



# HHS Public Access

Author manuscript

*Psychol Sci.* Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

*Psychol Sci.* 2018 January ; 29(1): 110–120. doi:10.1177/0956797617729816.

## Strengthening Causal Estimates for Links between Spanking and Children's Externalizing Behavior Problems

**Elizabeth Gershoff,**  
University of Texas

**Kierra Sattler,** and  
University of Texas

**Arya Ansari**  
University of Virginia

### Abstract

Establishing causal links when experiments are not feasible is an important challenge for psychology researchers. The question of whether parents' spanking causes children's externalizing behavior problems poses such a challenge because randomized experiments of spanking are unethical and correlational studies cannot rule out potential selection factors. This study used propensity score matching based on the lifetime prevalence and recent incidence of spanking in a large and nationally representative sample ( $N=12,112$ ) as well as lagged dependent variables to get as close to causal estimates outside an experiment as possible. Whether children were spanked at age 5 predicted increases in externalizing behavior problems by ages 6 and 8, even after the groups based on spanking prevalence or incidence were matched on a range of sociodemographic, family, and cultural characteristics and children's initial behavior problems. These statistically rigorous methods afford a conclusion that spanking predicts a deterioration of children's externalizing behavior over time.

### Keywords

spanking; externalizing behavior problems; propensity score matching; causal estimates

### Introduction

Whether parental spanking makes children better or worse behaved has been the focus of much debate. Several psychological theories predict that spanking should make children's behavior worse, not better. According to social learning theory (Bandura, 1977), children learn from their parents' use of spanking that aggression can be a useful way to get what

---

Address correspondence to: Elizabeth Gershoff, University of Texas - Department of Human Development and Family Sciences, 108 E. Dean Keeton St., Stop A2702 Austin Texas 78712, United States, Phone: 5124714800; Fax: 5124715630; liz.gershoff@austin.utexas.edu.

#### Author Contributions

E. T. Gershoff developed the study concept. All authors contributed to the study design. K. M. Sattler performed the data analysis under the supervision of A. Ansari. All authors interpreted the results. E. T. Gershoff drafted the manuscript, and K. M. Sattler and A. Ansari provided critical revisions. All authors approved the final version of the manuscript for submission.

they want and thus imitate their parents by acting aggressively with their peers. Attribution theory (Lepper, 1983; Grolnick, Deci, & Ryan, 1997) argues that parents' use of physical force to compel compliance interferes with children's internalization of reasons not to act in aggressive or self-interested ways, and therefore parenting behaviors like spanking lead to unregulated child behavior. Spanking is also problematic from the perspective of attachment theory (Bowlby, 1980): The experience of being spanked and thus physically hurt by their parents can interfere with children's closeness to and trust of their parents, which in turn undermines parents' socialization messages about appropriate behavior.

Consistent with these theories, research to date has consistently found that spanking is linked with more externalizing behavior problems, such as aggression and conduct disorder. In a recent meta-analysis of the association between spanking and externalizing behavior problems, all 14 studies found a statistically significant association and yielded a mean effect size of  $d = .41$ ,  $p < .001$  (Gershoff & Grogan-Kaylor, 2016). However, the majority of these studies are correlational, leading to a fundamental questions about causality: Does spanking predict increases in behavior problems? Or are children with behavior problems eliciting more spanking? Or are the factors that lead parents to spank the same that cause their children to have high levels of behavior problems?

Three criteria must be met for a causal conclusion to be reached: (1) spanking and behavior problems must be correlated; (2) spanking must precede behavior problems in time; and (3) the association between spanking and behavior problems cannot be attributed to a third factor (Shadish, Cook, & Campbell, 2001). Experiments are the preferred scientific method that meet all three criteria by design, through controlled protocols and randomization. Yet some aspects of human behavior, such as spanking, do not readily lend themselves to experimental designs. Although we cannot randomly assign a child to be spanked or not to study the effects of spanking on later behavior, we can improve our ability to make causal conclusions.

The first criterion for establishing causality, namely that spanking must be correlated with behavior problems, has already been met through several meta-analyses (Ferguson, 2013; Gershoff, 2002; Gershoff & Grogan-Kaylor, 2016). The second criterion, that spanking temporally precede behavior problems, has been met through longitudinal studies that have confirmed that spanking predicts deterioration, not improvement, in children's behavior problems over time (e.g., Grogan-Kaylor, 2005; Mulvaney & Mebert, 2007).

The third criterion of ruling out possible confounding variables can be met through a number of statistical procedures. One method is to statistically control for potential confounding variables (e.g., race, gender, socioeconomic status) that have been linked with both spanking (Gershoff, 2002) and behavior problems (Eisner & Malti, 2015); this method has been used routinely in studies of spanking and child behavior. Another more rigorous solution is the use of fixed effects, which eliminates any time-invariant factors that may account for links between spanking and child behavior. Fixed effects models have found that increases in spanking predict increases in children's problem behavior over time (Grogan-Kaylor, 2005).

The primary candidate for a confounding variable in the case of spanking is children's initial behavior problems. Children with difficult behaviors are likely to elicit harsher punishments from their parents. It is possible to include this potential child-elicitation effect at the same time as the parent-to-child effect in longitudinal cross-lagged panel designs. Such models have found that spanking continues to predict increases in behavior problems even when the child elicitation effects are included in the model (Berlin et al., 2009; Choe, Olson, & Sameroff, 2013; Gershoff, Lansford, Sexton, Davis-Kean, & Sameroff, 2012; Maguire-Jack, Gromoske, & Berger, 2012).

Each of these studies has moved the field closer to confidence in a causal link between spanking and children's behavior problems. However, without randomization, there remain parent and child characteristics that determine whether parents choose to spank or not, and these same selection factors may predict children's outcomes independent of spanking. In order to remove the influence of selection factors, we can use a statistical method that allows the closest possible approximation to a randomized experiment, namely the use of propensity score matching (PSM; Miller, Henry, & Votruba-Drzal, 2016; Rosenbaum & Rubin, 1983). PSM takes two groups that are thought of as "treatment" and "control" groups (in our case, spanked and not spanked groups) and matches the groups on a set of observed covariates, so that individuals in the study have the same propensity to be in the treatment group (Miller et al., 2016). When PSM is successful, it approximates randomization—individuals do not differ on any of several key covariates and only differ on whether they are in the treatment or control group (Stuart, 2010). For example, if boys have a higher propensity to be spanked, the PSM finds enough girls with the same propensity to be spanked and matches them to the boys so that there is no longer a gender difference in the propensity to be spanked or not. In this example, because there are fewer girls with high propensity to be spanked, PSM would allow a single girl to be matched to several boys (we allowed up to 4 such matches). PSM simultaneously conducts this matching procedure for all of the covariates, such that a child would have an equal propensity to be in the treatment or control groups regardless of their gender, race/ethnicity, socioeconomic background, and any other covariates included in the model. PSM is generally designed to be used with treatment and control groups, and thus the predictor of interest should be dichotomous.

In the current study, we examined whether spanking predicts changes in children's behavior problems 1 and 3 years later and took several steps to increase the strength of our causal estimates. To avoid broad categories such as "physical punishment" that may include potentially abusive methods (Baumrind, Larzelere, & Cowan, 2002), we focused on parents' use of spanking, a behavior that continues to be used by over 80% of parents (Gershoff et al., 2012). To increase the likelihood that our results would be generalizable to children across the U.S., we used data from a nationally representative study, the Early Childhood Longitudinal Study-Kindergarten Cohort 1998–1999 (ECLS-K); by using PSM, our study improves on the one existing study using the ECLS-K that examined spanking as a predictor of increases in behavior problems in a cross-lagged model (Gershoff et al., 2012). To rule out possible selection factors, we used PSM to create two groups of children who were identical on a range of child and family characteristics but who differed in whether their parents ever spanked or not. Although this comparison is a clear test of the overall lifetime prevalence of spanking, it necessarily ignores the incidence, or frequency, of spanking; in

order to assess spanking frequency, we conducted a second PSM with children who had been ever spanked, dichotomizing them into a group spanked in the last week and a group who had been spanked but not in the last week. To meet the requirement of time precedence, we used spanking at age 5 to predict children's behavior problems at ages 6 and 8. To guard against spurious associations, we included a robust set of covariates in all models and adjusted for children's initial levels of externalizing behavior problems to create lagged dependent variables (NICHD & Duncan, 2003). Finally, to avoid shared rater measurement error, we used parent ratings of spanking and teacher ratings of behavior problems. This is the most rigorous test to date of the hypothesis that spanking causes increases in children's behavior problems.

## Method

### Participants

Data were drawn from the ECLS-K 1998 Cohort, a study that followed a nationally representative sample of 21,409 kindergarten children from kindergarten entry through the end of the eighth grade year (for sampling information see: Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009). For the purposes of this study, we included the assessments from the age 5 (spring of kindergarten), age 6 (spring of first grade), and age 8 (third grade) waves of data collection. We restricted our sample to children and families who had a valid longitudinal weight, which were required to ensure that our models were nationally representative. In doing so, this resulted in final analytic sample of 12,112 children (49% female; 62% White, 11% Black, 16% Hispanic, 10% Asian or other). Because 92% of the parents surveyed in the ECLS-K were mothers or female guardians, we will refer to parents in the study as mothers. This sample size is appropriate for our research questions because it is large and because, when weighted, represents a nationally representative sample of children who were kindergartners in 1998.

### Measures

**Spanking**—When children were 5 years of age, mothers were asked, “Sometimes kids mind pretty well and sometimes they don’t. About how many times, if any, have you spanked [name of child] in the past week?” The response was open-ended (range: 0 to 30); if parents volunteered that they *never* spanked, the interviewer entered a -1 code. To indicate lifetime prevalence of spanking, responses were recoded into a dichotomous “ever spanked” variable, such that children whose mothers volunteered that they had *never* spanked their children received a score of 0 (unmatched *n*s = 2,478–2,491; sample sizes vary across the 50 imputed data sets) and children whose mothers reported that they had *ever* spanked them were coded as 1 (unmatched *n*s = 9,621–9,634). To indicate recent incidence of spanking, we took this latter group and further dichotomized it into a “spanked in the last week” group, with children whose mothers had not spanked them in the past week but who did not indicate they never spanked (score of 0; unmatched *n*s = 6,445–6,464) and a group of children whose mothers had spanked them one or more times in the previous week (score of 1; unmatched *n*s = 3,167–3,181).

**Children’s externalizing behavior problems**—Teachers reported on children’s externalizing behavior problems at ages 5, 6 and 8 using an adapted version of the Social Skills Rating Scale (SSRS: Gresham & Elliott, 1990). Teachers reported the frequency with which children argued, fought, got angry, acted impulsively, or disturbed ongoing activities using a 4-point Likert scale (1 = *never* to 4 = *very often*). The reliability of the externalizing scale (W1  $\alpha = .90$ ; W2  $\alpha = .86$ ; and W3  $\alpha = .89$ ) was strong across each wave of data collection.

**Variables used in matching and as control variables**—For the PSM, we matched the two sets of spanking comparison groups on 38 separate child and family characteristics designed to rule out potential alternative explanations for any links between spanking and later externalizing behavior problems: *child characteristics* (age, gender, overall health, disability status, level of externalizing behaviors at age 5); *parent characteristics* (education level, age, whether born in U.S., marital status [married, separated, divorced, widowed, never married], whether biological or adoptive parent); *family economic status* (household size, mother’s employment status [full time, part time, unemployed], family income, family food insecurity [Bickel, Nord, Price, Hamilton, & Cook, 2000]); *harshness of home environment* (parenting harshness [from the average of 4 0/1 potential responses to a vignette about how parents would respond to a child hitting them: hit child back, make fun of child, yell at child, and put child in time out], low parenting warmth [4 items, Home Observation for Measurement of the Environment (HOME) Scale: Caldwell & Bradley, 1984], parenting stress [7 items, Parenting Stress Index: Abidin & Abidin, 1990], intimate partner conflict [2 items (“argue heatedly” and “end up hitting or throwing things at each other”), Conflict Tactics Scale: Straus, 1990], mother’s depressive symptoms [12 items, Center for Epidemiological Studies-Depression Scale: Radloff, 1977]); *cultural background* (race [White, Black, Asian or Other race], Latino ethnicity, English as a home language, religiosity); and *geographic characteristics* (region [Northeast, Midwest, South, West], urbanicity [city, suburb, town/rural]). In addition to being used to match the spanking groups, these same variables were also included as covariates in all models.

### Analytic Strategy

All analyses were conducted in Stata 14 (StataCorp, 2015). To account for missing data (range by variable: 0% to 21%), we imputed 50 datasets through the chained equations method and report the averaged results. To account for non-independence of children’s outcomes as a result of sampling by schools, we clustered all models at the school level.

Before conducting the PSM procedures, we first examined how our two sets of spanking groups differed on the control variables. The left half of Table 1 presents the differences between the ever spank and never spank groups; the groups significantly differed on 68% (26/38) of the variables and thus were considered to be unbalanced. Similarly, as seen in Table 2, the “not in last week” and “in last week” groups differed on 71% (27/38) of the covariates. In both comparisons, differences were found in each of the covariate categories; in particular, initial teacher-rated externalizing behaviors at age 5 were higher for the ever spanked group (compared to never) and the spanked in the past week group (compared to not in the past week). These differences between groups across a range of covariate

categories demonstrate the need to eliminate these initial differences through PSM. To match the ever spanked treatment group with the never spanked control group across this set of covariates, we used all of the 38 covariates listed in Tables 1 and 2 to predict membership in the ever spanked versus the never spanked groups in logit models within each of the 50 imputed datasets. We repeated this procedure for the spanked in the last week and not in the last week groups. For each PSM, we used the nearest neighbor matching method with up to four matches and with a caliper of .01, thereby ensuring sufficient overlap between our two conditions. We assessed the quality of the matches in two ways. First, we checked the standardized mean differences between the two groups for all of the covariates to ensure that no differences exceeded 10% of a standard deviation. Second, we regressed each of the covariates, individually, on the indicator variable that distinguished children who were spanked or not within the matched samples. After matching, there were no longer any significant group differences on any of the covariates between the ever and never groups (see last column of Table 1) or between the in the last week and not in the last week groups (see last column of Table 2). Moreover, after matching, none of the covariates had a standardized mean difference exceeding 10% of a standard deviation, indicating that balance had been successfully achieved.

In order to determine whether spanking at age 5 predicted increases in children's externalizing behavior problems from age 5 to ages 6 and 8, we first estimated basic OLS regression models that were weighted with a longitudinal study weight to ensure that the sample was nationally representative. These OLS models adjusted for the covariates discussed above. We then ran a second set of OLS regressions using the propensity matched versions of the never versus ever and the in the last week versus not in the last week groups. Because our implementation of PSM allowed children to be matched up to four times, our OLS models within the matched samples were weighted to account for the number of times children were matched, which in turn prevented us from using the longitudinal sampling weight and thus the models with the matched spanking groups are not nationally representative. Finally, to guard against any remaining bias, our analyses within the matched samples controlled for all covariates listed in Tables 1 and 2 to achieve what has been called doubly robust estimation (Funk, Westreich, Wiesen, Stürmer, Brookhart, & Davidian, 2011).

## Results

In the OLS regressions with the unmatched samples, children who had ever been spanked by their parents at age 5 were reported by teachers to have significantly higher increases in externalizing behavior problems by age 6 than children who had never been spanked ( $\beta = .06, p = .038$ ; see Table 3). The association with teacher-rated externalizing behavior at age 8 was of similar, but slightly smaller, magnitude and failed to reach conventional levels of statistical significance ( $\beta = .04, p = .139$ ). These results were flipped for the models using the in the last week indicator of spanking, such that spanking in the past week did not significantly predict increases in externalizing behavior between age 5 and age 6 ( $\beta = .03, p = .059$ ) but did predict increases between age 5 and age 8 ( $\beta = .11, p = .001$ ).

We then repeated the OLS regressions with the PSM-matched samples; results from these models are presented in the right side of Table 3. Children who were ever spanked



experienced stronger increases in their externalizing behavior problems from age 5 to age 6 ( $\beta = .06, p = .023$ ) and to age 8 ( $\beta = .07, p = .014$ ) than children who were never spanked. Among children who had ever been spanked, children spanked in the past week at age 5 did not significantly increase in their behavior problems by age 6 ( $\beta = .04, p = .140$ ) but did by age 8 ( $\beta = .09, p < .001$ ). Thus, even when characteristics known to predict whether parents spank were removed from the model through PSM, ever spanking remained a significant predictor of children's later externalizing behaviors, both when spanking was dichotomized according to lifetime prevalence (ever vs. never) and according to recent incidence among those ever spanked (in last week vs. not in last week).

Upon finding that having ever been spanked was linked with increases in children's behavior problems at both age 6 and age 8 in the matched models, we wondered whether the association between spanking and increased externalizing behavior at age 8 (compared to age 5) was mediated by increased externalizing behavior problems at age 6. We estimated a regression in which age 8 externalizing behavior was regressed both on ever having been spanked and on age 6 externalizing behavior and then estimated a Sobel test (Sobel, 1982) of mediation. The test revealed that the link between ever having been spanked and the increase in externalizing behavior from age 5 to age 8 was indeed mediated through the increase in externalizing behavior from age 5 to age 6 ( $z = 2.23, p = .026$ ). We did not repeat this analysis of the two spanked groups, namely those spanked in the past week and those ever spanked but not in the past week, because spanked in the past week at age 5 did not significantly predict externalizing behavior at age 6.

## Discussion

The analyses presented in this paper tested whether spanking would continue to predict changes in children's externalizing behavior problems over time after groups of children who differed on their exposure to spanking were matched across a range of child, parent, and family demographic characteristics, including children's initial problem behaviors. If selection factors were the main explanation for the association between spanking and behavior problems (Larzelere, Kuhn, & Johnson, 2004), then any significant associations observed in the unmatched regressions would disappear in the matched models using propensity scores.

In models that contained both matched groups and lagged dependent variables, we did not find any support for the selection effects explanation, and specifically did not find support for a child elicitation explanation. Whether a child had ever been spanked at age 5 significantly predicted increases in their externalizing behavior problems by both age 6 and age 8 in the matched models, over and above children's initial levels of behavior problems and a range of covariates in our doubly robust regressions with PSM. In other words, with 38 different child, parent, and family characteristics held equal through PSM, having ever been spanked predicted increased externalizing behavior problems over time. We also found that part of the process by which ever spanking predicts behavior problems three years into the future is by increasing behavior problems one year later.

There are two other key conclusions of note. First, we found no evidence that either the lifetime prevalence or recent incidence of spanking is effective at reducing externalizing behavior problems over time, which is parents' goal when using spanking to control their children. Our findings are consistent with conclusions from a number of other longitudinal studies (Grogan-Kaylor, 2005; Olson, Lopez-Duran, Lunkenheimer, Chang, & Sameroff, 2011) and with findings from several meta-analyses (Ferguson, 2013; Gershoff, 2002; Gershoff & Grogan-Kaylor, 2016) that have linked spanking with more, rather than fewer, behavior problems in children. Taken together, these studies meet the three criteria (Shadish et al. 2001) for reaching a causal conclusion that spanking predicts more behavior problems in children. Researchers who continue to insist that spanking is effective in promoting better child behavior (Larzelere, Gunnoe, Roberts, & Ferguson, 2017) do so in defiance of accumulated research evidence.

Second, it is somewhat remarkable that just knowing whether a child had ever been spanked or knowing whether they had been spanked in the week before the survey, rather than knowing exactly how many spankings they were subjected to either over time or in any given instance, was a significant predictor of children's externalizing behavior. This is noteworthy given that we matched the two sets of spank groups on a range of factors that have been shown to predict whether parents spank or not. Thus, in the PSM models, the only observed characteristic on which the matched groups varied was spanking. It is also likely that this variable underestimated the number of parents who never spank because parents had to volunteer this information when asked how often they spanked, making it even harder to find differences between the spank and never spank groups. Yet while some critics have argued that children who are ever spanked are better off than children who are never spanked (Larzelere et al., 2017), our results contradict that claim: over and above a child's initial level of externalizing behavior problems, a child who is spanked even once is more likely to have behavior problems in the future than a peer who is never spanked.

Our conclusions are limited by the fact that we do not have every possible confounding variable in our matching models. For example, whether parents were spanked as children (Russa, Rodriguez, & Silvia, 2014) and their attitudes about spanking (Lansford, Deater-Deckard, Bornstein, Putnick, & Bradley, 2014; Taylor, Hamvas, Rice, Newman, & DeJong, 2011) are strong predictors of whether they spank their own children, but we did not have these variables in the dataset. However, this limitation is tempered by the fact that we included over three dozen covariates in the matching process and then again as controls in the matched regression models.

In summary, this study demonstrated that the links between both the lifetime prevalence and recent incidence of spanking at age 5 and children's externalizing behavior problems one and three years later are robust to a number of statistical methods designed to eliminate spurious findings. Because experiments on spanking are unethical, studies such as this one are crucial for enhancing causal inference about links between spanking and children's behavior problems.



## Acknowledgments

This research was supported by grants (R24HD042849, PI: Umberson; T32HD007081, PIs: Raley & Gershoff) awarded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, by a grant from the National Science Foundation (1519686, PIs: Gershoff & Crosnoe), and by a grant from the Institute of Education Sciences, U.S. Department of Education (R305B130013, PI: Rimm-Kaufman).

## References

- Abidin, RR., Abidin, RR. Parenting Stress Index (PSI). Charlottesville, VA: Pediatric Psychology Press; 1990.
- Bandura, A. Social learning theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
- Baumrind D, Larzelere RE, Cowan PA. Ordinary physical punishment: Is it harmful? Comment on Gershoff (2002). *Psychological Bulletin*. 2002; 128:580–589. DOI: 10.1037/0033-2909.128.4.580 [PubMed: 12081082]
- Berlin LJ, Ispa JM, Fine MA, Malone PS, Brooks-Gunn J, Brady-Smith C, ... Bai Y. Correlates and consequences of spanking and verbal punishment for low-income White, African American, and Mexican American toddlers. *Child Development*. 2009; 80:1403–1420. DOI: 10.1111/j.1467-8624.2009.01341.x [PubMed: 19765008]
- Bickel, G., Nord, M., Price, C., Hamilton, WL., Cook, JT. Guide to measuring household food security. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service; 2000.
- Bowlby, J. Attachment and loss, Vol. 3: Loss, sadness and depression. New York: Basic Books; 1980.
- Caldwell, BM., Bradley, RH. Home observation for measurement of the environment. Little Rock: University of Arkansas at Little Rock; 1984.
- Choe DE, Olson SL, Sameroff AJ. The interplay of externalizing problems and physical and inductive discipline during childhood. *Developmental Psychology*. 2013; 49:2029–2039. DOI: 10.1037/a0032054 [PubMed: 23458660]
- Eisner, MP., Malti, T. Aggressive and violent behavior. In: Lamb, ME., Lerner, RM., editors. *Handbook of child psychology and developmental science, Vol. 3: Social, emotional and personality development*. 7. New York, NY: Wiley; 2015. p. 795-884.
- Ferguson CJ. Spanking, corporal punishment and negative long-term outcomes: metaanalytic review of longitudinal studies. *Clinical Psychology Review*. 2013; 33:196–208. DOI: 10.1016/j.cpr.2012.11.002 [PubMed: 23274727]
- Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *American Journal of Epidemiology*. 2011; 173:761–767. DOI: 10.1093/aje/kwq439 [PubMed: 21385832]
- Gershoff ET. Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*. 2002; 128:539–579. DOI: 10.1037/0033-2909.128.4.539 [PubMed: 12081081]
- Gershoff ET, Grogan-Kaylor A. Corporal punishment by parents and its consequences for children: Old controversies and new meta-analyses. *Journal of Family Psychology*. 2016; 30:453–469. DOI: 10.1037/fam0000191 [PubMed: 27055181]
- Gershoff ET, Lansford JE, Sexton HR, Davis-Kean P, Sameroff AJ. Longitudinal links between spanking and children's externalizing behaviors in a national sample of White, Black, Hispanic, and Asian American families. *Child Development*. 2012; 83:838–843. DOI: 10.1111/j.1467-8624.2011.01732.x [PubMed: 22304526]
- Gresham, FM., Elliott, SN. *Social Skills Rating System Manual*. Circle Pines, MN: American Guidance Service; 1990.
- Grogan-Kaylor A. The effect of corporal punishment on antisocial behavior in children. *Social Work Research*. 2004; 28:153–162. DOI: 10.1093/swr/28.3.153
- Grogan-Kaylor A. Relationship of corporal punishment and antisocial behavior by neighborhood. *Archives of Pediatric and Adolescent Medicine*. 2005; 159:938–942. DOI: 10.1001/archpedi.159.10.938

- Grolnick, WS., Deci, EL., Ryan, RM. Internalization within the family: The self-determination theory perspective. In: Grusec, JE., Kuczynski, L., editors. Parenting and children's internalization of values: A handbook of contemporary theory. New York: Wiley; 1997. p. 135-161.
- Lansford JE, Deater-Deckard K, Bornstein MH, Putnick DL, Bradley RH. Attitudes justifying domestic violence predict endorsement of corporal punishment and physical and psychological aggression towards children: A study in 25 low-and middle-income countries. *Journal of Pediatrics*. 2014; 164:1208–1213. DOI: 10.1016/j.jpeds.2013.11.060 [PubMed: 24412139]
- Larzelere RE, Gunnoe ML, Roberts MW, Ferguson CJ. Children and parent deserve better parental discipline research: Critiquing the evidence for exclusively “positive” parenting. *Marriage & Family Review*. 2017; 53:24–35. DOI: 10.1080/01494929.2016.1145613
- Larzelere RE, Kuhn BR, Johnson B. The intervention selection bias: An underrecognized confound in intervention research. *Psychological Bulletin*. 2004; 130:289–303. DOI: 10.1037/0033-2909.130.2.289 [PubMed: 14979773]
- Lepper, MR. Social control processes and the internalization of social values: An attributional perspective. In: Higgins, ET, Ruble, DN., Hartup, WW., editors. Social cognition and social development. New York, NY: Cambridge University Press; 1983. p. 294-330.
- Maguire-Jack K, Gromoske AN, Berger LM. Spanking and child development during the first 5 years of life. *Child Development*. 2012; 83:1960–1977. DOI: 10.1111/j.1467-8624.2012.01820.x [PubMed: 22860622]
- Miller P, Henry D, Votruba-Drzal E. Strengthening causal inference in developmental research. *Child Development Perspectives*. 2016; 10:275–280. DOI: 10.1111/cdep.12202
- Mulvaney MK, Mebert CJ. Parental corporal punishment predicts behavior problems in early childhood. *Journal of Family Psychology*. 2007; 21:389–397. DOI: 10.1037/0893-3200.21.3.389 [PubMed: 17874924]
- Duncan GJ. National Institute of Child Health and Human Development Early Child Care Research Network. Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*. 2003; 74:1454–1475. DOI: 10.1111/1467-8624.00617 [PubMed: 14552408]
- Olson SL, Lopez-Duran N, Lunkenheimer ES, Chang H, Sameroff AJ. Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and Psychopathology*. 2011; 23:253–266. DOI: 10.1017/S0954579410000775 [PubMed: 21262052]
- Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401. DOI: 10.1177/014662167700100306
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55. DOI: 10.1093/biomet/70.1.41
- Russa MB, Rodriguez CM, Silvia PJ. Frustration influences impact of history and disciplinary attitudes on physical discipline decision making. *Aggressive Behavior*. 2014; 40:1–11. DOI: 10.1002/ab.21500 [PubMed: 23926046]
- Shadish, WR., Cook, TD., Campbell, DT. Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin; 2001.
- Sobel, ME. Asymptotic intervals for indirect effects in structural equations models. In: Leinhardt, S., editor. *Sociological methodology* 1982. San Francisco: Jossey-Bass; 1982. p. 290-312.
- StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; Station, TX: Author; 2015.
- Straus, MA. The Conflict Tactics Scales and its critics: An evaluation and new data on validity and reliability. In: Straus, MA., Gelles, RJ., editors. *Physical violence in American families: Risk factors and adaptations to violence in 8, 145 families*. New Brunswick, NJ: Transaction Publishing; 1990. p. 49-73.
- Stuart EA. Matching methods for causal inference: A review and a look forward. *Statistical Science*. 2010; 25:1–21. doi:0.1214/09-STS313. [PubMed: 20871802]
- Taylor CA, Hamvas L, Rice J, Newman DL, DeJong W. Perceived social norms, expectations, and attitudes toward corporal punishment among an urban community sample of parents. *Journal of Urban Health*. 2011; 88:254–269. DOI: 10.1007/s11524-011-9548-7

Tourangeau, K., Nord, C., Lê, T., Sorongon, AG., Najarian, M. Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User’s Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2009. NCES 2009–004

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Demographic Characteristics of the Never Spanked Group and the Ever Spanked Group Before and After Propensity Score Matching

Table 1

	Differences between groups before propensity score matching			Differences between groups after propensity score matching		
	Never spanked <i>ns</i> = 2,478–2,491	Ever spanked <i>ns</i> = 9,621–9,634	<i>p</i>	Never spanked <i>ns</i> = 2,469–2,488	Ever spanked <i>ns</i> = 9,585–9,617	<i>p</i>
<b>Child characteristics</b>						
Age at spring of kindergarten (in months)	74.62 (4.36)	74.65 (4.44)	> .250	74.70 (4.26)	74.72 (4.45)	> .250
Male	0.50	0.52	.044	0.52	0.52	> .250
Overall health	1.71 (0.85)	1.71 (0.84)	> .250	1.67 (0.82)	1.66 (0.81)	> .250
Diagnosed with a disability	0.16	0.15	> .250	0.14	0.14	> .250
Externalizing behavior at age 5	1.61 (0.62)	1.70 (0.66)	< .001	1.65 (0.64)	1.64 (0.62)	> .250
<b>Parent characteristics</b>						
Mother's education (in years)	14.89 (3.59)	14.44 (3.43)	< .001	15.02 (3.59)	14.94 (3.37)	> .250
Mother's age (in years)	34.12 (6.51)	32.92 (6.66)	< .001	33.76 (6.41)	33.66 (6.43)	> .250
Mother was born in U.S.	0.82	0.84	.098	0.82	0.83	> .250
Parents married	0.71	0.67	.015	0.74	0.75	> .250
Parents separated	0.06	0.05	.118	0.03	0.04	.229
Parents divorced	0.09	0.09	> .250	0.08	0.08	> .250
Parents widowed	0.00	0.01	.003	0.01	0.01	> .250
Parents never married	0.11	0.15	< .001	0.11	0.11	> .250
Non-biological/adoptive	0.03	0.03	.010	0.02	0.02	> .250
<b>Family economic status</b>						
Household size	4.46 (1.37)	4.51 (1.39)	.219	4.51 (1.38)	4.51 (1.34)	> .250
Mother employed full time	0.45	0.46	> .250	0.46	0.46	> .250
Mother employed part time	0.24	0.21	.003	0.22	0.23	> .250
Mother unemployed	0.31	0.33	.066	0.32	0.332	> .250
Income in dollars/1000	54.18 (39.49)	46.31 (35.58)	< .001	53.47 (38.00)	52.35 (36.72)	.207
Family food insecurity	0.50 (1.60)	0.73 (1.95)	< .001	0.54 (1.69)	0.56 (1.72)	> .250
<b>Harshness of home environment</b>						
Parenting harshness	0.19 (0.14)	0.18 (0.15)	< .001	0.19 (0.14)	0.19 (0.15)	> .250
Low parenting warmth	1.27 (0.37)	1.31 (0.38)	< .001	1.32 (0.40)	1.32 (0.38)	> .250

	Differences between groups before propensity score matching		Differences between groups after propensity score matching		
	Never spanked <i>ns</i> = 2,478–2,491	Ever spanked <i>ns</i> = 9,621–9,634	Never spanked <i>ns</i> = 2,469–2,488	Ever spanked <i>ns</i> = 9,585–9,617	
				<i>p</i>	
Parenting stress	1.54 (0.45)	1.64 (0.48)	1.61 (0.48)	1.61 (0.46)	> .250
Intimate partner conflict	0.49 (0.47)	0.56 (0.49)	0.57 (0.50)	0.55 (0.48)	.152
Mother's depressive symptoms	1.43 (0.47)	1.49 (0.47)	1.46 (0.49)	1.45 (0.44)	> .250
<b>Cultural background</b>					
White	0.63	0.56	0.61	0.62	> .250
Black	0.09	0.18	0.12	0.12	> .250
Asian/Other	0.10	0.07	0.10	0.10	> .250
Latino ethnicity	0.18	0.19	0.17	0.17	> .250
English home language	0.87	0.88	0.88	0.88	> .250
Mother's religiosity	2.83 (1.29)	2.89 (1.34)	2.96 (1.29)	2.96 (1.31)	> .250
<b>Geographic characteristics</b>					
Northeast	0.29	0.16	0.17	0.17	> .250
Midwest	0.27	0.22	0.26	0.26	> .250
South	0.24	0.40	0.35	0.35	> .250
West	0.20	0.22	0.22	0.22	> .250
City	0.36	0.37	0.37	0.37	> .250
Suburb	0.47	0.41	0.40	0.38	> .250
Town/Rural	0.17	0.22	0.23	0.24	.159

*Note:* Sample sizes vary across the 50 imputed datasets. Means and standard deviations or proportions are reported and are the averaged values across imputations

Demographic Characteristics of the Not Spanked in Last Week Group and the Spanked in Last Week Group Before and After Propensity Score Matching

	Differences between groups before propensity score matching		Differences between groups after propensity score matching	
	Not in last week <i>ns</i> = 6,445–6,464	<i>p</i>	Not in last week <i>ns</i> = 4,778–4,953	<i>p</i>
<b>Child characteristics</b>				
Age at spring of kindergarten (in months)	74.85 (4.45)	<.001	74.38 (4.39)	>.250
Male	0.51	<.001	0.54	>.250
Overall health	1.68 (0.83)	<.001	1.74 (0.84)	>.250
Diagnosed with a disability	0.15	.015	0.15	>.250
Externalizing behavior at age 5	1.64 (0.62)	<.001	1.76 (0.68)	>.250
<b>Parent characteristics</b>				
Mother's education (in years)	14.64 (3.42)	<.001	14.47 (3.47)	>.250
Mother's age (in years)	33.39 (6.63)	<.001	32.76 (6.39)	>.250
Mother was born in U.S.	0.84	>.250	0.81	>.250
Parents married	0.70	<.001	0.71	>.250
Parents separated	0.05	>.250	0.04	>.250
Parents divorced	0.09	.107	0.08	>.250
Parents widowed	0.01	>.250	0.01	>.250
Parents never married	0.12	<.001	0.14	>.250
Non-biological/adoptive	0.03	>.250	0.02	>.250
<b>Family economic status</b>				
Household size	4.51 (1.36)	>.250	4.53 (1.40)	>.250
Mother employed full time	0.46	>.250	0.46	>.250
Mother employed part time	0.22	.018	0.20	>.250
Mother unemployed	0.32	.128	0.34	>.250
Income in dollars/1000	50.01 (36.09)	<.001	45.03 (34.34)	>.250
Family food insecurity	0.59 (1.76)	<.001	0.77 (2.04)	>.250
<b>Harshness of home environment</b>				
Parenting harshness	0.19 (0.14)	<.001	0.18 (0.15)	>.250
Low parenting warmth	1.29 (0.36)	<.001	1.36 (0.41)	>.250

Table 2



	Differences between groups before propensity score matching			Differences between groups after propensity score matching		
	Not in last week <i>ns</i> = 6,445–6,464	In last week <i>ns</i> = 3,167–3,181	<i>p</i>	Not in last week <i>ns</i> = 4,778–4,953	In last week <i>ns</i> = 3,159–3,177	<i>p</i>
Parenting stress	1.59 (0.45)	1.73 (0.53)	< .001	1.70 (0.50)	1.70 (0.50)	> .250
Intimate partner conflict	0.53 (0.48)	0.62 (0.50)	< .001	0.61 (0.51)	0.61 (0.50)	> .250
Mother's depressive symptoms	1.45 (0.46)	1.56 (0.50)	< .001	1.52 (0.50)	1.51 (0.47)	> .250
<b>Cultural background</b>						
White	0.59	0.50	< .001	0.55	0.55	> .250
Black	0.15	0.22	< .001	0.16	0.16	> .250
Asian/Other	0.07	0.07	> .250	0.11	0.10	> .250
Latino ethnicity	0.18	0.20	.047	0.18	0.18	> .250
English home language	0.88	0.87	.120	0.86	0.87	> .250
Mother's religiosity	2.86 (1.32)	2.96 (1.36)	< .001	3.01 (1.32)	3.02 (1.33)	> .250
<b>Geographic characteristics</b>						
Northeast	0.16	0.14	.007	0.15	0.15	> .250
Midwest	0.25	0.18	< .001	0.20	0.21	> .250
South	0.36	0.48	< .001	0.42	0.43	> .250
West	0.23	0.20	.001	0.22	0.22	> .250
City	0.37	0.38	> .250	0.38	0.38	> .250
Suburb	0.42	0.39	.006	0.36	0.36	> .250
Town/Rural	0.21	0.24	.037	0.26	0.26	> .250

*Note:* Sample sizes vary across the 50 imputed datasets. Means and standard deviations or proportions are reported and are the averaged values across imputations.

**Table 3**  
 Results from Regressions Predicting Child Externalizing Behavior from the Unmatched and Matched Ever/Never and In Last Week/Not In Last Week Spanking Groups

	OLS regressions with unmatched samples <sup>a</sup>					OLS regressions with propensity score matched samples <sup>b</sup>										
	Age 6 externalizing behavior problems		Age 8 externalizing behavior problems		Age 6 externalizing behavior problems		Age 8 externalizing behavior problems		Age 6 externalizing behavior problems		Age 8 externalizing behavior problems					
	B	95% CI	$\beta$	p	B	95% CI	$\beta$	p	B	95% CI	$\beta$	p				
Ever spanked (vs. never spanked) by age 5	.04	.00-.12	.06	.038	.03	-.01-.10	.04	.139	.04	.01-.11	.06	.023	.04	-.01-.12	.07	.014
Spanked in last week (vs. not in the last week) at age 5	.03	-.00-.11	.05	.059	.07	.05-.16	.11	<.001	.02	-.01-.09	.04	.140	.06	.04-.15	.09	<.001

Note: Spanking was mother-rated; externalizing behavior problems was teacher-rated. OLS = ordinary least squares.

<sup>a</sup>Sample size for regressions with unmatched “ever” and “never” spanked groups was 12,112. Sample size for the regressions with unmatched “in the last week” and “not in the last week” groups was 9,629. These results are weighted to be nationally representative.

<sup>b</sup>Sample size for regressions with PSM matched “ever” and “never” spanked groups was 12,082. Sample size for the regressions with PSM matched “in last week” and “not in last week” groups was 8,008.