [0.22, 0.32] |
| | girls | -0.55 [-0.59, -0.51] | 0.49 [0.44, 0.53] | -0.29 [-0.35, 0.24] | 0.48 [0.42, 0.53] |
| SES high | boys | -0.60 [-0.64, -0.55] | 0.41 [0.37, 0.45] | -0.26 [-0.33, -0.20] | 0.48 [0.41, 0.54] |
| | girls | -0.37 [-0.41, -0.32] | 0.69 [0.65, 0.74] | -0.09 [-0.16, -0.03] | 0.65 [0.58, 0.71] |

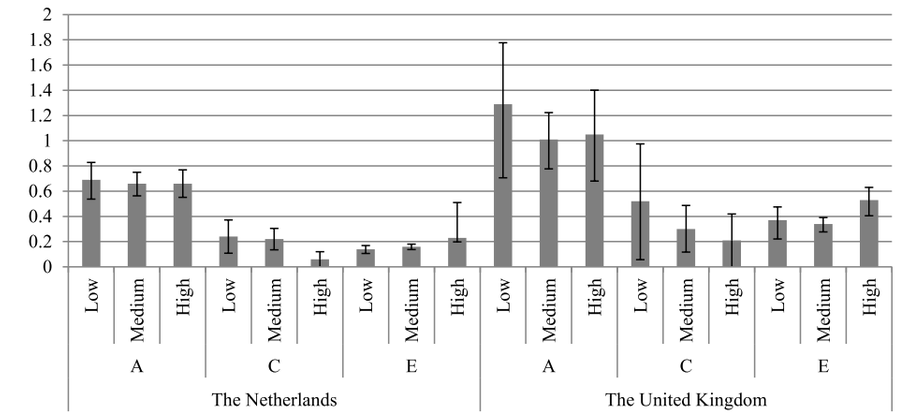
Based on the outcomes from the saturated model (model 4), we fitted the ACE model allowing thresholds to vary across sex, SES strata, and countries, and allowing A, C, and E to vary across SES strata (model 5). Because in the final saturated model, the MZ and DZ correlations did not differ significantly between the Netherlands and the United Kingdom, we constrained A, C, and E to be equal across country. The results of the ACE model fitting are displayed in the lower half of Table 4. We tested the moderating effect of SES by constraining A, C, and E to be equal across SES strata (model 6). This constraint significantly deteriorated model fit, indicating that the values of A, C, and E were significantly different for low, medium, and high SES. Between low and medium SES, A, C, and E were distributed similarly. For high SES compared to low and medium SES, A appeared to be higher, C lower, and E higher. To explore the moderating effect of SES, we performed pairwise comparisons on the A, C, and E estimates between SES strata to test which strata significantly differed from each other. The difference between low and medium SES (model 6.1) did not significantly decrease model fit. The difference between low and high SES (model 6.2), and medium and high SES (model 6.3) did significantly decrease model fit, indicating a significant difference in A, C, and E between low and high SES and medium and high SES. A appeared to be higher, C lower, and E higher, for high SES compared to both low SES and medium SES.

**Table 6.** Parameter estimates of the best fitting ACE model (model 4) with 95% confidence intervals from the model estimating the thresholds and separately for the Netherlands and the United Kingdom from the model estimating the absolute variance.

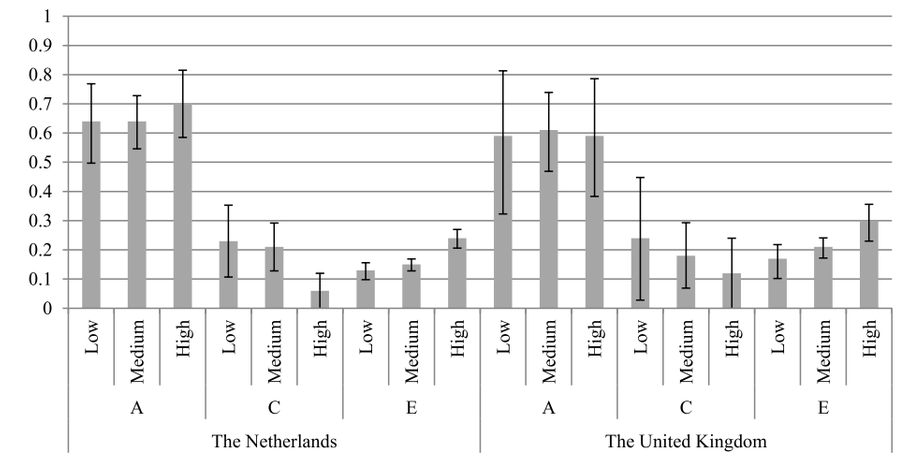
| SES category | A unstandardized | C unstandardized | E unstandardized | A | C | E |
|---|---|---|---|---|---|---|
| NL SES low | 0.69 [0.55, 0.84] | 0.24 [0.11, 0.37] | 0.14 [0.11, 0.18] | .64 [.51, .79] | .23 [.12, .35] | .13 [.11, .16] |
| NL SES medium | 0.66 [0.57, 0.75] | 0.22 [0.13, 0.30] | 0.16 [0.14, 0.18] | .64 [.55, .73] | .21 [.13, .29] | .15 [.13, .18] |
| NL SES high | 0.66 [0.55, 0.77] | 0.06 [0.00, 0.15] | 0.23 [0.20, 0.26] | .70 [.59, .82] | .06 [.00, .16] | .24 [.21, .27] |
| UK SES low | 1.29 [0.81, 1.88] | 0.52 [0.06, 0.98] | 0.37 [0.27, 0.52] | .59 [.37, .86] | .24 [.03, .45] | .17 [.12, .24] |
| UK SES medium | 1.01 [0.79, 1.24] | 0.30 [0.12, 0.49] | 0.34 [0.29, 0.41] | .61 [.48, .75] | .18 [.07, .30] | .21 [.18, .25] |
| UK SES high | 1.05 [0.70, 1.42] | 0.21 [0.00, 0.49] | 0.53 [0.43, 0.66] | .59 [.39, .79] | .12 [.00, .27] | .30 [.24, .37] |

**Figure 1.** Unstandardized and standardized estimates of the best fitting ACE model with 95% confidence intervals.

Unstandardized estimates




Standardized estimates






The additional analyses revealed absolute differences in variance across SES (Δ -2 LL = 13.58, Δ df = 4, $p < .01$), with the highest variance for low SES, lower variance for medium SES, and lowest variance for high SES. Furthermore, the absolute variance varied between the Netherlands and the United Kingdom (Δ -2 LL = 221.37, Δ df = 3, $p < .01$); the variance was larger in the United Kingdom. Table 6 presents the absolute and standardized estimates across countries and SES strata. Figure 1 graphically displays the unstandardized and standardized parameter estimates. The parameter estimates and their confidence intervals suggested that the moderating effect of SES was mainly driven by E. Constraining A and C to be equal across SES strata, however, revealed that there were significant differences in A and C across SES strata (Δ -2 LL = 32.92, Δ df = 8, $p < .01$).

## DISCUSSION

The purpose of the present study was to investigate whether the genetic architecture of childhood behavioral problems differs across low, medium, and high SES. In addition, we examined whether this effect varied between the Netherlands and the United Kingdom, two countries that differ in income disparity (The World Bank Group, 2017). For this, we used data from two large longitudinal prospective twin cohorts. We found more childhood behavioral problems in children from low SES compared to medium SES, and from medium SES compared to high SES. Furthermore, children from the Netherlands were more likely to be in the middle scoring group on childhood behavioral problems compared to children from the United Kingdom, whereas children from the United Kingdom were more likely to be in the high scoring group compared to children from the Netherlands, except for girls from medium and high SES. The variance of childhood behavioral problems was higher in the United Kingdom compared to the Netherlands.

In our study, SES did have a moderating effect on the contribution of genetic and environmental factors to childhood behavioral problems. Because we categorized SES into three strata, we were able to examine whether the moderating effect of SES on the genetic architecture of childhood behavioral problems may be due to the presence of advantageous factors (i.e., high vs. medium SES) or to the presence of disadvantageous factors (i.e., low vs. medium SES; Brumley & Jaffee, 2016). The difference in the contribution of genetic and environmental factors to childhood behavioral problems between low and medium SES was not significant, while the difference between low and high SES and the difference between medium and high SES were significant. These findings indicate that the moderating effect of SES appears to be due to growing up in a more advantageous environment. Children from high SES families tend to have less exposure to environmental factors that increase the risk of behavioral problems (e.g., parental stress) and

more exposure to environmental factors that may decrease the risk of behavioral problems (e.g., monitoring). The environment in high SES families may leave less room for individual differences in childhood behavioral problems, as evidenced in the non-significant contribution of the shared environment. Among children from high SES backgrounds, genetic and nonshared environmental factors appear to be more relevant in explaining variability in behavioral problems, compared to children from low or medium SES backgrounds, for whom the influence of the shared environment is larger. The total variance estimate was higher for low and medium SES, compared to high SES, indicating that children from low and medium SES were more heterogeneous in their levels of childhood behavioral problems. For the variance component estimates, the estimate of genetic variance was similar across SES strata, the estimate of the shared environment variance component was larger in low and medium SES than in high SES families, and the estimate of the nonshared environment variance component was higher for children from high SES families than for children from low and medium SES. This pattern occurred both in the Netherlands and the United Kingdom. Our results were in line with the theory that under advantageous circumstances (i.e., high SES) genetic influences are more fully realized compared to less advantageous or disadvantageous circumstances in which genetic influences are more suppressed and the environment contributes more strongly (Bronfenbrenner & Ceci, 1994).

A clinical implication from these results is that the etiology of childhood behavioral problems is different for children from low or medium SES families compared to children from high SES families. For children from low or medium SES families, the shared environment explains a larger proportion of the variance compared to children from high SES families, suggesting that these children could benefit from treatment aiming to ameliorate shared environmental factors (Burt, 2009). Examples of these factors could be healthy family functioning, less parental stress, positive school attachment, or housing quality (Brumley & Jaffee, 2016; Burt, Klahr, Neale, & Klump, 2013; Burt & Klump, 2014; Hudziak & Ivanova, 2016).

We did not find evidence for our hypothesized stronger moderating effect of SES on the genetic architecture of behavioral problems for children in the United Kingdom than in the Netherlands. Nevertheless, this should be interpreted with caution because the two countries measured behavioral problems with a different questionnaire. Porsch and colleagues (2016) found that the contribution of genetic and environmental factors to childhood behavioral problems is similar for children from the Netherlands and the United Kingdom. Extending this finding, our results showed that the moderating effect of SES was comparable in the Netherlands and the United Kingdom. Although there is more income inequality in the United Kingdom compared to the Netherlands (The World Bank Group, 2017), the United Kingdom also, for example, invests a larger percentage of the Gross National Product in family benefits (OECD Social Policy Division, 2016) and education (Eurostat, 2016), which may attenuate the effects of larger income inequality. A similar explanation is proposed for the different findings of the moderating effect of SES on the contribution of genetic and environmental factors to intelligence (Duncan, Morris, & Rodrigues, 2011; Tucker-Drob & Bates, 2015). Therefore, it would be promising for future research to investigate whether our findings replicate in other countries with different levels of income inequality and country investments in children and families.

## Strengths and Weaknesses

One strength of this study is that we applied thresholds to take the skewed distribution of childhood behavioral problems into account. Studies so far on the moderating effect of SES on the genetic architecture of childhood behavioral problems included childhood behavioral problems as a continuous variable (Middeldorp et al., 2014; Tuvblad, Eley, & Lichtenstein, 2005). Because of the non-normal distribution of childhood behavioral problems, analyzing behavioral problems continuously might lead to overestimated genetic variance and underestimated shared environmental variance. By incorporating thresholds, our analyses could have led to more precise parameter estimates (Derks et al., 2004). Nevertheless, although we used a method to obtain more precise estimates compared to previous articles, our results confirmed their findings regarding the moderating effect of SES on the contribution of genetic and environmental factors to individual differences in behavioral problems both in the Netherlands and the United Kingdom, indicating robustness. A second strength of this study is that we fitted all the estimates for the different SES strata and countries simultaneously, instead of separately for each country, allowing for formal comparison of parameter estimates across all groups.

Despite the strengths of our study, several limitations warrant mentioning. One limitation is that the measures of childhood behavioral problems differed between the Netherlands and the United Kingdom. For this reason, it is not certain whether the difference in means and variances between the Netherlands and the United Kingdom was due to true differences across countries or due to different measures. Furthermore, the different measures could lead to underestimation of the moderating effect of countries on the contribution of genetic and environmental factors to childhood behavioral problems. Although several studies found high comparability between the Child Behavior Checklist and the Strengths and Difficulties Questionnaire (Goodman & Scott, 1999; Mieloo et al., 2012), they assess different symptoms of behavioral problems. A recent study with data from the Netherlands and the United Kingdom found that the genetic architecture of childhood behavioral problems was similar for both instruments (Porsch et al., 2016). In this paper, we took a next step and allowed moderation of genetic architecture by SES. Here

also, the outcomes suggested very similar findings for both instruments, but we acknowledge that we cannot state to which extent the results may be affected by different measures of behavioral problems and different countries. For example, whereas the sample from the United Kingdom represents the population (Haworth et al., 2013), the sample from the Netherlands is on average more highly educated than the Dutch population (Hoekstra et al., 2010). Future research should investigate whether the genetic architecture of childhood behavioral problems varies across questionnaires in the same sample and whether our findings hold when employing the same questionnaire in the Netherlands and the United Kingdom.

A second limitation is that we used only parental education as measure of SES. Studies with TEDS data usually employ a measure of SES comprised of parental education, parental occupation, income, and sometimes maternal age at birth of the first child (Hanscombe et al., 2012; Krapohl & Plomin, 2016). However, because we used data from two different countries with different questionnaires, we decided to use a homogenous measure in order to optimize comparability between the data from the Netherlands and from the United Kingdom. An opportunity for future research would be to test whether our results regarding the moderating effect of SES on the genetic architecture of childhood behavioral problems across countries also apply to other measures of SES. Nevertheless, although the abovementioned limitations require more cautious interpretation of our findings, the present study does provide important insight in the genetic architecture of childhood behavioral problems across SES strata and countries.

A third limitation was that we only included a single age group. It is known that the contribution of genetic and environmental factors to individual differences in behavioral problems changes with age; the role of the shared environment disappears in adolescence (Van Beijsterveldt et al., 2003; Wesseldijk et al., 2017). Furthermore, the way adolescents perceive their SES differs from how children perceive it (Goodman, Huang, Schafer-Kalkhoff, & Adler, 2007). Also, the association between behavioral problems and SES decreases as children become older (Piotrowska et al., 2015). Therefore, it is likely that our findings cannot be generalized to other ages, and thus it would be useful for future research to examine the moderating effect of SES on the genetic architecture of behavioral problems across childhood and adolescence.



## Conclusion

The present study sought to gain insight in the etiology of childhood behavioral problems by investigating whether the contribution of genetic and environmental factors varies across SES strata in the Netherlands and the United Kingdom. Our results showed that both in the Netherlands and the UK, shared environmental factors have a stronger effect in behavioral problems in children from low SES families, while genetic factors are most prominent for behavioral problems in children from medium and high SES families. These findings have important implications for prevention and intervention goals.

Chapter 6

# Content, Diagnostic, Correlational, and Genetic Similarities Between Common Measures of Childhood Aggressive Behaviors and Related Psychiatric Traits.

Submitted as: Hendriks, A.M., Ip, H.F., Nivard, M.G., Finkenauer, C., Van Beijsterveldt, C.E.M., Bartels, M., & Boomsma, D.I. (2019). Content, diagnostic, correlation, and genetic similarities between common measures of childhood aggressive behaviors and related psychiatric disorders.

## ABSTRACT

Background: Given the role of childhood aggressive behavior as part of psychiatric disorders and in everyday child development, it is critical to precisely and accurately measure childhood aggressive behavior in clinical practice and research. The goal of our study is to quantify the agreement between widely used measures of childhood aggressive behaviors in terms of item content, clinical concordance, correlation, and the degree to which they measure a common genetic construct.

Methods: To this end, we analyzed a sample of 1254 Dutch twin pairs (age 8 to 10 years, 51.1 % boys) for whom the mothers and fathers filled in the A-TAC, the CBCL, and the SDQ to assess aggressive behavior.

Results: Results revealed variation in item content among the various measures of aggressive behaviors. The agreement with respect to diagnosis was very weak to weak, correlations between continuous scores were generally moderate, and polychoric correlations among measures were generally moderate to strong. In contrast, we found strong genetic overlap among the different measures for childhood aggressive behavior, suggesting that the different measures assessed a similar underlying genetic construct.

Conclusion: Our results demonstrated to what extent different measures of childhood aggressive behavior converge depends on the type (i.e., item content, clinical concordance, correlation, genetic overlap) of agreement considered. High genetic overlap suggests that the underlying construct of childhood aggression is consistent across measures.



**Keywords:**
childhood aggressive behaviors, item overlap, clinical concordance, genetic correlation.




Aggressive behaviors are disruptive and key features of childhood onset psychiatric disorders like conduct disorder (CD) and oppositional defiant disorder (ODD). Both CD and ODD have a 12-month prevalence of about 1% (Kessler, Chiu, Demler, & Walters, 2005), but the lifetime prevalence CD is estimated to be between 2.1% and 9.5 % and the lifetime prevalence of ODD is estimated to be between 3.6% and 10.2% (Kessler, Berglund, et al., 2005; Nock, Kazdin, Hiripi, & Kessler, 2007; Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). These prevalences are relatively stable across cultures (Canino, Polanczyk, Bauermeister, Rohde, & Frick, 2010). In addition to its role in the development of psychiatric disorders, aggressive behavior is associated with adverse outcomes, including high co-occurrence with other behavioral and emotional problems (Bartels et al., 2018), negative consequences for parents (Meltzer, Ford, Goodman, & Vostanis, 2011; Roberts, McCrory, Joffe, de Lima, & Viding, 2017), and high financial costs for society (Rivenbark et al., 2018; Romeo, Knapp, & Scott, 2006). Given the high impact of aggressive behaviors, reliable and valid assessment and measurement of aggressive behavior during childhood is of considerable interest for clinical screening, clinical referral, differential diagnosis, as a criterion for inclusion or exclusion from research, and as an outcome in its own right.

A variety of common screening instruments contains scales which measure aggressive behavior, or psychiatric disorders which are directly related to aggressive behavior, like CD and ODD. Among these instruments are the Autism - tics, attention-deficit hyperactivity disorder, and other comorbidities (A-TAC; Hansson et al., 2005), the Child Behavior Checklist (CBCL; Achenbach, Ivanova, & Rescorla, 2017; Achenbach & Rescorla, 2001), and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). The CBCL assess aggressive behavior in a broader sense, the SDQ mainly assesses CD, whereas the A-TAC CD and ODD scales aim to specifically assess CD and ODD (Hansson et al., 2005; Warnick, Bracken, & Kasl, 2008). The A-TAC is developed as a diagnostic interview to screen general populations and contains subscales to assess CD and ODD based on DSM-IV symptoms (Hansson et al., 2005). The CBCL has an 18-item Aggressive Behavior subscale, that is not tailored to a specific disorder, but broadly measures aggressive behavior. It was developed as an empirically based assessment tool, with items based on literature searches, expertise of mental health professionals, and pilot testing with parents, followed by factor analyses to create empirical scales (Achenbach et al., 2017; Achenbach & Rescorla, 2001). The SDQ is specifically developed to be a brief questionnaire and contains a 5-item scale to assess conduct problems (Goodman, 2001). Nosological concepts from the DSM-IV and the ICD-10 and factor analyses guided item selection for the SDQ (Goodman & Scott, 1999). Thus, although all instruments aim to assess childhood aggression and aggression-related disorders, they differ in their origin and approach to the assessment of aggressive behaviors, which may reflect in the content of items across instruments.

Despite their differences, these instruments have been used for similar clinical practices (e.g., screening) and to answer similar research questions, such as assessment of treatment effectiveness or validation of instruments (Warnick et al., 2008). They have furthermore been used together in several studies that combined data from different research groups, implicitly assuming that they measure the same underlying construct (Pappa et al., 2016; Porsch et al., 2016).

To test the validity of the assumption that data with different measures of aggressive behaviors can be combined we aim, in this study, to quantify the agreement of several measures of childhood aggressive behaviors in terms of item content, clinical concordance, correlation, and the degree to which they measure a common genetic construct.

## Item Content

The different assessment instruments vary by item content, mainly driven by the construct that they tend to assess. For instance, the DMS IV distinguishes between ODD and CD; CD comprises aggression to people and animals, destruction of property, deceitfulness or theft, and serious violation of rules, whereas ODD consists of losing temper, and hostile and defiant behavior. If children meet criteria for both CD and ODD, they receive a diagnosis of CD, thus diagnoses for CD and ODD are mutually exclusive (American Psychiatric Association, 1994). This division between CD and ODD is reflected in the A-TAC subscales. The SDQ claims to assess symptoms associated with CD, while the CBCL Aggressive Behavior syndrome scale is a broader concept, not directly related specifically to either CD or ODD.

Previous work that tested convergence between measures of childhood aggressive behaviors, focused on agreement between scores, but did not explicitly test the similarity (or dissimilarities) in item content (Goodman & Scott, 1999; Halleröd et al., 2010). In addition, although previous studies discussed similarity in item content between DSM-IV criteria of CD or ODD and measures of childhood aggressive behavior, they did not formally test this similarity (Gould, Bird, & Jaramillo, 1993; Hansson et al., 2005). The A-TAC CD and A-TAC ODD subscales measure two mutually exclusive disorders, directly based on the DSM-IV criteria and there is no overlap in item content between these subscales. The A-TAC CD scale and the SDQ conduct problems scale, however, both assess symptoms related to CD, which would imply a high overlap in item content. Finally, the CBCL Aggressive Behavior subscale assesses aggression in a broader sense, which means we have no strong prior expectation of the overlap (in terms of item content) with the other measures, we do expect neither a full absence of overlap nor a very high overlap. Nonetheless, it is undetermined whether the extent to which measures of childhood aggressive behaviors overlap in item content translates into a level of agreement with regards to clinical concordance and correlation between the measures.



## Clinical Concordance

The different measures of childhood aggressive behavior all have a scale specific threshold to distinguish between clinical and non-clinical levels of aggressive behavior. However, it is uncertain whether the same children would receive a diagnosis according to the different measures of childhood aggressive behavior. For instance, one study found that the CBCL and SDQ discriminated equally well between a sample of children aged 4 to 7 years collected at a dental clinic and children collected at psychiatric clinics who were referred for externalizing behaviors (N = 132; Goodman & Scott, 1999). Similarly, another study found that the CBCL and SDQ discriminated equally well between German children aged 4 to 16 years from a community sample and from psychiatric clinics (N = 273; Klasen et al., 2000). These findings suggest good clinical concordance between these measures of aggressive behavior. Although the different measures discriminated between different groups does not mean that they are equally suitable to distinguish between individual children. To our knowledge, clinical concordance between scales of the A-TAC and other measures of childhood aggressive behavior has not been examined. Information can be drawn, however, from research comparing clinical levels of aggressive behavior as assessed using the CBCL with DSM-III diagnoses of CD and ODD, which found point-biserial correlations of respectively .22 and .57 (Gould et al., 1993). This research suggests higher clinical concordance between the CBCL Aggressive Behavior scale and the DSM-based A-TAC ODD scale than between the CBCL Aggressive Behavior scale and the DSM-based A-TAC CD scale. Prior work on clinical concordance between the scales finds scales to discriminate equally well between cases and controls, but it does not explicitly compare clinical decisions between the scales. It is undetermined whether the extent to which measures of aggressive behaviors assess the same symptoms and rate the same children as clinical translates into a strong correlation between the different measures if they are all applied in one group.

## Correlations

Previous research has established correlations between the total scores of measures of childhood aggressive behavior. For instance, correlations between the A-TAC CD and ODD and the CBCL Aggression scale in a sample of 106 twin pairs aged 9 or 12 years were, respectively, .48 and .32, indicating moderate convergence (Halleröd et al., 2010). Between the CBCL Externalizing scale, which sums Aggressive Behavior and Rule Breaking, and the SDQ conduct scale, convergence was high (correlations from .71 to .84), in a sample of 132 children aged 4 to 7 years, a sample of 292 children in child welfare aged 3 to 12 years, and a sample of 287 children aged 8 to 16 years (Goodman & Scott, 1999; Janssens & Deboutte, 2009; Van Widenfelt,

Goedhart, Treffers, & Goodman, 2003). The moderate to high correlation across scales found in these prior studies suggests that children considered to score higher according to one measure will probably also score higher on another measure.

## Genetic Architecture

A wide body of literature reports childhood aggressive behavior to be a highly heritable trait (Burt, 2009; Tuvblad & Baker, 2011; Waltes, Chiocchetti, & Freitag, 2016). Here we want to assess whether, regardless of overlap in item content, clinical concordance, or correlation, different instruments measure the same underlying genetic construct. Several twin studies examined the heritability of childhood aggressive behaviors for the A-TAC, CBCL, and SDQ. Generally, these studies found in 9-year-old children that genetic factors explained at least 50% of individual differences in aggressive behaviors and shared environmental factors and that shared environmental factors explained between 0% and 36% percent (Kerekes et al., 2014; Porsch et al., 2016). In most studies different instruments were employed to assess aggression and although heritability estimates are rather similar, this does not formally establish whether different instruments reflect a common underlying genetic construct. We therefore consider, in addition to overlap in item content, clinical concordance, and correlations, also the genetic correlations between different measures of aggressive behavior. To this end we make use of the unique characteristics of a twin sample, where both parents reported on their children's aggressive behavior on three questionnaires (i.e., A-TAC, CBCL, SDQ) for a sample of 1254 twin pairs. These data allow us to employ a multivariate genetic model to estimate the heritability and genetic correlations among different scales.

# METHOD

## Sample

The sample comprised 2508 children (1254 twin pairs) aged 8 to 10 years old (51.1% boys) born between September 2005 and October 2008 sampled from the Netherlands Twin Register (NTR; Van Beijsterveldt et al., 2013). The NTR is a nation-wide population-based register founded in 1987. In 2016, mothers and fathers of these twin pairs were invited to complete the NTR survey that was collected around age 9 years and which included several measures of aggressive behavior (i.e., A-TAC, CBSL, SDQ). Mothers reported on aggressive behaviors on at least one measure for 2405 children, fathers for 1613 children. Some families, had multiple sets of twins; here we included one twin pair per family, yielding a sample of 1240 twin pairs of which 486 were monozygotic (MZ) and 754 dizygotic (DZ). The research was conducted according to the principles of the Declaration of Helsinki.



## Instruments

**A-TAC.** Two scales from the A-TAC assessed aggressive behaviors, namely the conduct disorder (A-TAC-CD) scale and oppositional defiant disorder (A-TAC-ODD) scale. Both scales consisted of five items; parents were asked whether their children displayed the problem behaviors more frequently than peers in any period of their life. Response categories were 0 = *"No"*, 0.5 = *"Yes, to some extent"*, or 1 = *"Yes"* (Hansson et al., 2005). Children with more than a single item missing were not included in the analyses. Scores higher than 1.5 on the A-TAC-CD and 2.5 on the A-TAC-ODD indicated clinical levels, respectively (Kerekes et al., 2014).

**CBCL.** The Aggressive Behavior syndrome (CBCL-AGG) subscale from the CBCL. consisted of 18 items, asking parents to report on their children's behaviors in the past six months. The response categories contained 0 = *"Not true"*, 1 = *"Sometimes or somewhat true"*, or 2 = *"Very true or often true"* (Achenbach & Rescorla, 2001). If more than three items were missing, children were excluded from analyses. Scores from the 98$^{th}$ percentile (i.e., above 13) are considered to be in the clinical range (Achenbach & Ruffle, 2000).

**SDQ.** The Conduct Problem subscale (SDQ-CP) consisted of five items asking parents to report on their children's behavior. Parents could respond with 0 = *"Not true"*, 1 = *"Somewhat true"*, or 2 = *"Certainly true"* (Goodman, 1997, 2001). Children with more than two items missing were excluded from analyses. Scores above 3 revealed clinical levels (Goodman, 1997).

## Analyses

**Item content.** We examined similarity in item content of the aggressive behavior measures using the Jaccard index and added the DSM-IV criteria for conduct disorder (DSM-IV-CD) and oppositional defiant disorder (DSM-IV-ODD) as a benchmark (American Psychiatric Association, 1994). Supplementary Table 1 displays all available items in the study, and the symptoms of DSM-IV diagnosis of ODD and CD. Together, the three scales comprised 55 items, which assessed 26 different symptoms of aggressive behavior. Symptoms in all scales were coded present (i.e., 1) or absent (i.e., 0). To examine to which extent the scales contained similar symptoms, we calculated the Jaccard index that ranges from 0 (i.e., not similar) to 1 (i.e., fully similar). This index calculates similarity by dividing the overlap in symptoms between two measures by the number of non-overlapping symptoms between two measures. If multiple items tapped the same symptom, we considered them as a single item. In line with Fried (Fried, 2017), we used the following rules of thumb to interpret item overlap: very weak = 0.00 – 0.19, weak = 0.20 – 0.39, moderate = 0.40 – 0.59, strong = 0.60 – 0.79, and very strong = 0.80 – 1.00.

**Clinical concordance and correlation.** To investigate the extent to which clinical decisions based on different measures of childhood aggressive behaviors would be similar, we evaluated the number of children with clinical levels across scales, separately in the data from mothers and fathers, and boys and girls. We tested agreement (i.e., no clinical level vs. clinical level) between scales within rater with a Cohen's Kappa (Landis & Koch, 1977) and, to analyze the full range of scores while taking the skewed distribution of aggressive behavior into account, we calculated Spearman's rank correlations to assess the association between the scales (Spearman, 1904). Through bootstrapping with 1000 repetitions, we obtained 95% confidence intervals for the Spearman's rank correlations with the RVAideMemoire-package (Herv, 2018). To compute polychoric correlations (i.e., correlations for the underlying normally distributed liability distribution) we categorized the aggressive behavior measures into three categories; 0, 0.5/1 (i.e., 0.5 for A-TAC-CD and A-TAC-ODD, and 1 for CBCL-AGG and SDQ-CP), and higher (frequencies of the categories are in Supplementary Table 3). Based on these categories, we estimated polychoric correlations with the polycor package (Fox, 2016). Confidence intervals were computed based on the standard errors. We interpreted the Cohens's Kappa, Spearman's rank correlation, and polychoric correlation according to the following rules of thumb: very weak = 0.00 – 0.19, weak = 0.20 – 0.39, moderate = 0.40 – 0.59, strong = 0.60 – 0.79, and very strong = 0.80 – 1.00 (Landis & Koch, 1977; Spearman, 1904). In addition, we calculated weighted mean correlations for interpretation. We performed these analyses separately for boys and girls.

**Genetic analyses.** Twin analyses allow for disentangling the extent to which individual differences in a trait can be attributed to additive genetic factors (A), shared environment (C) common to children from the same family, or nonshared environment (E). This is done by comparing the resemblance on a trait between MZ and DZ twins. An extension is a multivariate model, in which cross-trait cross-twin correlations provide the information needed to disentangle the extent to which the correlation between two traits is explained by genetic or environmental factors (Boomsma, Busjahn, & Peltonen, 2002; Kendler, Neale, Kessler, Heath, & Eaves, 1992). If multiple traits are influenced by the same genetic factors, there is a non-zero genetic correlation among the different traits. The same reasoning applies to environmental correlations among variables. To estimate genetic and environmental correlations among scales, multivariate analyses of twin data were carried out in R version 3.5.1, using the OpenMx package (version 2.11.5; Neale et al., 2016) specifying NPSOL optimizer. Confidence intervals were calculated using MxCI in OpenMx. We fitted a four-variate model with the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ-CP with two groups (MZ and DZ twins), since previous research found a similar genetic architecture for boys and girls (Porsch et al., 2016; Vink et al., 2012). We allowed for mean differences between boys and girls. Analyses were carried out for mother and father ratings of aggressive behavior. The variance of each scale and the covariance among scales were partitioned into components explained by additive genetic factors (A), shared environmental factors (C), and nonshared environmental factors (E). Because the model assumes the data follow a multivariate normal distribution we expected some bias in the parameter estimates when the model is applied to the (skewed) measures of aggressive behavior (Derks, Dolan, & Boomsma, 2004). For this reason, we performed multivariate genetic models on the ordinal data as a sensitivity check. A description of the ordinal data methods can be found in Supplement 1.

# RESULTS

## Item Content

Table 1 summarizes the outcomes of the Jaccard analyses. All measures of agreement are displayed in Figure 1. Consistent with mutual exclusivity of diagnoses of CD and ODD (American Psychiatric Association, 1994), there was no overlap in content between DSM-CD and A-TAC-ODD, DSM-ODD and A-TAC-CD, and A-TAC-CD and A-TAC-ODD. Because the A-TAC scales were based directly on the DSM-IV descriptions of CD and ODD, we expected high overlap. Between DSM-CD and A-TAC-CD, however, overlap was weak; between DSM-ODD and A-TAC-ODD the overlap was moderate. The SDQ-CP assessed symptoms of CD, suggesting a stronger overlap with the DSM-CD and A-TAC-CD; which also reflected in the moderate overlap between the SDQ-CP and A-TAC-CD but not in the very weak overlap between the SDQ-CP and DSM-CD. Interestingly, overlap between DSM-ODD and SDQ-CP was higher (i.e., weak), but the overlap between the A-TAC-ODD and SDQ-CP was clearly lower (i.e., very weak) than the association between the SDQ-CP and the A-TAC-CD. As expected, the broader CBCL-AGG had a very weak to weak overlap with all other scales (i.e., lowest overlap with the A-TAC-CD and highest overlap with the DSM-ODD), which confirmed that it measured no specific disorder but aggressive behavior in a broader sense.

**Table 1.** Jaccard index for item overlap.

| | DSM-CD | DSM-ODD | A-TAC-CD | A-TAC-ODD | CBCL-AGG | SDQ-CP |
|---|---|---|---|---|---|---|
| DSM-CD | | | | | | |
| DSM-ODD | 0 | | | | | |
| A-TAC-CD | 0.29 | 0 | | | | |
| A-TAC-ODD | 0 | 0.57 | 0 | | | |
| CBCL-AGG | 0.17 | 0.31 | 0.12 | 0.29 | | |
| SDQ-CP | 0.19 | 0.20 | 0.50 | 0.12 | 0.19 | |

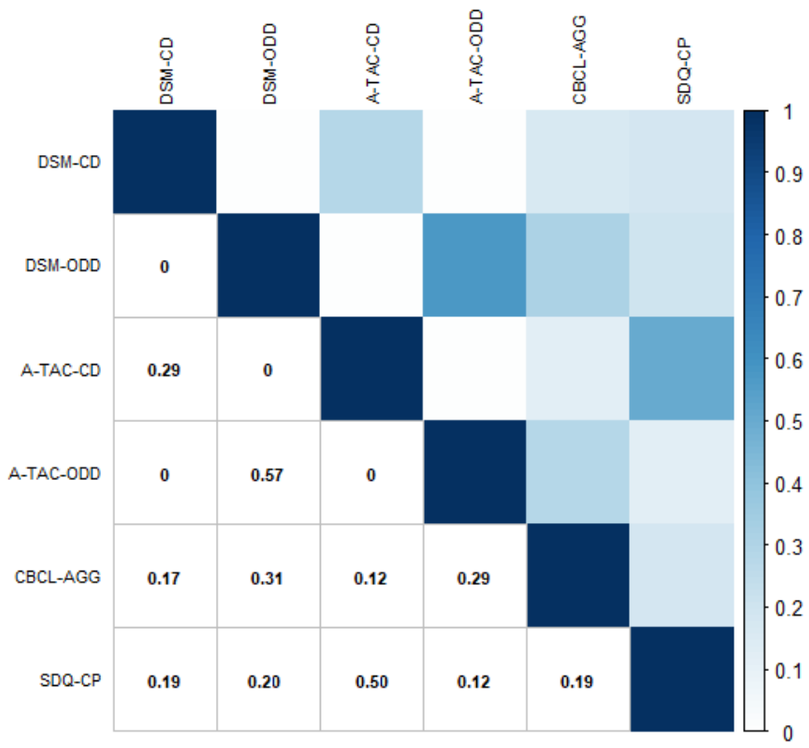
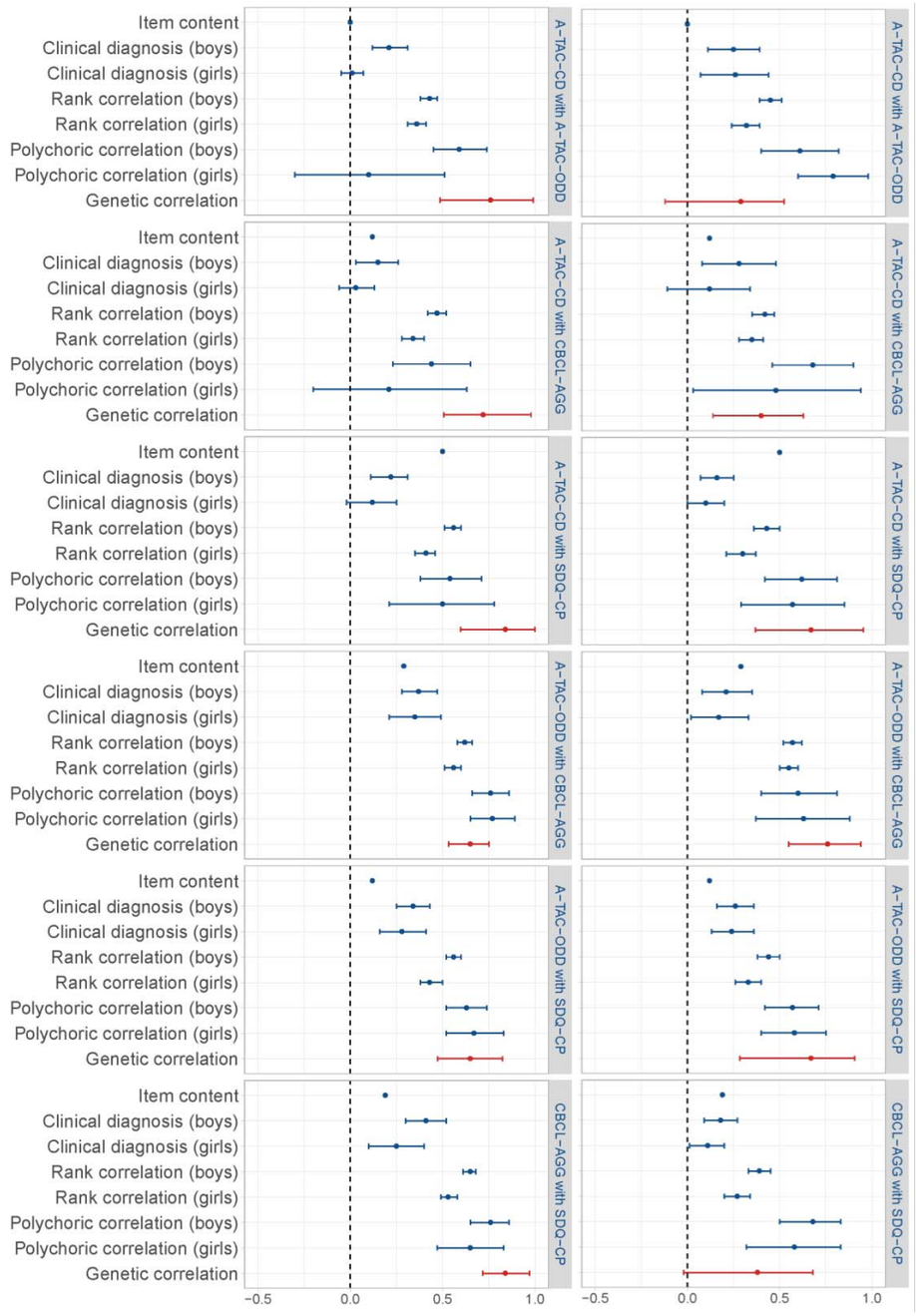





**Figure 1.** Agreement between aggressive behavior measures, the left panel is for mother reports, the right panel is for fathers.



## Clinical Concordance

Supplementary Table 2 displays the absolute counts of the clinical diagnoses across measures separately for mother- and father-report and for boys and girls. Prevalences of clinical aggression for boys as reported by mothers ranged from 3% (A-TAC-CD) to 10% (A-TAC-ODD); as reported by fathers they ranged from 2% (A-TAC-CD) to 13% (SDQ-CP). For girls, prevalences for mother-report ranged from 1% (A-TAC-CD) to 5% (A-TAC-ODD). For father-report on girls, prevalences ranged from 1% (A-TAC-CD) to 9% (SDQ-CP).

Table 2 summarizes the clinical concordance as assessed with Cohen's kappa's. Cohen's kappa assessed agreement between measures on whether or not children scored above a clinical threshold. Although diagnoses for CD and ODD were mutually exclusive by definition, there was very weak to weak agreement on clinical diagnosis between the A-TAC-CD and A-TAC-ODD (0.01 for mother report on girls to .26 for father report on girls), with a mean agreement of .10. In line with higher point-biserial correlations between CBCL-AGG and ODD compared to CD in previous research, clinical concordance between the A-TAC-ODD and CBCL-AGG was higher than between the A-TAC-CD and CBCL-AGG. Clinical concordance between the A-TAC-CD and CBCL-AGG ranged from .03 (mother report on girls) to .28 (father report on boys) with a mean of .10; between the A-TAC-ODD and CBCL-AGG it ranged from .17 (father report on girls) to .37 (mother report on boys) with a mean of .32. Although SDQ-CP mostly assessed symptoms of CD, clinical concordance with the A-TAC-ODD was higher than with the A-TAC-CD. Clinical concordance between the A-TAC-CD and SDQ-CP ranged from .10 (father report on girls) to .22 (mother report on boys) with a mean of .15. Clinical concordance between the A-TAC-ODD and SDQ-CP ranged from .24 (father report on girls) to .34 (mother report on boys) with a mean of .29. Finally, although in previous research CBCL-AGG and SDQ-CP discriminated equally well between clinical and general population samples, clinical concordance ranged from .11 (father report on girls) to .41 (mother report on boys) with a mean of .22, indicating that agreement generally was weak on who received a score above the clinical threshold. These findings indicated that there was limited (i.e., very weak to weak) agreement on which children received a clinical diagnosis according to these measures of aggressive behavior.

**Table 2.** Cohen's Kappa with 95% confidence intervals.

| **Boys, mother report** | | | | |
|---|---|---|---|---|
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .21 [.12, .31] | 1.00 | | |
| CBCL-AGG | .15 [.03, .26] | .37 [.28, .47] | 1.00 | |
| SDQ-CP | .22 [.11, .32] | .34 [.25, .43] | .41 [.30, .52] | 1.00 |
| **Boys, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .25 [.11, .39] | 1.00 | | |
| CBCL-AGG | .28 [.08, .48] | .21 [.08, .35] | 1.00 | |
| SDQ-CP | .16 [.07, .25] | .26 [.16, .36] | .18 [.09, .27] | 1.00 |
| **Girls, mother report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .01 [-.05, .07] | 1.00 | | |
| CBCL-AGG | .03 [-.06, .13] | .35 [.21, .49] | 1.00 | . |
| SDQ-CP | .12 [-.02, .25] | .28 [.16, .41] | .25 [.10, .40] | 1.00 |
| **Girls, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .26 [.07, .44] | 1.00 | | |
| CBCL-AGG | .12 [-.11, .34] | .17 [.02, .33] | 1.00 | |
| SDQ-CP | .10 [.00, .20] | .24 [.13, .36] | .11 [.01, .20] | 1.00 |

## Correlation

Spearman's rank correlations assessed agreement between measures based on continuous scores, while taking the skewed distributions into account. Although the A-TAC-CD and A-TAC-ODD assessed mutually exclusive psychiatric disorders, their correlations were weak (.32; father report on girls) to moderate (.45; father report on boys) with a mean of .40. Previous research found an association of .48 between the A-TAC-CD and the CBCL-AGG and of .32 between the A-TAC-ODD and CBCL-AGG. In the present study, rank correlations between the A-TAC-CD and CBCL-AGG were slightly lower, ranging from .34 (mother report on girls) to .47 (mother report on boys) with a mean of .41. Rank correlations between the A-TAC-ODD and CBCL-AGG were in contrast slightly higher than in previous research; they ranged from .55 (father report on girls) to .62 (mother report on boys) with a mean of .58. Although the A-TAC-CD and SDQ-CP were more similar by means of content compared to the A-TAC-ODD and the SDQ-CP, the correlations were very similar. For the A-TAC-CD and SDQ-CP, correlations ranged from .30 (father report on girls) to .56 (mother report on boys) with a mean of .46; for the A-TAC-ODD and SDQ-CP, correlations ranged from .33 (father report on girls) to .56 (mother report on boys), also with a mean of .46. Previous research found strong to very strong correlations between the CBCL-AGG

and SDQ-CP, however the rank correlations in the present study were lower. They ranged from .27 (father report on girls) to .65 (mother report on boys) with a mean of .55 (as can be seen in Table 3). Altogether, the rank correlations indicated moderate agreement between the measures of aggressive behavior.

**Table 3.** Spearman's rho with 95% confidence intervals.

| **Boys, mother report** | | | | |
|---|---|---|---|---|
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .43 [.38, .47] | 1.00 | | |
| CBCL-AGG | .47 [.42, .52] | .62 [.58, .66] | 1.00 | |
| SDQ-CP | .56 [.51, .60] | .56 [.52, .60] | .65 [.61, .68] | 1.00 |
| **Boys, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .45 [.39, .51] | 1.00 | | |
| CBCL-AGG | .42 [.35, .47] | .57 [.52, .62] | 1.00 | |
| SDQ-CP | .43 [.36, .50] | .44 [.38, .50] | .39 [.33, .45] | 1.00 |
| **Girls, mother report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .36 [.31, .41] | 1.00 | | |
| CBCL-AGG | .34 [.28, .40] | .56 [.51, .60] | 1.00 | |
| SDQ-CP | .41 [.35, .46] | .43 [.38, .48] | .53 [.49, .58] | 1.00 |
| **Girls, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .32 [.24, .39] | 1.00 | | |
| CBCL-AGG | .35 [.28, .41] | .55 [.50, .60] | 1.00 | |
| SDQ-CP | .30 [.21, .37] | .33 [.26, .40] | .27 [.20, .34] | 1.00 |



**Table 4.** Polychoric correlations with 95% confidence intervals.

| **Boys, mother report** | | | | |
|---|---|---|---|---|
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .59 [.45, .74] | 1.00 | | |
| CBCL-AGG | .44 [.23, .65] | .76 [.66, .86] | 1.00 | |
| SDQ-CP | .54 [.38, .71] | .63 [.52, .74] | .76 [.65, .86] | 1.00 |
| **Boys, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .61 [.40, .82] | 1.00 | | |
| CBCL-AGG | .68 [.46, .90] | .60 [.40, .81] | 1.00 | |
| SDQ-CP | .62 [.42, .81] | .57 [.42, .71] | .68 [.50, .83] | 1.00 |
| **Girls, mother report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .10 [-.30, .51] | 1.00 | | |
| CBCL-AGG | .21 [-.20, .63] | .77 [.65, .89] | 1.00 | |
| SDQ-CP | .50 [.21, .78] | .67 [.52, .83] | .65 [.47, .83] | 1.00 |
| **Girls, father report** | | | | |
| **Scale** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A-TAC-CD | 1.00 | | | |
| A-TAC-ODD | .79 [.60, .98] | 1.00 | | |
| CBCL-AGG | .48 [.03, .94] | .63 [.37, .88] | 1.00 | |
| SDQ-CP | .57 [.29, .85] | .58 [.40, .75] | .58 [.32, .83] | 1.00 |



Results from the polychoric correlations are displayed in Table 4. Polychoric correlations took into account the skewed distributions of the aggressive behavior measures by estimating the correlations on the underlying liability scale. Agreement between the A-TAC-CD and A-TAC-ODD ranged from .10 (mother report on girls) to .79 (father report on girls) with a mean of .64. For the A-TAC-CD with the CBCL-AGG, polychoric correlations ranged from .21 (mother report on girls) to .68 (father report on boys) with a mean of .53. Associations between the A-TAC-ODD and CBCL-AGG ranged between .60 (father report on boys) and .77 (mother report on girls) with a mean of .74. Between the A-TAC-CD and SDQ-CP associations ranged from .50 (father report on girls) to .62 (father report on boys) with a mean of .56. Between the A-TAC-ODD and SDQ-CP, polychoric correlations ranged from .57 (father report on boys) to .67 (mother report on girls) with a mean of .62. Finally, polychoric correlations between the CBCL-AGG and SDQ-CP ranged from .58 (father report on girls) to .76 (mother report on boys) with a mean of .71. Overall, polychoric correlations indicated moderate to strong agreement between measures of aggressive behavior.

## Genetic Analyses

Cross-twin cross-instrument correlations for MZ and DZ twins between the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ-CP for mother- and father-report are presented in Supplementary Table 4. Supplementary Table 5 contains the means and variances, and slopes for the sex differences on the means; for all measures, the mean score was higher for boys than for girls. The contribution of A for the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ was respectively .34, .42, .61, and .42; C explained, respectively, .04, .19, .13, and .09; E contributed, respectively, .62, .39, .26, and .50 to the variances. The covariance between the different measures was moderately to strongly accounted for by genetic factors, namely 50% (A-TAC-ODD and SDQ-CP) to 83% (A-TAC-CD and CBCL-AGG). Shared environmental factors explained up to 22% of the covariance (A-TAC-ODD and CBCL-AGG) between the different aggressive behavior measures. The nonshared environment weakly explained covariance between different measures, namely between 20% (A-TAC-CD and CBCL-AGG) and 32% (A-TAC-ODD and SDQ-CP; see Table 5). Genetic correlations are provided in Figure 1 and in Table 6. For mother-report, genetic correlations ranged from .65 (95% CI = .53 - .75; A-TAC-ODD and CBCL-AGG) to .84 (95% CI = .60 – 1.00; CBCL-AGG and SDQ-CP), with a mean of .75. These genetic correlations indicated a strong to very strong association of underlying genes between different measures of childhood aggressive behaviors. Correlations between the shared environmental influences on aggression scale scores for mother reports ranged from -.16 (95% CI = -1.00 – 1.00; A-TAC-CD with CBCL-AGG) to .90 (95% CI = .63 – 1.00; A-TAC-ODD with the CBCL-AGG), with a mean of .50. The large fluctuations in these correlations reflect the relatively small estimates of the shared environment (C) variance components; when the contribution of C to variance is small, its contribution to the covariance between measures is also small and estimates of correlations may become unstable. Correlations between the nonshared environmental influences on the scale scores were very weak to moderate; estimates varied between .19 (95% CI = .09 - .28; A-TAC-CD with A-TAC-ODD) and .55 (95% CI = .48 - .61; CBCL-AGG with SDQ-CP; see Table 6), the mean was .36.

For father-report, the variance decomposition for the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ-CP were as follows: A contributed, respectively, .40, .35, .45, and .31; C explained, respectively, .11, .25, .25, and .14; E contributed, respectively, .49, .40, .30, and .56 to the variances. Genetic factors explained between 25% (A-TAC-CD and A-TAC-ODD) and 58% (A-TAC-CD and SDQ-CP) of the covariance between the different measures. The shared environment explained 11% (A-TAC-CD and SDQ-CP) to 38% (A-TAC-CD and A-TAC-ODD) of the covariance among measures. Nonshared environmental factors influenced 23% (A-TAC-ODD and CBCL-AGG) to 37% (A-TAC-CD and A-TAC-ODD) of the covariance between measures of aggressive behaviors (see Table 5) Genetic correlations for father reports were weak to strong, ranging from .29 (95% CI = -.12 - .52; A-TAC-CD with A-TAC-ODD) to .76 (95% CI = .55 - .94; A-TAC-ODD with CBCL-AGG), with a mean of .55, indicating that all aggressive behavior measures assessed overlapping genetic factors. Correlations between the shared environment were weak to strong, with estimates between .35 (95% CI = -.57 - .94; A-TAC-CD with SDQ-CP) and 1.00 (95% CI = .73 – 1.00; CBCL-AGG with SDQ-CP), the mean was .79, which indicated overlap in shared environmental factors across aggressive behavior measures. Correlations between the nonshared environment indicated weak to moderate overlap between aggressive behavior measures, ranging from .24 (95% CI = .13 - .34; A-TAC-CD with SDQ-CP) to .43 (95% CI = .34 - .51; A-TAC-ODD with CBCL-AGG; see Table 6), the mean was .33. To check for bias, we also conducted categorical twin analyses; results are presented in Supplement 1 and Supplementary Tables 6 to 8. Generally, genetic correlations were of a similar strength or stronger compared to the continuous analyses.

**Table 5.** Standardized variance and covariance decomposition into contribution of genetic factors, shared environmental factors, and unshared environmental factors. Mother-reports in the top half and father-reports in the lower half.

| **Mothers** | | | | | |
|---|---|---|---|---|---|
| | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | .34 [.15, .46] | | | |
| | A-TAC-ODD | .74 [.40, 1.04] | .42 [.27, .58] | | |
| | CBCL-AGG | .83 [.56, 1.07] | .51 [.34, .68] | .61 [.49, .69] | |
| | SDQ-CP | .70 [.42, .93] | .50 [.28, .72] | .63 [.47, .77] | .42 [.26, .54] |
| C | A-TAC-CD | .04 [.00, .19] | | | |
| | A-TAC-ODD | .03 [-.19, .29] | .19 [.07, .32] | | |
| | CBCL-AGG | -.03 [-.19, .18] | .22 [.08, .37] | .13 [.05, .24] | |
| | SDQ-CP | .04 [-.10, .24] | .19 [.05, .36] | .07 [-.03, .21] | .09 [.01, .21] |
| E | A-TAC-CD | .62 [.55, .89] | | | |
| | A-TAC-ODD | .23 [.11, .36] | .39 [.34, .44] | | |
| | CBCL-AGG | .20 [.10, .30] | .27 [.22, .33] | .26 [.23, .30] | |
| | SDQ-CP | .26 [.16, .38] | .32 [.25, .40] | .30 [.25, .36] | .50 [. 45, .56] |
| **Fathers** | | | | | |
| | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | .40 [.22, .55] | | | |
| | A-TAC-ODD | .25 [-.07, .55] | .35 [.17, .53] | | |
| | CBCL-AGG | .42 [.11, .74] | .47 [.27, .69] | .45 [.29, .62] | |
| | SDQ-CP | .58 [.25, .89] | .50 [.16, .84] | .32 [-.01, .65] | .31 [.09, .50] |
| C | A-TAC-CD | .11 [.02, .24] | | | |
| | A-TAC-ODD | .38 [.14, .63] | .25 [.11, .40] | | |
| | CBCL-AGG | .30 [.05, .55] | .30 [.11, .48] | .25 [.11, .39] | |
| | SDQ-CP | .11 [-.10, .35] | .18 [-.09, .46] | .33 [.06, .61] | .14 [.01, .30] |
| E | A-TAC-CD | .49 [.42, .58] | | | |
| | A-TAC-ODD | .37 [.26, .40] | .40 [.34, .46] | | |
| | CBCL-AGG | .27 [.17, .40] | .23 [.17, .30] | .30 [.26, .35] | |
| | SDQ-CP | .31 [.16, .48] | .32 [.20, .33] | .35 [.24, .47] | .56 [.48, .65] |

**Table 6.** Correlations between underlying genetics. shared environmental factors, and unshared environmental factors are presented on the off-diagonal. Mother-reports in the top half and father-reports in the lower half.

| **Mothers** | | | | | |
|---|---|---|---|---|---|
| | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .76 [.49, .99] | 1.00 | | |
| | CBCL-AGG | .72 [.50, .97] | .65 [.53, .75] | 1.00 | |
| | SDQ-CP | .84 [.60, .99] | .65 [.47, .82] | .84 [.72, .97] | 1.00 |
| C | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .12 [-1.00, 1.00] | 1.00 | | |
| | CBCL-AGG | -.16 [-1.00, 1.00 | .90 [.63, 1.00] | 1.00 | |
| | SDQ-CP | .31 [-1.00, 1.00] | .78 [.13, 1.00] | .45 [-.42, .95] | 1.00 |
| E | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .19 [.09, .27] | 1.00 | | |
| | CBCL-AGG | .19 [.10, .28] | .55 [.48, .61] | 1.00 | |
| | SDQ-CP | .21 [.13, .30] | .40 [.32, .47] | .55 [.49, .61] | 1.00 |
| **Fathers** | | | | | |
| | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .29 [-.12, .52] | 1.00 | | |
| | CBCL-AGG | .40 [.14, .63] | .76 [.55, .94] | 1.00 | |
| | SDQ-CP | .67 [.37, .95] | .67 [.28, .91] | .38 [-.02, .68] | 1.00 |
| C | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | 1.00 [.73, 1.00] | 1.00 | | |
| | CBCL-AGG | .71 [.19, .96] | .75 [.44, .96] | 1.00 | |
| | SDQ-CP | .35 [-.57, .94] | .41 [-.44, .94] | .78 [.23, 1.00] | 1.00 |
| E | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .38 [.28, .47] | 1.00 | | |
| | CBCL-AGG | .29 [.18, .39] | .43 [.34, .51] | 1.00 | |
| | SDQ-CP | .24 [.13, .34] | .29 [.20, .39] | .37 [.27, .46] | 1.00 |

## DISCUSSION

We aimed to quantify the agreement among different measures of childhood aggressive behaviors. To this end, we compared the different measures on convergence of their item content, concordance at the recommended clinical cutoff, correlation among the scores at the different scales, and the extent to which they measure the same underlying genetic mechanisms.

Overlap in item content across aggressive behavior measures ranged from absent (i.e. mutually exclusive) to moderate. For conduct disorder (CD) as assessed by the A-TAC scale there was more overlap with DSM-IV criteria than for ODD, while the A-TAC ODD scale, CBCL aggressive behaviors, and the SDQ conduct problems overlapped more with the DSM-IV criteria for ODD. The absence of overlap between the A-TAC CD and ODD scales confirmed mutual exclusivity of these psychiatric disorders. Between the different measures, overlap was highest (i.e., moderate; .50) between the A-TAC CD scale and the SDQ conduct problems; the SDQ conduct problem scale had weak overlap with the other measures. (i.e., .12 - .19). As expected, the CBCL aggressive behavior scale weakly overlapped with all other measures (i.e., .12 to .29) with the strongest overlap for the A-TAC ODD scale. This indicated that based on their content, different measures of aggressive behavior cannot be used interchangeably.

Clinical concordance for the different aggressive behavior measures was very weak to weak. Although the CBCL and SDQ, in prior research, discriminated equally well between children from general population samples and clinical samples (Goodman & Scott, 1999; Klasen et al., 2000), their clinical concordance in the present study was weak (i.e., .22). In prior work (Gould et al., 1993), clinical concordance between the CBCL aggressive behavior scale and the A-TAC ODD scale was higher (i.e., .32; weak agreement) than between the CBCL aggressive behavior scale and the A-TAC CD scale (i.e., .10; very weak agreement). Similarly, clinical concordance was higher between the SDQ conduct problem scale and the A-TAC ODD scale (i.e., .29; weak agreement) than between the SDQ conduct problems scale and the A-TAC CD scale (i.e., .15; very weak agreement). Despite the mutual exclusivity of the A-TAC CD scale and the A-TAC ODD scale, there was very weak clinical concordance between these measures (i.e., .10). These findings revealed that the different aggressive behavior measures tend to classify different children as potential clinical cases. As a result, different aggressive behavior instruments may result in different clinical decisions with respect to inclusion, exclusion, referral, or treatment.

Rank correlations among continuous scores of the aggressive behavior measures were moderate, suggesting stronger agreement when considering the continuous scores compared to considering clinical cut-offs. The association between the A-TAC ODD scale and the CBCL aggressive behavior was the highest (i.e., .58, moderate agreement); the association between the A-TAC CD scale and the A-TAC ODD was the lowest (i.e., .40, moderate agreement), yet not absent. Thus, there is some overlap between the different aggressive behavior measures, but they also provide distinct information. Polychoric correlations, which estimate the correlation on an underlying scale of liability, revealed moderate to strong agreement between the different aggressive behavior measures. Strongest agreement was between the A-TAC ODD scale and the CBCL aggressive behavior scale, namely .74. Agreement between the A-TAC CD scale and the CBCL aggressive behavior was weakest, but still moderate, namely .53. These results reveal that agreement between the different measures of aggressive behavior based on continuous scores, while correcting for skewness, yield higher agreement than clinical cut-off scores.

Our results demonstrate that when a child is assessed for aggressive behaviors, it largely depends on the measure whether a diagnosis is given, whereas the measures converge moderately to strongly on who receives a higher score on aggressive

behavior. There are several other arguments in favor of a continuous or dimensional approach for the assessment of childhood aggressive behavior. For instance, children may vary across development in scores, fluctuating above and below clinical thresholds, which may cause them to not receive treatment although they might score above-threshold at another age (Biederman, Mick, Faraone, & Burback, 2001). In addition, similar to clinical aggressive behavior, subthreshold aggressive behavior is associated with adverse outcomes and high costs and therefore it is beneficial to detect heightened levels of aggressive behavior that not meet clinical levels (Fatori et al., 2018). An earlier diagnosis is associated with better outcomes later in life, which suggests additional benefits from detection of children with subthreshold levels of aggressive behavior (Campbell, Lundstrom, Larsson, Lichtenstein, & Lubke, 2018). Therefore, we propose that the use of continuous scores should be considered to assess childhood aggressive behavior, especially for the purpose of combining data with different measures of aggressive behavior.

Genetic correlations were strong to very strong (mean = .75) for mother reports and weak to strong (mean = .55) for father reports, indicating a substantial overlap in underlying genetic liability among the different aggressive behavior measures. A sensitivity check, where the genetic model was fitted to the underlying liability distributions, confirmed the genetic correlations among aggressive behavior measures to be moderate to high. Unlike observed correlations, such as the analyses of clinical concordance and correlation in the present study, genetic correlations are not influenced by measurement error. Therefore, the high genetic correlation may suggest that if we were to account for measurement error, the constructs that underlie the different measures of aggressive behavior we evaluate are highly consistent. These findings suggest that the measures of aggressive behavior that we examined, can readily be used in research relying on multiple instruments to assess the same construct, especially for genetic analyses.

### Strengths and Limitations

One of the strengths of this study was the availability of a unique sample in which multiple aggressive behavior measures were administered. Because all participants were twins, the correlations among measures could be decomposed into parts explained by genetic and environmental factors. Additionally, we considered the possible bias induced by the skewness of the aggressive behavior variables by carrying out sensitivity analyses.

The results of our analyses also come with some limitations. For instance, the order of the items was the same for all participants (Brace, 2008). Parents may interpret questions in light of prior questions, which may cause them to structurally respond more positively or negatively in the beginning of the questionnaire compared to the end of the questionnaire. Prior work found covariances among sets of items (e.g., measures) to vary when assessed in different orders (Weinberger, Darkes, Del Boca, Greenbaum, & Goldman, 2006). Changing order, however, might induce random error. Nonetheless, analyses in the present study may under- or overestimate agreement between aggressive behavior measures due to the same order of items for all participants. An additional limitation is that we only collected data for a single age-group and cannot make inferences on all of childhood. Because aggressive behavior expresses itself differently at different ages, similarity between aggressive behavior measures may vary across development. Nonetheless, stability in the underlying genetic mechanisms of aggressive behaviors (Porsch et al., 2016; Wichers et al., 2013) suggests that the genetic correlations between the different aggressive behavior measures are likely to remain similar across development.

## CONCLUSION

We have used a variety of definitions of similarity to compare several measures of aggressive behavior. Based on what definition one prefers, conclusions as to whether the instruments measure the same construct differ. If we, for example, consider item content, one may conclude the overlap to be very limited, whereas our genetic analyses point to shared etiology for the constructs measured by the different instruments. Whether researchers consider the agreement between instruments as satisfactorily, depends on their application. It is highly recommended to consider multiple metrics of similarity to decide whether differ measures tap into the same underlying construct. By leveraging a genetically informative design, multiple raters, and several commonly used instruments, we attempted to provide a holistic perspective on the nuances involved in the measurement of aggressive behavior in childhood.

# SUPPLEMENT 1

## Description sensitivity analysis

A shortcoming of the continuous genetic analyses was that the data violated the assumption of twin models of multivariate normality because of skewness, possibly leading to biased results (Derks et al., 2004). One way to correct for skewness is by categorizing the data into ordinal data. For this reason, we performed analyses with the different aggressive behavior measures divided into three categories; 0, 0.5/1 (i.e., 0.5 for A-TAC-CD and A-TAC-ODD, and 1 for CBCL-AGG and SDQ-CP), and higher (frequencies of the categories can be found in Supplementary Table 3). Because the analyses were highly computationally demanding, we performed them using Weighted Least Squares (WLS), in which the data input consisted of a variance-covariance matrix instead of raw data. This analysis did not allow for covariates, therefore, the model did not include sex differences on the means (as were included in the continuous model). As in the continuous analyses, we ran the ACE model, separately for mothers and fathers.

The matrix containing the within pair correlations for MZ and DZ twins between the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ-CP for mother-report is presented in Supplementary Table 6. The contribution of A for the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ was respectively .53, .31, .41, and .52; C explained, respectively, .20, .42, .46, and .11; E contributed, respectively, .27, .27, .13, and .37 to the variances (Supplementary Table 7). The multivariate ACE model revealed that for mother-report, correlations between the genetic effects on aggression measures (i.e., genetic correlations) ranged from .49 (A-TAC-ODD and CBCL-AGG) to .97 (A-TACCD with SDQ-CP), with a mean of .80. These genetic correlations indicated a strong to very strong association of underlying genes between different measures of childhood aggressive behaviors. Correlations between the shared environmental influences on aggression scale scores for mother reports ranged from .43 (A-TAC-CD with CBCL-AGG) to .96 (A-TAC-ODD with the SDQ-CP), with a mean of .74. These correlations indicated a moderate to very strong positive association, suggesting that the same shared environmental factors influence the aggressive behavior measures. Correlations between the nonshared environmental influences on the scale scores were very weak to moderate; estimates varied between .20 (A-TAC-CD with A-TAC-ODD) and .60 (CBCL-AGG with SDQ-CP; see Supplementary Table 8), the mean was .42.

For father-report MZ and DZ twin correlations can be seen in Supplementary Table 5. The variance decomposition for the A-TAC-CD, A-TAC-ODD, CBCL-AGG, and SDQ-CP were as following: A contributed, respectively, .49, .37, .40, and .48; C explained, respectively, .23, .37, .48, and .15; E contributed, respectively, .29, .26, .13, and .37 to the variances (see Supplementary Table 7). Genetic correlations for father reports were moderate to very strong, ranging from .53 (A-TAC-ODD with SDQ-CP) to .81 (A-TAC-CD with SDQ-CP), with a mean of .67, indicating that all aggressive behavior measures assessed overlapping genetic factors. Correlations between the shared environment were very strong negative to strong, with estimates between -.91 (CBCL-AGG with SDQ-CP) and .72 (A-TAC-CD with CBCL-AGG), the mean was -.04, which indicated that overlap in shared environmental factors across aggressive behavior measures strongly varied, possibly due to the sample size. Correlations between the nonshared environment indicated weak negative to moderate overlap between aggressive behavior measures, ranging from -.38 (CBCL-AGG with SDQ-CP) to .46 (A-TAC-CD with CBCL-AGG; see Supplementary Table 8), the mean was .21.

**Supplementary Table 1.** Item content

| Key word | A-TAC-ODD | A-TAC-CD | CBCL-AGG | SDQ-CP | DSM-IV-CD | DSM-IV-ODD |
|---|---|---|---|---|---|---|
| Argues | Does he/she often argue with adults? | | Argues a lot | | | Often argues with adults |
| Breaking in | | | | | Has broken into someone else's house, building, or car | |
| Cruelty | | Has he/she ever deliberately been physically cruel to anybody? | Cruelty, bullying, or meanness to others | | Has been physically cruel to people; Has been physically cruel to animals | |
| Demands attention | | | Demands a lot of attention | | | |
| Destroys | | | Destroys his/her own things; Destroys things belonging to his/her family or others | | Has deliberately destroyed others' property (other than by fire setting) | |
| Disobedient | | | Disobedient at home; Disobedient at school | Generally obedient, usually does what adults request | | Often actively defies or refuses to comply with adults' requests of rules |
| Fights | | Does he/she often get into fights? | Gets in many fights; Physically attacks people | Often fights with other children or bullies them | Often initiates physical fights | |
| Fire setting | | | | | Has deliberately engaged in fire setting with the intention of causing serious damage | |
| Forced sexual activity | | | | | Has forced someone into sexual activity | |
| Irritability | Is he/she easily offended, or disturbed by others?; Is he/she easily teased? | | Stubborn, sullen, or irritable | | | Often touchy or easily annoyed ny others |
| Lying | | Does he/she often lie or cheat? | | Often lies or cheats | Often lies to obtain goods or favors or to avoid obligations (e.g., "cons" others) | |

**Supplementary Table 1.** Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Mood changes | | | Sudden changes in mood or feelings | | | |
| Resentful | | | | | | Often angry and resentful |
| Run away | | | | | Has run away from home overnight at least twice while living in parental or parental surrogate home (or once without returning home for a lengthy period) | |
| Screams | | | Screams a lot; Unusually loud | | | |
| Spiteful | | | | | | Often spiteful or vindictive |
| Stays out at night | | | | | Often stays out at night despite parental prohibitions, beginning before age 13 years | |
| Stealing | | Has he/she ever engaged in shoplifting?; Does he/she steal things at home or outside home? | | Steals from home, school or elsewhere | Has stolen items of nontrivial value without confronting a victim (e.g., shoplifting, but without breaking and entering; forgery) | |
| Stealing with victim | | | | | Has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, or armed robbery) | |
| Sulks | | | Sulks a lot | | | |
| Suspicious | | | Suspicious | | | |

**Supplementary Table 1.** Continued

| | | | | | |
|---|---|---|---|---|---|
| Teases | Does he/she often tease others by deliberately doing things that are perceived as provocative? | Teases a lot | Often deliberately annoys people | | |
| | | | Often loses temper | | |
| Temper | Has there ever been a time when he/she would be angry to the extent that he/she cannot be reached? | Temper tantrums or hot temper | Often has temper tantrums or hot tempers | | |
| Threaten | | Threatens people | Often bullies, threatens, or intimidates others | | |
| Truancy | | | Is often truant from school, beginning before age 13 | | |
| Weapon use | | | Has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun) | | |
| Total number of items | 5 | 18 | 7 | 15 | 5 |
| Number of symptoms | 4 | 14 | 7 | 14 | 5 |



**Supplementary Table 2**. Prevalences

| Rater | Measure | Boys | Girls |
|---|---|---|---|
| Mothers | A-TAC CD | 3% | 1% |
| | A-TAC ODD | 10% | 5% |
| | CBCL AGG | 4% | 2% |
| | SDQ CP | 7% | 3% |
| Fathers | A-TAC CD | 2% | 1% |
| | A-TAC ODD | 6% | 4% |
| | CBCL AGG | 2% | 1% |
| | SDQ CP | 13% | 9% |

**Supplementary table 3.** Frequencies of categorical responses

| Scale | Response | Count | % | Scale | Response | Count | % |
|---|---|---|---|---|---|---|---|
| A-TAC CD mother | | | | A-TAC CD father | | | |
| | 0 | 1754 | 78.8 | | 0 | 1232 | 81.8 |
| | 0.5 | 251 | 11.3 | | 0.5 | 170 | 11.3 |
| | > 0.5 | 220 | 9.9 | | > 0.5 | 105 | 7.0 |
| | Total | 2225 | 100 | | Total | 1507 | 100 |
| A-TAC ODD mother | | | | A-TAC ODD father | | | |
| | 0 | 645 | 26.9 | | 0 | 520 | 32.3 |
| | 0.5 | 499 | 20.8 | | 0.5 | 346 | 21.5 |
| | > 0.5 | 1254 | 52.3 | | > 0.5 | 742 | 46.1 |
| | Total | 2398 | 100 | | Total | 1608 | 100 |
| CBCL AGG mother | | | | CBCL AGG father | | | |
| | 0 | 719 | 30.6 | | 0 | 566 | 35.3 |
| | 1 | 369 | 15.7 | | 1 | 265 | 16.5 |
| | > 1 | 1265 | 53.8 | | > 1 | 774 | 48.2 |
| | Total | 2353 | 100 | | Total | 1605 | 100 |
| SDQ CP mother | | | | SDQ CP father | | | |
| | 0 | 1252 | 52.1 | | 0 | 66 | 4.1 |
| | 1 | 522 | 21.7 | | 1 | 113 | 7.0 |
| | > 1 | 631 | 26.2 | | > 1 | 1434 | 88.9 |
| | Total | 2405 | 100 | | Total | 1613 | 100 |

**Supplementary Table 4**. MZ and DZ twin correlations of the saturated model for mother report on the top half and for father report on the lower half with 95% confidence intervals

| **Mothers** | | | | |
|---|---|---|---|---|
| **MZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .39 [.30, .48] | | | |
| A-TAC-ODD twin 1 | .29 [.19, .37] | .62 [.54, .69] | | |
| CBCL-AGG twin 1 | .37 [.28, .45] | .52 [.44, .60] | .75 [.68, .82] | |
| SDQ-CP twin 1 | .37 [.29, .46] | .45 [.37, .53] | .52 [.44, .59] | .52 [.44, .60] |
| **DZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .20 [.13, .27] | | | |
| A-TAC-ODD twin 1 | .16 [.09, .22] | .40 [.33, .47] | | |
| CBCL-AGG twin 1 | .13 [.06, .19] | .28 [.21, .35] | .42 [.35, .49] | |
| SDQ-CP twin 1 | .14 [.07, .20] | .23 [.16, .31] | .24 [.17, .31] | .27 [.20, .35] |
| **Fathers** | | | | |
| **MZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .54 [.43, .64] | | | |
| A-TAC-ODD twin 1 | .28 [.18, .38] | .61 [.52, .70] | | |
| CBCL-AGG twin 1 | .28 [.17, .40] | .51 [.40, .61] | .71 [.62, .80] | |
| SDQ-CP twin 1 | .28 [.17, .38] | .34 [.24, .45] | .30 [.21, .40] | .45 [.35, .56] |
| **DZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .25 [.16, .34] | | | |
| A-TAC-ODD twin 1 | .23 [.15, .31] | .42 [.34, .51] | | |
| CBCL-AGG twin 1 | .20 [.12, .28] | .36 [.27, .44] | .47 [.39, .56] | |
| SDQ-CP twin 1 | .13 [.05, .21] | .19 [.11, .28] | .17 [.07, .26] | .29 [.20, .37] |

**Supplementary Table 5**. Means, variances, and betas from the saturated model reported by mothers in the top half and reported by fathers in the lower half.

| **Mothers** | | | |
|---|---|---|---|
| **Measure** | **Mean** | **Variance** | **Beta** |
| A-TAC-CD | 0.43 | 0.27 | -0.15 |
| A-TAC-ODD | 1.54 | 1.01 | -0.34 |
| CBCL-AGG | 4.87 | 17.38 | -1.04 |
| SDQ-CP | 1.45 | 1.69 | -0.34 |
| **Fathers** | | | |
| **Measure** | **Mean** | **Variance** | **Beta** |
| A-TAC-CD | 0.29 | 0.23 | -0.08 |
| A-TAC-ODD | 1.09 | 0.89 | -0.14 |
| CBCL-AGG | 3.70 | 13.50 | -0.65 |
| SDQ-CP | 2.59 | 1.05 | -0.18 |



**Supplementary Table 6**. MZ and DZ correlations from the saturated categorical twin models, reported by mothers on the top half and by fathers on the lower half, with 95% confidence intervals

| **Mothers** | | | | |
|---|---|---|---|---|
| **MZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .72 [.63, .82] | | | |
| A-TAC-ODD twin 1 | .52 [.40, .64] | .71 [.63, .78] | | |
| CBCL-AGG twin 1 | .62 [.51, .74] | .46 [.36, .56] | .88 [.84, .91] | |
| SDQ-CP twin 1 | .65 [.55, .75] | .47 [.37, .57] | .53 [.43, .62] | .64 [.56, .72] |
| **DZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .46 [.33, .59] | | | |
| A-TAC-ODD twin 1 | .36 [.25, .47] | .56 [.48, .63] | | |
| CBCL-AGG twin 1 | .30 [.17, .42] | .38 [.29, .48] | .66 [.59, .73] | |
| SDQ-CP twin 1 | .26 [.15, .38] | .33 [.24, .43] | .30 [.20, .40] | .33 [.24, .42] |
| **Fathers** | | | | |
| **MZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .70 [.56, .85] | | | |
| A-TAC-ODD twin 1 | .50 [.35, .66] | .70 [.60, .80] | | |
| CBCL-AGG twin 1 | .56 [.38, .75] | .50 [.35, .64] | .85 [.79, .92] | |
| SDQ-CP twin 1 | .28 [-.06, .62] | .04 [-.23, .30] | .02 [-.23, .28] | .65 [.43, .87] |
| **DZ twins** | **A-TAC-CD twin 2** | **A-TAC-ODD twin 2** | **CBCL-AGG twin 2** | **SDQ-CP twin 2** |
| A-TAC-CD twin 1 | .48 [.29, .68] | | | |
| A-TAC-ODD twin 1 | .36 [.20, .52] | .55 [.44, .66] | | |
| CBCL-AGG twin 1 | .42 [.26, .59] | .39 [.25, .52] | .65 [.54, .75] | |
| SDQ-CP twin 1 | .14 [-.13, .41] | -.04 [-.23, .15] | -.09 [-.29, .11] | .36 [.14, .57] |

**Supplementary table 7**. Results from the genetic modeling on categorical data for mother-reports in the top half and father-reports in the lower half. Standardized variance and covariance decomposition into contribution of genetic factors, shared environmental factors, and unshared environmental factors

| **Mothers** | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | .53 | | | |
| | A-TAC-ODD | .56 | .31 | | |
| | CBCL-AGG | .66 | .29 | .41 | |
| | SDQ-CP | .73 | .48 | .52 | .52 |
| C | A-TAC-CD | .20 | | | |
| | A-TAC-ODD | .34 | .42 | | |
| | CBCL-AGG | .22 | .53 | .46 | |
| | SDQ-CP | .09 | .36 | .27 | .11 |
| E | A-TAC-CD | .27 | | | |
| | A-TAC-ODD | .09 | .27 | | |
| | CBCL-AGG | .13 | .18 | .13 | |
| | SDQ-CP | .18 | .16 | .20 | .27 |
| **Fathers** | | | | | |
| **Parameter** | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | .49 | | | |
| | A-TAC-ODD | .50 | .37 | | |
| | CBCL-AGG | .43 | .37 | .40 | |
| | SDQ-CP | 1.31 | .99 | 32.13 | .48 |
| C | A-TAC-CD | .23 | | | |
| | A-TAC-ODD | .32 | .37 | | |
| | CBCL-AGG | .42 | .52 | .48 | |
| | SDQ-CP | -.44 | -.41 | -23.43 | .15 |
| E | A-TAC-CD | .29 | | | |
| | A-TAC-ODD | .18 | .26 | | |
| | CBCL-AGG | .15 | .11 | .13 | |
| | SDQ-CP | .13 | .42 | -7.70 | .37 |



**Supplementary table 8**. Results from the genetic modeling on categorical data for mother-reports on the top half and for father-reports on the lower half. Correlations between underlying genetics, shared environmental factors, and unshared environmental factors are on the off-diagonal.

| **Mothers** | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .78 | 1.00 | | |
| | CBCL-AGG | .84 | .49 | 1.00 | |
| | SDQ-CP | .97 | .69 | .73 | 1.00 |
| C | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .66 | 1.00 | | |
| | CBCL-AGG | .43 | .74 | 1.00 | |
| | SDQ-CP | .44 | .96 | .78 | 1.00 |
| E | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .20 | 1.00 | | |
| | CBCL-AGG | .41 | .59 | 1.00 | |
| | SDQ-CP | .40 | .28 | .60 | 1.00 |
| **Fathers** | | | | | |
| **Parameter** | **Measure** | **A-TAC-CD** | **A-TAC-ODD** | **CBCL-AGG** | **SDQ-CP** |
| A | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .70 | 1.00 | | |
| | CBCL-AGG | .56 | .57 | 1.00 | |
| | SDQ-CP | .81 | .53 | .80 | 1.00 |
| C | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .65 | 1.00 | | |
| | CBCL-AGG | .72 | .71 | 1.00 | |
| | SDQ-CP | -.70 | -.39 | -.91 | 1.00 |
| E | A-TAC-CD | 1.00 | | | |
| | A-TAC-ODD | .38 | 1.00 | | |
| | CBCL-AGG | .46 | .34 | 1.00 | |
| | SDQ-CP | .12 | .30 | -.37 | 1.00 |

Chapter 7

# Summary and Discussion

The aim of this dissertation was to further our understanding of risk factors associated with childhood aggression and the assessment of childhood aggression. To this end, Chapter 2 comprised an overview of meta-analyses and systematic reviews on treatment effectiveness and its moderators for childhood aggression. In addition, Chapter 3 to 5 examined macro-level and micro-level predictors of childhood and adolescent aggression and moderation on the contribution of genetic and environmental factors to individual differences in childhood aggression. Finally, Chapter 6 tested the agreement between different instruments commonly used to assess aggressive behavior. Table 1 briefly describes the aims and highlights of each chapter. The next paragraphs provide a more elaborate summary of each chapter.

The goal of Chapter 2 was to enhance our understanding of treatment effectiveness for childhood aggression. Therefore, Chapter 2 presented a literature synthesis of 72 meta-analyses and systematic reviews that examined effectiveness of treatments for childhood aggression. The study reviewed the characteristics of the meta-analyses and systematic reviews, effect sizes across types of treatments, and effects of various moderators (i.e., participant variables, treatment variables, and methodological variables). Treatments included psychosocial (non-pharmacological) universal prevention, selective prevention, indicated prevention, and intervention. The conclusion was that for universal and selective prevention, effects were mostly absent or small; for indicated prevention and intervention, effects were mostly small to medium. Furthermore, most moderators of treatment effectiveness had no effect in the majority of studies (i.e., child age, child gender, implementation to individuals or groups, person implementing the treatment, different treatment programs, and session related factors or treatment intensity) or mixed effects (i.e., socioeconomic status, type of treatment, informant, research quality). The only two significant moderators comprised of pre-treatment levels of aggression and parental involvement. Treatment effectiveness was higher for children with higher levels of aggression before treatment and when parents were involved in the treatment.

The discussion elaborated on two patterns that emerged within the results and on the implications of those patterns for research and clinical practice. First, the results identified similarities between universal and selective prevention compared to indicated prevention and intervention, respectively. Second, results revealed that based on existing research it is not yet possible to distinguish subgroups of children that would benefit more from treatment for aggression than others. The positive moderating effect of parental involvement on treatment effectiveness for childhood aggression suggests that an opportunity for future research may be to focus more on parental influences as possible moderators of treatment effectiveness. In addition, more systematic research attention for the association between individual factors and treatment effectiveness for childhood aggression would be promising.



**Table 1.** Research aim and highlights of each chapter.

| Chapter | Research aim | Highlights |
|---|---|---|
| 2 | Create an overview of overall treatment effectiveness and its moderators for childhood aggression. | Effect sizes for treatments for childhood aggression were mostly small.<br>Promising distinction between treating aggression vs. treating associated factors.<br>Treatment might benefit from a stronger emphasis on individual differences. |
| 3 | Examine the association between national-level policies for child and adolescent mental health (CAMH) and adolescent mental health. | The association between policies for CAMH and adolescent aggressive behaviors was negative; aggressive behaviors were higher in countries with less policies.<br>This association held when controlling for other national-level variables.<br>There was no association between policies for CAMH and adolescent life satisfaction or psychosomatic symptoms. |
| 4 | Predict childhood aggression based on a large sample with a broad set of predictor variables. | Regression coefficients were in line with previous research, yet weaker, probably due to simultaneous inclusion.<br>Most important predictors were externalizing, non-aggressive behaviors such as arguing, being easily distracted, and hyperactivity.<br>These behaviors may function as salient targets for early detection and prevention of childhood aggression. |
| 5 | Investigate the moderating effect of socioeconomic status (SES) on the genetic architecture of childhood aggressive behavior. | SES moderated the contribution of genetic and environmental factors to childhood aggressive behavior.<br>Heritability was higher, the contribution of the shared environment was lower, and the contribution of the nonshared environment was higher for children from high SES families compared to children from low or medium SES families.<br>This pattern was similar in the Netherlands and the United Kingdom. |
| 6 | Assess the agreement between different measures of childhood aggressive behavior. | Convergence in item content was low.<br>Concordance between diagnoses was low.<br>Correlations between measures were moderate to high.<br>Genetic overlap was moderate to high.<br>The extent to which different measures of childhood aggressive behavior converge depends on the type (i.e., item content, clinical concordance, correlation, genetic overlap) of agreement considered |

To examine the extent to which national-level variables explain variance in aggression, Chapter 3 assessed the association between national-level policies for child and adolescent mental health (CAMH) and individual-level adolescent mental health. Data were from 172,829 adolescents aged eleven to fifteen years, from 30 European countries in the 2013/14 Health Behaviour in School-aged Children (HBSC) study. Adolescent mental health indicators comprised aggressive behavior, life satisfaction, and psychosomatic symptoms. Information on national-level policies for CAMH was gathered from renowned statistical institutes and included availability of epidemiological data, the number of CAMH facilities, investment in family benefits, and investment in education. In addition, to ascertain that the association between

CAMH policies and adolescent mental health was not overestimated, analyses controlled for national-level adult violence, adult well-being, and income inequality. Multi-level analyses revealed that adolescent aggressive behavior was lower in countries with more CAMH policies, even when taking other national-level variables into account. There was no association between CAMH policies and adolescent life satisfaction and psychosomatic symptoms, respectively. More research is needed to understand how and why policies for child and adolescent mental health associate with adolescent mental health and might be deployed for better adolescent mental health.

Chapter 4 focused on identifying more proximal predictors for childhood aggression. In this chapter, data were analyzed from the Child and Adolescent Twins Study in Sweden (CATSS) and the Netherlands Twin Register (NTR; combined sample size was 62,227 children) to find a model to predict childhood overt/physical aggression with a large set of predictor variables using a novel methodological approach. Overt/physical aggression, as assessed around age 9, was psychometrically harmonized across multiple European cohorts including CATSS and NTR. The large set of predictor variables encompassed demographics, prenatal characteristics, physical development, parental education level, life events, and mother-reported behavioral symptoms. To avoid capitalization of chance, data were partitioned in four parts for the different analysis steps. These included 1) exploratory data analysis and tuning meta-parameters for the data mining, 2) fitting increasingly complex data mining models to assess which predictors had which types of effects (i.e., linear, nonlinear, interaction), 3) assessment of model performance and importance of the predictor variables, and 4) fitting a confirmatory prediction model of aggression that integrated the results of the data mining analyses. The resulting multi-group model accounted for interactions with sex and cohort and confirmed linear main effects of variables measuring behavioral symptoms (e.g., related to non-physical aggression, attention-deficit/hyperactivity disorder, and conduct disorder), maternal smoking during pregnancy, parenting, and proportion of life events. The most important predictors comprised behavioral symptoms such as arguing, being easily distracted, and hyperactivity. Findings were in line with previous research (e.g., Burke, Pardini, & Loeber, 2008; Malanchini et al., 2018; Mcknight, Huebner, & Suldo, 2002; Piotrowska, Stride, Croft, & Rowe, 2015; Racz & McMahon, 2011), yet weaker, likely due to the simultaneous analyses of many predictors. These easily observable predictive behaviors may act as targets for early detection and prevention of childhood aggression.

Chapter 5 aimed to examine whether the contribution of genetic and environmental factors to individual differences in childhood aggression varied in different environmental circumstances. To this end, Chapter 5 tested the moderating effect of socioeconomic status (SES) on the genetic architecture of childhood aggressive behavior in 7-year-old children. Data were from the Netherlands Twin Register (N = 24,112) and from the Twins Early Development Study (N = 19,644) from the United Kingdom. Results revealed that SES moderated the contribution of genetic and environmental factors. For the standardized variance components, the contribution of genetic factors was higher, the contribution of the shared environment was lower, and the contribution of the nonshared environment was higher for children from a high SES background compared to children from a low or medium SES background. The unstandardized variance components revealed that the contribution of genetic factors was similar across SES strata, the contribution of shared environmental factors was lower, and the contribution of nonshared environmental factors was higher for children from a high SES background compared to children from a low or medium SES background. This pattern was similar for children from the Netherlands and the United Kingdom. The total variance was higher for low and medium SES, compared to high SES, indicating that children from low and medium SES were more heterogeneous in their levels of childhood aggressive behavior. Further work is required to examine whether these findings replicate in other countries, because the moderating effect of SES on the contribution of genetic and environmental factors to childhood aggression may vary as a result of differences in income inequality or a countries' investment in children and families. Moreover, aggressive behavior was assessed differently in the samples. It is an important issue for future research and collaboration projects to examine the extent to which the use of different measures has implications for outcomes.

Therefore, the purpose of Chapter 6 was to assess the extent to which heterogeneity in measures of aggressive behavior has implications for studies that combine data from research groups. To this end, Chapter 6 examined the level of agreement of different measures childhood aggressive behavior on item content, clinical concordance, correlation, and the degree to which they measure a common genetic construct. The sample consisted of 1,254 twin pairs aged 8 to 10 years from the Netherlands Twin Register. Mothers and fathers filled in multiple aggressive behavior measures for these children. These measures included the Autism - tics, attention-deficit hyperactivity disorder, and other comorbidities (A-TAC; Hansson et al., 2005), the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). Overall, the findings revealed that agreement between different measures of childhood aggressive behaviors depended on the metric of agreement under consideration (i.e., item content, clinical concordance, correlation, underlying genetics). For instance, the overlap between the item content of the aggressive behavior measures was absent to moderate. Concordance on who received a score above the clinical cut-off was very weak to weak. Associations between the different measures of aggressive behavior based on continuous scores, while correcting for skewness,

yielded higher agreement (i.e., moderate to strong) than clinical cut-off scores. Genetic correlations ranged from weak to very strong, which generally indicated high overlap in underlying genetics between the different measures of aggressive behavior. Unlike observed correlations, such as the analyses of clinical concordance and correlation in the present study, genetic correlations were not influenced by measurement error. Therefore, the high genetic correlations may suggest that if we were to account for measurement error, the constructs that underlie the different measures of aggressive behavior in the present study were highly consistent.

## Implications for Treatment, Prediction, and Assessment

This section discusses implications with regards to improvement of development and implementation of treatment and prevention programs for childhood aggression. In addition, this section will translate the outcomes of this dissertation into implications with regards to the assessment and prediction of childhood aggression.

Chapter 2 concluded that general treatment effects on childhood aggression were weak. Nevertheless, the treatments may have positively affected children or their families on other aspects than aggressive symptoms. Many of the treatments, especially the prevention programs, focused on skills and risk factors associated with childhood aggression such as social and emotional learning, academic performance, or classroom management for teachers (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Oliver, Wehby, & Reschly, 2011; Park-Higgerson, Perumean-Chaney, Bartolucci, Grimley, & Singh, 2008). In addition, found treatment effects were larger for social and emotional learning and academic performance than for childhood aggression (e.g., Durlak et al., 2011). Because Chapter 2 did not consider such outcomes of treatment effectiveness for childhood aggression, the effectiveness of treatments for childhood aggression might be more promising than our findings suggest. Childhood aggression is associated with many adversities and other psychosocial problems (Bartels et al., 2018), burden for parents, and high financial costs for society (Meltzer, Ford, Goodman, & Vostanis, 2011; Rivenbark et al., 2018; Roberts, McCrory, Joffe, de Lima, & Viding, 2017). Although treatments may not have been effective in reducing aggression, they may have been effective for comorbid problems, indirectly preventing the aggravation of childhood aggression. Future studies examining the direct and indirect effects of treatments for childhood aggression would be promising.

In addition, Chapter 2 revealed that a significant moderator of treatment effectiveness for childhood aggression was parental involvement. Parents are important because they shape the environment in which their children grow up. If parents suffer from emotional or behavioral problems, this may disrupt their parenting behaviors (Belsky, Hsieh, & Crnic, 1998; Berg-Nielsen, Vikan, & Dahl, 2002). Prior research suggests that treatments that solely focus on the parents' emotional and behavioral health, in which children are not involved, positively affect their children (Hudziak & Ivanova, 2016; Ivanova, Dewey, Swift, Weinberger, & Hudziak, 2019). Parental psychopathology is associated with higher child psychopathology (Wesseldijk, Dieleman, van Steensel, Bleijenberg, et al., 2018). Moreover, children with behavioral problems and parents with a psychiatric disorder have worse longitudinal outcomes than children with behavioral problems with parents that do not have a psychiatric disorder (Roetman et al., 2019). Conversely, parents of children with behavior problems are at higher risk for a psychiatric disorder (Wesseldijk, Dieleman, van Steensel, Bartels, et al., 2018). Additionally, because of the strong contribution of genetic factors to individual differences in aggression (Burt, 2009; Tuvblad & Baker, 2011; Waltes, Chiocchetti, & Freitag, 2016), it is not unlikely that parents of children with (symptoms of) aggression show aggression-related symptoms themselves (Frick et al., 1992). Altogether, these findings support the importance of parental involvement in treatment for childhood aggression. This suggests that it would be beneficial to screen for parental psychiatric disorders for the treatment of childhood behavior problems (Roetman et al., 2019; Wesseldijk, Dieleman, van Steensel, Bleijenberg, et al., 2018)**.** A better understanding of direct and indirect treatment effects and the contribution of parental influences to treatment effectiveness for childhood aggression would be promising and could be informative for policy making.

Chapter 3 revealed an association between national-level policies for child and adolescent mental and adolescent mental health. For further advancement of our understanding of this association, two approaches might be beneficial.

First, longitudinal research which monitors the implementation of policies in countries and levels of adolescent mental health before and after the implementation of a specific CAMH policy may potentially reveal information about the direction of effect (e.g., better adolescent mental health as a result of more policies for child and adolescent mental health or vice versa). Longitudinal research also allows assessment of changes over time in the association between policies for child and adolescent mental health and adolescent mental health (e.g., some policies might become increasingly effective over time, or policies may lose their impact over time). To illustrate, it may take time before an increase in child and adolescent mental health services affects child and adolescent mental health, because it may take time for services to accommodate to the needs of children and adolescents and overcome barriers related to availability, accessibility, acceptability, and equity (Tylee, Haller, Graham, Churchill, & Sanci, 2007).

Second, research on more intermediate geographical levels, such as provinces or smaller regions, in addition to our country level analyses, might reveal a stronger association between policies for child and adolescent mental health and adolescent mental health. Indeed, the implementation of policies for child and mental health

may vary across regions (Braddick, Carral, Jenkins, & Jané-Llopis, 2009; Signorini et al., 2017). For example, in Slovenia the number of facilities for child and adolescent mental health increased in the past years, but there are still regions without child and adolescent mental health services (Kumperscak, 2019). This implies that information at the national-level is not fully representative. Moreover, only 3.1% of variance in adolescent aggressive behavior was explained by country differences, which indicates more heterogeneity in adolescent aggressive behaviors within countries than between countries. Therefore, a more fine-grained examination of the association between policies for child and adolescent mental health and adolescent mental is needed to better understand this association.

The importance of behavioral symptoms for the prediction of childhood aggression reported in Chapter 4 confirms the high co-occurrence of childhood aggression with other behavioral and emotional problems reported in existing research (Bartels et al., 2018; Harvey, Breaux, & Lugo-Candelas, 2016; Marshall, Arnold, Rolon-Arroyo, & Griffith, 2015). These findings need to be interpreted taking two issues into account. First, many behavioral symptoms are found to genetically overlap with childhood aggression (Dick, Viken, Kaprio, Pulkkinen, & Rose, 2005; Latvala, Kuja-Halkola, Almqvist, Larsson, & Lichtenstein, 2015). It is likely that due to genetic overlap, associations between childhood aggression and predictor variables in Chapter 4 were overestimated and that controlling for this overlap would yield weaker associations. Second, the literature reveals important associations between childhood aggression and family factors that were not available for analysis for Chapter 4. For instance, growing up with parental harsh control, psychological control, or neglectful parenting is consistently related with childhood aggression (Larsson, Viding, Rijsdijk, & Plomin, 2008; Oliver, 2015; Pinquart, 2017). Exposure to interparental violence also contributes to childhood aggression (Buehler et al., 1997; Ehrensaft & Cohen, 2012), as does exposure to parental psychopathology (Goodman et al., 2011). Factors outside the family also associate with aggression, such as neighborhood disadvantage (Burt, Klump, Gorman-Smith, & Neiderhiser, 2016; Leventhal & Brooks-Gunn, 2000), and national-level policies for child and adolescent mental, as demonstrated in Chapter 3. To paint a more complete picture of childhood aggression, future research requires to pay more attention to risk factors at different levels of context.

The results from Chapter 5 indicated that the shared environment contributes more strongly to individual differences in childhood aggression in children from low and medium SES backgrounds compared to children from high SES backgrounds. This finding suggests that children from low or medium SES backgrounds would benefit more from treatment to improve shared environmental factors. Examples of such shared environmental factors are healthy family functioning, less parental stress, housing quality, or positive school attachment (Brumley & Jaffee, 2016; Burt, Klahr, Neale, & Klump, 2013; Hudziak & Ivanova, 2016; Klahr & Burt, 2014). To some extent, policies for child and adolescent mental health such as investment in family benefits and investment in education as examined in Chapter 3 attempt to already do so through alleviation of factors associated with low SES (Piotrowska et al., 2015; Reiss, 2013). The results from Chapter 5 suggest that such early prevention measures on the shared environment indeed may benefit children at risk for heightened levels of aggression. Furthermore, due to the large genetic component in childhood aggression, there might be intergenerational transmission of aggression from parents to children, both through genetic factors and environmental factors (D'Onofrio et al., 2007) such as heightened levels of interparental violence (Ehrensaft & Cohen, 2012). The findings of Chapter 5 underline the importance of parental involvement in treatments for childhood aggression as found in Chapter 2, and suggest that this would be most beneficial for children from low and medium SES backgrounds.

Chapter 6 revealed that the level of agreement between measures of aggressive behavior depends on the type of agreement under scrutiny (i.e., item content, clinical concordance, correlation, and genetic overlap). Genetic correlations between the measures of aggressive behavior were strong to very strong. This suggests that the same genetic factors influence the different measures, despite the differences in the purpose, construct of interest, and item content. Therefore, the findings indicate that different measures of aggressive behavior can readily be combined in future collaboration studies on the genetics of childhood aggressive behavior.

In addition, Chapter 6 revealed that agreement between measures of childhood aggressive behaviors was stronger for continuous scores than for agreement on clinical levels. What this implies for future collaboration research is that the different measures of aggression cannot be used interchangeably when making decisions based on a clinical cut-off score, rather, decision-making based on continuous scores may improve reliability. Continuous scores may be more robust across measures, and perhaps also across development. It is not uncommon that children differ in their expression of childhood aggression with age (i.e., heterotypic continuity; Bolhuis et al., 2017; Lubke, Mcartor, Boomsma, & Bartels, 2017), which implies that children's aggression scores may fluctuate above and below the clinical threshold in assessment. As demonstrated in Chapter 4, children with heightened, yet not necessarily clinical, levels of aggression likely display other behavioral symptoms of disorders strongly associated with childhood aggression. Prevention or intervention would possibly diminish the likelihood that children with subclinical levels of aggression would develop clinical aggression or a disorder strongly associated with aggression and, consequently, mitigate the adverse outcomes associated with later detection (Campbell, Lundstrom, Larsson, Lichtenstein, & Lubke, 2018). In addition, Chapter 2 showed that children with subthreshold levels of aggression are as likely

to benefit from treatment as children with above-threshold levels of aggression. Altogether, these findings suggest that a dimensional approach to the assessment of childhood aggression would be promising, especially when combining data that use different measures.

The research in this dissertation, especially in Chapters 3, 4, 5, and 6, focused on the more overt and physical types of aggression. Nonetheless, childhood aggression may take other forms, such as indirect or relational aggression. These types of aggression do not occur in direct confrontation, as is the case for fighting or bullying, but comprise behaviors such as manipulation or social exclusion (Björkqvist, Lagerspetz, & Kaukiainen, 1992; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003). Overt/physical aggression may differ in etiology from indirect or relational aggression, as suggested by the partial genetic correlations (i.e., .54 for boys and .43 for boys; Ligthart, Bartels, Hoekstra, Hudziak, & Boomsma, 2005). This suggests that it is uncertain whether the results from this dissertation would apply to other types of aggression. Future research that, for example, examines whether the predictors for childhood aggression found in Chapter 4, such as arguing, being easily distracted, and hyperactivity, also predict indirect or relational aggression would reveal whether or not the same targets may be useful for early detection and prevention for different types of aggression.

The results from Chapter 3 to Chapter 6 need to be interpreted as cross-sectional. The conclusions are not causal, however, future longitudinal research may reveal the direction of effect, such as whether the predictors in Chapter 3 and Chapter 4 are a cause or an effect of higher levels of aggression. In addition, the chapters did not allow for an examination of developmental trajectories (e.g., which risk factors best predict the onset and development of childhood aggression at which age). Longitudinal research is needed to uncover whether predictors such as parental education level, maternal smoking during pregnancy, or hyperactivity as found in Chapter 4 are equally predictive across development. A reason to suspect differential predictor effects across age is that child development is marked by different developmental stages with their accompanying landmarks for development, such as the formation of attachment around the 6- to 12-month period or development of the ability to inhibit aggressive outbursts, which develops in children aged 4 to 7 years old (Bakermans-kranenburg, Ijzendoorn, & Juffer, 2003; Wachs, Georgieff, Cusick, & Mcewen, 2014). These different stages suggest that children vary in sensitivity for risk factors across development. To illustrate, a meta-analysis revealed that the concurrent association between parental emotion socialization behaviors and childhood aggression decreases with age (Johnson, Hawes, Eisenberg, Kohlhoff, & Dudeney, 2017). In addition, prior work revealed that the contribution of genetic and environmental factors to childhood aggression varies across age; the shared environment explains around 44% of individual differences in childhood, but this influence disappears in adolescence (Porsch et al., 2016; Wesseldijk et al., 2017). It would be useful to examine at what age childhood aggression can be best predicted by which risk factors to optimally detect children most likely to become aggressive early enough to prevent worse outcomes from later diagnosis and treatment (e.g., Campbell, Lundstrom, Larsson, Lichtenstein, & Lubke, 2018).



## GENERAL CONCLUSION

Within ACTION, the goal was to inform the development of prevention and treatment strategies. The wealth of available data and expertise within ACTION permitted examination of new research questions. The studies in this dissertation highlighted the complexities in the etiology of childhood aggression. Childhood aggression is found to be associated with a broad range of factors, from country-level policies to more proximal factors as the family environment, and individual level factors such as behavior and genetics, which may also interact. Assessing whether children score above or below a clinical cut-off for inclusion for treatment may lead to children to be excluded from treatment from which they would benefit as much as children who score above a clinical cut-off. Moreover, measures of aggressive behavior agree only to a small extent on which children display clinical levels of aggression, which may cause children to miss out on treatment not because of their level of aggression, but because of the measure selected to assess their aggression.

Although the influence of the broad range of factors discussed in this dissertation on childhood aggression adds complexity to the etiology of childhood aggression, it also provides opportunities to improve prevention and intervention strategies for childhood aggression. For example, inclusion of parental characteristics (i.e., a family based approach) in diagnosis and treatment might improve treatment effectiveness for childhood aggression. Additionally, more policies for child and adolescent mental health were associated with lower levels of aggressive behaviors, which suggests merit in employment of policies as early prevention efforts. Moreover, differences in etiology of aggression as a result of socioeconomic background highlight that it is promising to distinguish subgroups of children more likely to develop childhood aggression and children more likely to benefit from treatment. The research in this dissertation contributes to previous work to advance our understanding of treatment, prediction, and assessment of childhood aggression and provided directions for future research working towards a more personalized approach to childhood aggression.

# Nederlandse samenvatting

Agressief gedrag bij kinderen wordt onderzocht als een probleem op zichzelf of als symptoom van psychiatrische diagnoses zoals oppositioneel-opstandige gedragsproblemen of normoverschrijdende gedragsproblemen. In deze dissertatie ligt de focus voornamelijk op openlijke (dus niet heimelijke) en fysieke uitingen van agressie bij kinderen. Voorbeelden zijn vechten of ongehoorzaamheid. Agressie bij kinderen gaat gepaard met negatieve uitkomsten zoals lasten voor de ouders en financiële kosten voor de maatschappij. Om agressie bij kinderen beter te begrijpen heeft het onderzoek in deze dissertatie zich gericht op behandeling, voorspelling en meting van agressie bij kinderen.

Hoewel er veel onderzoek gedaan is naar behandeling voor agressie bij kinderen, zijn de effecten over het algemeen klein. Daarnaast suggereert de genetische stabiliteit van agressie dat kinderen die agressief gedrag vertonen dat ook zullen blijven doen en dat later ingrijpen geassocieerd is met meer nadelige uitkomsten op volwassen leeftijd. Daarom streefden we met Hoofdstuk 2 om beter inzicht te krijgen in voor wie behandeling beter aanslaat en onder welke omstandigheden.

Voor vroege preventie zou het ook zinvol zijn om beter te kunnen voorspellen welke kinderen een verhoogde kans hebben om agressief gedrag te vertonen. Zowel genetische- als omgevingsfactoren spelen een belangrijke rol bij agressie in kinderen. Het onderzoek in deze dissertatie heeft gekeken naar biologische factoren op het niveau van het individu, maar ook op meer distale niveaus zoals kenmerken van de familie en landsvariabelen. Daarnaast is er gekeken of de bijdrage van genetische factoren en omgevingsfactoren verschilt tussen kinderen die opgroeien in verschillende achtergronden om de complexiteit van agressie bij kinderen in acht te nemen.

Verschillende hoofdstukken in deze dissertatie combineerden data van meerdere onderzoeksgroepen, die verschillende instrumenten gebruikten om agressie bij kinderen te meten. Het is nog niet vastgesteld in hoeverre heterogeniteit in instrumenten om agressie bij kinderen te meten invloed heeft op vergelijkbaarheid van resultaten en hoe hier het beste mee omgegaan kan worden. Daarom heeft onderzoek in deze dissertatie gekeken naar overeenstemming tussen verschillende meetinstrumenten om agressie bij kinderen mee te beoordelen.

Het doel van Hoofdstuk 2 was om een overzicht te creëren van de effectiviteit van behandelingen voor agressie bij kinderen en om in kaart te brengen of er moderatoren (variabelen die de effectiviteit beïnvloeden) zijn die structureel samenhangen met verhoogde effectiviteit van behandelingen. Hiervoor hebben we 72 systematische reviews en meta-analyses verzameld die de effectiviteit van behandelingen voor agressie bij kinderen en mogelijke moderatoren bespraken. Behandelingen omvatten in dit hoofdstuk universele preventie, selectieve preventie, geïndiceerde preventie en interventie. Universele preventie is ontwikkeld voor alle kinderen. Selectieve preventie is speciaal voor kinderen die een verhoogd risico hebben om agressief gedrag te vertonen. Geïndiceerde preventie is voor kinderen met een verhoogde agressie score, maar niet boven een diagnostische cut-off. Interventie is behandeling voor kinderen die een score hebben op agressief gedrag boven een diagnostische cut-off. We beoordeelden de kenmerken van de systematische reviews en meta-analyses, effectgroottes van de effectiviteit van behandelingen voor agressie bij kinderen en effecten van verschillende moderatoren (participant kenmerken, behandelingskenmerken en methodologische kenmerken). Voor universele preventie en selectieve preventie waren effecten voornamelijk afwezig of klein. Voor geïndiceerde preventie en interventie waren effecten voornamelijk klein of matig. De meeste moderatoren hadden in de meerderheid van de studies geen invloed op de effectiviteit van behandelingen voor agressie bij kinderen. Deze moderatoren waren leeftijd van het kind, sekse van het kind, of de behandeling geïmplementeerd werd bij individuen of groepen, wie de behandeling implementeerde, behandelprogramma's, en kenmerken met betrekking tot het aantal behandelingen of de intensiteit van behandelingen. Er waren gemende resultaten voor de volgende moderatoren: sociaaleconomische status, type behandeling, informant en kwaliteit van de onderzoeken. De twee moderatoren die in een meerderheid van studies samenhingen met verhoogde effectiviteit van behandelingen waren een hogere agressiescore bij het kind voorafgaand aan de behandeling en betrokkenheid van de ouders bij de behandeling. Alles samengenomen laten de uitkomsten van de resultaten twee patronen zien. Ten eerste lijken er overeenkomsten te zijn tussen universele preventie en selectieve preventie en tussen geïndiceerde preventie en interventie. Ten tweede suggereerden de resultaten dat op basis van bestaand onderzoek het nog net mogelijk is om onderscheid te maken tussen subgroepen van kinderen die meer baat zouden hebben bij behandeling voor agressie. Op basis van het positieve effect van betrokkenheid van ouders bij behandelingen voor agressie bij kinderen stellen wij dat verder onderzoek naar het modererende effect van oudereigenschappen op de effectiviteit van behandelingen voor agressie bij kinderen veelbelovend kan zijn om meer inzicht te krijgen in welke kinderen baat zouden hebben bij behandeling voor agressie.

Het doel van Hoofdstuk 3 was om te onderzoeken of er een associatie is tussen de hoeveelheid getroffen beleidsmaatregelen in een land welke specifiek gericht zijn op het verbeteren van de mentale gezondheid van kinderen en adolescenten en hoe hoog adolescenten in een land gemiddeld scoren op verschillende indicatoren van mentale gezondheid. Er was informatie beschikbaar van 172,829 adolescenten met een leeftijd tussen 11 en 15 jaar uit 30 Europese landen. De geselecteerde indicatoren van mentale gezondheid bij adolescenten waren agressief gedrag,

tevredenheid met leven en psychosomatische symptomen. Informatie over beleid omtrent de mentale gezondheid van adolescenten en kinderen was afkomstig van gevestigde statistische instituten en betrof informatie over de aanwezigheid van epidemiologische data, het aantal faciliteiten specifiek voor de mentale gezondheid van kinderen en adolescenten, investeringen in gezinsbijslagen, en investeringen in onderwijs. Om overschatting van de associatie tussen beleid voor de mentale gezondheid voor kinderen en adolescenten en indicatoren van de mentale gezondheid van adolescenten te voorkomen hebben we in de analyses ook gecorrigeerd voor andere landsvariabelen, waaronder hoe hoog landen scoren op geweldplegingen door volwassenen, welbevinden van volwassenen en inkomensongelijkheid. Analyses lieten zien dat agressief gedrag van adolescenten lager was in landen met meer beleid voor de mentale gezondheid van kinderen en adolescenten, ook als er gecorrigeerd werd voor de andere landsvariabelen. Deze associatie werd niet gevonden voor tevredenheid met leven en psychosomatische symptomen bij adolescenten. Er is meer onderzoek nodig om te begrijpen waarom en op welke manier beleid voor de mentale gezondheid van kinderen en adolescenten samenhangt met de mentale gezondheid van adolescenten en hoe het ingezet zou kunnen worden om de mentale gezondheid van adolescenten te verbeteren.

Na de focus op distale factoren in Hoofdstuk 3, richt Hoofdstuk 4 zich op meer proximale voorspellers van agressie bij kinderen. Het doel van Hoofdstuk 4 was om op basis van een grote dataset (62,227 kinderen) met tweelingen uit Nederland en Zweden te zoeken naar sterke voorspellers van fysieke agressie bij kinderen. Omdat de data afkomstig waren uit meerdere datasets was de agressiescore geharmoniseerd zodat deze vergelijkbaar was tussen de verschillende sets. Voorspellers bestonden uit demografische kenmerken, prenatale kenmerken, fysieke ontwikkeling, familie omgeving, oudergedrag, opleidingsniveau van de ouders, levensgebeurtenissen en gedragsmatige symptomen. Door het grote aantal kinderen in de data was het mogelijk om de data op te delen in vier onafhankelijke datasets en geavanceerde methodes toe te passen op deze onafhankelijke datasets. Stap 1 bestond uit exploratieve data analyses en het afstemmen van de metaparameters voor data mining. Stap 2 bestond uit het draaien van steeds complexere data mining modellen om te testen of de voorspellers lineaire, nonlineaire of interactie-effecten hadden of agressie bij kinderen. Stap 3 bestond uit het beoordelen van hoe goed de modellen agressie bij kinderen voorspellen en het verkrijgen van het relatieve belang van alle predictoren. Stap 4 bestond uit een bevestigend model waarin alle resultaten van data mining geïntegreerd waren om agressie bij kinderen te voorspellen. De resultaten lieten zien dat sekse en land interacties hadden met sommige voorspellers, dus daarom werd het uiteindelijke model in Stap 4 apart gedraaid voor jongens en meisjes en voor Nederland en Zweden. De meest belangrijke voorspellers van agressie bij kinderen waren externaliserende gedragingen zoals tegenspreken, snel afgeleid zijn en hyperactiviteit. De voorspeller effecten waren vergelijkbaar met



resultaten van eerder onderzoek, maar zwakker. Dat kan verklaard worden door het feit dat in deze studie meer variabelen tegelijkertijd beschouwd werden. De voorspellers uit deze studie kunnen dienen als makkelijk waarneembare symptomen om te identificeren welke kinderen mogelijk later agressief gedrag zullen vertonen en baat kunnen hebben bij vroege preventie.

Het doel van Hoofdstuk 5 was om te onderzoeken of de bijdrage van genetische factoren en omgevingsfactoren aan individuele verschillen in agressie bij kinderen verschillend is voor verschillende niveaus van sociaaleconomische status (SES). Om dit te testen hebben we gekeken naar het agressieve gedrag van zevenjarige tweelingen uit gezinnen met een lage, middelhoge of hoge SES uit Nederland (24,112 kinderen) en het Verenigd Koninkrijk (19,644 kinderen). Doordat we naar tweelingen keken was het mogelijk om te kijken in hoeverre individuele verschillen in agressie verklaard konden worden door genetische factoren, gedeelde omgevingsfactoren en unieke omgevingsfactoren. De analyses toonden aan dat de bijdrage van genetische factoren, de gedeelde omgeving en de unieke omgeving verschillend was tussen niveaus van SES, specifiek tussen lage en middelhoge SES vergeleken met hoge SES. De gestandaardiseerde resultaten lieten zien dat voor kinderen uit een hoge SES familie de bijdrage van genetische factoren hoger, de bijdrage van de gedeelde omgeving lager en de bijdrage van de unieke omgeving hoger was in verhouding tot kinderen uit lage of middelhoge SES families. De ongestandaardiseerde resultaten lieten zien dat de absolute bijdrage van genetische factoren gelijk was tussen SES niveaus, maar dat de bijdrage van de gedeelde omgeving lager en de bijdrage van de unieke omgeving hoger was voor kinderen uit families met een hoge SES vergeleken met kinderen uit families met een lage of middelhoge SES. Dit patroon was aanwezig zowel in kinderen uit Nederland als in kinderen uit het Verenigd Koninkrijk. Deze resultaten suggereren dat kinderen met een lage of middelhoge SES meer baat hebben bij een behandeling die zich richt op het verbeteren van de gedeelde omgeving dan kinderen met een hoge SES. Voorbeelden van gedeelde omgevingsfactoren om te verbeteren zijn het functioneren binnen een gezin, stress bij de ouders, kwaliteit van huisvesting en gehechtheid aan school.

Het doel van Hoofdstuk 6 was om te kijken in hoeverre verschillende instrumenten om agressie bij kinderen mee te meten met elkaar overeenstemmen. Om de mate van overeenstemming te meten hebben we gekeken naar in inhoud van items, klinische concordantie (scoren dezelfde kinderen boven een diagnostische cut-off), correlatie en genetische correlatie (hangt een genetische aanleg voor een verhoogde score volgens een instrument samen met een genetische aanleg voor een verhoogde score volgens een ander instrument). Hiervoor onderzochten we scores van vragenlijsten ingevuld door vaders en moeders van 1,254 tweelingparen met een leeftijd van acht tot tien jaar. De ouders hebben verschillende vragenlijsten ingevuld die agressie bij kinderen meten, waaronder de A-TAC (Autism - tics, attention-deficit hyperactivity

disorder, and other comorbidities), CBCL (Child Behavior Checklist) en de SDQ (Strengths and Difficulties Questionnaire). De resultaten lieten zien dat de mate van overeenstemming tussen instrumenten om agressie te meten afhankelijk is van welke vorm van overeenstemming er beschouwd werd. De overlap in item inhoud varieerde van geen overlap tot matige overlap. De mate waarin dezelfde kinderen boven een diagnostische cut-off scoorden was zeer zwak tot zwak. De correlaties tussen schalen waren matig tot sterk. De genetische correlaties varieerden van zwak tot zeer sterk, maar over het algemeen vonden wij een sterke overlap in de onderliggende genetische factors tussen de verschillende instrumenten om agressie bij kinderen te meten.. Deze resultaten duiden aan dat verschillende instrumenten om agressie bij kinderen te meten gecombineerd kunnen worden in toekomstige samenwerkingsprojecten om de genetische mechanismen te onderzoeken die samenhangen met agressie bij kinderen.

# REFERENTIES

References marked with an asterisk indicate studies included in the literature synthesis of Chapter 2.


Achenbach, T. M., Ivanova, M. Y., & Rescorla, L. A. (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. ***Comprehensive Psychiatry, 79***, 4–18. https://doi.org/10.1016/j.comppsych.2017.03.006

Achenbach, T. M., & Rescorla, L. (2001). ***Manual for the ASEBA School-Age Forms & Profiles***. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. ***Pediatrics in Review, 21***(8), 265–271. https://doi.org/10.1542/pir.21-8-265

Althoff, R. R., Verhulst, F. C., Rettew, D. C., Hudziak, J. J., & Van Der Ende, J. (2010). Adult outcomes of childhood dysregulation: A 14-year follow-up study. ***Journal of the American Academy of Child and Adolescent Psychiatry, 49***(11), 1105-1116.e1. https://doi.org/10.1016/j.jaac.2010.08.006

American Psychiatric Association. (1994). ***DSM-IV Diagnostic and Statistical Manual of Mental Disorder.*** American Psychiatric Organization (Fourth Edi, Vol. 33). Washington, D.C. https://doi.org/10.1073/pnas.0703993104

Anckarsäter, H., Lundström, S., Kollberg, L., Kerekes, N., Palm, C., Carlström, E., … Lichtenstein, P. (2011). The child and adolescent twin study in Sweden (CATSS). ***Twin Research and Human Genetics, 14***(6), 495–508. https://doi.org/10.1375/twin.14.6.495

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. ***Review of General Psychology, 8***(4), 291–322. https://doi.org/10.1037/1089-2680.8.4.291

Baker, K. (2009). Conduct disorders in children and adolescents. ***Paediatrics and Child Health, 19***(2), 73–78. https://doi.org/10.1016/j.paed.2008.10.008

Bakermans-Kranenburg, M. J., Ijzendoorn, M. H. Van, & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. ***Psychological Bulletin, 129***(2), 195–215. https://doi.org/10.1037/0033-2909.129.2.195

*Bakker, M. J., Greven, C. U., Buitelaar, J. K., & Glennon, J. C. (2017). Practitioner Review: Psychological treatments for children and adolescents with conduct disorder problems - a systematic review and meta-analysis. ***Journal of Child Psychology and Psychiatry, 58***(1), 4–18. https://doi.org/10.1111/jcpp.12590

*Barlow, J., & Stewart-Brown, S. (2000). Behavior problems and group-based parent education programs. ***Journal of Developmental and Behavioral Pediatrics, 21***(5), 356–370. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=11064964

*Barnes, T. N., Smith, S. W., & Miller, M. D. (2014). School-based cognitive-behavioral interventions in the treatment of aggression in the United States: A meta-analysis. ***Aggression and Violent Behavior, 19***(4), 311–321. https://doi.org/10.1016/j.avb.2014.04.013

Barry, M. M. (2009). Addressing the Determinants of Positive Mental Health: Concepts, Evidence and Practice. ***International Journal of Mental Health Promotion, 11***(3), 4–17. https://doi.org/10.1080/14623730.2009.9721788

Bartels, M., Cacioppo, J. T., Van Beijsterveldt, T. C. E. M., & Boomsma, D. I. (2013). Exploring the association between well-being and psychopathology in adolescents. ***Behavior Genetics, 43***(3), 177–190. https://doi.org/10.1007/s10519-013-9589-7

Bartels, M., Hendriks, A., Mauri, M., Krapohl, E., Whipp, A., Bolhuis, K., … Boomsma, D. I. (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project. ***European Child and Adolescent Psychiatry, 27***(9), 1105–1121. https://doi.org/10.1007/s00787-018-1169-1

*Battagliese, G., Caccetta, M., Luppino, O. I., Baglioni, C., Cardi, V., Mancini, F., & Buonanno, C. (2015). Cognitive-behavioral therapy for externalizing disorders: A meta-analysis of treatment effectiveness. ***Behaviour Research and Therapy, 75***, 60–71. https://doi.org/10.1016/j.brat.2015.10.008





Belsky, J., Hsieh, K. H., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: differential susceptibility to rearing experience? ***Development and Psychopathology, 10***(2), 301–319. https://doi.org/10.1017/S095457949800162X

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. ***Psychological Bulletin, 135***(6), 885–908. https://doi.org/10.1037/a0017376

*Bennett, D. S., & Gibbons, T. A. (2000). Efficacy of child cognitive-behavioral interventions for antisocial behavior: A meta-analysis. ***Child & Family Behavior Therapy, 22***(1), 1–15. https://doi.org/10.1300/J019v22n01_01

Berg-Nielsen, T. S., Vikan, A., & Dahl, A. A. (2002). Parenting related to child and parental psychopathology: A descriptive review of the literature. ***Clinical Child Psychology and Psychiatry, 7***(4), 529–552. https://doi.org/10.1177/1359104502007004006

Berntsson, L. T., Köhler, L., & Gustafsson, J. E. (2001). Psychosomatic complaints in schoolchildren: a Nordic comparison. ***Scandinavian Journal of Public Health, 29***, 44–54. https://doi.org/10.1177/14034948010290011001

Biederman, J., Mick, E., Faraone, S. V., & Burback, M. (2001). Patterns of remission and symptom decline in conduct disorder: A four-year prospective study of an ADHD sample. ***Journal of the American Academy of Child and Adolescent Psychiatry, 40***(3), 290–298. https://doi.org/10.1097/00004583-200103000-00008

Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. ***Annals of Statistics, 41***(3), 1111–1141. https://doi.org/10.1214/13-AOS1096

Bien, J., & Tibshirani, R. (2014). hierNet: A lasso for hierarchical interactions. R package version 1.7. Retrieved from https://cran.r-project.org/package=hierNet

Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. ***Aggressive Behavior, 18***, 117–127. https://doi.org/10.1002/1098-2337

Bodner, T. E. (2008). What improves with increased missing data imputations? ***Structural Equation Modeling, 15***(4), 651–675. https://doi.org/10.1080/10705510802339072

Bolhuis, K., Lubke, G. H., van der Ende, J., Bartels, M., van Beijsterveldt, C. E. M., Lichtenstein, P., … Tiemeier, H. (2017). Disentangling heterogeneity of childhood disruptive behavior problems Into dimensions and subgroups. ***Journal of the American Academy of Child and Adolescent Psychiatry, 56***(8), 678–686. https://doi.org/10.1016/j.jaac.2017.05.019

*Bond, C., Woods, K., Humphrey, N., Symes, W., & Green, L. (2013). Practitioner review: The effectiveness of solution focused brief therapy with children and families: A systematic and critical evaluation of the literature from 1990-2010. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 54***(7), 707–723. https://doi.org/10.1111/jcpp.12058

Bonthuis, M., Van Stralen, K. J., Verrina, E., Edefonti, A., Molchanova, E. A., Hokken-Koelega, A. C. S., … Jager, K. J. (2012). Use of national and international growth charts for studying height in european children: Development of up-to-date european height-for-age charts. ***PLoS ONE, 7***(8), 1–11. https://doi.org/10.1371/journal.pone.0042506

Boomsma, D. I. (2015). Aggression in Children: Unravelling the interplay of genes and environment through (epi)genetics and metabolomics. ***Journal of Paediatric and Neonatal Individualized Medicine, 4***(2).

Boomsma, D. I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. ***Nature Reviews Genetics, 3***(11), 872–882. https://doi.org/10.1038/nrg932

Bornstein, M. H., Putnick, D. L., Suwalsky, J. T. D., & Gini, M. (2006). Maternal chronological age, prenatal and perinatal history, social support, and parenting of infants. ***Child Development, 77***(4), 875–892. https://doi.org/https://doi.org/10.1111/j.1467-8624.2006.00908.x

Boyle, M. H., Offord, D. R., Racine, Y., Szatmari, P., Fleming, J. E., & Sanford, M. (1996). Identifying thresholds for classifying childhood psychiatric disorder: Issues and prospects. ***Journal of the American Academy of Child & Adolescent Psychiatry, 35***(11), 1440–1448. https://doi.org/10.1097/00004583-199611000-00012

Brace, I. (2008). ***Questionnaire design: How to plan, structure and write survey material for effective market research***. Applied Ergonomics (2nd ed., Vol. 6). London & Philadelphia: Kogan Page Limited. https://doi.org/10.1016/0003-6870(75)90299-9

Braddick, F., Carral, V., Jenkins, R., & Jané-Llopis, E. (2009). Child and adolescent mental health in Europe: Infrastructures, policy and programmes, 1–212. Luxembourg: European Communities.

*Bradley, M. C., & Mandell, D. (2005). Oppositional defiant disorder: A systematic review of evidence of intervention effectiveness. ***Journal of Experimental Criminology, 1***(3), 343–365. https://doi.org/10.1007/s11292-005-0062-3

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. ***Annual Review of Psychology, 53***, 371–399.

Breiman, L. (1984). ***Classification and regression trees.*** Belmont, California: Wadsworth International Group.

Brener, N. D., Collins, J. L., Kann, L., Warren, C. W., & Williams, B. I. (1995). Reliability of the Youth Risk Behavior Survey Questionnaire. ***American Journal of Epidemiology, 141***(6), 575–580.

*Briggs, H. E., Cox, W. H., Sharkey, C. N., Briggs, A. C., & Black, M. (2015). A review of the research on Pinkston's single-parent group training program. ***Research on Social Work Practice, 26***(1), 128–144. https://doi.org/10.1177/1049731515592033

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: a bioecological model. ***Psychological Review, 101***(4), 568–586. https://doi.org/10.1037/0033-295X.101.4.568

Brumley, L. D., & Jaffee, S. R. (2016). Defining and distinguishing promotive and protective effects for childhood externalizing psychopathology: A systematic review. ***Social Psychiatry and Psychiatric Epidemiology, 51***(6), 803–815. https://doi.org/10.1007/s00127-016-1228-1

*Buchanan-Pascall, S., Gray, K. M., Gordon, M., & Melvin, G. A. (2017). Systematic review and meta-analysis of parent group interventions for primary school children aged 4–12 years with externalizing and/or internalizing problems. ***Child Psychiatry and Human Development,*** 1–24. https://doi.org/10.1007/s10578-017-0745-9

Buehler, C., Anthony, C., Krishnakumar, A., Stone, G., Gerard, J., & Pemberton, S. (1997). Interparental conflict and youth problem behaviors: A meta-analysis. ***Journal of Child and Family Studies, 6***(2), 233–247.

*Bunge, E. L., Dickter, B., Jones, M. K., Alie, G., Spear, A., & Perales, R. (2016). Behavioral intervention technologies and psychotherapy with youth: A review of the literature. ***Current Psychiatry Reviews, 12***(1), 14–28. https://doi.org/10.2174/1573400511666150930232254

Burke, J. D., Pardini, D. A., & Loeber, R. (2008). Reciprocal relationships between parenting behavior and disruptive psychopathology from childhood through adolescence. ***Journal of Abnormal Child Psychology, 36***(5), 679–692. https://doi.org/10.1007/s10802-008-9219-7

Burt, S. A. (2009). Are there meaningful etiological differences within antisocial behavior? Results of a meta-analysis. ***Clinical Psychology Review, 29***(2), 163–178. https://doi.org/10.1016/j.cpr.2008.12.004

Burt, S. A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. ***Psychological Bulletin, 135***(4), 608–637. https://doi.org/10.1037/a0015702

Burt, S. A. (2013). Do etiological influences on aggression overlap with those on rule breaking? A meta-analysis. ***Psychological Medicine, 43***(9), 1801–1812. https://doi.org/10.1017/S0033291712001894

Burt, S. A., Klahr, A. M., Neale, M. C., & Klump, K. L. (2013). Maternal warmth and directiveness jointly moderate the etiology of childhood conduct problems. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 54***(10), 1030–1037. https://doi.org/10.1111/jcpp.12095

Burt, S. A., & Klump, K. L. (2014). Parent-child conflict as an etiological moderator of childhood conduct problems: an example of a "bioecological" gene-environment interaction. ***Psychological Medicine, 44***(5), 1065–1076. https://doi.org/10.1017/S0033291713001190

Campbell, I., Lundstrom, S., Larsson, H., Lichtenstein, P., & Lubke, G. (2018). The relation between the age at diagnosis of problem behaviors related to aggression and distal outcomes in Swedish children. ***European Child and Adolescent Psychiatry,*** 1–13. https://doi.org/10.1007/s00787-018-1250-9

*Candelaria, A. M., Fedewa, A. L., & Ahn, S. (2012). The effects of anger management on children's social and emotional outcomes: A meta-analysis. ***School Psychology International, 33***(6), 596–614. https://doi.org/10.1177/0143034312454360

Canino, G., Polanczyk, G., Bauermeister, J. J., Rohde, L. A., & Frick, P. J. (2010). Does the prevalence of CD and ODD vary across cultures? ***Social Psychiatry and Psychiatric Epidemiology, 45***(7), 695–704. https://doi.org/10.1007/s00127-010-0242-y



Cantril, H. (1965). ***The Pattern of Human Concerns.*** New Brunswick, NJ: Rutgers University Press.

Card, N., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta analytic review of gender differences, intercorrelations, and relations to maladjustments. ***Child Development, 79***(5), 1185–1229. https://doi.org/10.1111/j.1467-8624.2008.01184.x

Carral Bielsa, V., Braddick, F., Jané-Llopis, E., Jenkins, R., & Puras, D. (2010). Child and adolescent mental health policies, programmes and infrastructures across Europe. ***International Journal of Mental Health Promotion, 12***(4), 10–26. https://doi.org/http://dx.doi.org/10.1080/14623730.2010.9721822

Caspi, A., Houts, R. M., Belsky, D. W., Harrington, H., Hogan, S., Ramrakha, S., … Moffitt, T. E. (2016). Childhood forecasting of a small segment of the population with large economic burden. ***Nature Human Behaviour, 1***, 1–10. https://doi.org/10.1038/s41562-016-0005

Cavallo, F., Zambon, A., Borraccino, A., Raven-Sieberer, U., Torsheim, T., Lemma, P., … Kiaer, T. (2006). Girls growing through adolescence have a higher risk of poor health. ***Quality of Life Research, 15***(10), 1577–1585. https://doi.org/10.1007/s11136-006-0037-5

Centers for Disease Control and Prevention, Eaton, D. K., Kann, L., Kinchen, S., Ross, J., Hawkins, J., … Wechsler, H. (2006). Youth Risk Behavior Surveillance - United States, 2005. ***MMWR Surveillance Summaries, 55***(SS-5), 1–112. https://doi.org/10.1016/S0002-9343(97)89440-5

Chan, E., Fogler, J. M., & Hammerness, P. G. (2016). Treatment of attention-deficit/hyperactivity disorder in adolescents: A systematic review. ***JAMA - Journal of the American Medical Association, 315***(18), 1997–2008. https://doi.org/10.1001/jama.2016.5453

Choi, N. H., Li, W., & Zhu, J. (2010). Variable selection with the strong heredity constraint and its oracle property. ***Journal of the American Statistical Association, 105***(489), 354–364. https://doi.org/10.1198/jasa.2010.tm08281

*Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., … Starace, N. (2011). Evidence-based treatment of children and adolescents: An updated review of indicators of efficacy and effectiveness. ***Clinical Psychology: Science and Practice, 18***, 154–172.

*Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., Mcgee, C., … Morelli, P. (2002). Toward large-scale implementation of empirically supported treatments for children : A review and observations by the Hawaii Empirical Basis to Services Task Force. ***Psychological Science, 9***, 165–190. https://doi.org/10.1111/j.1468-2850.2002.tb00504.x

Cohn, M. A., Fredrickson, B. L., Brown, S. L., Mikels, J. A., & Conway, A. M. (2009). Happiness unpacked: Positive emotions increase life satisfaction by building resilience. ***Emotion, 9***(3), 361–368. https://doi.org/10.1037/a0015952

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., … Long, B. (1993). The science of prevention. ***American Psychologist, 48***(10), 1013–1022. https://doi.org/10.1037/0003-066X.48.10.1013

*Comer, J. S., Chow, C., Chan, P. T., Cooper-Vince, C., & Wilson, L. A. S. (2013). Psychosocial treatment efficacy for disruptive behavior problems in very young children: A meta-analytic examination. ***Journal of the American Academy of Child and Adolescent Psychiatry, 52***(1), 26–36. https://doi.org/10.1016/j.jaac.2012.10.001

Connell, A. M., & Goodman, S. H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: A meta-analysis. ***Psychological Bulletin, 128***(5), 746–770. https://doi.org/10.1037//0033-2909.128.5.746

*Connor, D. F., Carlson, G. A., Chang, K. D., Daniolos, P. T., Ferziger, R., Findling, R. L., … Steiner, H. (2006). Juvenile maladaptive aggression: A review of prevention, treatment, and service configuration and a proposed research agenda. ***Journal of Clinical Psychiatry, 67***(5), 808–820. https://doi.org/http://dx.doi.org/10.4088/JCP.v67n0516

Copeland, W. E., Wolke, D., Shanahan, L., & Costello, J. (2015). Adult functional outcomes of common childhood psychiatric problems: A prospective, longitudinal study. ***JAMA Psychiatry, 72***(9), 892–899. https://doi.org/10.1001/jamapsychiatry.2015.0730

Coppens, E., Vermet, I., Knaeps, J., De Clerck, M., De Schrijver, I., Matot, J. P., & Van Audenhove, C. (2015). Adolescent mental health care in Europe: state of the art, recommendations, and guidelines by the ADOCARE* network. Brussels.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. ***Developmental Psychology, 33***, 579–588.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. ***Psychological Methods, 14***(2), 81–100. https://doi.org/10.1037/a0015914

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. ***Developmental Psychology, 44***(2), 365–380. https://doi.org/10.1037/0012-1649.44.2.365

Currie, C., & Alemán-Díaz, A. Y. (2015). Building knowledge on adolescent health: Reflections on the contribution of the Health Behaviour in School-aged Children (HBSC) study. ***European Journal of Public Health, 25***, 4–6. https://doi.org/10.1093/eurpub/ckv017

Currie, C., Inchley, J., Molcho, M., Lenzi, M., Veselska, Z., Wild, F., & 1. (2014). Health Behaviour in School-aged Children (HBSC) Study Protocol: Background, Methodology and Mandatory items for the 2013/14 Survey. Retrieved from http://www.hbsc.org

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. ***Social Science and Medicine, 66***(6), 1429–1436. https://doi.org/10.1016/j.socscimed.2007.11.024

Currie, C., Zanotti, C., Morgan, A., Currie, D., de Looze, M., Roberts, C., … Barnekow, V. (2012). Social determinants of health and well-being among young people. Health Behaviour in School-aged Children (HBSC) study: International report from the 2009/2010 survey. World Health Organisation Health Policy for Children and Adolescents, No 6, (6), 1–272. https://doi.org/ISBN 987 92 890 1423 6

D'Onofrio, B. M., Slutske, W. S., Turkheimer, E., Emery, R. E., Harden, K. P., Heath, A. C., … Martin, N. G. (2007). Intergenerational transmission of childhood conduct problems: a Children of Twins Study. ***Archives of General Psychiatry, 64***(7), 820–829. https://doi.org/10.1001/archpsyc.64.7.820

*De Graaf, I., Speetjens, P., Smit, F., De Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple P Positive Parenting Program on behavioral problems in children: A meta-analysis. ***Behavior Modification, 32***, 714–735. https://doi.org/10.1111/j.1741-3729.2008.00522.x

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. ***Twin Research, 7***(6), 659–669. https://doi.org/10.1375/1369052042663832

Dick, D. M., Viken, R. J., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2005). Understanding the covariation among childhood externalizing symptoms: Genetic and environmental influences on conduct disorder, attention deficit hyperactivity Ddsorder, and oppositional defiant disorder symptoms. ***Journal of Abnormal Child Psychology, 33***(2), 219–229. https://doi.org/10.1007/s10802-005-1829-8

*Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., … Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct problems: A systematic review of randomised controlled trials. ***Child and Adolescent Psychiatry and Mental Health, 3***(1), 7. https://doi.org/10.1186/1753-2000-3-7

*Dretzke, J., Frew, E., Davenport, C., Barlow, J., Stewart-Brown, S., Sandercock, J., … Taylor, R. (2005). The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children. ***Health Technology Assessment, 9***(50), 1–250. https://doi.org/10.3310/hta9500

Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does Money Really Matter? Estimating Impacts of Family Income on Young Children's Achievement With Data From Random-Assignment Experiments. ***Developmental Psychology, 47***(5), 1263–1279. https://doi.org/10.1037/a0023875

*Durlak, J. A., & Weissberg, R. P. (2007). The impact of after-school programs that promote personal and social skills. ***Learning***, 1–47. https://doi.org/10.3102/0034654308325693

*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***(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

*Dymnicki, A. B., Weissberg, R. P., & Henry, D. B. (2011). Understanding how programs work to prevent overt aggressive behaviors: A meta-analysis of mediators of elementary school–based programs. ***Journal of School Violence, 10***(4), 315–337. https://doi.org/10.1080/15388220.2011.602599

Ehrensaft, M. K., & Cohen, P. (2012). Contribution of family violence to the intergenerational transmission of externalizing behavior. Prevention Science, 13(4), 370–383. https://doi.org/10.1007/s11121-011-0223-8



Elgar, F. J., Gariépy, G., Torsheim, T., & Currie, C. (2017). Early-life income inequality and adolescent health and well-being. ***Social Science & Medicine, 174***, 197–208. https://doi.org/10.1016/j.socscimed.2016.10.014

Elgar, F. J., Pförtner, T. K., Moor, I., De Clercq, B., Stevens, G. W. J. M., & Currie, C. (2015). Socioeconomic inequalities in adolescent health 2002-2010: A time-series analysis of 34 countries participating in the Health Behaviour in School-aged Children study. ***The Lancet, 385***(9982), 2088–2095. https://doi.org/10.1016/S0140-6736(14)61460-4

Emck, C., Bosscher, R., Beek, P., & Doreleijers, T. (2009). Gross motor performance and self-perceived motor competence in children with emotional, behavioural, and pervasive developmental disorders: A review. ***Developmental Medicine and Child Neurology, 51***(7), 501–517. https://doi.org/10.1111/j.1469-8749.2009.03337.x

*Epstein, R. A., Fonnesbeck, C., Potter, S., Rizzone, K. H., & McPheeters, M. (2015). Psychosocial interventions for child disruptive behaviors: A meta-analysis. ***Pediatrics, 136***(5), 947–960. https://doi.org/10.1542/peds.2015-2577

*Erford, B. T., Paul, L. E., Oncken, C., Kress, V. E., & Erford, M. R. (2014). Counseling outcomes for youth with oppositional behavior: A meta-analysis. ***Journal of Counseling and Development, 92***(1), 13–24. https://doi.org/10.1002/j.1556-6676.2014.00125.x

Erskine, H. E., Baxter, A. J., Patton, G., Moffitt, T. E., Patel, V., Whiteford, H. A., & Scott, J. G. (2017). The global coverage of prevalence data for mental disorders in children and adolescents. ***Epidemiology and Psychiatric Sciences, 26***(04), 395–402. https://doi.org/10.1017/S2045796015001158

Evans, S. E., Davies, C., & DiLillo, D. (2008). Exposure to domestic violence: A meta-analysis of child and adolescent outcomes. ***Aggression and Violent Behavior, 13***(2), 131–140. https://doi.org/10.1016/j.avb.2008.02.005

Eurostat. (2016). Public expenditure on education - Percent of GDP. Retrieved November 27, 2017, from http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc510&plugin=1

Eurostat. (2017). Eurostat Database: Recorded offences by offence category - police data. Retrieved July 18, 2017, from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=crim_off_cat&lang=en

Evans, G. (2004). The environment of childhood poverty. ***American Psychologist, 59***(2), 77–92. https://doi.org/10.1037/0003-066X.59.2.77

*Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. ***Journal of Clinical Child and Adolescent Psychology, 37***(1), 215–237. https://doi.org/10.1080/15374410701820117

Fanti, K. A. (2016). Understanding heterogeneity in conduct disorder: A review of psychophysiological studies. ***Neuroscience and Biobehavioral Reviews.*** https://doi.org/10.1016/j.neubiorev.2016.09.022

*Farahmand, F. K., Grant, K. E., Polo, A. J., & Duffy, S. N. (2011). School-based mental health and behavior programs for low-income, urban youth: A systematic and meta-analytic review. ***Clinical Psychology Science and Practice, 18,*** 372–390. https://doi.org/10.1111/j.1468-2850.2011.01265.x

Faraway, J. J. (2016). Does data splitting improve prediction? ***Statistics and Computing, 26***(1–2), 49–60. https://doi.org/10.1007/s11222-014-9522-9

*Farmer, E. M. Z., Compton, S. N., Bums, B. J., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Externalizing disorders. ***Journal of Consulting and Clinical Psychology, 70***(6), 1267–1302. https://doi.org/10.1037/0022-006X.70.6.1267

Fatori, D., Salum, G., Itria, A., Pan, P., Alvarenga, P., Rohde, L. A., … Graeff-Martins, A. S. (2018). The economic impact of subthreshold and clinical childhood mental disorders. ***Journal of Mental Health,*** 1–7. https://doi.org/10.1080/09638237.2018.1466041

Fedko, I. O., Wesseldijk, L. W., Nivard, M. G., Hottenga, J.-J., Van Beijsterveldt, C. E. M., Middeldorp, C. M., … Boomsma, D. I. (2016). Heritability of behavioral problems in 7-year olds based on shared and unique aspects of parental views. ***Behavior Genetics***. https://doi.org/10.1007/s10519-016-9823-1

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 46***(8), 837–849. https://doi.org/10.1111/j.1469-7610.2004.00387.x

Fergusson, D. M., & Woodward, L. J. (2002). Mental health, educational, and social role outcomes of adolescents with depression. ***Archives of General Psychiatry, 59***(3), 225. https://doi.org/10.1001/archpsyc.59.3.225

*Fossum, S., Handegård, B. H., Adolfsen, F., Vis, S. A., & Wynn, R. (2016). A meta-analysis of long-term outpatient treatment effects for children and adolescents with conduct problems. ***Journal of Child and Family Studies, 25***(1), 15–29. https://doi.org/10.1007/s10826-015-0221-8

*Fossum, S., Handegård, B. H., Martinussen, M., & Mørch, W. T. (2008). Psychosocial interventions for disruptive and aggressive behaviour in children and adolescents: a meta-analysis. ***European Child & Adolescent Psychiatry, 17***(7), 438–51. https://doi.org/10.1007/s00787-008-0686-8

Foster, E. M., & Jones, D. E. (2005). The high costs of aggression: Public expenditures resulting from conduct disorder. ***American Journal of Public Health, 95***(10), 1767–1772. https://doi.org/10.2105/AJPH.2004.061424

Fox, J. (2016). Polychoric and polyserial correlations: polycor. Version 0.7-9. Retrieved from http://cran.r-project.org/package=polycor%0D

*Franklin, C., Kim, J. S., Beretvas, T. S., Zhang, A., Guz, S., Park, S., … Maynard, B. R. (2017). The effectiveness of psychosocial interventions delivered by teachers in schools: A systematic review and meta-analysis. ***Clinical Child and Family Psychology Review, 20***(3), 333–350. https://doi.org/10.1007/s10567-017-0235-4

Frick, P. J. (2001). Effective interventions for children and adolescents with conduct disorder. ***The Canadian Journal of Psychiatry, 46***(7), 597–608.

Frick, P. J., & Dickens, C. (2006). Current perspectives on conduct disorder. ***Current Psychiatry Reports, 8***(1), 59–72. https://doi.org/10.1007/s11920-006-0082-3

Frick, P. J., Lahey, B. B., Loeber, R., Stouthamer-Loeber, M., Christ, M. A. G., & Hanson, K. (1992). Familial risk factors to conduct disorder and oppositional defiant disorder: Parental psychopathology and maternal parenting. ***Journal of Consulting and Clinical Psychology, 60***(1), 49–55.

Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Van Horn, Y., Christ, M. A. G., … Hanson, K. (1993). Oppositional defiant disorder and conduct disorder: A meta-analytic review of factor analyses and cross-validation in a clinic sample. ***Clinical Psychology Review, 13***(4), 319–340. https://doi.org/10.1016/0272-7358(93)90016-F

Fried, E. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. ***Journal of Affective Disorders, 208***, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. ***The Annals of Statistics, 29***(5), 1189–1232. https://doi.org/https://www.jstor.org/stable/2699986

Friedman, J. H., Hastie, T., Tibshirani, R., Simon, N., Narasimhan, B., & Qian, J. (2018). glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 2.0-16.

Fuller, D., Buote, R., & Stanley, K. (2017). A glossary for big data in population and public health: Discussion and commentary on terminology and research methods. ***Journal of Epidemiology and Community Health, 71***(11), 1113–1117. https://doi.org/10.1136/jech-2017-209608

*Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnely, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years (review). ***The Cochrane Database of Systematic Reviews,*** (2), 1–326. https://doi.org/10.1002/ebch.1904

Gallup Healthways Well-Being Index. (2014). 2014 Country Well-Being Rankings. Retrieved from http://info.healthways.com/hubfs/Well-Being_Index/2014_Data/Gallup-Healthways_State_of_Global_Well-Being_2014_Country_Rankings.pdf

*Gansle, K. A. (2005). The effectiveness of school-based anger interventions and programs: A meta-analysis. ***Journal of School Psychology, 43***(4), 321–341. https://doi.org/10.1016/j.jsp.2005.07.002

*Gavita, O., & Joyce, M. (2008). A review of the effectiveness of group cognitively enhanced behavioral based parent programs designed for reducing disruptive behavior in children. ***Journal of Cognitive and Behavioral Psychotherapies, 8***(2), 185–199.

Goodman, E., Huang, B., Schafer-Kalkhoff, T., & Adler, N. E. (2007). Perceived socioeconomic status: A new type of identity that influences adolescents' self-rated health. ***Journal of Adolescent Health, 41***(5), 479–487. https://doi.org/10.1016/j.jadohealth.2007.05.020

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. ***Journal of Child Psychology and Psychiatry, 38***(5), 581–586. https://doi.org/https://doi.org/10.1111/j.1469-7610.1997.tb01545.x



Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. ***Journal of the American Academy of Child and Adolescent Psychiatry, 40***(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? ***Journal of Abnormal Child Psychology, 27***(1), 17–24. https://doi.org/10.1023/A:1022658222914

Goodman, S. H., Rouse, M. H., Connell, A. M., Broth, M. R., Hall, C. M., & Heyward, D. (2011). Maternal depression and child psychopathology: A meta-analytic review. ***Clinical Child and Family Psychology Review, 14***(1), 1–27. https://doi.org/10.1007/s10567-010-0080-1

Gould, M. S., Bird, H., & Jaramillo, B. S. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. ***Journal of Abnormal Child Psychology, 21***(3), 287–313. https://doi.org/10.1007/BF00917536

Granic, I. (2014). The role of anxiety in the development, maintenance, and treatment of childhood aggression. ***Development and Psychopathology, 26***(4 Pt 2), 1515–1530. https://doi.org/10.1017/S0954579414001175

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. ***Journal of Clinical Child and Adolescent Psychology, 33***(2), 412–425. https://doi.org/https://doi.org/10.1207/s15374424jccp3302_23

*Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. ***Prevention & Treatment, 4***(1), 1–62. https://doi.org/10.1037/1522-3736.4.1.41a

Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2019). Generalized Boosted Regression Models. R package version 2.1.5. Retrieved from https://github.com/gbm-developers/gbm

*Grove, A. B., Evans, S. W., Pastor, D. A., & Mack, S. D. (2008). A meta-analytic examination of follow-up studies of programs designed to prevent the primary symptoms of oppositional defiant and conduct disorders. ***Aggression and Violent Behavior, 13***(3), 169–184. https://doi.org/10.1016/j.avb.2008.03.001

Guerra, N. G., Huesmann, L. R., Tolan, P. H., Van Acker, R., & Eron, L. D. (1995). Stressful events and individual beliefs as correlates of economic disadvantage and aggression among urban children. ***Journal of Consulting and Clinical Psychology, 63***(4), 518–528. https://doi.org/10.1037/0022-006X.63.4.518

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. ***BMC Psychiatry, 10***(1), 113. https://doi.org/1471-244X-10-113 [pii]\r10.1186/1471-244X-10-113 [doi]

*Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., … Dahlberg, L. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior. ***American Journal of Preventive Medicine, 33***(2), S114–S129. https://doi.org/10.1016/j.amepre.2007.04.012

*Hale, D. R., Fitzgerald-Yau, N., & Viner, R. M. (2014). A systematic review of effective interventions for reducing multiple health risk behaviors in adolescence. ***American Journal of Public Health, 104***(5), 19–41. https://doi.org/10.2105/AJPH.2014.301874

Halleröd, S. L. H., Larson, T., Ståhlberg, O., Carlström, E., Gillberg, C., Anckarsäter, H., … Gillberg, C. (2010). The autism-Tics, AD/HD and other comorbidities (A-TAC) telephone interview: Convergence with the child behavior checklist (CBCL). ***Nordic Journal of Psychiatry, 64***(3), 218–224. https://doi.org/10.3109/08039480903514443

Hannigan, L., Walaker, N., Waszczuk, M., McAdams, T., & Eley, T. (2017). Aetiological influences on stability and change in emotional and behavioural problems across development: a systematic review. ***Psychopathology Review, 4***(1), 1–57. https://doi.org/10.5127/pr.038315

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. ***PLoS ONE, 7***(2), e30320. https://doi.org/10.1371/journal.pone.0030320

Hansson, S. L., Röjvall, A. S., Rastam, M., Gillberg, C., Gillberg, C., & Anckarsäter, H. (2005). Psychiatric telephone interview with parents for screening of childhood autism - Tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): Preliminary reliability and validity. ***British Journal of Psychiatry, 187***, 262–267. https://doi.org/10.1192/bjp.187.3.262

Haris, A., Witten, D., & Simon, N. (2016). Convex modeling of interactions with strong heredity. *Journal of Computational and Graphical Statistics, 25*(4), 981–1004. https://doi.org/10.1080/10618600.2015.1067217

Harvey, E. A., Breaux, R. P., & Lugo-Candelas, C. I. (2016). Early development of comorbidity between symptoms of attention-deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD). *Journal of Abnormal Psychology, 125*(2), 154–167. https://doi.org/10.1037/abn0000090

*Harwood, A., Lavidor, M., & Rassovsky, Y. (2017). Reducing aggression with martial arts: A meta-analysis of child and youth studies. *Aggression and Violent Behavior, 34*, 96–101. https://doi.org/10.1016/j.avb.2017.03.001

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer. https://doi.org/10.1198/jasa.2004.s339

Hatch, S. L., Woodhead, C., Frissa, S., Fear, N. T., Verdecchia, M., Stewart, R., … Hotopf, M. (2012). Importance of thinking locally for mental health: Data from cross-sectional surveys representing south east London and England. *PLoS ONE, 7*(12). https://doi.org/10.1371/journal.pone.0048012

Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A Genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16*(01), 117–125. https://doi.org/10.1017/thg.2012.91

Hendriks, A. M., Bartels, M., Colins, O. F., & Finkenauer, C. (2018). Childhood aggression: A synthesis of reviews and meta-analyses to reveal patterns and opportunities for prevention and intervention strategies. *Neuroscience and Biobehavioral Reviews, 91*(December 2017), 278–291. https://doi.org/10.1016/j.neubiorev.2018.03.021

Herv, M. (2018). RVAideMemoire: Testing and plotting procedures for biostatistics. Version 0.9-69-3. Retrieved August 22, 2018, from https://rdrr.io/cran/RVAideMemoire/

Hoagwood, K. (2002). Making the translation from research to its application: The je ne sais pas of evidence-based practices. *Clinical Psychology: Science and Practice, 9*(2), 210–213. https://doi.org/10.1093/clipsy/9.2.210

Hoekstra, C., van Beijsterveldt, C. E. M. T., Lambalk, C. B., Hoekstra, C., Boomsma, D. I., Willemsen, G., & Montgomery, G. W. (2010). Body composition, smoking, and spontaneous dizygotic twinning. *Fertility and Sterility, 93*(3), 885–893. https://doi.org/10.1016/j.fertnstert.2008.10.012

Hofvander, B., Ossowski, D., Lundström, S., & Anckarsäter, H. (2009). Continuity of aggressive antisocial behavior from childhood to adulthood: The question of phenotype definition. *International Journal of Law and Psychiatry, 32*(4), 224–234. https://doi.org/10.1016/j.ijlp.2009.04.004

Hong, M., Jacobucci, R., & Lubke, G. H. (2019). *Deductive data mining*. Manuscript submitted for publication.

Hudziak, J. J., & Bartels, M. (2008). Genetic and environmental influences on wellness, resilience, and psychopathology: A family-based approach for promotion, prevention, and intervention. In Developmental Psychopathology and Wellness: Genetic and Environmental Influences. (pp. 267–286). Arlington: American Psychiatric Publishing.

Hudziak, J. J., & Ivanova, M. Y. (2016). The Vermont Family Based Approach: Family based health promotion, illness prevention, and intervention. *Child and Adolescent Psychiatric Clinics of North America, 25*(2), 167–178. https://doi.org/10.1016/j.chc.2015.11.002

Hudziak, J. J., Van Beijsterveldt, C. E. M., Bartels, M., Rietveld, M. J. H., Rettew, D. C., Derks, E. M., & Boomsma, D. I. (2003). Individual differences in aggression: Genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins. *Behavior Genetics, 33*(5), 575–589. https://doi.org/10.1023/A:1025782918793

Huesmann, L. R., Dubow, E. F., & Boxer, P. (2009). Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: Implications for the adolescent-limited and life-course-persistent models. *Aggressive Behavior, 35*(2), 136–149. https://doi.org/10.1002/ab.20300

Hunter, L. (2003). School psychology: A public health framework. III. Managing disruptive behavior in schools: The value of a public health and evidence-based perspective. *Journal of School Psychology, 41*(1), 39–59. https://doi.org/10.1016/S0022-4405(02)00143-7

ILO. (2014). World Social Protection Report 2014-2015: Building economic recovery, inclusive development and a social justice. Geneva.

Institute for Health Metrics and Evaluation. (2018). Global Burden of Disease. Retrieved February 5, 2018, from http://www.healthdata.org/gbd



Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L., Almqvist, F., Weintraub, S., … Verhulst, F. C. (2007). Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child & Adolescent Psychology, 36*(3), 405–417. https://doi.org/10.1080/15374410701444363

Ivanova, M. Y., Dewey, L., Swift, P., Weinberger, S., & Hudziak, J. (2019). Health promotion in primary care pediatrics: Initial results of a randomized clinical trial of the Vermont Family Based Approach. *Child and Adolescent Psychiatric Clinics of North America, 28*, 237–246. https://doi.org/10.1016/j.chc.2018.11.005

Jaffee, S. R., Strait, L. B., & Odgers, C. L. (2012). From correlates to causes: Can quasi-experimental studies and statistical innovations bring us closer to identifying the causes of antisocial behavior? *Psychological Bulletin, 138*(2), 272–295. https://doi.org/10.1037/a0026020

Janssens, A., & Deboutte, D. (2009). Screening for psychopathology in child welfare: The Strengths and Difficulties Questionnaire (SDQ) compared with the Achenbach System of Empirically Based Assessment (ASEBA). *European Child and Adolescent Psychiatry, 18*(11), 691–700. https://doi.org/10.1007/s00787-009-0030-y

Jenkins, R. (2003). Making psychiatric epidemiology useful: The contribution of epidemiology to government policy. *International Review of Psychiatry, 15*(1–2), 188–200. https://doi.org/10.1080/0954026021000046164

Johnson, A. M., Hawes, D. J., Eisenberg, N., Kohlhoff, J., & Dudeney, J. (2017). Emotion socialization and child conduct problems: A comprehensive review and meta-analysis. *Clinical Psychology Review, 54*(April), 65–80. https://doi.org/10.1016/j.cpr.2017.04.001

Johnson, M. H., George, P., Armstrong, M. I., Lyman, D. R., Dougherty, R. H., Daniels, A. S., … Delphin-Rittmon, M. E. (2014). Behavioral management for children and adolescents: Assessing the evidence. *Psychiatric Services, 65*(5), 580–590. https://doi.org/10.1176/appi.ps.201300253

*Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*(4), 567–589. https://doi.org/10.1007/s10802-007-9201-9

Kaprio, J. (2013). The Finnish Twin Cohort Study: An update. *Twin Research and Human Genetics, 16*(1), 157–162. https://doi.org/10.1017/thg.2012.142

Kendler, K., Neale, M., Kessler, R., Heath, A., & Eaves, L. (1992). Major depression and generalized anxiety disorder: Same genes, (partly) different environments? *Archives of General Psychiatry, 49*(9), 716–722. https://doi.org/10.1001/archpsyc.1992.01820090044008

Kerekes, N., Lundström, S., Chang, Z., Tajnia, A., Jern, P., Lichtenstein, P., … Anckarsäter, H. (2014). Oppositional defiant- and conduct disorder-like problems: Neurodevelopmental predictors and genetic background in boys and girls, in a nationwide twin study. *PeerJ,* 2, e359. https://doi.org/10.7717/peerj.359

Kerig, P. K., & Stellwagen, K. K. (2010). Roles of callous-unemotional traits, narcissism, and machiavellianism in childhood aggression. *Journal of Psychopathology and Behavioral Assessment, 32*(3), 343–352. https://doi.org/10.1007/s10862-009-9168-7

Kessler, R. C., Angermeyer, M., Anthony, J. C., DE Graaf, R., Demyttenaere, K., Gasquet, I., … Ustün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA), 6*(3), 168–176. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(June), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

King, S., & Waschbusch, D. A. (2010). Aggression in children with attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics, 10*(10), 1581–1594. https://doi.org/10.1586/ERN.10.146

Klahr, A. M., & Burt, S. A. (2014). Practitioner Review: Evaluation of the known behavioral heterogeneity in conduct disorder to improve its assessment and treatment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 55*(12), 1300–1310. https://doi.org/10.1111/jcpp.12268

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., … Goodman, R. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. ***European Child and Adolescent Psychiatry, 9***(4), 271–276. https://doi.org/10.1007/s007870070030

Kleinert, S. (2007). Adolescent health: an opportunity not to be missed. ***Lancet, 369***(9575), 1788–1789. https://doi.org/10.1016/S0140- 6736(07)60374-2

Knapp, M. R. J., Scott, S., & Davies, J. (1999). The cost of antisocial behaviour in younger children. ***Clinical Child Psychology and Psychiatry, 4***(4), 457–473. https://doi.org/10.1177/1359104599004004003

Kontopantelis, E., White, I. R., Sperrin, M., & Buchan, I. (2017). Outcome-sensitive multiple imputation: A simulation study. ***BMC Medical Research Methodology, 17***(1), 1–13. https://doi.org/10.1186/s12874-016-0281-5

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. ***Molecular Psychiatry, 21***(3), 437–443. https://doi.org/10.1038/mp.2015.2

*Kremer, K. P., Maynard, B. R., Polanin, J. R., Vaughn, M. G., & Sarteschi, C. M. (2014). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: A systematic review and meta-analysis. ***Journal of Youth and Adolescence, 44***(3), 616–636. https://doi.org/10.1007/s10964-014-0226-4

Kuhn, M. (2018). Classification and regression training. R package version 6.0-80. Retrieved from https://cran.r-project.org/package=caret

Kuja-Halkola, R., Lichtenstein, P., D'Onofrio, B. M., & Larsson, H. (2015). Codevelopment of ADHD and externalizing behavior from childhood to adulthood. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 56***(6), 640–647. https://doi.org/10.1111/jcpp.12340

Kumperscak, H. G. (2019). Communications of the European Society for Child and Adolescent Psychiatry. ***European Child and Adolescent Psychiatry, 28***, 147–151. https://doi.org/10.1007/s00787-018-1232-y

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. ***Psychological Bulletin, 143***(2), 142–186. https://doi.org/10.1037/bul0000069

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. ***Biometrics, 33***(1), 159–174. https://doi.org/10.1109/ICDMA.2010.328

LaPrairie, J. L., Schechter, J. C., Robinson, B. A., & Brennan, P. A. (2011). Perinatal risk factors in the development of aggression and violence. Advances in Genetics (1st ed., Vol. 75). Elsevier Inc. https://doi.org/10.1016/B978-0-12-380858-5.00004-6

Larsson, H., Viding, E., Rijsdijk, F. V., & Plomin, R. (2008). Relationships between parental negativity and childhood antisocial behavior over time: A bidirectional effects model in a longitudinal genetically informative design. ***Journal of Abnormal Child Psychology, 36***(5), 633–645. https://doi.org/10.1007/s10802-007-9151-2

Latvala, A., Kuja-Halkola, R., Almqvist, C., Larsson, H., & Lichtenstein, P. (2015). A longitudinal study of resting heart rate and violent criminality in more than 700000 men. ***JAMA Psychiatry, 72***(10), 971–978. https://doi.org/10.1001/jamapsychiatry.2015.1165

*Lee, C. M., Horvath, C., & Hunsley, J. (2013). Does it work in the real world? The effectiveness of treatments for psychological problems in children and adolescents. ***Professional Psychology: Research and Practice, 44***(2), 81–88. https://doi.org/10.1037/a0031133

*Leijten, P., Raaijmakers, M. A. J., De Castro, B. O., & Matthys, W. (2013). Does socioeconomic status matter ? A meta-analysis on parent training effectiveness for disruptive child behavior. ***Journal of Clinical Child & Adolescent Psychiatry, 42***(3), 37–41. https://doi.org/10.1080/15374416.2013.769169

Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. ***Journal of Emotional and Behavioral Disorders, 21***(3), 211–224. https://doi.org/10.1177/1063426611421007

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. ***Psychological Bulletin, 126***(2), 309–337. https://doi.org/10.1037/0033-2909.126.2.309



Levin, K. A., & Currie, C. (2014). Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples. ***Social Indicators Research, 119***(2), 1047–1063. https://doi.org/10.1007/s11205-013-0507-4

Levin, K. A., Torsheim, T., Vollebergh, W., Richter, M., Davies, C. A., Schnohr, C. W., … Currie, C. (2011). National income and income inequality, family affluence and life satisfaction among 13 year old boys and girls: A multilevel study in 35 countries. ***Social Indicators Research, 104***(2), 179–194. https://doi.org/10.1007/s11205-010-9747-8

Lewinsohn, P. M., Shankman, S. a, Gau, J. M., & Klein, D. N. (2004). The prevalence and co-morbidity of subthreshold psychiatric conditions. ***Psychological Medicine, 34***, 613–622. https://doi.org/10.1017/S0033291703001466

Lichtenstein, P., Tuvblad, C., Larsson, H., & Carlström, E. (2007). The Swedish Twin study of CHild and Adolescent Development: The TCHAD-study. ***Twin Research and Human Genetics, 10***(1), 67–73. https://doi.org/10.1375/twin.10.1.67

Ligthart, L., Bartels, M., Hoekstra, R. A., Hudziak, J. J., & Boomsma, D. I. (2005). Genetic contributions to subtypes of aggression. ***Twin Research and Human Genetics, 8***(5), 483–491. https://doi.org/10.1375/twin.8.5.483

Link, B. G., Phelan, J. C., Miech, R., Westin, E. L., Journal, S., Behavior, S., … Phelan, J. O. C. (2008). The resources that matter: Fundamental social causes of health disparities and the challenge of intelligence. ***Journal of Health and Social Behavior, 49***(1), 72–91.

Lipsey, M. W., & Wilson, D. B. (2000). ***Practical meta-analysis*** (49th ed.). Thousand Oaks: SAGE Publications.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. ***Annual Review of Psychology, 48***(1997), 371–410. https://doi.org/10.1146/annurev.psych.48.1.371

*Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. ***The ANNALS of the American Academy of Political and Social Science, 587***(1), 84–109. https://doi.org/10.1177/0002716202250793

Lubke, G. H., & Campbell, I. (2016). Inference based on the best-fitting model can contribute to the replication crisis: assessing model selection uncertainty using a bootstrap approach. ***Structural Equation Modeling, 23***(4), 479–490. https://doi.org/10.1080/10705511.2016.1141355

Lubke, G. H., Mcartor, D. B., Boomsma, D. I., & Bartels, M. (2017). Genetic and environmental contributions to the development of childhood aggression. ***Developmental Psychology***. https://doi.org/http://dx.doi.org/10.1037/dev0000403

*Lundahl, B., Risser, H. J., & Lovejoy, M. C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. ***Clinical Psychology Review, 26***(1), 86–104. https://doi.org/10.1016/j.cpr.2005.07.004

Luningham, J. M., Hendriks, A. M., Krapohl, E., Ip, H. ., van Beijsterveldt, C. E. M., Nivard, M. G., … Lubke, G. H. (2019). ***Harmonizing behavioral outcomes across studies, raters, and countries: An application to the genetic analysis of aggression in the ACTION Consortium***. Manuscript submitted for publication.

Lynam, D., Caspi, A., Moffitt, T. E., Wikström, P.-O. H., Loeber, R., & Novak, S. (2000). The interaction between impulsivity and neighborhood context on offending: The effects of impulsivity are stronger in poorer neighborhoods. ***Journal of Abnormal Psychology, 109***(4), 695–704. https://doi.org/10.1037//0021-843X.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? ***Psychological Bulletin, 131***(6), 803–855. https://doi.org/10.1037/0033-2909.131.6.803

Malanchini, M., Smith-Woolley, E., Ayorech, Z., Rimfeld, K., Krapohl, E., Vuoksimaa, E., … Plomin, R. (2018). Aggressive behaviour in childhood and adolescence: the role of smoking during pregnancy, evidence from four twin cohorts in the EU-ACTION consortium. ***Psychological Medicine***, 1–9. https://doi.org/10.1017/S0033291718001344

Marshall, N. A., Arnold, D. H., Rolon-Arroyo, B., & Griffith, S. F. (2015). The association between relational aggression and internalizing symptoms: A review and meta-analysis. ***Journal of Social and Clinical Psychology, 34***(2), 135–160. https://doi.org/10.1521/jscp.2015.34.2.135

*Maughan, D. R., Christiansen, E., Jenson, W. R., Olympia, D., & Clark, E. (2005). Behavioral parent training as a treatment for externalizing behaviors and disruptive behavior disorders: A meta-analysis. ***School Psychology Review, 34***(3), 267–286. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-26644436946&partnerID=40&md5=9302eec84d358650b88180ff263ddc66

*McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: A meta-analysis. ***Journal of Abnormal Child Psychology, 34***(4), 527–543. https://doi.org/10.1007/s10802-006-9031-1

Mcknight, C. G., Huebner, E. S., & Suldo, S. (2002). Relationships among stressful life events, temperament, problem behavior, and global life satisfaction in adolescents. ***Psychology in the Schools, 39***(6), 677–687. https://doi.org/10.1002/pits.10062

Mcleroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An Ecological Perspective on Health Promotion Programs. ***Health Education & Behavior, 15***(4), 351–377. https://doi.org/10.1177/109019818801500401

Mcmahon, S. D., Grant, K. E., Compas, B. E., Thurm, A. E., & Ey, S. (2003). Stress and psychopathology in children and adolescents: Is there evidence of specificity? ***Journal of Child Psychology and Psychiatry, 44***(1), 107–133. https://doi.org/10.1111/1469-7610.00105

Meltzer, H., Ford, T., Goodman, R., & Vostanis, P. (2011). The burden of caring for children with emotional or conduct disorders. ***International Journal of Family Medicine, 2011***, 1–8. https://doi.org/10.1155/2011/801203

*Menting, A. T. A., Orobio de Castro, B., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: A meta-analytic review. ***Clinical Psychology Review, 33***(8), 901–913. https://doi.org/10.1016/j.cpr.2013.07.006

Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. ***Dialogues in Clinical Neuroscience, 11***, 7–20.

*Michelson, D., Davenport, C., Dretzke, J., Barlow, J., & Day, C. (2013). Do evidence-based interventions work when tested in the "real world?" A systematic review and meta-analysis of parent management training for the treatment of child disruptive behavior. ***Clinical Child and Family Psychology Review, 16***(1), 18–34. https://doi.org/10.1007/s10567-013-0128-0

Middeldorp, C. M., Lamb, D. J., Vink, J. M., Bartels, M., Van Beijsterveldt, C. E. M., & Boomsma, D. I. (2014). Child care, socio-economic status and problem behavior: A study of gene-environment interaction in young Dutch twins. ***Behavior Genetics, 44***(4), 314–325. https://doi.org/10.1007/s10519-014-9660-z

Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the strengths and Difficulties Questionnaire in 5-6 year olds: Differences by gender or by parental education? ***PLoS ONE, 7***(5). https://doi.org/10.1371/journal.pone.0036805

Miller, P. J., Lubke, G. H., McArtor, D. B., & Bergeman, C. S. (2016). Finding structure in data using multivariate tree boosting. ***Psychological Methods, 21***(4), 583–602. https://doi.org/10.1037/met0000087

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. ***Annuals of Internal Medicine, 151***(4), 264–269. https://doi.org/10.1371/journal.pmed1000097

*Montgomery, P., Bjornstad, G. J., & Dennis, J. A. (2006). Media-based behavioural treatments for behavioural problems in children. ***Cochrane Database of Systematic Reviews***, (1). https://doi.org/10.1002/14651858.CD002206.pub3

*Montgomery, P., & Maunders, K. (2015). The effectiveness of creative bibliotherapy for internalizing, externalizing, and prosocial behaviors in children: A systematic review. ***Children and Youth Services Review, 55***, 37–47. https://doi.org/10.1016/j.childyouth.2015.05.010

Mrazek, P. J., & Haggerty, R. J. (1994). Reducing risks for mental disorders: Frontiers for preventive intervention research. (I. of M. Committee on Prevention of Mental Disorders, Ed.). Washington, D.C.: National Academy Press.

Nagin, D. S., & Tremblay, R. E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. ***Archives of General Psychiatry, 58***(4), 389–394. https://doi.org/10.1001/archpsyc.58.4.389

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., … Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. ***Psychometrika, 81***(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

NICE. (2013). CG158: Antisocial behaviour and conduct disorders in children and young people: recognition and management. NICE Clinical Guideline, (March).



Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. ***Psychological Medicine, 36***(5), 699. https://doi.org/10.1017/S0033291706007082

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 48***(7), 703–713. https://doi.org/10.1111/j.1469-7610.2007.01733.x

*Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of triple P-positive parenting program using hierarchical linear modeling: Effectiveness and moderating variables. ***Clinical Child and Family Psychology Review, 11***(3), 114–144. https://doi.org/10.1007/s10567-008-0033-0

OECD Social Policy Division. (2016). PF1.1 Public spending on family benefits. Retrieved from http://www.oecd.org/els/soc/PF1_1_Public_spending_on_family_benefits.pdf

Oliver, B. R. (2015). Unpacking externalising problems: Negative parenting associations for conduct problems and irritability. ***British Journal of Psychiatry Open, 1***(1), 42–47. https://doi.org/10.1192/bjpo.bp.115.000125

*Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom: Management practices: Effects on disruptive or aggressive student behavior. ***The Campbell Systematic Reviews***, 44, 55. https://doi.org/10.4073/csr.2011.4

Olweus, D. (1992). Bullying among schoolchildren: Intervention and prevention. In R. D. Peters, R. J. McMahon, & V. L. Quinsey (Eds.), Aggression and Violence Throughout the Lifespan (pp. 127–139). Newbury Park, California: SAGE Publications.

Ottova, V., Erhart, M., Vollebergh, W., Koekoenyei, G., Morgan, A., Gobina, I., … Grp, P. H. F. (2012). The role of individual- and macro-level social determinants on young adolescents' psychosomatic complaints. ***Journal of Early Adolescence, 32***(1, SI), 126–158. https://doi.org/10.1177/0272431611419510

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M. G., … Tiemeier, H. (2016). A genome-wide approach to children's aggressive behavior: The EAGLE consortium. ***American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 171***(5), 562–572. https://doi.org/10.1002/ajmg.b.32333

*Park-Higgerson, H. K., Perumean-Chaney, S. E., Bartolucci, A. A., Grimley, D. M., & Singh, K. P. (2008). The evaluation of school-based violence prevention programs: A meta-analysis. ***Journal of School Health, 78***(9), 420–465. https://doi.org/JOSH332 [pii]\r10.1111/j.1746-1561.2008.00332.x

Patton, G. C., Coffey, C., Romaniuk, H., Mackinnon, A., Carlin, J. B., Degenhardt, L., … Moran, P. (2014). The prognosis of common mental disorders in adolescents: A 14-year prospective cohort study. ***The Lancet, 383***(9926), 1404–1411. https://doi.org/10.1016/S0140-6736(13)62116-9

Patton, G. C., Olsson, C. A., Skirbekk, V., Saffery, R., Wlodek, M. E., Azzopardi, P. S., … Sawyer, S. M. (2018). Adolescence and the next generation. ***Nature, 554***(7693), 458–466. https://doi.org/10.1038/nature25759

Phelan, J. C., Link, B. G., Diez-Roux, A., Kawachi, I., & Levin, B. (2004). "Fundamental Causes" of social inequalities in mortality: A test of the theory. ***Journal of Health and Social Behavior, 45***(3), 265–285.

Phelan, J. C., Link, B. G., & Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. ***Journal of Health and Social Behavior, 51***(1_suppl), S28–S40. https://doi.org/10.1177/0022146510383498

Pickett, K. E., & Wilkinson, R. G. (2007). Child wellbeing and income inequality in rich countries: ecological cross sectional study. ***British Medical Journal, 335***. https://doi.org/10.1136/bmj.39377.580162.55

Pickett, K. E., & Wilkinson, R. G. (2010). Inequality: An underacknowledged source of mental illness and distress. ***British Journal of Psychiatry, 197***(6), 426–428. https://doi.org/10.1192/bjp.bp.109.072066

Pickett, W. (2005). Cross-national study of fighting and weapon carrying as determinants of adolescent injury. ***Pediatrics, 116***(6), e855–e863. https://doi.org/10.1542/peds.2005-0607

Pickett, W., Molcho, M., Elgar, F. J., Brooks, F., de Looze, M., Rathmann, K., … Currie, C. (2013). Trends and socioeconomic correlates of adolescent physical fighting in 30 countries. ***Pediatrics, 131***, e18-26. https://doi.org/10.1542/peds.2012-1614

Piko, B. F. (2007). Self-perceived health among adolescents: The role of gender and psychosocial factors. ***European Journal of Pediatrics, 166***(7), 701–708. https://doi.org/10.1007/s00431-006-0311-0

Pinquart, M. (2017). Associations of parenting dimensions and styles with externalizing problems of children and adolescents: An updated meta-analysis. ***Developmental Psychology, 53***(5), 873–932. https://doi.org/10.1037/dev0000295

Pinquart, M., & Shen, Y. (2011). Anxiety in children and adolescents with chronic physical illnesses: A meta-analysis. ***Acta Paediatrica, 100***(8), 1069–1076. https://doi.org/10.1111/j.1651-2227.2011.02223.x

Piotrowska, P. J., Stride, C. B., Croft, S. E., & Rowe, R. (2015). Socioeconomic status and antisocial behaviour among children and adolescents: A systematic review and meta-analysis. ***Clinical Psychology Review, 35***, 47–55. https://doi.org/10.1016/j.cpr.2014.11.003

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 56***(3), 345–365. https://doi.org/10.1111/jcpp.12381

Porsch, R. M., Middeldorp, C. M., Cherny, S. S., Krapohl, E., van Beijsterveldt, C. E. M., Loukola, A., … Bartels, M. (2016). Longitudinal heritability of childhood aggression. ***American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics,*** (January), 697–707. https://doi.org/10.1002/ajmg.b.32420

Proctor, C. L., Linley, P. A., & Maltby, J. (2009). Youth life satisfaction: A review of the literature. ***Journal of Happiness Studies, 10***(5), 583–630. https://doi.org/10.1007/s10902-008-9110-9

R Core Team. (2018). The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Racz, S. J., & McMahon, R. J. (2011). The relationship between parental knowledge and monitoring and child and adolescent conduct problems: A 10-year update. ***Clinical Child and Family Psychology Review, 14***(4), 377–398. https://doi.org/10.1007/s10567-011-0099-y

Ragozin, A. S., Basham, R. B., Crnic, K. A., Greenberg, M. T., & Robinson, N. (1982). Effects of maternal age on parenting role. ***Developmental Psychology, 18***(4), 627–634. https://doi.org/10.1037/0012-1649.18.4.627

Raine, A., Dodge, K., Loeber, R., Gatzke-kopp, L., Lynam, D., Stouthamer-loeber, M., & Liu, J. (2006). The Reactive–Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. ***Aggressive Behavior, 32***(2), 159–171. https://doi.org/10.1002/ab.20115.The

Raine, A., Reynolds, C., & Venables, P. H. (1998). Fearlessness, stimulation-seeking, and large body size at age 3 years as early predispositions to childhood aggression at age 11 years. ***Archives of General Psychiatry, 55***, 745–751.

Ravens-Sieberer, U., Erhart, M., Torsheim, T., Hetland, J., Freeman, J., Danielson, M., … Overpeck, M. (2008). An international scoring system for self-reported health complaints in adolescents. ***European Journal of Public Health, 18***(3), 294–299. https://doi.org/10.1093/eurpub/ckn001

Ravens-Sieberer, U., Torsheim, T., Hetland, J., Vollebergh, W., Cavallo, F., Jericek, H., … Erginoz, E. (2009). Subjective health, symptom load and quality of life of children and adolescents in Europe. ***International Journal of Public Health, 54***(SUPPL. 2). https://doi.org/10.1007/s00038-009-5406-8

*Ray, D. C., Armstrong, S. A., Balkin, R. S., & Jayne, K. M. (2015). Child-centered play therapy in the school: Review and meta-analysis. ***Psychology in the Schools, 52***(2), 107–123. https://doi.org/10.1002/pits.21798

Reijntjes, A., Kamphuis, J. H., Prinzie, P., Boelen, P. A., Van Der Schoot, M., & Telch, M. J. (2011). Prospective linkages between peer victimization and externalizing problems in children: A meta-analysis. ***Aggressive Behavior, 37***(3), 215–222. https://doi.org/10.1002/ab.20374

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. ***Social Science and Medicine, 90***, 24–31. https://doi.org/10.1016/j.socscimed.2013.04.026

*Reyno, S. M., & McGrath, P. J. (2006). Predictors of parent training efficacy for child externalizing behavior problems - A meta-analytic review. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 47***(1), 99–111. https://doi.org/10.1111/j.1469-7610.2005.01544.x

Rhee, S. H., & Waldman, I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. ***Psychological Bulletin, 128***(3), 490–529. https://doi.org/10.1037//0033-2909.128.3.490

Rhee, S. H., Willcutt, E. G., Hartman, C. A., Pennington, B. F., & DeFries, J. C. (2008). Test of alternative hypotheses explaining the comorbidity between attention-deficit/hyperactivity disorder and conduct disorder. ***Journal of Abnormal Child Psychology, 36***(1), 29–40. https://doi.org/10.1007/s10802-007-9157-9



Rivenbark, J. G., Odgers, C. L., Caspi, A., Harrington, H. L., Hogan, S., Houts, R. M., … Moffitt, T. E. (2018). The high societal costs of childhood conduct problems: Evidence from administrative records up to age 38 in a longitudinal birth cohort. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 59***(6), 703–710. https://doi.org/10.1111/jcpp.12850

Roberts, R., McCrory, E., Joffe, H., de Lima, N., & Viding, E. (2017). Living with conduct problem youth: Family functioning and parental perceptions of their child. ***European Child and Adolescent Psychiatry, 27***(5), 1–10. https://doi.org/10.1007/s00787-017-1088-6

Rocha, T. B. M., Graeff-Martins, A. S., Kieling, C., & Rohde, L. A. (2015). Provision of mental healthcare for children and adolescents: A worldwide view. ***Current Opinion in Psychiatry, 28***(4), 330–335. https://doi.org/10.1097/YCO.0000000000000169

Roetman, P. J., Lundström, S., Finkenauer, C., Vermeiren, R. R. J. M., Lichtenstein, P., & Colins, O. F. (2019). Children with early-onset disruptive behavior: Parental mental disorders predict poor psychosocial functioning in adolescence. ***Journal of the American Academy of Child & Adolescent Psychiatry***, (2019). https://doi.org/10.1016/j.jaac.2018.10.017

Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children--and who pays it. ***The British Journal of Psychiatry: The Journal of Mental Science, 188***, 547–553. https://doi.org/10.1192/bjp.bp.104.007625

*Rosato, N. S., Correll, C. U., Pappadopulos, E., Chait, A., Crystal, S., & Jensen, P. S. (2012). Treatment of maladaptive aggression in youth: CERT guidelines II. Treatments and ongoing management. ***Pediatrics, 129***(6), e1577–e1586. https://doi.org/10.1542/peds.2010-1361

Royal College of Pediatrics and Child Health. (2012). UK-World Health Organisation growth charts - 2-18 years. Retrieved August 9, 2018, from https://www.rcpch.ac.uk/resources/uk-world-health-organisation-growth-charts-2-18-years

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR. ***Journal of the American Statistical Association, 82***(398), 543–546. https://doi.org/10.2307/2289460

Sabina, C., & Banyard, V. (2015). Moving toward well-being: The role of protective factors in violence research. ***Psychology of Violence, 5***(4), 337–342. https://doi.org/10.1037/a0039686

Salekin, R. T. (2017). Research review: What do we know about psychopathic traits in children? ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 58***(11), 1180–1200. https://doi.org/10.1111/jcpp.12738

Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. ***Child Development, 81***(1), 6–22. https://doi.org/10.1111/j.1467-8624.2009.01378.x

Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S. J., Dick, B., Ezeh, A. C., & Patton, G. C. (2012). Adolescence: A foundation for future health. ***The Lancet, 379***(9826), 1630–1640. https://doi.org/10.1016/S0140-6736(12)60072-5

*Sawyer, A. M., Borduin, C. M., & Dopp, A. R. (2015). Long-term effects of prevention and treatment on youth antisocial behavior: A meta-analysis. ***Clinical Psychology Review, 42***, 130–144. https://doi.org/10.1016/j.cpr.2015.06.009

Scott, S., Knapp, M., Henderson, J., & Maughan, B. (2001). Financial cost of social exclusion: Follow-up study of antisocial children into adulthood. ***British Medical Journal (Clinical Research Ed.), 323***(7306), 191–194. https://doi.org/10.1136/bmj.323.7306.191

Shanahan, M. J., & Hofer, S. M. (2005). Social Context in Gene – Environment Interactions : Retrospect and Prospect. ***Journals of Gerontology: SERIES B, 60***(I), 65–76.

Shankman, S. A., Lewinsohn, P. M., Klein, D. N., Small, J. W., Seeley, J. R., & Altman, S. E. (2009). Subthreshold conditions as precursors for full syndrome disorders: A 15-year longitudinal study of multiple diagnostic classes. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 50***(12), 1485–1494. https://doi.org/10.1111/j.1469-7610.2009.02117.x

Shatkin, J. P., & Belfer, M. L. (2004). The global absence of child and adolescent mental health policy. ***Child & Adolescent Mental Health, 9***(3), 104–108. https://doi.org/10.1111/j.1475-3588.2004.00090.x

*Shelleby, E. C., & Shaw, D. S. (2014). Outcomes of parenting interventions for child conduct problems: A review of differential effectiveness. ***Child Psychiatry and Human Development, 45***(5), 628–45. https://doi.org/10.1007/s10578-013-0431-5

Signorini, G., Singh, S. P., Boricevic-Marsanic, V., Dieleman, G., Dodig-Ćurković, K., Franic, T., … de Girolamo, G. (2017). Architecture and functioning of child and adolescent mental health services: a 28-country survey in Europe. ***The Lancet Psychiatry, 0366***(17). https://doi.org/10.1016/S2215-0366(17)30127-X

*Smedler, A.-C., Hjern, A., Wiklund, S., Anttila, S., & Pettersson, A. (2015). Programs for prevention of externalizing problems in children: Limited evidence for effect beyond 6 months post intervention. ***Child & Youth Care Forum, 44***(2), 251–276. https://doi.org/10.1007/s10566-014-9281-y

*Smeets, K. C., Leeijen, A. A. M., Van der Molen, M. J., Scheepers, F. E., Buitelaar, J. K., & Rommelse, N. N. J. (2015). Treatment moderators of cognitive behavior therapy to reduce aggressive behavior: A meta-analysis. ***European Child & Adolescent Psychiatry, 24***(3), 255–264. https://doi.org/10.1007/s00787-014-0592-1

Sousa, S., Correia, T., Ramos, E., Fraga, S., & Barros, H. (2010). Violence in adolescents: social and behavioural factors. ***Gaceta Sanitaria, 24***(1), 47–52. https://doi.org/10.1016/j.gaceta.2009.08.002

South, S. C., Hamdi, N. R., & Krueger, R. F. (2015). Biometric modeling of gene-environment interplay: The intersection of theory and method and applications for social inequality. ***Journal of Personality,*** (February). https://doi.org/10.1111/jopy.12231

Spearman, C. (1904). The proof and measurement of association between two things. ***The American Journal of Psychology, 15***(1), 72–101. https://doi.org/https://doi.org/10.1037/11491-005

Stattin, H., & Kerr, M. (2000). Parental monitoring: A reinterpretation. ***Child Development, 71***(4), 1072–1085. https://doi.org/10.1111/1467-8624.00210

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. ***Britisch Medical Journal, 339***(7713), 157–160. https://doi.org/10.1136/bmj.b2393

*Stoltz, S., Londen, M. V., Dekovic, M., Castro, B. O. D., & Prinzie, P. (2012). Effectiveness of individually delivered indicated school-based interventions on externalizing behavior. ***International Journal of Behavioral Development, 36,*** 381–388. https://doi.org/10.1177/0165025412450525

*Sukhodolsky, D. G., Kassinove, H., & Gorman, B. S. (2004). Cognitive-behavioral therapy for anger in children and adolescents: A meta-analysis. ***Aggression and Violent Behavior, 9***(3), 247–269. https://doi.org/10.1016/j.avb.2003.08.005

Svensson, O., Sörman, K., Durbeej, N., Lichtenstein, P., Anckarsäter, H., Kerekes, N., & Nilsson, T. (2018). Associations Between Conduct Disorder, Neurodevelopmental Problems and Psychopathic Personality Traits in a Swedish Twin Youth Population. ***Journal of Psychopathology and Behavioral Assessment, 40***(4), 586–592. https://doi.org/10.1007/s10862-018-9689-z

Sweeting, H., & West, P. (2003). Sex differences in health at ages 11, 13 and 15. ***Social Science & Medicine (1982), 56***(1), 31–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12435549

*Tarver, J., Daley, D., Lockwood, J., & Sayal, K. (2014). Are self-directed parenting interventions sufficient for externalising behaviour problems in childhood? A systematic review and meta-analysis. ***European Child & Adolescent Psychiatry, 1123***–1137. https://doi.org/10.1007/s00787-014-0556-5

Tearne, J. E., Robinson, M., Jacoby, P., Li, J., Newnham, J., & McLean, N. (2015). Does late childbearing increase the risk for behavioural problems in children? A longitudinal cohort study. ***Paediatric and Perinatal Epidemiology, 29***(1), 41–49. https://doi.org/10.1111/ppe.12165

The World Bank Group. (2017). GINI index (World Bank estimate). Retrieved July 4, 2017, from https://data.worldbank.org/indicator/SI.POV.GINI

*Thomas, R., Abell, B., Webb, H. J., Avdagic, E., & Zimmer-Gembeck, M. J. (2017). Parent-child interaction therapy: A meta-analysis. ***Pediatrics, 140***(3).

*Thomas, R., & Zimmer-Gembeck, M. J. (2007). Behavioral outcomes of parent-child interaction therapy and triple P-positive parenting program: A review and meta-analysis. ***Journal of Abnormal Child Psychology, 35***(3), 475–495. https://doi.org/10.1007/s10802-007-9104-9

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. ***Journal of the Royal Statistical Society. Series B (Methodological), 58***(1), 267–288.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. ***Journal of the Royal Statistical Society. Series B (Methodological), 73***(Part 3), 273–282. https://doi.org/10.2307/2346101

Tielbeek, J. J., Johansson, A., Polderman, T. J. C., Rautiainen, M. R., Jansen, P., Taylor, M., … Posthuma, D. (2017). Genome-wide association studies of a broad spectrum of antisocial behavior. ***JAMA Psychiatry, 74***(12), 1242–1250. https://doi.org/10.1001/jamapsychiatry.2017.3069



Torsheim, T., Cavallo, F., Levin, K. A., Schnohr, C., Mazur, J., Niclasen, B., & Currie, C. (2016). Psychometric validation of the revised Family Affluence Scale: a latent variable approach. ***Child Indicators Research, 9***(3), 771–784. https://doi.org/10.1007/s12187-015-9339-x

Torsheim, T., Currie, C., Boyce, W., Kalnins, I., Overpeck, M., & Haugland, S. (2004). Material deprivation and self-rated health: A multilevel study of adolescents from 22 European and North American countries. ***Social Science & Medicine, 59***(1), 1–12. https://doi.org/10.1016/j.socscimed.2003.09.032

Tremblay, R. E. (2000). The development of agressive behaviour during childhood: What have we learned in the past century? ***International Journal of Behavioral Development, 24***(2), 129–141. https://doi.org/10.1080/016502500383232

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: The "original sin" hypothesis, epigenetics and their consequences for prevention. ***Journal of Child Psychology and Psychiatry and Allied Disciplines, 51***(4), 341–367. https://doi.org/10.1111/j.1469-7610.2010.02211.x

Tucker-Drob, E. M., & Bates, T. C. (2015). Large cross-national differences in gene × socioeconomic status interaction on intelligence. ***Psychological Science,*** 1–12. https://doi.org/10.1177/0956797615612727

*Tully, L. A., & Hunt, C. (2016). Brief parenting interventions for children at risk of externalizing behavior problems: A systematic review. ***Journal of Child and Family Studies, 25***, 705–719. https://doi.org/10.1007/s10826-015-0284-6

Tuvblad, C., & Baker, L. A. (2011). Human aggression across the lifespan: Genetic propensities and environmental moderators. ***Advances in Genetics, 75***, 171–214. https://doi.org/10.1016/b978-0-12-380858-5.00007-1

Tuvblad, C., Eley, T. C., & Lichtenstein, P. (2005). The development of antisocial behaviour from childhood to adolescence: A longitudinal twin study. ***European Child and Adolescent Psychiatry, 14***(4), 216–225. https://doi.org/10.1007/s00787-005-0458-7

Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: gene-environment interaction. ***Journal of Child Psychology and Psychiatry, 47***(7), 734–743. https://doi.org/10.1111/j.1469-7610.2005.01552.x

Tylee, A., Haller, D. M., Graham, T., Churchill, R., & Sanci, L. A. (2007). Youth-friendly primary-care services: How are we doing and what more needs to be done? ***Lancet, 369***(9572), 1565–1573. https://doi.org/10.1016/S0140-6736(07)60371-7

UNICEF. (2016). MODULE 1: What are the Social Ecological Model (SEM), Communication for Development (C4D)? Retrieved from www.unicef.org/cbsc/files/Module_1_SEM-C4D.docx

Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? ***Child Development***, ***74***(1), 1628–1638.

Van Beijsterveldt, C. E. M., Bartels, M., Hudziak, J. J., & Boomsma, D. I. (2003). Causes of stability of aggression from early childhood to adolescence: A longitudinal genetic analysis in Dutch twins. ***Behavior Genetics, 33***(5), 591–605. https://doi.org/10.1023/A:1025735002864

Van Beijsterveldt, C. E. M., Groen-Blokhuis, M., Hottenga, J. J., Franić, S., Hudziak, J. J., Lamb, D., … Boomsma, D. I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. ***Twin Research and Human Genetics, 16***(01), 252–267. https://doi.org/10.1017/thg.2012.118

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice : Multivariate imputation by chained equations in r. ***Journal of Statistical Software, 45***(3), 1–68. https://doi.org/10.1109/TMAG.2010.2043064

Van der Valk, J. C., Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2003). Using shared and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems. ***Behavior Genetics, 33***(4), 409–420. https://doi.org/10.1023/A:1025369525924

Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). ***European Child and Adolescent Psychiatry, 12***(6), 281–289. https://doi.org/10.1007/s00787-003-0341-3

Veldkamp, S. A. M., Boomsma, D. I., De Zeeuw, E., van Beijsterveldt, C. E. M., Bartels, M., Dolan, C. V., & Van Bergen, E. (2019). ***Genetic and environmental influences on different forms of bullying perpetration, bullying victimization, and their co-occurrence.*** Manuscript submitted for publication.

Viner, R. M., Ozer, E. M., Denny, S., Marmot, M., Resnick, M., Fatusi, A., & Currie, C. (2012). Adolescence and the social determinants of health. ***The Lancet, 379***(9826), 1641–1652. https://doi.org/10.1016/s0140-6736(12)60149-4

Vink, J. M., Bartels, M., van Beijsterveldt, T. C. E. M., van Dongen, J., van Beek, J. H. D. A., Distel, M. A., … Boomsma, D. I. (2012). Sex differences in genetic architecture of complex phenotypes? ***PloS One, 7***(12), e47371. https://doi.org/10.1371/journal.pone.0047371

Vinkhuyzen, A. A. E., Van Der Sluis, S., De Geus, E. J. C., Boomsma, D. I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. ***Genes, Brain and Behavior, 9***(3), 276–287. https://doi.org/10.1111/j.1601-183X.2009.00554.x

*Von Sydow, K., Retzlaff, R., Beher, S., Haun, M. W., & Schweitzer, J. (2013). The efficacy of systemic therapy for childhood and adolescent externalizing disorders: A systematic review of 47 RCT. ***Family Process, 52***(4), 576–618. https://doi.org/10.1111/famp.12047

Wachs, T. D., Georgieff, M., Cusick, S., & Mcewen, B. S. (2014). Issues in the timing of integrated early interventions: Contributions from nutrition, neuroscience, and psychological research. ***Annals of the New York Academy of Sciences, 1308***, 89–106. https://doi.org/10.1111/nyas.12314

Walsh, S. D., Bruckauf, Z., & Gaspar, T. (2016). Adolescents at Risk: Psychosomatic health complaints, low life satisfaction, excessive sugar consumption and their relationship with cumulative risks (Innocenti Working Paper No. 2016–13). Florence.

Walsh, S. D., Molcho, M., Craig, W., Harel-Fisch, Y., Huynh, Q., Kukaswadia, A., … Pickett, W. (2013). Physical and mmotional health problems experienced by youth engaged in physical fighting and weapon carrying. ***PLoS ONE, 8***(2). https://doi.org/10.1371/journal.pone.0056403

Waltes, R., Chiocchetti, A. G., & Freitag, C. M. (2016). The neurobiological basis of human aggression: A review on genetic and epigenetic mechanisms. ***American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 171***(5), 650–675. https://doi.org/10.1002/ajmg.b.32388

Warnick, E. M., Bracken, M. B., & Kasl, S. (2008). Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: A systematic review. ***Child and Adolescent Mental Health, 13***(3), 140–147. https://doi.org/10.1111/j.1475-3588.2007.00461.x

Weinberger, A. H., Darkes, J., Del Boca, F. K., Greenbaum, P. E., & Goldman, M. S. (2006). Items as context: Effects of item order and ambiguity on factor structure. ***Basic and Applied Social Psychology, 28***(1), 17–26. https://doi.org/10.1207/ s15324834basp2801_2

Weissman, M. M., Pilowsky, D. J., Wickramaratne, P. J., Talati, A., Wisniewski, S. R., Fava, M., … STAR*D-Child Team. (2006). Remissions in maternal depression and child psychopathology: a STAR*D-child report. ***JAMA : The Journal of the American Medical Association, 295***(12), 1389–1398. https://doi.org/10.1001/jama.295.12.1389

*Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care. ***JAMA Psychiatry, 70***(7), 750. https://doi.org/10.1001/jamapsychiatry.2013.1176

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., … Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. ***American Psychologist, 72***(2), 79–117. https://doi.org/10.1037/a0040360

Welsh, J., Strazdins, L., Ford, L., Friel, S., O'Rourke, K., Carbone, S., & Carlon, L. (2015). Promoting equity in the mental wellbeing of children and young people: A scoping review. ***Health Promotion International, 30***, ii36-ii76. https://doi.org/10.1093/heapro/dav053

Wesseldijk, L. W., Bartels, M., Vink, J. M., van Beijsterveldt, C. E. M., Ligthart, L., Boomsma, D. I., & Middeldorp, C. M. (2018). Genetic and environmental influences on conduct and antisocial personality problems in childhood, adolescence, and adulthood. ***European Child and Adolescent Psychiatry, 27***(9), 1123–1132. https://doi.org/10.1007/s00787-017-1014-y

Wesseldijk, L. W., Fedko, I. O., Bartels, M., Nivard, M. G., van Beijsterveldt, C. E. M., Boomsma, D. I., & Middeldorp, C. M. (2016). Psychopathology in 7-year-old children: Differences in maternal and paternal ratings and the genetic epidemiology. ***American Journal of Medical Genetics Part B: Neuropsychiatric Genetics***. https://doi.org/10.1002/ajmg.b.32500

Wichers, M., Gardner, C., Maes, H. H., Lichtenstein, P., Larsson, H., & Kendler, K. S. (2013). Genetic innovation and stability in externalizing problem behavior across development: A multi-informant twin study. ***Behavior Genetics, 43***, 191–201. https://doi.org/10.1007/s10519-013-9586-x

*Wilson, D. B., Gottfredson, D. C., & Najaka, S. S. (2001). School-based prevention of problem behaviors: A meta-analysis. ***Journal of Quantitative Criminology, 17***(3), 247–272. https://doi.org/10.1023/A:1011050217296

*Wilson, S. J., & Lipsey, M. W. (2006). The effects of school-based social information processing interventions on aggressive behavior, Part II: Selected/indicated pull-out programs. ***Campbell Systematic Reviews, 6,*** 37. https://doi.org/10.4073/csr.2006.6

*Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. ***American Journal of Preventive Medicine, 33***(2), S130–S143. https://doi.org/10.1016/j.amepre.2007.04.011

*Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: a meta-analysis. ***Journal of Consulting and Clinical Psychology, 71***(1), 136–149. https://doi.org/10.1037/0022-006X.71.1.136

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. American ***Journal of Public Health, 82***(6), 816–820. https://doi.org/10.2105/AJPH.82.6.816

Wittchen, H. U., & Jacobi, F. (2005). Size and burden of mental disorders in Europe - A critical review and appraisal of 27 studies. ***European Neuropsychopharmacology, 15***(4), 357–376. https://doi.org/10.1016/j.euroneuro.2005.04.012

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., … Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. ***European Neuropsychopharmacology, 21***(9), 655–679. https://doi.org/10.1016/j.euroneuro.2011.07.018

World Health Organization. (2005). Mental Health Action Plan for Europe: Facing the Challenges , Building Solutions. WHO European Ministerial Conference on Mental Health, (12–15 January), 1–12. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0013/100822/edoc07.pdf

World Health Organization. (2007). Growth reference 5-19 years: Weight-for-age (5-10 years). Retrieved August 9, 2018, from http://www.who.int/growthref/who2007_weight_for_age/en/

World Health Organization. (2015). The European mental health action plan 2013-2020. The European Mental Health Action Plan. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0020/280604/WHO-Europe-Mental-Health-Acion-Plan-2013-2020.pdf

World Health Organization. (2016). Growing up unequal: Gender and socioeconomic differences in young people's health and well-being. https://doi.org/ISBN 987 92 890 1423 6

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. ***Genome Research, 17***, 1520–1528. https://doi.org/10.1101/gr.6665407