Research Article

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

Elizabeth T. Gershoff[1], Kierra M. P. Sattler[1], and Arya Ansari[2]
[1]Department of Human Development and Family Sciences, University of Texas, and [2]Curry School of Education, 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 the age of 5 years 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 yield the conclusion that spanking predicts a deterioration of children's externalizing behavior over time.

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 they want and thus imitate their parents by acting aggressively with their peers. Attribution theory (Grolnick, Deci, & Ryan, 1997; Lepper, 1983) posits 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; therefore, parenting behaviors such as 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 (d) of .41 (p < .001; Gershoff & Grogan-Kaylor, 2016). However, the majority of these studies are correlational, leading to 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

Corresponding Author:
Elizabeth T. Gershoff, University of Texas, Department of Human Development and Family Sciences, 108 E. Dean Keeton St., Stop A2702, Austin, TX 78712
E-mail: liz.gershoff@austin.utexas.edu

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: (a) Spanking and behavior problems must be correlated, (b) spanking must precede behavior problems in time, and (c) the association between spanking and behavior problems cannot be attributed to a third factor (Shadish, Cook, & Campbell, 2001). Experimentation is the preferred scientific method that meets 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 spanked just 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 precedes 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, 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 but differ only on whether they are in the treatment or control group (Stuart, 2010). For example, if boys have a high propensity to be spanked, 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 a high propensity to be spanked, PSM would allow a single girl to be matched to several boys (we allowed up to four 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 group regardless of his or her 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; 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 more than 80% of parents (Gershoff et al., 2012). To increase the likelihood that our results would be generalizable to children across the United States, we used data from a nationally representative study, the Early Childhood Longitudinal Study–Kindergarten Cohort 1998–1999 (ECLS-K; Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009); 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 them 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; to assess spanking frequency, we conducted a second PSM with children who had been spanked, dichotomizing them into a group spanked in the week prior to the study and a group who had been spanked but not in the prior week. To meet the requirement of time precedence, we used spanking at the age of 5 years 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 (National Institute of Child Health and Human Development Early Child Care Research Network & 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 children from kindergarten entry through the end of the eighth-grade year (for sampling information, see Tourangeau et al., 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 was required to ensure that our models were nationally representative. In doing so, this resulted in a 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 was appropriate for our research questions because it is large and because, when weighted, it 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 = 0–30); if parents volunteered that they never spanked their child, the interviewer entered a −1 code. To indicate lifetime prevalence of spanking, we recoded responses 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 ns = 2,478–2,491; sample sizes varied across the 50 imputed data sets), and children whose mothers reported that they had spanked them at least once were coded as 1 (unmatched ns = 9,621–9,634). To indicate recent incidence of spanking, we took this latter group and further dichotomized it into a group that had been spanked in the past week, with children whose mothers had not spanked them in the past week but who did not indicate they never spanked them (score of 0; unmatched ns = 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 ns = 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 System (Gresham & Elliott, 1990). Teachers reported the frequency with which children argued, fought, got angry, acted impulsively, and disturbed ongoing activities using a 4-point Likert-type scale (1 = *never* to 4 = *very often*). The reliability of the externalizing scale (Wave 1: α = .90; Wave 2: α = .86; and Wave 3: α = .89) was strong across each wave of data collection.

Variables used in matching and as control variables. For 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. The first set of variables was related to child characteristics: age, gender, overall health, disability status, and level of externalizing behaviors at age 5 years. The second set covered parent characteristics: education level, age, place of birth (born in the United States or not), marital status (married, separated, divorced, widowed, never married), and whether they were biological or adoptive parents. Family economic status was assessed with household size, mother's employment status (full time, part time, unemployed), family income, and family food insecurity (Bickel, Nord, Price, Hamilton, & Cook, 2000). Harshness of the home environment covered parenting harshness (from the average of four 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 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 Scales; Straus, 1990), and mother's depressive symptoms (12 items, Center for Epidemiological Studies-Depression Scale; Radloff, 1977). Cultural background consisted of race (White, Black, Asian, other race, Latino ethnicity), English as a home language, and mother's religiosity. Finally, we assessed geographic characteristics with region (Northeast, Midwest, South, West) and 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%–21%), we imputed 50 data sets through the chained-equations method and report the averaged results. To account for nonindependence 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. Table 1 presents the differences between the spanked and never-spanked groups; the groups significantly differed on 68% (26/38) of the variables and thus were considered to be unbalanced. Similarly, as Table 2 shows, the not-spanked-in-past-week and spanked-in-past-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 spanked group (compared with the never-spanked group) and the spanked-in-the-past-week group (compared with the not-spanked-in-the-past-week group). These differences between groups across a range of covariate categories demonstrate the need to eliminate these initial differences through PSM. To match the 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 group membership in logit models within each of the 50 imputed data sets. We repeated this procedure for the spanked-in-the-past-week and not-spanked-in-the-past-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 spanked within the matched samples. After matching, there were no longer any significant group differences on any of the covariates between the spanked and never-spanked groups (see last column of Table 1) or between the past-week and not-in-the-past-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.

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 ordinary-least-squares (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-spanked versus spanked and the past-week versus not-in-the-past-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; thus, the models with the matched spanking groups were 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 et al., 2011).

Results

In the OLS regressions with the unmatched samples, children who had 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 ($\beta = 0.06$, $p = .038$; 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 ($\beta = 0.04$, $p = .139$). These results were flipped for the models using the in-the-past-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 ($\beta = 0.03$, $p = .059$) but did predict increases between age 5 and age 8 ($\beta = 0.11$, $p = .001$).

Table 1. Demographic Characteristics of the Never-Spanked Group and the Spanked Group Before and After Propensity Score Matching (PSM)

	Before PSM			After PSM		
	Never spanked	Spanked		Spanked	Spanked	
Variable	*n*s = 2,478–2,491	*n*s = 9,621–9,634	*p*	*n*s = 2,469–2,488	*n*s = 9,585–9,617	*p*
Child characteristics						
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
Proportion male	.50	.52	.044	.52	.52	> .250
Overall health	1.71 (0.85)	1.71 (0.84)	> .250	1.67 (0.82)	1.66 (0.81)	> .250
Proportion diagnosed with a disability	.16	.15	> .250	.14	.14	> .250
Externalizing behavior at age 5 years	1.61 (0.62)	1.70 (0.66)	< .001	1.65 (0.64)	1.64 (0.62)	> .250
Parent characteristics						
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
Proportion of mothers born in U.S.	.82	.84	.098	.82	.83	> .250
Proportion of parents married	.71	.67	.015	.74	.75	> .250
Proportion of parents separated	.06	.05	.118	.03	.04	.229
Proportion of parents divorced	.09	.09	> .250	.08	.08	> .250
Proportion of parents widowed	.00	.01	.003	.01	.01	> .250
Proportion of parents never married	.11	.15	< .001	.11	.11	> .250
Proportion of parents who were nonbiological/ adoptive	.03	.03	.010	.02	.02	> .250
Family economic status						
Household size	4.46 (1.37)	4.51 (1.39)	.219	4.51 (1.38)	4.51 (1.34)	> .250
Proportion of mothers employed full time	.45	.46	> .250	.46	.46	> .250
Proportion of mothers employed part time	.24	.21	.003	.22	.23	> .250
Proportion of mothers unemployed	.31	.33	.066	.32	.332	> .250
Income (in thousands of dollars)	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
Harshness of home environment						
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
Parenting stress	1.54 (0.45)	1.64 (0.48)	< .001	1.61 (0.48)	1.61 (0.46)	> .250
Intimate partner conflict	0.49 (0.47)	0.56 (0.49)	< .001	0.57 (0.50)	0.55 (0.48)	.152
Mother's depressive symptoms	1.43 (0.47)	1.49 (0.47)	< .001	1.46 (0.49)	1.45 (0.44)	> .250

(continued)

Table 1. *(continued)*

	Before PSM			After PSM		
	Never spanked	Spanked		Spanked	Spanked	
Variable	ns = 2,478–2,491	ns = 9,621–9,634	p	ns = 2,469–2,488	ns = 9,585–9,617	p
Cultural background						
Proportion White	.63	.56	< .001	.61	.62	> .250
Proportion Black	.09	.18	< .001	.12	.12	> .250
Proportion Asian/other	.10	.07	< .001	.10	.10	> .250
Proportion Latino	.18	.19	> .250	.17	.17	> .250
Proportion who spoke English at home	.87	.88	.119	.88	.88	> .250
Mother's religiosity	2.83 (1.29)	2.89 (1.34)	.044	2.96 (1.29)	2.96 (1.31)	> .250
Geographic characteristics						
Proportion Northeast	.29	.16	< .001	.17	.17	> .250
Proportion Midwest	.27	.22	< .001	.26	.26	> .250
Proportion South	.24	.40	< .001	.35	.35	> .250
Proportion West	.20	.22	.011	.22	.22	> .250
Proportion city	.36	.37	> .250	.37	.37	> .250
Proportion suburb	.47	.41	< .001	.40	.38	> .250
Proportion town/rural	.17	.22	< .001	.23	.24	.159

Note: Sample sizes varied across the 50 imputed data sets. The table presents means, unless otherwise indicated (standard deviations are given in parentheses); values were averaged across imputations.

We then repeated the OLS regressions with the PSM-matched samples; results from these models are presented in Table 3. Children who were spanked experienced stronger increases in their externalizing behavior problems from age 5 to age 6 (β = 0.06, p = .023) and to age 8 (β = 0.07, p = .014) than children who were never spanked. Among children who had been spanked, children spanked in the past week at age 5 did not significantly increase in their behavior problems by age 6 (β = 0.04, p = .140) but did by age 8 (β = 0.09, p < .001). Thus, even when characteristics known to predict whether parents spank their children were removed from the model through PSM, 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 spanked (in past week vs. not in past week).

On finding that having 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 with age 5) was mediated by increased externalizing behavior problems at age 6. We estimated a regression in which externalizing behavior at age 8 was regressed both on 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 for the two spanked groups, namely, those spanked in the past week and those spanked but not in the past week, because having been spanked in the past week at age 5 did not significantly predict externalizing behavior at age 6.

Discussion

The analyses presented in this article 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,

Table 2. Demographic Characteristics of the Not-Spanked-in-Past-Week Group and the Spanked-in-Past-Week Group Before and After Propensity Score Matching (PSM)

	Before PSM			After PSM		
	Not spanked in past week	Spanked in past week		Not spanked in past week	Spanked in past week	
Variable	ns = 6,445–6,464	ns = 3,167–3,181	p	ns = 4,778–4,953	ns = 3,159–3,177	p
Child characteristics						
Age at spring of kindergarten (in months)	74.85 (4.45)	74.28 (4.40)	< .001	74.38 (4.39)	74.35 (4.49)	> .250
Proportion male	.51	.55	< .001	.54	.54	> .250
Overall health	1.68 (0.83)	1.77 (0.85)	< .001	1.74 (0.85)	1.74 (0.84)	> .250
Proportion diagnosed with a disability	.15	.17	.015	.15	.15	> .250
Externalizing behavior at age 5 years	1.64 (0.62)	1.81 (0.71)	< .001	1.76 (0.68)	1.76 (0.67)	> .250
Parent characteristics						
Mother's education (in years)	14.64 (3.42)	14.05 (3.40)	< .001	14.47 (3.47)	14.48 (3.37)	> .250
Mother's age (in years)	33.39 (6.63)	32.04 (6.64)	< .001	32.76 (6.39)	32.74 (6.50)	> .250
Proportion of mothers born in U.S.	.84	.83	> .250	.81	.81	> .250
Proportion of parents married	.70	.62	< .001	.71	.71	> .250
Proportion of parents separated	.05	.05	> .250	.04	.04	> .250
Proportion of parents divorced	.09	.10	.107	.08	.08	> .250
Proportion of parents widowed	.01	.01	> .250	.01	.01	> .250
Proportion of parents never married	.12	.19	< .001	.14	.14	> .250
Proportion of parents who were nonbiological/ adoptive	.03	.03	> .250	.02	.02	> .250
Family economic status						
Household size	4.51 (1.36)	4.51 (1.44)	> .250	4.53 (1.40)	4.52 (1.38)	> .250
Proportion of mothers employed full time	.46	.47	> .250	.46	.46	> .250
Proportion of mothers employed part time	.22	.19	.018	.20	.20	> .250
Proportion of mothers unemployed	.32	.34	.128	.34	.33	> .250
Income (in thousands of dollars)	50.01 (36.09)	39.36 (33.51)	< .001	45.03 (34.34)	44.84 (34.73)	> .250
Family food insecurity	0.59 (1.76)	0.98 (2.24)	< .001	0.77 (2.04)	0.76 (1.95)	> .250
Harshness of home environment						
Parenting harshness	0.19 (0.14)	0.18 (0.15)	< .001	0.18 (0.15)	0.18 (0.15)	> .250
Low parenting warmth	1.29 (0.36)	1.35 (0.40)	< .001	1.36 (0.41)	1.35 (0.40)	> .250
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

(continued)

Table 2. *(continued)*

	Before PSM			After PSM		
	Not spanked in past week	Spanked in past week		Not spanked in past week	Spanked in past week	
Variable	ns = 6,445–6,464	ns = 3,167–3,181	p	ns = 4,778–4,953	ns = 3,159–3,177	p
Mother's depressive symptoms	1.45 (0.46)	1.56 (0.50)	< .001	1.52 (0.50)	1.51 (0.47)	> .250
Cultural background						
Proportion White	.59	.50	< .001	.55	.55	> .250
Proportion Black	.15	.22	< .001	.16	.16	> .250
Proportion Asian/ other	.07	.07	> .250	.11	.10	> .250
Proportion Latino ethnicity	.18	.20	.047	.18	.18	> .250
Proportion who spoke English at home	.88	.87	.120	.86	.87	> .250
Mother's religiosity	2.86 (1.32)	2.96 (1.36)	< .001	3.01 (1.32)	3.02 (1.33)	> .250
Geographic characteristics						
Proportion Northeast	.16	.14	.007	.15	.15	> .250
Proportion Midwest	.25	.18	< .001	.20	.21	> .250
Proportion South	.36	.48	< .001	.42	.43	> .250
Proportion West	.23	.20	.001	.22	.22	> .250
Proportion city	.37	.38	> .250	.38	.38	> .250
Proportion suburb	.42	.39	.006	.36	.36	> .250
Proportion town/rural	.21	.24	.037	.26	.26	> .250

Note: Sample sizes varied across the 50 imputed data sets. The table presents means, unless otherwise indicated (standard deviations are given in parentheses); values were averaged across imputations.

did not find support for a child-elicitation explanation. Whether children had ever been spanked by 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 been spanked predicted increased externalizing behavior problems over time. We also found that part of the process by which spanking predicts behavior problems 3 years into

Table 3. Results From Regressions Predicting Child Externalizing Behavior From the Unmatched and Matched Groups

	OLS regressions with unmatched samples[a]								OLS regressions with matched samples[b]							
	Age 6 externalizing behavior problems				Age 8 externalizing behavior problems				Age 6 externalizing behavior problems				Age 8 externalizing behavior problems			
	b	95% CI	β	p	b	95% CI	β	p	b	95% CI	β	p	b	95% CI	β	p
Spanked (vs. never spanked) by age 5	0.04	[0.00, 0.12]	0.06	.038	0.03	[–0.01, 0.10]	0.04	.139	0.04	[0.01, 0.11]	0.06	.023	0.04	[0.01, 0.12]	0.07	.014
Spanked in past week (vs. not spanked in past week) at age 5	0.03	[–0.00, 0.11]	0.05	.059	0.07	[0.05, 0.16]	0.11	< .001	0.02	[–0.01, 0.09]	0.04	.140	0.06	[0.04, 0.15]	0.09	< .001

Note: Spanking was mother rated; externalizing behavior problems was teacher rated. OLS = ordinary least squares, CI = confidence interval.
[a]Sample size for the regressions with unmatched spanked and never-spanked groups was 12,112. Sample size for the regressions with unmatched spanked-in-the-past-week and not-spanked-in-the-past-week groups was 9,629. These results were weighted to be nationally representative.
[b]Sample size for the regressions with matched spanked and never-spanked groups was roughly 12,082 (sample sizes varied across imputations). Sample size for the regressions with matched spanked-in-the-past-week and not-spanked-in-the-past-week groups was roughly 8,008 (sample sizes varied across imputations).

the future is by increasing behavior problems 1 year later.

There are two other 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 he or she had been spanked in the week before the survey, rather than knowing exactly how many spankings he or she was subjected to either over time or in any given instance, was a significant predictor of that child's externalizing behavior. This is noteworthy given that we matched the two sets of spanked groups on a range of factors that have been shown to predict whether parents spank their children 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 spanked their children because parents had to volunteer this information when asked how often they spanked them, making it even harder to find differences between the spanked and never-spanked groups. Yet although some critics have argued that children who have been spanked even once 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 did 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 data set.

However, this limitation is tempered by the fact that we included more than 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 1 and 3 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.

Action Editor

Ian H. Gotlib served as action editor for this article.

Author Contributions

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

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

References

Abidin, R. R., & Abidin, R. R. (1990). *Parenting Stress Index (PSI)*. Charlottesville, VA: Pediatric Psychology Press.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Baumrind, D., Larzelere, R. E., & Cowan, P. A. (2002). Ordinary physical punishment: Is it harmful? Comment on Gershoff (2002). *Psychological Bulletin*, *128*, 580–589. doi:10.1037/0033-2909.128.4.580

Berlin, L. J., Ispa, J. M., Fine, M. A., Malone, P. S., Brooks-Gunn, J., Brady-Smith, C., . . . Bai, Y. (2009). Correlates and consequences of spanking and verbal punishment for low-income White, African American, and Mexican American toddlers. *Child Development*, *80*, 1403–1420. doi:10.1111/j.1467-8624.2009.01341.x

Bickel, G., Nord, M., Price, C., Hamilton, W. L., & Cook, J. T. (2000). *Guide to measuring household food security*. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss, sadness and depression*. New York, NY: Basic Books.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas at Little Rock.

Choe, D. E., Olson, S. L., & Sameroff, A. J. (2013). The interplay of externalizing problems and physical and inductive discipline during childhood. *Developmental Psychology*, *49*, 2029–2039. doi:10.1037/a0032054

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In M. E. Lamb (Vol. Ed.) & R. M. Lerner (Series Ed.), *Handbook of child psychology and developmental science: Vol. 3. Socioemotional processes* (7th ed., pp. 794–841). New York, NY: Wiley.

Ferguson, C. J. (2013). Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies. *Clinical Psychology Review*, *33*, 196–208. doi:10.1016/j.cpr.2012.11.002

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, *173*, 761–767. doi:10.1093/aje/kwq439

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, *128*, 539–579. doi:10.1037/0033-2909.128.4.539

Gershoff, E. T., & Grogan-Kaylor, A. (2016). Corporal punishment by parents and its consequences for children: Old controversies and new meta-analyses. *Journal of Family Psychology*, *30*, 453–469. doi:10.1037/fam0000191

Gershoff, E. T., Lansford, J. E., Sexton, H. R., Davis-Kean, P., & Sameroff, A. J. (2012). Longitudinal links between spanking and children's externalizing behaviors in a national sample of White, Black, Hispanic, and Asian American families. *Child Development*, *83*, 838–843. doi:10.1111/j.1467-8624.2011.01732.x

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System manual*. Circle Pines, MN: American Guidance Service.

Grogan-Kaylor, A. (2005). Relationship of corporal punishment and antisocial behavior by neighborhood. *Archives of Pediatric & Adolescent Medicine*, *159*, 938–942. doi:10.1001/archpedi.159.10.938

Grolnick, W. S., Deci, E. L., & Ryan, R. M. (1997). Internalization within the family: The self-determination theory perspective. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 135–161). New York, NY: Wiley.

Lansford, J. E., Deater-Deckard, K., Bornstein, M. H., Putnick, D. L., & Bradley, R. H. (2014). 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*, *164*, 1208–1213. doi:10.1016/j.jpeds.2013.11.060

Larzelere, R. E., Gunnoe, M. L., Roberts, M. W., & Ferguson, C. J. (2017). Children and parents deserve better parental discipline research: Critiquing the evidence for exclusively "positive" parenting. *Marriage & Family Review*, *53*, 24–35. doi:10.1080/01494929.2016.1145613

Larzelere, R. E., Kuhn, B. R., & Johnson, B. (2004). The intervention selection bias: An underrecognized confound in intervention research. *Psychological Bulletin*, *130*, 289–303. doi:10.1037/0033-2909.130.2.289

Lepper, M. R. (1983). Social control processes and the internalization of social values: An attributional perspective. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development* (pp. 294–330). New York, NY: Cambridge University Press.

Maguire-Jack, K., Gromoske, A. N., & Berger, L. M. (2012). Spanking and child development during the first 5 years of life. *Child Development*, *83*, 1960–1977. doi:10.1111/j.1467-8624.2012.01820.x

Miller, P., Henry, D., & Votruba-Drzal, E. (2016). Strengthening causal inference in developmental research. *Child Development Perspectives*, *10*, 275–280. doi:10.1111/cdep.12202

Mulvaney, M. K., & Mebert, C. J. (2007). Parental corporal punishment predicts behavior problems in early childhood. *Journal of Family Psychology*, *21*, 389–397. doi:10.1037/0893-3200.21.3.389

National Institute of Child Health and Human Development Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*, 1454–1475. doi:10.1111/1467-8624.00617

Olson, S. L., Lopez-Duran, N., Lunkenheimer, E. S., Chang, H., & Sameroff, A. J. (2011). Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and Psychopathology*, *23*, 253–266. doi:10.1017/S0954579410000775

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401. doi:10.1177/014662167700100306

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. doi:10.1093/biomet/70.1.41

Russa, M. B., Rodriguez, C. M., & Silvia, P. J. (2014). Frustration influences impact of history and disciplinary attitudes on physical discipline decision making. *Aggressive Behavior*, *40*, 1–11. doi:10.1002/ab.21500

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290–312. doi:10.2307/270723

StataCorp. (2015). Stata Statistical Software: Release 14 [Computer software]. College Station, TX: Author.

Straus, M. A. (1990). The Conflict Tactics Scales and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in*

American families: Risk factors and adaptations to violence in 8,145 families (pp. 49–73). New Brunswick, NJ: Transaction Publishing.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1–21. doi:10.1214/09-STS313

Taylor, C. A., Hamvas, L., Rice, J., Newman, D. L., & DeJong, W. (2011). Perceived social norms, expectations, and attitudes toward corporal punishment among an urban community sample of parents. *Journal of Urban Health, 88*, 254–269. doi:10.1007/s11524-011-9548-7

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *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 (NCES 2009–004).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.