

Sex Differences in the Effects of Parents' Use of Corporal Punishment on Children's Aggressive Behavior

Tuba Topçuoğlu*
Manuel P. Eisner**
Denis Ribeaud***

Author Note

Asst. Prof. Dr. Tuba Topçuoğlu, Faculty of Law, Istanbul University, Istanbul, Turkey; Prof. Dr. Manuel P. Eisner, Institute of Criminology, University of Cambridge, Cambridge, UK; Dr. Denis Ribeaud, Department of Sociology, Swiss Federal Institute of Technology, Zurich, Switzerland.

We wish to acknowledge financial support for the study by the Swiss National Science Foundation, the Jacobs Foundation, the Swiss Federal Office of Public Health, the Canton of Zurich Ministry of Education, and the Julius Baer Foundation. We also thank all the parents, children and teachers who participated in the Zurich Project on the Social Development of children.

Correspondence concerning this article should be addressed to Tuba Topçuoğlu. E-mail: tuba.topcuoglu@gmail.com

* Istanbul University Faculty of Law, Turkey

** University of Cambridge, United Kingdom

*** Swiss Federal Institute of Technology, Zurich, Switzerland

Abstract

The current study investigates the sex differences in the link between parents' use of corporal punishment (CP) and the subsequent change in children's aggressive behaviour using the propensity score matching technique—a statistical tool used to correct selection bias in observational studies. Analyses based on data from 697 school children drawn from the first four waves of the Zurich Project on the Social Development of Children (z-proso) indicate significant sex differences: Boys who experience CP at age 9 display increased levels of aggressive behaviour problems during the following two years, whereas no significant association is found for girls. Findings suggest that etiological factors for aggressive behaviour may not be common in boys and girls, and thus somewhat different intervention and prevention strategies for child aggressive behaviour might be needed for boys and girls.

Key words: Sex differences, parental use of corporal punishment, aggressive behaviour, propensity score matching, European sample

Aggressive behavior problems that are developmentally normative in toddlers and pre-schoolers are problematic in elementary school-age children. Numerous longitudinal studies have shown that the frequency of aggressive behaviour peaks at school entry and then starts decreasing during the school years (Cairns, Cairns, Neckerman, Ferguson & Garipey, 1989; Nagin & Tremblay, 1999). However, children who display persistent behaviour problems before age 12 are at higher risk for child and juvenile delinquency than children who do not (Fergusson & Horwood, 1995; Loeber & Farrington, 2001a; 2001b; Offord, Lipman & Duku, 2001). It is thus crucial to identify causal risk factors which lead to aggressive behaviour problems during middle childhood. Parenting risk factors have consistently been highlighted by previous research as significant childhood predictors of aggressive behavior (Loeber & Stouthamer-Loeber, 1986). One of the most important child-rearing variables often linked to aggressive behavior relates to parents' use of corporal punishment (CP) (Gershoff, 2002). In the present study, parental use of CP is differentiated from physical abuse, and is defined as the "use of

physical force with the intention of causing a child to experience pain but not injury, for purposes of correction or control of the child's behavior" (Straus & Donnelly, 2001, p.4). Boys, in particular, might be differentially more vulnerable than girls to the adverse effects of CP. Moffitt (1993), in her widely-accepted dual taxonomy of antisocial behavior, argues that aggressive behavior is an early-onset, life-course persistent (LCP) type of antisocial behavior and that its causes are rooted in early childhood in children's inherited or acquired neuro-developmental risk factors (initially manifested as difficult temperament, low verbal ability, poor self-control, hyperactivity, impulsive personality, cognitive deficits) and poor parenting, as well as in cumulative interactions between difficult children and adverse child-rearing context. It is well-documented that males are differentially more exposed than females to risk factors for LCP type of antisocial behavior (Eme & Kavanaugh, 1995; Moffitt, Caspi, Rutter, & Silva, 2001). Because boys tend to have both more neurological deficits in executive and verbal functioning and are more likely to experience CP, there is a greater likelihood of interaction between boys' neurological deficits and adverse familial factors. As a result, they might have more difficulty in adjusting to adverse familial factors and be more vulnerable than females to the effects of CP for aggressive behavior (Murray, 2002).

Second, CP might also have stronger effects on boys due to greater male vulnerability to psychological stress. The male organism already suffers from greater exposure to biological deficits and hence is more vulnerable to a number of pre-, peri-, and post-natal factors and developmental disorders from the moment of conception (Eme & Kavanaugh, 1995; Moffitt et al., 2001). Because boys are genetically and biologically more vulnerable than girls, psycho-social stress due to the experience of CP represents an additional stressor for boys, leading to higher stress reactivity and lower ability to tolerate and cope with distressed states (Eme & Kavanaugh, 1995; Murray, 2002).

Finally, differential parental socialisation of aggression in boys may add to boys' biological weaknesses and partially account for boys' differential reactions to CP (Eme & Kavanaugh, 1995; Keenan & Shaw, 1997; Murray, 2002). In a review on parents' socialisation of boys and

girls, Keenan and Shaw (1997) found moderate support for the hypothesis that parents display qualitatively different responses to the same temperamental characteristics and externalising behaviors in boys and girls during early childhood. Parents' differential treatment of boys and girls may originate in part from parental beliefs about gender, as well as from differences in children's behavior problems. Because expression of anger is inconsistent with stereotypical beliefs about women, parents tend to respond negatively to girls' expression of anger while they tend to encourage aggression in boys (Dodge, Pettit & Bates, 1994). As a result, differential reinforcement of aggression may amplify existing behavioral tendencies in boys and exacerbate the adverse effect of CP on aggressive behavior.

Despite the interest amongst researchers, sex differences in the link between child-rearing factors, and particularly CP, and children's resultant aggressive behavior are not well-researched. Gershoff (2002), in her meta-analysis of the studies on the link between CP and aggressive and delinquent behavior found that the association between CP and the outcome got stronger as more boys were included in a study sample. However, most studies included in this review were cross-sectional. Cross-sectional studies on the effects of CP on children's adjustment often assume that the direction is flowing from the parent to the child. This assumption made about the direction of the causal relationship between CP and children's resultant behaviour—assumed to be moderated by child sex—may be incorrect. It is equally likely that the effect can run in the opposite direction, with more aggressive children eliciting more frequent use of CP from their parents.

Even if researchers establish a time order between the CP and the outcome variable using longitudinal data (e.g., Dornfeld & Kruttchnitt, 1992; Eamon, 2001; Lahey, Van Hulle, Keenan et al., 2008; Lengua, 2008), the actual causal direction cannot be safely established (Benjet & Kazdin, 2003). Judd and Kenny (2010) have shown that a lack of control for the direction of the causal relationship between the predictor and the criterion variable can have a dramatic effect on the sign and the magnitude of the estimated moderator effect. One way to establish the

causal direction between parenting and children's behavior adjustment is to control for children's prior behavior problems (Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000).

Few longitudinal studies attempted to rule out reverse causation by controlling for children's prior aggressive behavior in the investigation of the sex differences in the link between CP and later aggression, and these yielded mixed findings. Straus, Sugarman, and Giles-Sims (1997), for example, used data on 807 children aged six to nine years old in the 1988 assessment of the National Longitudinal Study of Youth, and found significant sex differences in the link between spanking and antisocial behavior. For both sexes, spanking frequency at time 1 was associated with higher levels of antisocial behavior two years later, but the link was stronger and more linear for boys than for girls.

Gunnoe and Mariner (1997) used data on 1,112 European American and African American children (aged 4 to 11 years old) drawn from the first two waves of the National Survey of Families and Households, and found that for girls, spanking predicted significantly *less* aggressive behavior five years later, whereas for boys the path was positive and non-significant. The multi-group analysis, however, indicated no significant sex difference in the link between spanking and aggressive behavior. However, both studies were based on ethnically mixed samples with significant numbers of African American children. Prior research on the effects of parents' use of CP on children's aggressive behaviour points to significant ethnic differences in American samples. More specifically, they suggest that CP might adversely affect behavior problems in European American children, and that it might have either a null or a deterrent effect in African American children (e.g., Gunnoe & Mariner, 1997; Lansford, Deater-Deckard, Dodge, Bates & Pettit, 2004). Gunnoe and Mariner further examined the sex differences in the link between CP and child behavior outcome separately within each ethnic group in their sample. They could not formally examine the child sex by ethnicity interaction due to limited sample size; instead they reported the covariate-adjusted effect of CP on child aggression from 16 sub-group regression analyses. Results indicated that sex differences were evident

in the European American group only, with stronger adverse effects for boys. Given the limited amount of research in this area, together with the fact that prior research mainly focuses on the U.S. samples, there is a need for continued research on the sex differences in children's developmental processes with regard to the effects of CP in other socio-cultural contexts. The primary aim in this paper was therefore to examine the sex differences in the link between CP and children's aggressive behavior using longitudinal data on a European sample, and to see whether findings observed for European American children similarly apply to children of European origin once children's prior aggressive behavior is accounted for.

Another important limitation of prior research relates to inadequate control for confounding. Because it is not ethical to randomly assign children to different levels of CP and then examine the differences in their behavior outcome, research on the effects of CP on children relies on observational studies in which children's experience of CP is not random but rather occurs in an ecological context of other risk factors (Belsky, 1980; Bronfenbrenner, 1986). Since CP is related to a wide range of disadvantageous individual, family, school and socio-economic risk factors which might also contribute to children's aggressive behavior (Xu, Tung & Dunaway, 2000), children who are physically punished will systematically differ from those who are not with respect to these co-occurring risk factors. Unless all significant differences in children's baseline characteristics are adequately accounted for by means of statistical controls, the change in children's aggressive behavior may not be due to CP per se but rather due to the effects of other factors correlated with CP.

Prior studies on the effects of CP rely mostly on regression-based models for covariate adjustment, and hence they control for a limited number of covariates (Topçuoğlu, 2011). An inadequate control of the differences between these children will lead to biased estimates of the overall effect of CP (i.e., selection bias). Limited control for confounding may also lead to biased estimates of the interaction effects, in the sense that, even if the hypothesis on the sex differences between the slopes may be correct, both the sign and the magnitude of the estimated slopes being

compared may be wrong. Further, selection bias might lead to incorrect conclusions about the moderation hypothesis if the source of this potential bias varies across the groups (i.e., differential selection bias). Thus, a second purpose of this paper was to investigate the sex differences in the prospective link between parents' use of CP and the subsequent change in children's aggressive behavior using an alternative technique which corrects for selection bias and allows for a larger number of covariates to be controlled than regression-based techniques.

The present study

The present study uses data on 697 school children drawn from the first four waves of the *Zurich Project on the Social Development of Children (z-proso)*, and investigates the sex differences in the risk status of CP. Based on prior research, we expect to find sex differences in the link between CP and child aggressive behavior and hypothesise that the relationship between CP and the change in aggressive behaviour will be stronger and more adverse for boys than for girls.

To test these hypotheses, we use a different analytic approach that addresses some of the limitations of the past research. We apply propensity score matching (PSM) approach to minimise bias due to selection. The PSM approach is conceptualised within the counterfactual framework and borrows the logic and language of experiments (Harding, 2003). The central idea in PSM is to reproduce the conditions of an experiment such that the treatment variable, in the present context the use of CP, can be treated as though it occurred at random, and children under analysis become similar on all baseline characteristics except the treatment itself (Rosenbaum & Rubin, 1983). This technique corrects for selection bias by creating a counterfactual between a treatment (i.e., CP) and a control (i.e., no CP) group. More specifically, this approach uses baseline observed covariates of a treatment variable to estimate each child's propensity to receive the treatment. Given that all baseline differences between the treatment groups are captured by their observables, matching on the propensity scores balances all the observed covariates

between different CP groups so that the treatment assignment can be considered as a random event (Rosenbaum & Rubin, 1983; 1985). As a consequence, the missing counterfactual mean (i.e., the mean outcome for a child who experienced CP had he or she not received CP) can be constructed from the outcomes of the non-treated children (i.e., no CP children) and the treatment effect can be computed by directly comparing the outcomes of the treated and the control children (Heckman, Ichimura & Todd, 1997).

This approach allows us to control for a large number of covariates including prior child behaviour problems (Harding, 2003). Unlike the regression-based models which ‘assume’ that the initial differences amongst children are properly accounted for once they are statistically controlled within the regression model, the PSM provides transparency on the selection bias before the estimation of the average treatment effects of experiencing CP (Brand & Halaby, 2006; Harding, 2003). Thus, children with similar baseline characteristics are compared, and any difference in children’s behaviour outcomes can be safely attributed to the CP experienced (Harding, 2003). Another advantage of matching over regression adjustment relates to the fact that while regression adjustment assumes linearity to account for selection bias, matching estimators are not model-based and hence do not depend on any arbitrary assumption regarding the functional form of the relationship between the covariates and the outcome variable (Brand & Halaby, 2006; Harding, 2003). This approach has been developed nearly three decades ago by Rosenbaum and Rubin (1983) as an alternative statistical tool to correct for selection bias in observational studies and has been effectively used across many disciplines (DeLisi, Barnes, Beaver & Gibson, 2009; Gibson, Miller, Jennings, Swatt & Gover, 2009; Haviland, Nagin & Rosenbaum, 2007; Theobald & Farrington, 2009).

Method

Participants

The *Zurich Project on the Social Development of Children* (*z-proso*) is an on-going prospective longitudinal study of 1,235 children who entered public primary school in the City of Zurich in year 2004. (For details of the study, see Eisner & Ribeaud, 2005). According to the school statistics, more than half of the parents were born outside Switzerland. Because of the ethnic diversity of the sample, extensive measures were taken to maximize the response rate. The questionnaires were translated into eight languages that were most frequently spoken in Zurich (e.g., Albanian, Serbo-Croatian, Portuguese, Italian, Turkish and Spanish) and respondents who were not competent in German have been interviewed in their own language. Bilingual interviewers from the same nationality of the respondents were provided for each language group.

Data were collected annually from the primary caregiver (mostly the mother) and the child. At the beginning of each school year, computer-assisted face-to-face interviews were initiated with the parents. Data from the children were obtained by means of computer-assisted personal face-to-face interviews in their schools. Every six months, teachers completed their assessments on children's academic achievement, social behaviour in the classroom, and on school problems by means of paper-and-pencil questionnaires. In the current study, teacher assessments that coincided with the timing of the annual parent and child assessments were used only. A written informed consent was obtained from the parents for the first three years of the project and needed to be renewed at wave 4. For their participation, parents were offered an incentive of about €25 per interview, whereas children were given a small present worth €3.

The first wave of the interviews was conducted in 2004 when children were about seven years of age. The second, third and the fourth waves were held in 2005 ($M_{\text{age}} = 8$ years), 2006 ($M_{\text{age}} = 9$ years), and 2008 ($M_{\text{age}} = 11$

years), respectively. The average retention rate was about 93 per cent between the first and the fourth waves.

The sub-sample used in the present study consisted of 701 children (1) whose parents participated in all four waves and responded in German at the parent interview, and (2) who had complete data on the child outcome measure at waves 3 and 4. The final sample reduced to 697 because four participants were not the primary caregiver of the child in at least one of four parent interviews. These participants were contrasted with the non-participants (i.e., those parents who participated in wave 1 and responded in German but had one or more data points missing) on child gender, family socio-economic status, baseline child aggression and CP. Only one significant difference emerged. Participants who had complete records in all waves had relatively higher socio-economic status ($M = 51.5$, $SD = 15.9$) than those missing one or more data points ($M = 48$, $SD = 16.7$), $t(810) = 2.186$, $p < .05$. Of the 697 children included in the final sample, 369 (52.9 per cent) were male, 150 (21.5 per cent) were from single-parent homes, and the index of socio-economic status based on wave 1 measurement (Ganzeboom, De Graaf & Treiman, 1992) ranged from 16 to 88, with a mean of 51.5 ($SD = 15.9$).

Dependent Variable: Change in Aggressive Behavior

Children's social behaviours were measured using a recent version of the *Social Behaviour Questionnaire* (SBQ; Tremblay et al., 1991). This instrument is repeatedly administered to primary caregivers, teachers and to the target child in each wave. The parent-SBQ is a 55-item questionnaire used to assess children's anxiety-depression, attention-deficit hyperactivity disorder, non-aggressive problem behaviour, direct and indirect aggressive behaviour and pro-social behaviour. For teachers, the parent-SBQ has been adapted in such a way that the 12 items which did not fit teachers' perspective were eliminated and the final questionnaire consisted of 43 items. Both caregivers and teachers responded on a 5-point Likert scale (1 "never", 2 "rarely", 3 "sometimes", 4 "often", and

5 “very often”) to indicate how often each child displayed each social behaviour within the last 12 months.

Children also self-reported on their behaviour problems using an adapted computer-based multimedia version of Tremblay's Social Behaviour Questionnaire – *Tom and Tina*. This instrument consists of a number of drawings which depict specific social behaviours of a child and, for each drawing, a voice recorded on the computer asks the child whether he/she happens to do what is shown on the drawing. The child responds using the “Yes” and “No” buttons at the bottom of each screen. At each assessment, children responded to 54 drawings, and their responses were coded as “1” if they pushed the ‘Yes’ button and “0” if they chose the ‘No’ button.

All the child behaviour measures used in the current study were computed using information from all three informants (i.e., parent, teacher and child). Initially, for each informant, mean scores were computed by averaging all the items belonging to the particular concept. These mean scores were converted into z scores first, and then the standardised mean scores were averaged across three informants. *Non-aggressive and aggressive behaviour* problems measured in wave 2 were used as covariates in the PSM analysis. *Aggressive behaviour* measured in waves 3 and 4 were used as the child outcome variables when examining the effect of CP. *Aggressive behaviour* scale measured children's physical, proactive- and reactive aggression and was measured with 12 items in both parent interviews and child assessments, and it was measured with 11 items in teacher assessments. The item examples for this scale are: gets into fights, physically attacks, cruel, bullies or is mean to others, threatens, tries to dominate other children, and reacts in an aggressive manner when teased/ something taken/ contradicted. The SBQ factors have been used in various studies with kindergarten, elementary school-aged children, and early to mid-adolescents demonstrating excellent reliability and validity (Nagin & Tremblay, 2001; Tremblay et al., 1991).

Predictor Variable: CP Group Membership

The CP groups were determined based on the parents' self-reported use of CP in wave 3 (mean age 9 years). In each wave, parents' use of CP was measured using the three items from the Alabama Parenting Questionnaire (Shelton, Frick & Wootton, 1996). Parents reported how often they "spanked", "slapped" or "hit their child with an object" during the previous year, using a five-point Likert scale (1 "never", 2 "rarely", 3 "sometimes", 4 "often" and 5 "always"). Children whose parents used any of these forms of CP were assigned to the "CP" group, whereas the remainder were assigned to the "no CP" group. Overall, 279 (40 per cent) children had experienced at least one form of CP during the preceding year. The prevalence rate for any CP was higher for boys (43 per cent) than for girls (37 per cent) but the difference was not statistically significant ($\chi^2(1) = 3.06, p > .05, OR = 1.3, 95\% CI [.968, 1.78]$).

Matching Covariates Used in the PSM Analysis

To minimise bias due to selection and reduce all systematic differences between children who received CP and those who did not, the PSM analysis included 20 matching covariates. Selection of these variables were based on the following criteria: (a) previous theoretical and empirical research have suggested that these variables might be associated with children's experience of CP (Belsky, 1980; Xu et al., 2000) and their aggressive behaviour (Farrington, 2002; Loeber, Farrington, Stouthamer-Loeber & Van Kammen, 1998), and therefore, if uncontrolled, might confound the effect of CP on child aggressive behaviour; and (b) they were all measured prior to the measurement of the treatment variable (i.e., CP) in wave 3, and hence were not an intermediate step in the causal path between children's exposure to CP and their behavioral outcome (Rothman, 1986).

Family socioeconomic status. Parents' occupations were initially classified according to the International Standard Classification of Occupations (ISCO) (Elias, 1997) using data obtained in wave 1 on parents'

education and occupation. These ISCO scores were then translated into ISEI-scores (International Socio-Economic Index of Occupational Status) (Ganzeboom et al., 1992), and family socioeconomic status was then defined as the highest occupational status in the household.

Intact family. Caregivers were asked whether they were living with a partner in wave 2. Children were coded as “1” if they had an intact family in wave 2 and “0” otherwise.

Triple P & PATHS. As part of the Zurich project, two prevention programmes were implemented during the first and the second years of the study: a family-based parenting skills programme—Positive Parenting Program, Triple P (Sanders, 1999)—and a school-based social skills programme—Promoting Alternative Thinking Strategies, PATHS (Greenberg & Kusche, 1998). Schools were randomly allocated to one of the four treatment conditions: (1) Parents were offered to participate in Triple P, (2) PATHS was implemented in all the classes, (3) both parents were offered to participate in Triple P and children participated in PATHS, and (4) no prevention programme was implemented (see Eisner, Ribeaud, Junger & Meidert, 2007 for details). Overall, of the 1,235 parents who participated in the parent interview in wave 1, 568 (33.9%) were assigned to the Triple P condition. Of the 1,361 children who participated in the child interview in wave 1, 666 (48.9%) were assigned to the PATHS condition.

These treated families were not excluded in the present study mainly because prior research has not revealed a clear evidence that overall the Triple P and PATHS improved parental behaviour and children's resultant problem behaviours (Eisner et al., 2007). Nevertheless, two additional controls were introduced in order to take account of the possible effects of these interventions on children and their parents. Children were coded as “1” if the child's primary caregiver participated in Triple P and completed at least three sessions out of four and “0” otherwise. Similarly, children were coded as “1” if the child participated in the school-based social skills program PATHS and “0” otherwise.

Parental psychological well-being. Information on parental mental health problems was assessed using six negatively worded items from the 12-item General Health Questionnaire (GHQ-12; Goldberg, 1978) administered in wave 2. The scale has been constructed by taking the average of all six items. Higher scores indicate higher levels of psychological strain (Cronbach's alpha = 0.81).

Social networks. Caregivers' social networks and the level of everyday interaction with neighbours have been measured with five items in wave 1. Parents were asked to indicate how often they "helped a neighbour with a minor problem (e.g., repair something, help out with some food, etc.)", "watched their neighbour's property when they were out of town", "had lunch or dinner with a neighbour", "talked to neighbours about personal things", and "have taken care of neighbours' children while the neighbours were away from home". Parents responded on a 4-point Likert scale (1 "never", 2 "rarely", 3 "sometimes", and 4 "often"). The scale has been constructed by taking the average of all five items; higher mean scores indicate better social networks (Cronbach's alpha = 0.82).

Parental low self-control. Information on primary caregiver's self-control, as formulated by Gottfredson and Hirshi in their *General Theory of Crime* (1990), was assessed in wave 2 using a modified version of the Grasmick, Tittle, Bursik and Arneklev (1993) scale that consisted of 12 items. The scale has been constructed by taking the average of all the 12 items, and higher mean values indicate lower levels of parental self-control (Cronbach's alpha = .69).

Traditional parenting values. Caregivers' values which underlie their parenting goals have been measured in wave 1 with eight items, partially adapted from the World Values Survey (Inglehart, Basanez & Moreno, 1998) and developed by the z-proso project team. The scale has been constructed by taking the average of all the eight items; higher mean scores indicate more traditional parenting values (Cronbach's alpha = 0.82).

Parenting practices of parents. Caregivers' parenting practices were assessed using the 40-item *Alabama Parenting Questionnaire* (APQ; Shelton et al., 1996). In the present study, five aspects of parenting were used: positive parenting (five items), inconsistent discipline (six items), poor monitoring (10 items), yelling (one item) and reasoning (one item) in disciplining the child. At each parent interview, parents were asked to indicate how often they used each parenting practice during the last 12 months using a 5-point Likert scale (1 "never", 2 "almost never", 3 "sometimes", 4 "often", and 5 "always"). Except for yelling and reasoning, the scales were computed as the average of the items that comprised each scale. (Cronbach's alpha was 0.62 for positive parenting, 0.58 for inconsistent discipline and 0.68 for poor monitoring in wave 2). Numerous studies examined the psychometric properties of the APQ scale, and results indicate that overall the APQ scale yields factors that are reliable and valid in diverse samples from different countries (Clerkin, Marks, Policaro & Halperin, 2007; Essau, Sasagawa & Frick, 2006; Shelton et al., 1996).

Having sibling(s) and sibling's physical aggression. In wave 2, parents indicated whether the target child had a sibling who were born before 2003 (that is, older than two years old at the time of the interview) and were living at home. *Having sibling(s)* was coded as "1" if there was at least one child at home other than the target child in wave 2 and "0" otherwise. Then, in a series of follow-up questions, parents who replied that there was at least one other child at home also responded on the physical aggression of each sibling, using the three items from the Physical aggression subscale of the SBQ (Tremblay et al., 1991). Up to seven siblings, for each sibling at home, parents were asked to indicate how often 1) the sibling gets into fights, 2) kicks, bites, and hits other children, and 3) physically attacks other people. They responded on a five-point Likert scale (1 "never", 2 "rarely", 3 "sometimes", 4 "often", and 5 "very often"). Physical aggression of each sibling is computed as the mean of these three items. Internal consistency of the scale was adequate (Cronbach's alpha was 0.77 for the first sibling). *Sibling's physical aggression* was measured with the first sibling's mean physical aggression score

and if the child had more than one sibling living at home, the sibling who had the highest level of physical aggression was considered.

Child's bullying victimisation (prevalence) at school. In wave 2, children were asked whether they have been physically and materially victimised by other students at school. Each question was supported with a verbal description and a picture depicting the specific bullying situations provided by the "Pathways to Victimisation in Kindergarten" study (Alsaker & Valkanover, 2001). If children chose "no", then they were further asked whether they were really sure about their answer. Initially, children's responses were coded as "1" when they spontaneously answered "yes" to the first question, "2" when they initially answered "no" but then replied as "yes" upon inquiry, and "3" when they answered "no" on both occasions. Both items were dichotomised by collapsing the first two values (i.e., "1" and "2"). *Physical violence victimisation* was coded as "1" if the child reported that he/she had physical violence victimisation at school, and "0" otherwise. Similarly, *property victimisation* was coded as "1" if the child reported that his/her property was stolen or destroyed at school and "0" otherwise. Both questions referred to the period since the last summer holidays which covered approximately the previous 2-3 months.

Child's verbal competence. Children's language proficiency in Swiss German was measured by teachers' ratings in wave 2 on a single item. Teachers reported on a 5-point Likert-type scale and indicated whether the child's language competency was "much worse" ("1"), "somewhat worse" ("2"), "equal" ("3"), "somewhat better" ("4") or "much better" ("5") when compared with other children in class.

Prior aggressive and non-aggressive child behaviour. Children's prior aggressive and non-aggressive behaviour problems were measured using the SBQ (Tremblay et al., 1991) in wave 2. *Non-aggressive behaviour* problems measured children's non-aggressive conduct disorder and oppositional defiant behaviour and was measured with nine items in both parent interviews and child assessments, while it was measured with six items in teacher assessments. The item examples referring to this scale

are: ignores, tells lies and cheats, steals and disobedient at school. *Aggressive behaviour* was measured identically as the child outcome measure in waves 3 and 4.

The Analytic Plan

The main analyses consist of three parts. In the first part, using a large number of covariates (including children's baseline aggression), we apply PSM and remove all systematic differences between children who experienced CP and those who did not. For this purpose, we first group children by sex, and then within each sex group a propensity score for each child is estimated by means of a binary logistic regression where the dependent variable indicates whether or not the child experience any CP at Grade 3, and where the independent variables are the observed covariates measured at Grades 1 and 2. Once this procedure generates a propensity score for each child that is bound between zero and one, then a matching algorithm¹ pairs the treated (i.e., CP children) and the control (i.e., no CP) children within each sex group based on their probability of experiencing CP. Matching was carried out using the PSMATCH2 program in Stata developed by Leuven and Sianesi (2003). Once matching is completed, we evaluate the success of the matching procedure and check whether any systematic differences between the treated and the control children remain after matching. If balance is achieved on the observed covariates, this suggests that these confounding variables can no longer bias the effect of CP on child aggressive behaviour (Rosenbaum & Rubin, 1983). Removing all systematic differences between children who experience CP and those who do not in this way produces groups of

¹ For the purpose of matching, a local linear matching was carried out with a normal kernel function and a bandwidth value of 0.4. The local linear regression technique applies non-parametric regression to calculate a weighting function and then performs one-to-many matching and matches each treated child to a weighted average of all the possible control children within a span predetermined by the bandwidth parameter (Guo and Fraser, 2010). A bandwidth value of 0.4 indicates that, in the calculation of the weighted average of the control children for each treated child, only information from 40 per cent of the control children is used. In so doing, this algorithm uses more information from the control children whose propensity scores are closer to that of the treated child and less information from more distant control children.

children statistically similar with respect to key factors that might otherwise confound the effect of CP. Children's experience of CP can then be considered as though it occurred at random, and because these children differ only in their experience of CP, any difference in children's behaviour problems can be attributed to the CP they experienced. Therefore, in the second part, after correcting for selection separately for boys and girls, we compare the average treatment effect for the treated (CP) children in boys and girls to see whether CP leads to a similar change in child aggressive behaviour in boys and girls. Finally, we carry out a moderated regression analysis on the pooled matched data set and formally test whether the observed sex differences hold statistically.

Results

Propensity scores conditioning on 20 covariates were estimated by means of two separate binary logistic regressions (i.e., one for boys and one for girls) where the outcome variable indicated whether or not the child experienced CP at Grade 3. After the estimation of the propensity scores in each model, treated (CP) boys and treated (CP) girls were matched to the control (no CP) boys and control (no CP) girls, respectively. Table 1 displays the standardised bias statistics before and after matching separately for each covariate within each sex group (see Table 1). For each covariate, the standardised difference in per cent is calculated as the mean difference between the CP and the no CP groups as a percentage of the average standard deviation (Rosenbaum & Rubin, 1985). Rosenbaum and Rubin (1985) suggest that standardised mean differences² (d) greater than 20 indicate significant selection. Accordingly, these results suggest substantial initial differences between the control and the treatment groups in boys and girls. Matching, however, successfully removed all the bias in the covariates for both boys and girls since treated children were comparable to the control children on each covariate (see Table 1).

² The standardised difference in per cent is the mean difference as a percentage of the average standard deviation: $100(\bar{x}_t - \bar{x}_c) / [(s_t^2 + s_c^2)/2]^{1/2}$, where for each covariate, \bar{x}_t and \bar{x}_c are the sample means in the treated and the control groups, respectively, and s_t^2 and s_c^2 are the corresponding sample variances (Rosenbaum and Rubin, 1985).

Table 1 Standardised bias statistics[†] before and after matching

	BOYS		GIRLS	
	No CP vs. CP		No CP vs. CP	
	Before	After	Before	After
Property victimisation	13.4	13.7	-7.4	-5.6
Family SES	-29.8	-1.8	-13.6	-3.6
Child's language competency	-12.2	-1.3	-26.2	1.1
Physical violence victimisation (<i>no</i> = 0, <i>yes</i> = 1)	17.3	-2.1	8.2	2.3
Child participated PATHS	-5.6	-14.3	-28.5	-4.2
Child has sibling at wave 2 (<i>no</i> = 0, <i>yes</i> = 1)	25.6	-1.0	1.7	6.7
Sibling's physical aggression	23.6	3.4	17.8	5.2
Positive parenting	19.2	12.1	33.9	13.7
Inconsistent discipline	23.3	4.9	43.0	-18.4
Poor monitoring	-8.2	-1.9	-16.2	-3.8
Yelling/ screaming	39.2	0.1	48.7	-2.2
Reasoning	-17.3	-0.7	-17.7	4.1
Non-aggressive behaviour problem (wave 2)	22.4	12.2	26.5	2.5
Aggressive behaviour (wave 2)	23.0	14.2	36.0	0.9
Parent's social networks	0.2	3.9	-29.2	5.7
Traditional parenting values	28.5	14.5	38.6	-1.2
Parent has partner (<i>no</i> = 0, <i>yes</i> = 1)	27.4	-3.4	10.4	-12.6
Parental low self-control	3.0	5.3	13.4	-4.5
Parent's psychological well-being	8.2	8.1	20.0	4.0
Parent participated triple P	-4.9	-8.5	-3.2	6.2

† The standardised difference in per cent is the mean difference as a percentage of the average standard deviation: $100(\bar{x}_t - \bar{x}_c) / [(s_t^2 + s_c^2)/2]^{1/2}$, where for each covariate, \bar{x}_t and \bar{x}_c are the sample means in the treated and the control groups, respectively, and s_t^2 and s_c^2 are the corresponding sample variances (Rosenbaum and Rubin, 1985).

In the next step, we examine the average treatment effect for the treated (i.e., CP) children separately in boys and girls to see whether CP experienced at age 9 was associated with a similar change in boys' and girls' aggressive behaviour between ages 9 and 11. Here, we conduct a difference-in-differences analysis (Heckman et al., 1997) and use the differences in child aggression scores between Grade 3 and 4 as the outcome measure. Within each sex group, the average treatment effect for the treated is thus calculated as the difference between the average change in

aggression scores between the two grades for the treated (CP) children and the weighted average change in aggression scores for the matched control (no CP) children (Guo & Fraser, 2010).

Figure 1 displays the change in aggression scores between Grades 3 and 4 for CP and no CP children separately for boys and girls. The mean aggression score for CP boys increased from Grade 3 to Grade 4 by 0.090 units, and the mean score for the no CP boys decreased from Grade 3 to Grade 4 by 0.111 units. The mean difference between these groups was 0.201, meaning that the average change for aggressive behaviour for CP boys was 0.201 units higher (or worse) than that for no CP boys. These results were statistically significant as indicated by the 95 per cent confidence intervals (CI) (95% CI = 0.017, 0.326). As can be seen in Figure 1, results were different for girls. The mean aggression score for CP girls decreased from Grade 3 to Grade 4 by 0.093 units, and the mean score for no CP girls decreased from Grade 3 to Grade 4 by 0.028 units. The mean difference between these groups was -0.065, meaning that the average change for aggressive behaviour for CP girls 0.065 units lower (or better) than that for no CP girls. These results, however, were not statistically significant (95 per cent CI = -0.193, 0.094).

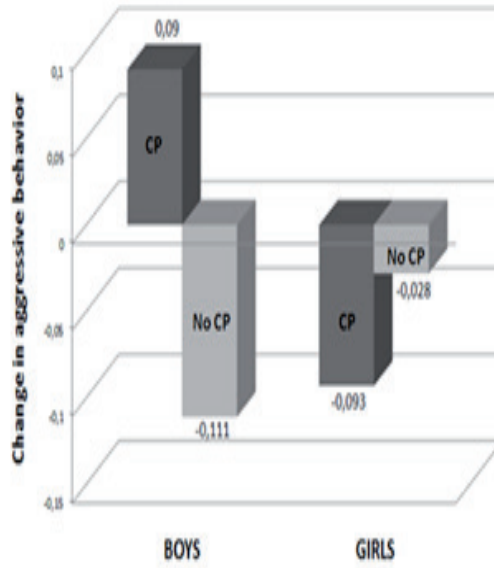


Figure 1. Mean change in children's aggressive behaviour scores from Grade 3 to Grade 5 for the CP and the matched No CP children by sex. CP = Corporal punishment, No CP = No corporal punishment.

Finally, to test formally whether the observed sex differences hold statistically, in a follow-up analysis, we examined whether the effect of CP on the change in child aggressive behaviour between grades 3 and 4 significantly differed between boys and girls. The matched data set for girls and boys were combined, and then the interaction test was carried out by means of a moderated regression analysis on the matched sample (Aguinis, 2004). The analysis was carried out with 558 children, and this included 159 CP boys and 120 CP girls, together with their matched controls (i.e., 159 no CP boys and 120 no CP girls). Differences in children's aggressive behaviour between grades 3 and 4 were regressed on child gender (0/1), CP group membership (CP vs. no CP), and an interaction term between child gender and CP group membership. Results indicated that the effect of experiencing CP at grade 3 on the change in children's aggressive behaviour between the grades 3 and 4 significantly differed

between boys and girls (B for the interaction term = 0.266, $p < .001$). CP was significantly positively associated with an increase in aggressive behaviour for boys ($B = 0.210$, $p < 0.01$), but this association was negative and not significant for girls ($B = -0.065$, $p > 0.05$).

Discussion

One of the most robust findings in deviance research relates to significant sex differences in aggression (Keenan & Shaw, 1997; Moffitt et al., 2001). Sex differences in aggressive behaviour may in part result from children's 'differential susceptibility' to certain risk factors (Keenan & Shaw, 1997). That is, children's developmental processes might be different, and some risk factors might have a differential effect on later aggressive behaviour depending on the child's sex (Deater-Deckard & Dodge, 1997; Gershoff 2002; Keenan, Loeber & Green, 1999). The present study examined the sex differences in the prospective link between parents' use of CP and the change in child aggressive behaviour in a European sample using the PSM technique to correct for selection bias. We hypothesised that there would be significant sex differences in the risk status of CP for aggressive child behaviour, and that CP would have a stronger and a more adverse effect on boys' aggressive behaviour. Our findings provided support for these hypotheses. Overall, these findings indicated significant sex differences in children's responses to parental use of CP. While there was evidence for a deterrent but insignificant effect of CP for girls, CP had an adverse and a stronger effect on boys' aggressive behaviour between ages 9 and 11.

Our findings regarding the sex differences are consistent with other longitudinal studies based on American children—with European Americans comprising the whole or the majority of the sample—which show that boys display a greater association between various aspects of parental caregiving and later antisocial behavior even after controlling for children's initial behavior problems. Boys' vulnerability has been found with respect to parents' physically negative behavior (Webster-Stratton, 1996), coercive parenting which included the use of physical punishment

(McFadyen-Ketchum, Bates, Dodge, & Pettit, 1996), and attachment relationship (Renken, Egeland, Marvinney, Mangelsdorf, & Sroufe, 1989). Our findings are also consistent with the previous longitudinal findings on the sex differences in the effect of CP on European American children's aggressive behavior (Gunnoe & Mariner, 1997).

It is, however, important to note that our findings do not parallel those revealed by the only study which investigated the effect of CP using the same analytical technique we employed here. Using data from the Project on Human Development in Chicago Neighbourhoods study, Morris and Gibson (2011) compared children who were physically punished with those who were not after matching them on their background characteristics by means of a PSM analysis. Results indicated that CP was not associated with later aggression regardless of children's sex or ethnic background after accounting for selection bias. This might in part be due to the fact that European Americans comprised less than one fifth of the sample and no further analysis was carried out to examine the Sex^x Ethnicity interaction. Also, one limitation of this study was that there was no clear time precedence between the covariates used in the matching analysis, including children's prior behavior problems, and the predictor (i.e., CP); therefore, it is possible that the covariates might have mediated some of the effect of CP on later aggression.

Given that we found significant sex differences in the link between CP and child aggression in our sample, it is important to clarify the nature of these differences. Although we hypothesised to find stronger effects of CP on aggressive behavior in boys—due to boys' differential exposure to neuropsychological deficits and poor parenting, and hence a greater likelihood of having multiplicative effects between the two, due to differential reaction to psychosocial stress because of extra male biological vulnerability, and due to sex differences in parental socialisation of aggression in boys—we have not examined what actually moderates the effect of CP in these children. Because boys' early vulnerability may escalate into a wider range of criminal behavior and a disproportionate degree of violence, further research is needed to untangle the causal me-

chanism between CP and aggressive behavior and those factors which increase the likelihood of an aggressive reaction to CP in boys.

Our results indicate that CP does not seem to be a risk factor for female aggressive behavior. It is still possible that girls are as just as vulnerable to the adverse effects of CP, but they have different manifestations of aversive outcomes (Eme & Kavanaugh, 1995; Zahn-Waxler, 1993). Further, in certain theories of female aggression, it is suggested that females face greater barriers to aggression and therefore must experience worse levels of the risk factors than males before becoming aggressive (Broidy & Agnew, 1997; Moffitt et al., 2001). Thus, future research should examine the effect of CP on other child behavior outcomes in girls and also the dose-response relationship to see whether there is a curvilinear relationship between CP and aggressive behavior in girls, with CP leading to aggressive behavior at very high and severe levels.

With these findings in mind, it is important to point out a number of limitations which might constrain the conclusions that can be drawn from these findings. First, in the present study, the sample was restricted to school-age children, and these findings may not generalise to adolescents or pre-school children. Socialization practices might have differential effects on children at different developmental periods (Frick, Christian, & Wootton, 1999). As parents use CP less frequently when the child becomes an adolescent, the meaning of CP to the child might change such that it might have stronger effects on adolescents' aggressive behavior (Loeber et al., 2000). This may be particularly important for girls because even if they are exposed to similar early risk factors, they might have a 'delayed onset' (Silverthorn & Frick, 1999). Thus, future research should examine other developmental stages and see whether the sex differences found here can be generalised to other developmental periods.

Second, although our sub-sample included parents who responded in German during the interviews, according to school data in 2002, in 27 per cent (190) of these families the primary caregiver did not have German as their mother language. However, an important aspect of minority

parents' parenting practices relates to their acculturation levels (Dumka, Roosa & Jackson, 1997; Kim, Cain & McCubbin, 2006). Thus, taking high competency in German as an indicator of the level of acculturation, it is plausible to expect these parents to adopt and use parenting practices in a way similar to Swiss parents.

A third limitation is related to the problem of hidden bias. Although the moderated regression analysis was based on the pooled matched sample and therefore the slopes were adjusted for all 20 covariates, our findings are based on the assumption that all systematic differences between the CP groups were accounted for (Rosenbaum & Rubin, 1985). Since there are variables that are unavailable and unobservable to the researcher, hidden bias is always a potential problem. The PSM analysis can account for unobserved covariates only to the extent that they are highly correlated with the observed ones (Rosenbaum & Rubin, 1985).

Finally, although the present analysis was longitudinal in nature, CP was measured at one point in time. A more causally rigorous way of estimating the effect of CP would necessitate measuring *both* CP and child behavior outcome at two points in time and then examining how a within-individual change in CP would be related to a subsequent within-individual change in child aggression (Farrington, Loeber, Yin, & Anderson, 2002). Future research should replicate these findings relying on within-individual lagged correlations.

Despite these limitations, however, the current study has a number of strengths. A majority of the previous studies on the effect of CP on children were conducted in the U.S. and hence included samples of American children. This study focused on an urban European city and added to the dearth of longitudinal research evidence on the sex differences in the effect of CP on children's resultant aggressive behavior in other socio-cultural contexts. Second, the sex differences in the link between CP and aggressive behavior have been examined using an analytical tool that has not been employed by previous research within this context (but also see Morris & Gibson, 2011). Despite its limitations, PSM provides a useful analytical procedure by creating experiment-like conditions in

the absence of experimentation. Unlike the normal statistical equating tools, this procedure ‘transparently’ minimised selection bias in each group prior to estimating the effect of CP (Rosenbaum & Rubin, 1983). This procedure also allowed us to use a wider set of covariates than used by prior research, and hence provided better control for selection bias. We also included children’s baseline aggression scores as covariate and hence partially controlled for reverse causation. Unlike the majority of the previous studies on the subject, we also established a clear time order between the risk factor and each covariate. Therefore, the possibility that the covariates can mediate some of the effect of CP on aggressive behavior has been ruled out.

In conclusion, the current study extended prior research by prospectively examining the sex differences in the link between parental use of CP and the change in school children’s aggressive behavior within the European context. Our findings reveal that boys and girls show important differential association between CP and later aggressive behaviour. For boys, CP at age 9 was associated with an increase in aggressive behavior over the following two years. No significant association emerged for girls however. Therefore, aggressive behavior in boys and girls may require entirely different explanatory models and somewhat different intervention strategies might be needed for boys and girls.

References

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: The Guilford Press.
- Alsaker, F. D. & Valkanover, S. (2001). Early diagnosis and prevention of victimisation in kindergarten. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 175-195). New York, NY: Guilford Press.
- Belsky, J. (1980). Child maltreatment: An ecological integration. *American Psychologist*, 35(4), 320-335.

- Benjet, C. & Kazdin, A. E. (2003). Spanking children: The controversies, findings and new directions. *Clinical Psychology Review, 23*, 197-224.
- Brand, J. E. & Halaby, C. N. (2006). Regression and matching estimates of the effects of elite college attendance on educational and career achievement. *Social Science Research, 35*, 749-770.
- Broidy, L. M. & Agnew, R. (1997). Gender and crime: A general strain theory perspective. *Journal of Research in Crime and Delinquency, 34*(3), 275-306.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*(6), 723-742.
- Cairns, R. B., Cairns, B. D., Neckerman, H. J., Ferguson, L. L. & Garipey, J. L. (1989). Growth and aggression, I: childhood to early adolescence. *Developmental Psychology, 25*(2), 320-330.
- Clerkin, S. M., Marks, D. J., Policaro, K. L. & Halperin, J. M. (2007). Psychometric properties of the Alabama parenting questionnaire—Presechool revision. *Journal of Clinical Child and Adolescent Psychology, 36*(1), 19-28.
- Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M. & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist, 55*(2), 218-232.
- Deater-Deckard, K. & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender. *Psychological Inquiry, 8*(3), 161-175.
- DeLisi, M., Barnes, J. C., Beaver, K. M. & Gibson, C. L. (2009). Delinquent gangs and adolescent victimization revisited: A propensity score matching approach. *Criminal Justice and Behavior, 36*(8), 808-823.
- Dodge, K. A., Pettit, G. S. & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*, 649-665.

- Dornfeld, M. & Kruttschnitt, C. (1992). Do the stereotypes fit? Mapping gender-specific outcomes and risk factors. *Criminology*, 30(3), 397-419.
- Dumka, L. E., Roosa, M. W. & Jackson, K. M. (1997). Risk, conflict, mothers' parenting and children's adjustment in low-income, Mexican immigrant and Mexican American families. *Journal of Marriage and Family*, 59(2), 309-323.
- Eamon, M. K. (2001). Poverty, parenting, peer and neighbourhood influences on young adolescent antisocial behavior. *Journal of Social Service Research*, 28(1), 1-23.
- Eisner, M. P. & Ribeaud, D. (2005). A randomised field experiment to prevent violence: The Zurich intervention and prevention project at schools, Zipps. *European Journal of Crime, Criminal Law and Criminal Justice*, 13(1), 27-43.
- Eisner, M. P., Ribeaud, D., Junger, R. & Meidert, U. (2007). *Frühprävention von Gewalt und Aggression. Ergebnisse des Zürcher Interventions und Präventionsprojektes an Schulen*. Zürich, Switzerland: Rüegger Verlag.
- Elias, M. J. (1997). Expanding our conceptualization of parent-child interrelatedness and the pathways toward 'externalizing' behaviors. *Psychological Inquiry*, 8(3), 186-191.
- Eme, R. F. & Kavanaugh, L. (1995). Sex differences in conduct disorder. *Journal of Clinical Child Psychology*, 24(4), 406-426.
- Essau, C. A., Sasagawa, S. & Frick, P. J. (2006). Psychometric properties of the Alabama parenting questionnaire. *Journal of Child and Family Studies*, 15(5), 597-616.
- Farrington, D. P. (2002). Developmental criminology and risk-focused prevention. In M. Maguire, R. Morgan, & R. Reiner (Eds.), *The Oxford handbook of criminology* (3rd ed., pp. 657-701). New York, NY: Oxford University Press.
- Farrington, D. P., Loeber, R., Yin, Y. & Anderson, S. J. (2002). Are within-individual causes of delinquency the same as between-individual causes? *Criminal Behavior and Mental Health*, 12, 53-68.

- Fergusson, D. M. & Horwood, L. J. (1995). Early disruptive behaviour, IQ, and later school achievement and delinquent behaviour. *Journal of Abnormal Child Psychology*, 23(2), 183-199.
- Frick, P. J., Christian, R. E. & Wootton, J. M. (1999). Age trends in the association between parenting practices and conduct problems. *Behavior Modification*, 23(1), 106-128.
- Ganzeboom, H. B. G., De Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1-56.
- Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, 128(4), 539-611.
- Gibson, C., Miller, J. M., Jennings, W. G., Swatt, M. & Gover, A. (2009). Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note. *Justice Quarterly*, 26(4), 625-643.
- Goldberg, D. P. (1978). *Manual of the general health questionnaire*. Windsor, UK: NFER.
- Gottfredson, M. R. & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Grasmick, H. G., Tittle, C. R., Robert J. Bursik, J. & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency*, 30(1), 5-29.
- Greenberg, M. T. & Kusche, C. A. (1998). *Promoting alternative thinking strategies*. Institute of Behavioral Sciences, University of Colorado.
- Gunnoe, M. L. & Mariner, C. L. (1997). Toward a developmental-contextual model of the effects of parental spanking on children's aggression. *Archives of Pediatrics and Adolescent Medicine*, 151(8), 768-775.
- Guo, S. & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage.

- Harding, D. J. (2003). Counterfactual models of neighbourhood effects: The effect of neighbourhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, 109(3), 676-719.
- Haviland, A., Nagin, D. S. & Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods*, 12(3), 247-267.
- Heckman, J., Ichimura, H. & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies*, 64(4), 605-654.
- Inglehart, R., Basanez, M. & Moreno, A. (1998). *Human values and beliefs: A cross-cultural sourcebook*. Ann Arbor, MI: The University of Michigan Press.
- Judd, C. M. & Kenny, D. A. (2010). Data analysis in social psychology: Recent and recurring issues. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 115-139). New York, NY: Wiley.
- Keenan, K., Loeber, R. & Green, S. (1999). Conduct disorder in girls: A review of the literature. *Clinical Child and Family Psychology Review*, 2(1), 3-19.
- Keenan, K. & Shaw, D. S. (1997). Developmental and social influences on young girls' early problem behavior. *Psychological Bulletin*, 121(1), 95-113.
- Kim, E., Cain, K. & McCubbin, M. (2006). Maternal and paternal parenting, acculturation and young adolescents' psychological adjustment in Korean American families. *Journal of Child and Adolescent Psychiatric Nursing*, 19(3), 112-129.
- Lahey, B. B., Van Hulle, C. A., Keenan, K., Rathouz, P. J., D'Onofrio, B. M., Rodgers, J. L. & Waldman, I. D. (2008). Temperament and parenting during the first year of life predict future child conduct problems. *Journal of Abnormal Child Psychology*, 36(8), 1139-1158.

- Lansford, J. E., Deater-Deckard, K., Dodge, K. A., Bates, J. E. & Pettit, G. S. (2004). Ethnic differences in the link between physical discipline and later adolescent externalizing behaviors. *Journal of Child Psychology and Psychiatry*, 45(4), 801-812.
- Lengua, L. J. (2008). Anxiousness, frustration, and effortful control as moderators of the relation between parenting and adjustment in middle-childhood. *Social Development*, 17(3), 554-577.
- Leuven, E. & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing (Version 3.1.5). Retrieved from <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Loeber, R. & Farrington, D. P. (2001a). Executive summary. In R. Loeber and D. P. Farrington (eds), *Child delinquents: Development, intervention and service needs*. Thousand Oaks, CA: Sage Publications.
- Loeber, R. & Farrington, D. P. (2001b). The significance of child delinquency. In R. Loeber and D. P. Farrington (eds), *Child delinquents: Development, intervention and service needs*. Thousand Oaks, CA: Sage Publications.
- Loeber, R. & Stouthamer-Loeber, M. (1986). Family factors as correlates and predictors of juvenile conduct problems and delinquency. *Crime and Justice*, 7, 29-150.
- Loeber, R., Farrington, D. P., Stouthamer-Loeber, M. & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Loeber, R., Drinkwater, M., Yin, Y., Anderson, S. J., Schmidt, L. C. & Crawford, A. (2000). Stability of family interaction from ages 6 to 18. *Journal of Abnormal Child Psychology*, 28(4), 353-369.

- McFadyen-Ketchum, S. A., Bates, J. E., Dodge, K. A. & Pettit, G. S. (1996). Patterns of change in early childhood aggressive-disruptive behavior: Gender differences in predictions from early coercive and affectionate mother-child interactions. *Child Development*, 67(5), 2417-2433.
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100(4), 674-701.
- Moffitt, T. E., Caspi, A., Rutter, M. & Silva, P. A. (2001). *Sex differences in antisocial behavior: Conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. Cambridge, UK: Cambridge University Press.
- Morris, S. Z. & Gibson, C. L. (2011). Corporal punishment's influence on children's aggressive and delinquent behavior. *Criminal Justice and Behavior*, 38, 818-839.
- Murray, J. (2002). *Explaining sex differences in offending behavior* (Unpublished M.Phil. thesis). University of Cambridge, UK.
- Nagin, D. S. & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, 70(5), 1181-1196.
- 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.
- Offord, D. R., Lipman, E. L. & Duku, E. K. (2001). Epidemiology of problem behavior up to age 12 years. In R. Loeber and D. P. Farrington (eds), *Child delinquents: Development, intervention and service needs*. Thousand Oaks, CA: Sage Publications.
- Renken, B., Egeland, B., Marvinney, D., Mangelsdorf, S. & Sroufe, L. A. (1989). Early childhood antecedents of aggression and passive-withdrawal in early elementary school. *Journal of Personality*, 57(2), 257-281.

- Rosenbaum, P. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenbaum, P. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.
- Rothman, K. J. (1986). *Modern epidemiology*. Boston: Little, Brown and Company.
- Sanders, M. R. (1999). Triple P—Positive parenting program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review*, 2(2), 71-90.
- Shelton, K. K., Frick, P. J. & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, 25(3), 317-329.
- Silverthorn, P. & Frick, P. J. (1999). Developmental pathways to antisocial behavior: The delayed onset pathway in girls. *Development and Psychopathology*, 11(1), 101-126.
- Straus, M. & Donnelly, D. A. (2001). *Beating the devil out of them: Corporal punishment in American families and its effects on children* (2nd ed.). New Brunswick, NJ: Transactions.
- Straus, M. A., Sugarman, D. B. & Giles-Sims, J. (1997). Spanking by parents and subsequent antisocial behavior of children. *Archives of Pediatrics and Adolescent Medicine*, 151(8), 761-767.
- Theobald, D. & Farrington, D. P. (2009). Effects of Getting Married on Offending: Results from a Prospective Longitudinal Survey of Males. *European Journal of Criminology*, 6, 496-516.
- Topçuoğlu, T. (2011). *Parents' use of corporal punishment and children's externalizing behaviour problems: A cross-cultural assessment* (Unpublished doctoral dissertation). University of Cambridge, UK.

- Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S. & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, 19(3), 285-300.
- Xu, X., Tung, Y.-Y. & Dunaway, R. G. (2000). Cultural, human and social capital as determinants of corporal punishment: Toward an integrated theoretical model. *Journal of Interpersonal Violence*, 15(6), 603-630.
- Zahn-Waxler, C. (1993). Warriors and worriers: Gender and psychopathology. *Development and Psychopathology*, 5, 79-89. doi: 10.1017/S0954579400004272
- Webster-Stratton, C. (1996). Early-onset conduct problems: Does gender make a difference? *Journal of Consulting and Clinical Psychology*, 64(3), 540-551.