

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Research article

Using Bayesian analysis to examine associations between spanking and child externalizing behavior across race and ethnic groups



Andrew Grogan-Kaylor^{a,*}, Julie Ma^b, Shawna J. Lee^a, Berenice Castillo^{a,c},
Kaitlin P. Ward^{a,c}, Sacha Klein^d

^a University of Michigan-Ann Arbor, School of Social Work, United States

^b University of Michigan-Flint, Department of Social Work, United States

^c University of Michigan, Department of Psychology, United States

^d Michigan State University, School of Social Work, United States

ARTICLE INFO

Keywords:

Corporal punishment
Physical punishment
Bayesian statistics

ABSTRACT

While corporal punishment is widely understood to have undesirable associations with children's behavior problems, there remains controversy as to whether such effects are consistent across different racial or ethnic groups. We employed a Bayesian regression analysis, which allows for the estimation of both similarities and differences across groups, to study whether there are differences in the relationship of corporal punishment and children's behavior problems using a diverse, urban sample of U.S. families ($n = 2653$). There is some moderation of the relationship between corporal punishment and child behavior by race or ethnicity. However, corporal punishment is associated with increases in behavior problems for all children. Thus, our findings add evidence from a new analytical lens that corporal punishment is consistently linked to increased externalizing behavior across African American, White, or Hispanic children, even after earlier externalizing behavior is controlled for. Our findings suggest that corporal punishment has detrimental consequences for all children and that all parents, regardless of their racial or ethnic background, should be advised to use alternatives to corporal punishment.

1. Introduction

1.1. Corporal punishment and child development

One of the long-standing issues of interest in the child development literature has related to the strength and rigor of research showing that corporal punishment is associated with negative outcomes for children. Corporal punishment (CP) is the use of physical force, including hitting, spanking, smacking, whipping or slapping a child, “with the intention of causing a child to experience pain, but not injury, for the purpose of correcting or controlling the child's behavior” (Straus & Donnelly, 2005, pg. 3). Corporal punishment is lawful in the home in all U.S. states, and it is lawful in public schools in 19 U.S. states (Gershoff & Font, 2016).

To briefly summarize the nature of the debate on corporal punishment and child development, scholars responded to an early meta-analysis that showed that parental spanking of children was associated with a host of negative outcomes (Gershoff, 2002) with a critique suggesting that many studies on which the meta-analysis was based had serious methodological limitations (Baumrind, Larzelere, & Cowan, 2002). Furthermore, researchers argued that many of the studies included in the initial meta-analysis failed to

* Corresponding author.

E-mail address: agrogan@umich.edu (A. Grogan-Kaylor).

<https://doi.org/10.1016/j.chiabu.2018.10.009>

Received 11 May 2018; Received in revised form 9 October 2018; Accepted 18 October 2018

Available online 28 October 2018

0145-2134/ © 2018 Elsevier Ltd. All rights reserved.

account for the context in which corporal punishment occurs. From the *conditional corporal punishment* perspective, the effects of spanking are “not necessarily negative or positive, but may be either or both depending on many other conditions” that characterize the parent-child relationship (Kazdin & Benjet, 2003). Commonly cited conditional corporal punishment arguments include the notion that parental corporal punishment may not be harmful when it is done in the context of an otherwise warm and loving parent-child relationship (Deater-Deckard, Ivy, & Petrill, 2006; McLoyd & Smith, 2002), or in subgroups of the U.S. population where use of such behavior is normative (Deater-Deckard, Bates, Dodge, & Pettit, 1996; Deater-Deckard & Dodge, 1997). A meta-analysis examining conditional corporal punishment suggested that, to reduce child noncompliance and antisocial behavior, conditional spanking was favored over most alternative disciplinary methods (Larzelere & Kuhn, 2005).

Subsequent to these critiques, the rigor and sophistication of the empirical evidence regarding the associations of parental corporal punishment to negative consequences for children improved considerably. Recognizing the quantity and quality of the studies published, in 2016 a meta-analysis was published incorporating over 50 years of research (Gershoff and Grogan-Kaylor, 2016a). The updated meta-analysis was based on studies that employed a more restrictive definition of “spanking” as a disciplinary tactic that does not intend to harm the child, many of them newer studies than those included in the original Gershoff, Sattler, and Ansari (2018) review, and many of them studies that used sophisticated analytical techniques (e.g., longitudinal analysis, rigorous control variables, etc.) of large, diverse samples. In sum, the results provided strong support based on analysis of 111 effect sizes from different studies for the supposition that physical punishment of children predicts increased aggressive and antisocial behavior (Gershoff & Grogan-Kaylor, 2016a). Indeed, other studies using fixed effects regression have found the link between physical punishment and aggressive behavior even after accounting for potential confounds inherent in the child and the parent such as the child’s initial level of aggression, parent’s temperament, and genetic heritage in the family (Grogan-Kaylor, 2004; Ma, Grogan-Kaylor, & Lee, 2018).

Research has also provided little support for conditional corporal punishment arguments. Two longitudinal studies showed no evidence of an interaction between maternal warmth and spanking on child aggression (Berlin et al., 2009; Stacks, Oshio, Gerard, & Roe, 2009). Another study showed that high levels of maternal warmth did not buffer children against the negative effects of maternal spanking (Lee, Altschul, & Gershoff, 2013). In addition, research has examined whether associations linking spanking and negative child outcomes were moderated by race and ethnicity. Apparent differences in parental attitudes and behavior concerning the use of corporal punishment for child discipline exist across race and ethnicity in the U.S. such that Black parents use corporal punishment most frequently followed by White and Hispanic parents (Berlin et al., 2009; Ma & Klein, 2018; Smith, Davern, Freese, & Hout, 2017). Cultural processes and socio-economic and contextual conditions that are confounded with race and ethnicity in the U.S. may also alter the associations between corporal punishment and negative child behavior (García Coll et al., 1996). Yet, contrary to the cultural normativeness perspective, current research suggests that spanking is associated with detrimental outcomes for White and Black children (Gershoff & Grogan-Kaylor, 2016b) and for White and Hispanic children in the U.S. (Ma & Klein, 2018). International studies have also shown that even in contexts in which corporal punishment is normative, in other words, is culturally accepted and used often or by the majority of parents, corporal punishment remains a significant predictor of negative outcomes for children (Gershoff et al., 2010).

However, despite the publication of the two comprehensive meta-analyses supporting associations between maternal use of corporal punishment and negative outcomes for children (Gershoff & Grogan-Kaylor, 2016a, 2016b), researchers continue to voice claims that the research base has not been sufficiently strong to make causal arguments (Larzelere, Gunnoe, Roberts, & Ferguson, 2017). Thus, debate continues within research communities and beyond regarding the strength of the research base on this topic (e.g., Holden, Grogan-Kaylor, Durrant, & Gershoff, 2017; Larzelere et al., 2017). In the current study, we examine the association between maternal corporal punishment and child behavior problems using a Bayesian approach, with particular attention to questions of whether effects of corporal punishment differ across race and ethnic groups. As described below, Bayesian analysis allows estimation of these relationships in a way that may be more accurate than more traditional frequentist approaches. Bayesian estimation may aid researchers in determining whether corporal punishment does, indeed, lead to negative outcomes for most children, or whether outcomes vary depending on the race or ethnicity of children and parents.

1.2. Bayes theorem

When analyzing data using frequentist data analysis methods such as an ordinary least squares regression in which p values are calculated to assess statistical significance, we are estimating the $P(D|H)$, or the probability that we would observe the relationship in the data that we did, given a particular hypothesis about the data. More specifically, we are estimating $P(D|H_0)$, the probability that if the null hypothesis were true, that we would observe data as extreme, or more extreme, than the data we observed. This significance level also reflects the probability of the null hypothesis being rejected simply due to random variation (i.e., Type I error). For example, in a study of the relationship of corporal punishment with undesired child outcomes, a significant outcome would suggest that, given a null hypothesis (H_0) of no relationship between physical punishment and undesired child outcomes, we would observe an association at least as strong as the one we observed in less than 5% of studies due to random sampling error. We call this probability $P(D|H_0)$ the p value, and by convention, we say that our results are statistically significant whenever $p = P(D|H_0) < 0.05$. For purposes of our discussion, it is worth noting especially that this p value says nothing *directly* about whether physical punishment is associated with undesired child outcomes, but instead is a statement about the probability of the data under an assumption that the two quantities, spanking and children’s behavior problems, are *not* related (Nuzzo, 2014; CF Ioannidis, 2018).

Additionally, when an estimate and its respective p value are provided under a frequentist method, the researcher is also provided with a confidence interval. The frequentist confidence interval assumes a symmetric distribution, and is computed by calculating *Estimate* \pm 1.96 \times SE. The frequentist 95% confidence interval means that, if an infinite number of samples were taken from the

given population, 95% of those samples would contain the true population value.

In contrast, when analyzing data using Bayesian data analysis, we are looking for $P(H|D)$, the probability of different parameter values supporting our hypothesis, given the data. Additionally, instead of the frequentist confidence interval, Bayesian estimates provide a credible interval that has an interpretation that tends to be more intuitive. A 95% credible interval means that there is a 95% probability that the estimated value lies within the limits of the given interval. The difference between the two approaches ($D|H$) vs. $P(H|D)$, and confidence intervals versus credible intervals, may seem arcane and difficult to discern. However, under the assumptions that we use in this paper, Bayesian methods are likely to provide a more stringent test of hypotheses than frequentist estimates (CF Sterne, Smith, & Cox, 2001).

Further, one of the main advantages of Bayesian statistics lies in the interpretation of and conclusions drawn from results, as the Bayesian paradigm allows us to interpret results (including parameter estimates and their distributions) in a different and potentially more meaningful manner than results interpreted under the frequentist paradigm. Whereas frequentist statistics provide a researcher with a parameter estimate and its standard error, Bayesian estimation provides the researcher with an estimate of the entire distribution. Notably, as will become more clear in the presentation of our results, Bayesian methods provide information about the full distribution of a parameter, while frequentist analysis would only provide a simple point estimate for that parameter (StataCorp, 2017a; StataCorp, 2017b). Fig. 1 presents some hypothetical Bayesian posterior parameter distributions, which may provide an illustration of some of the differences between frequentist and Bayesian analyses. While frequentist analysis would only provide a single best point estimate (analogous to the dashed line in Fig. 1), Bayesian analysis would provide a full posterior distribution for the parameter, and thus would provide a more qualitatively distinct picture of the parameter estimate than the simple point estimate provided by frequentist analysis. Therefore, Bayesian estimation may potentially allow researchers to gain a more comprehensive picture of patterns, such as harsh parenting, through obtaining the entire distribution estimate of the parameter.

There is a sometimes unappreciated consequence of interpreting results from standard frequentist analyses that John Kruschke (2011) has pointed out. Usually the null hypothesis is phrased as a statement about a mean being equal to 0 ($H_0: \bar{x} = 0$), or a regression parameter being equal to 0 ($H_0: \beta_k = 0$). We then look for enough evidence to reject this null hypothesis based on the p value that measures the probability that we found the observed evidence against the null hypothesis by random chance. An important caveat of frequentist analysis is that we can either reject or fail to reject the null hypothesis, but are unable to accept the null hypothesis, nor can the p value determine the probability that the null hypothesis is true. Thus, frequentist results evaluate the strength of evidence in the data to conclude that the relationship between the variables does not equal to 0 by rejecting the null hypothesis, or that the relationship between the variables is 0 ($H_0: \bar{x} = 0$ or $H_0: \beta_k = 0$) by failing to reject the null hypothesis (Nuzzo, 2014; Sterne & Smith, 2001). As these results cannot ascertain the degree to which the null hypothesis is true, the true relationship between the variables in the null hypothesis remains unknown. Also, p values are strongly dependent on statistical power in the data to reject the null hypothesis and detect a meaningful effect (Figueiredo Filho et al., 2013). With enhanced power in studies with large enough sample sizes, frequentist analysis will ultimately be able to find statistically significant results even for marginal effects (Kruschke, Aguinis, & Joo, 2012).

In contrast, Bayesian analysis is not based upon the idea of rejecting null hypotheses. By directly testing the probability that various hypotheses are true and making inferences about the underlying relationships between the variables, Bayesian analysis is able to conclude that the null hypothesis is the most likely result given the data at hand. In essence, Bayesian statistics allows researchers to accept the null hypothesis from a probability standpoint. Another important distinction between frequentist and Bayesian analysis is that Bayesian models allow the researcher to build in existing knowledge on the variables in the data (prior knowledge). Bayesian analysis adds evidence to the prior constructed by the researcher and then draws the conclusion given the data (posterior). Put another way, the posterior distribution, or conclusion about parameter estimates, is a mixture of the priors and the data (Van de Schoot et al., 2014). When prior information about the given parameters within a particular sample is weak or unavailable, it is preferable to allow the data to inform the probability distribution (Depaoli & van de Schoot, 2017; Van de Schoot et al., 2014). Therefore, in situations where priors are not sufficiently informative, “uninformative” or “default” priors can be used, which distributions are typically centered at 0 and allow for greater variance.

1.3. Bayes theorem applied to the study of child development

There is a considerable amount of empirical work on the association between parenting behaviors and child outcomes (for reviews see Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; Hovee et al., 2009; Kawabata, Alink, Tseng, van Ijzendoorn, & Crick, 2011). As reviewed previously, parental use of physical punishment is shown to be associated with increases in undesirable outcomes in children (Gershoff & Grogan-Kaylor, 2016b). At the same time, less is known about whether such associations of parenting and children’s outcomes are universal across different groups.

Consider the following practical example: Imagine a situation in which we are testing the link between a parenting behavior like spanking or expressing warmth, and an outcome like internalizing or externalizing behavior, among children of different racial or ethnic identities. Conceptually, we might test whether the relationship of parenting and child outcome is equivalent in these groups. Imagine further, that based upon the prior literature, we suspect that this relationship is somewhat universal, and therefore, that the relationship between a particular parenting behavior, such as physical punishment, and a particular child outcome, such as aggressive behavior, is consistent across groups.

A regression model might take the following form:

$$\text{child outcome}_i = \beta_0 + \beta_1 \text{spanking} + \beta_2 \text{group2} + \beta_3 \text{spanking} * \text{group2} + e_i$$

Possible Parameter Distributions For Bayesian Regression Parameters

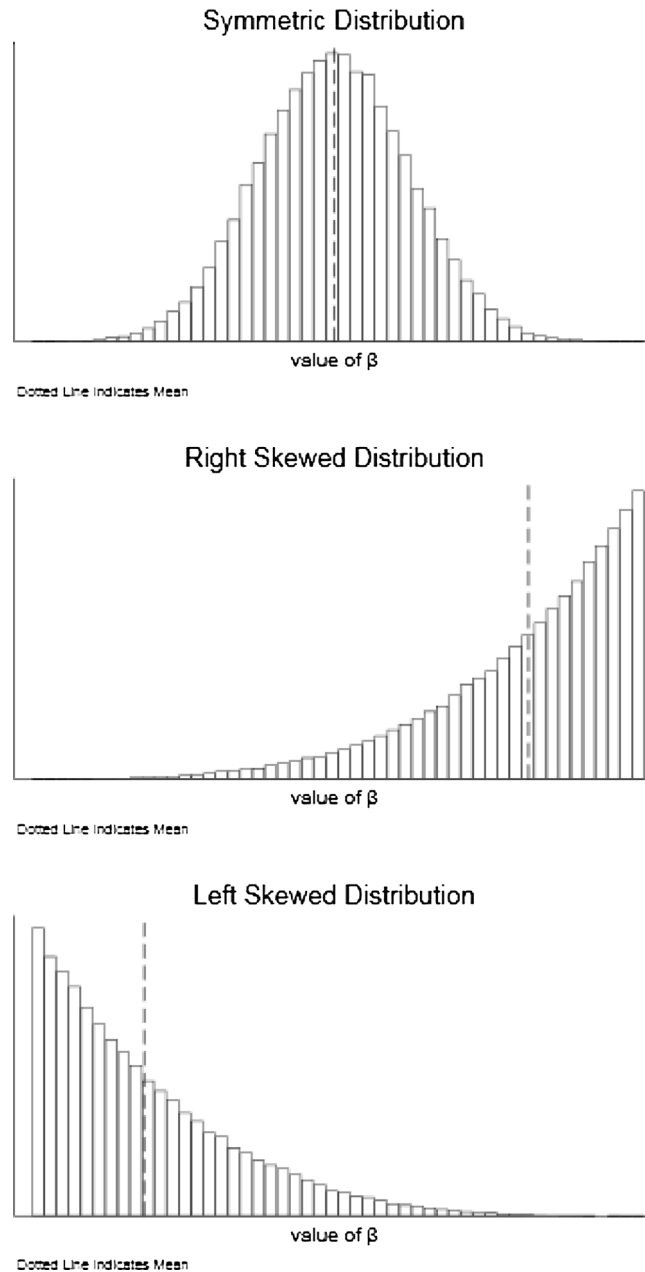


Fig. 1. Hypothetically Possible Distributions for a Regression Parameter.

The test of group differences would revolve around the β coefficients attached to the interaction terms, namely:

$$\beta_3 \text{ spanking} * \text{group2}$$

In respect to the outcomes of this study (and other child development studies), a primary advantage of Bayesian statistics over other statistical methods is the way in which results are tested and interpreted. From a frequentist perspective, the two null hypotheses to be tested are that the β_3 coefficients in the equations immediately above are equal to 0: $H_0: \beta_3 = 0$. Therefore, in interpreting the results, we would only be able to reject, or fail to reject, the null hypothesis. We could neither accept the null

hypothesis, nor interpret the degree to which the null or alternative hypotheses are true. In contrast, a Bayesian perspective would be able to conclude that the most likely result is that there are differences between the two groups, or that the most likely result is that there are not any differences between the two groups. Thus, in this study, using Bayesian analysis allows us to directly accept the idea that the effects of spanking are consistent across groups, or, if the evidence suggests, that the effects of spanking are different across groups.

2. Method

2.1. Sample

Our sample is derived from the Fragile Families and Child Wellbeing Study (FFCWS; $N = 4898$), a nationally representative sample of families with a focal child born in 20 large U.S. cities with populations of 200,000 or more between years 1998 to 2000 (Reichman, Teitler, Garfinkel, & McLanahan, 2001). The FFCWS oversampled nonmarital births such that single mothers constitute approximately three fourth of the full sample. Participants were recruited at the time of their child's birth in hospitals, where they completed the baseline (Wave 1; $n = 4789$) in-person core interviews. Follow-up core interviews were conducted over the phone at child age 1 (Wave 2; $n = 4,270$), age 3 (Wave 3; $n = 4140$), and age 5 (Wave 4; $n = 4,055$). At Wave 3 and Wave 4, a collaborative In-Home Longitudinal Study of Pre-School Aged Children (In-Home study hereafter) assessed various domains of child development using a parent interview and interviewer observation during a home visit. Participants who refused a home visit completed only the parent interview of the In-Home survey over the phone. Our analysis sample includes families who participated in both the Wave 3 ($n = 3,288$) and Wave 4 ($n = 3,024$) In-Home interviews, either in-person or over the phone.

2.2. Measures

2.2.1. Dependent variable

Externalizing Behavior Problems at age 5. Consistent with prior studies that examined externalizing behavior in the FFCWS (e.g., Ma & Grogan-Kaylor, 2017; Maguire-Jack, Gromoske, & Berger, 2012), we used the entire Child Behavior Checklist (CBCL) Aggressive Behavior subscale to approximate externalizing behavior. During the Wave 4 In-Home study, mothers responded to 20 items in the CBCL 4–18 Aggressive Behavior subscale (Achenbach, 1991) describing behaviors that the focal child may have engaged in during the past two weeks (0 = *not true*, 1 = *somewhat or sometimes true*, or 2 = *very true or often true*). The mean score of the CBCL 4–18 Aggressive Behavior subscale (e.g., “Child destroys his/her own things”, “Child physically attacks people”) assessed externalizing behavior.

2.3. Covariates

2.3.1. Externalizing behavior problems at wave 3

We controlled for focal children's mean raw CBCL Externalizing Behavior scores at Wave 3. At Wave 3, 15 items from the CBCL 2–3 Aggressive Behavior subscale (Achenbach, 1992), age appropriate for 2 to 3-year olds, such as “Child is defiant” and “Child hits others”, were assessed.

2.3.2. Maternal spanking at age 5

During the Wave 4 core interview, mothers were asked the following question about their use of spanking in the past month: “Sometimes children behave pretty well and sometimes they don't. In the past month, have you spanked (child) because (he/she) was misbehaving or acting up?” If the mother responded ‘yes’ to this question, a subsequent question asked how frequently she spanked the child in the past month. We collapsed responses to these questions to create an ordinal scale indicative of the frequency of maternal spanking at age 5 (0 = *never*, 1 = *only once or twice*, 2 = *a few times this past month*, 3 = *a few times a week or every day or nearly every day*).

2.3.3. Income of family when child age 3

Annual family income was measured during the Wave 3 core interview. Mothers reported the income of all individuals in the household in the prior year.

2.3.4. Race or ethnicity

Mothers reported their race or ethnicity (1 = *non-Hispanic White*, = *non-Hispanic Black*, 3 = *Hispanic*, 4 = *other*) during the baseline core interview.

2.3.5. Child sex

Child's sex (1 = *boy*, 2 = *girl*) was assessed during the baseline core interview.

2.4. Analysis

In this analysis, the measure of child externalizing behavior problems at age 5 was the dependent variable, while the above

mentioned covariates were entered as independent variables into the model. In order to facilitate interpretation, family income was divided by \$10,000 in the regression model so that the regression coefficient for income represented the association of a \$10,000 increase in annual income. We also entered interactions of maternal use of spanking with the race or ethnicity of the mother reported in the FFCWS. Consequently, we entered two interaction terms into the model: the interaction of spanking and identifying as Black, and the interaction of spanking and identifying as Hispanic.

We employed a Bayesian regression analysis using Stata 15.1 (StataCorp, 2017a; StataCorp, 2017b). Because we did not have sufficient prior information about the distribution of our parameters for our sample, we used the default “uninformative” regression priors provided by Stata, which were normal distributions centered at 0, with standard deviations of 10,000. This allowed the data to inform our probability distribution. Bayesian estimation was accomplished through the implementation of the Markov Chain Monte Carlo (MCMC) estimation using the Metropolis-Hastings Algorithm (StataCorp, 2017a; StataCorp, 2017b) with 2500 burn-in iterations, and 10,000 subsequent MCMC iterations. There were some missing data. Missing data were handled by using complete case analysis.

3. Results

3.1. Descriptive statistics

The analytic sample for this study, $n = 2653$ families, was drawn from the participants with complete data on all variables who completed the FFCWS In-Home surveys at Wave 3 and Wave 4. The majority of the sample identified as non-Hispanic Black (51.84%), followed by Hispanic (25.11%), and non-Hispanic White (23.05%). The FFCWS oversampled unmarried, low-income families, and the mean annual income for the families in this study was approximately \$34,481. Descriptive statistics indicate that average maternal use of spanking was 0.75 at child’s age 5, indicating a relatively low level of spanking overall. Children’s externalizing behavior decreased from a mean of 0.61 at age 3 to 0.44 at age 5. More detailed descriptive statistics are shown in Table 1.

3.2. Bayesian regression

Results from the Bayesian regression analysis are shown in Table 2. We base our interpretation of the Bayesian regression results primarily upon the mean of the parameter estimate and the 95% credible interval for that parameter estimate (StataCorp, 2017a; StataCorp, 2017b). For a number of factors, the equal tailed credible interval excluded 0, suggesting that there was a 95% probability that the parameter estimate was not 0. For example, externalizing behavior of the child at age 3 appeared to have an effect on externalizing behavior at age 5, in that higher levels of earlier externalizing behavior were associated with higher levels of later child externalizing behavior. Maternal use of spanking at age 5 was also associated with higher levels of child externalizing behavior at age 5, after controlling for prior externalizing behavior. The full posterior distribution of the regression coefficient for spanking is shown in Fig. 2. While there was variation in possible estimates of the regression coefficient for spanking, the credible interval clearly excluded 0. Lastly, increases in family income appeared to be associated with decreases in child externalizing behavior. A number of other factors did not appear to have an association with child externalizing behavior. Child sex did not have an association with externalizing behavior. Black families did not differ from White families in terms of the child’s level of externalizing behavior. Children in Hispanic families had lower levels of externalizing behavior than children in White families, as indicated by a credible interval that excluded zero. For Black families, there was no interaction of racial or ethnic identity and the use of spanking. However, in Hispanic families, there was a noticeable interaction of race or ethnicity and spanking, such that maternal spanking was associated with child’s externalizing behavior problems to a greater degree than in White families.

4. Discussion

Data from the most recent General Social Survey show that just over 68% of Americans endorse the statement that “it is sometimes necessary to discipline a child with a good, hard spanking” (Smith et al., 2017). Not surprisingly, spanking remains one of the most common discipline strategies used by U.S. parents. Most U.S. parents spank their children, often beginning when children

Table 1
Descriptive Statistics.

Measure (Range)	Total Sample $n = 2653$ Mean (SD)	White $n = 622$ (23.45%) Mean	Black $n = 1382$ (52.09%) Mean	Hispanic $n = 649$ (24.46%) Mean
Externalizing behavior at child age 5 (0-2)	0.44 (0.28)	0.43	0.45	0.43
Externalizing behavior at child age 3 (0-1.91)	0.61 (0.35)	0.57	0.63	0.62
Maternal spanking at child age 5 (0-4) ^a	0.75 (0.91)	0.76	0.84	0.56
Maternal spanking at child age 5 (Y/N) ^b	50.36%	47.88%	56.08%	40.52%
Family income at child age 3 (0-999,999)	\$34,481.03 (43,751.26)	\$61,441.64	\$25,581.11	\$28,048.67

^aContinuous variable employed in regression analysis.

^bCategorical variable included for descriptive purposes.

Table 2
Bayesian Regression Results.

Measure	Mean	Standard Deviation	Equal Tailed 95% Credible Interval
Child's sex (reference group: male)	-0.006	0.009	-0.024 - 0.012
Income at child age 3	-0.004	0.001	-0.006 - -0.002
Externalizing behavior at child age 3	0.391	0.014	0.365 - 0.418
Maternal spanking at child age 5	0.024	0.010	0.004 - 0.043
Race (reference group: White)			
Black	-0.026	0.016	-0.057 - 0.004
Hispanic	-0.044	0.017	-0.077 - -0.010
Maternal Spanking at age 5 X Race or Ethnicity is Black	0.011	0.011	-0.013 - 0.034
Maternal Spanking at age 5 X Race or Ethnicity is Hispanic	0.032	0.015	0.002 - 0.061
Constant	0.219	0.021	0.176 - 0.261

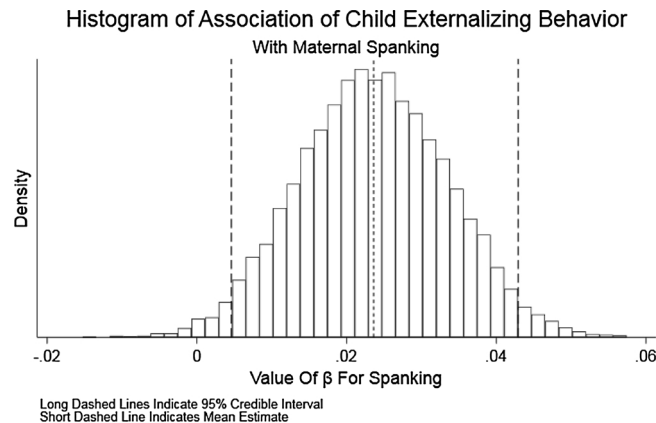


Fig. 2. Parameter Distribution for Association of Spanking and Externalizing Behavior.

are very young. About 30% of mothers report spanking their 1-year-old child in the past month (Lee, Grogan-Kaylor, & Berger, 2014), and 70% of mothers say they have spanked their child at least once by the time he or she is 2-years old (< 24 months) (Zolotor, Robinson, Runyan, Barr, & Murphy, 2011). Often, children are spanked frequently and by more than one caregiver. In one study, among those mothers who reported any spanking, 10% said they spanked their 2-year old 20 times or more in the past year (Zolotor et al., 2011). Our findings add evidence to the literature from a Bayesian lens that this common parenting practice used by the majority of parents of young children, is associated with elevated levels of child behavior problems. Furthermore, answering our main research question, the degree to which spanking is associated with children's behavior problems does not differ between White and African American families, while this association is stronger for Hispanic families than White families.

A substantial body of literature on corporal punishment concludes that corporal punishment has harmful effects on child behavior (e.g., Gershoff & Grogan-Kaylor, 2016b). Despite this research base, one of the most contentious debates in this literature centers around the conditional corporal punishment perspective, which suggests that race and ethnicity are a social and cultural context that moderates the effects of corporal punishment on children. Prior research that tested this hypothesis using frequentist methods has not shown that corporal punishment has differential effects on children as a function of their race and ethnicity, in other words, suggesting a "universal" undesirable association of parents' use of corporal punishment with heightened child aggression.

However, these earlier findings are largely based on frequentist analyses that failed to reject the null hypothesis that the effects of corporal punishment are consistent across racial and ethnic groups (e.g., Gershoff et al., 2010; Gershoff & Grogan-Kaylor, 2016a, Lansford et al., 2005; Ma & Klein, 2018). An important caveat in this line of research using traditional frequentist approaches is that these studies are designed to detect differences between groups; yet cannot establish similarity between groups (2014, Kruschke, 2011). Thus, existing frequentist studies have not been able to directly conclude that the null hypothesis was likely to be true, rather such studies have only failed to detect differential effects of corporal punishment across racial and ethnic groups. Using Bayesian methods, our results extend current knowledge by directly testing and finding statistical support for our null hypothesis that corporal punishment affects Black and White children in the U.S. similarly, an interpretation that cannot be drawn directly from existing frequentist analyses. Our results also find some differences between groups. While the associations of mother's spanking with child externalizing behavior remained positive across all three groups, the slope of this association was steeper for families who identified their ethnicity as Hispanic. The varying level of normativeness of spanking across racial and ethnic groups is a possible explanation for the stronger relationship between spanking and externalizing behavior among Hispanic families. Prior studies suggest that, compared with White families, the use of corporal punishment is less normative in Hispanic culture (Altschul & Lee, 2011; Berlin et al., 2009; Ma & Klein, 2018) and that perceived normativeness of corporal punishment within a culture moderates the link between corporal punishment and child aggression (Gershoff et al., 2010). Indeed, a recent study using the FFCWS data shows spanking to be

least prevalent in Hispanic families followed by White and African American families (Ma & Klein, 2018). Congruent with the findings of Gershoff et al. (2010) our findings indicate that Hispanic children, who are raised in a culture in which corporal punishment is less commonly used to correct child misbehavior, are more strongly affected by mother's spanking than White children, who may be more likely to accept spanking as a normative parenting practice.

Such a statistical perspective dovetails nicely with an emerging conversation around caregiving for children. A large part of this conversation is rooted in various arms of the United Nations, such as the United Nations Study of Violence Against Children (Pinheiro, 2006), resolutions of the United Nations General Assembly (2007), and reports by UNICEF (2014). With regard to the issue of corporal punishment of children, this conversation recognizes that there may be different contextual challenges faced by different families around the world. As a result, parents in various cultural contexts may engage in different levels of physical punishment. However, despite varying levels of use of corporal punishment as a function of cultural context, as noted previously, research has shown little support for the hypothesis that its effects on children may have distinguishable consequences for children depending on cultural values and norms around corporal punishment (Gershoff & Grogan-Kaylor, 2016a, 2016b; Lansford et al., 2005). Indeed, the United Nations increasingly recognizes universality in the fact that physical punishment of children is associated with undesirable outcomes, such as increases in child aggression and mental health problems (Pinheiro, 2006). With an ability to recognize differences and also to affirm similarities, the use of Bayesian statistical methods to study parenting and child development might be seen to be more in line with an emergent global discourse around caring for children than are standard frequentist methods.

Relatedly, amid renewed recognition of the persistent inequities and discrimination of American life, there have been calls for a new focus on what is being called “diversity science” (Plaut, 2010). Broadly construed, diversity science is focused on empirical research that pays special attention to the disparities in risk and protective factors that lead to disparities in human well-being, with consequent further focus on developing policy and practice strategies to ameliorate those disparities. For example, one line of research in child development emphasizes the diversity in risks and inequities that are more pronounced in minority communities (García Coll et al., 1996). A long history of social stratification has exposed minority children in the U.S. to disproportionately higher rates of contextual risk factors including poverty, crime, and violence (Sampson, Morenoff, & Gannon-Rowley, 2002) and negative social mechanisms such as racism, discrimination, and stereotypes (García Coll et al., 1996). Thus, developmental science proposes the need to consider the diverse negative influences that minority children are exposed at disproportionately higher rates both conceptually and analytically in integrative models.

Logically, such a diversity science perspective would then require empirical research that can identify both differences and similarities in levels of risk and protective factors across social groups, and equally as importantly, differences and similarities across groups in the mechanisms and processes that connect those risk and protective factors to well-being outcomes (Plaut, 2010). Again, the Bayesian method's ability to equally identify both similarity and difference seems well-suited to the development of such a diversity science agenda, and could have significant policy and practice implications. For example, if practices and resources were incorporated into a community to alleviate scientific disparities found at the 0.05 level using frequentist approaches, and these disparities happened to be due to Type I error, valuable time and money may have been spent to alleviate erroneous disparities. Because Bayesian estimation does not inherently search for differences among groups, Bayesian estimation may help researchers find the most salient and impactful disparities, which can help practitioners concentrate their time, funds, and resources in the most influential areas. Along a similar vein, Bayesian estimation may also help researchers to discover universal protective factors that reduce economic or social disparities - allowing practitioners to target these protective factors among multiple populations instead of a very specific community.

4.1. Limitations

There are some limitations to this research. The FFCWS purposely oversampled unmarried parents, thus generalizability of the study findings must be considered in reference to the fact that a larger portion of FFCWS families were single-parent families that were also socially and economically disadvantaged. Analytically, this research examined the association of spanking and children's externalizing behavior, while controlling for a number of covariates. That being said, a direct causal connection between parental spanking and children's externalizing behavior cannot be made, since there may be other unmeasured covariates which explain this relationship. While these results certainly suggest an association between spanking and behavior problems, and one that controls for a number of covariates, direct causal connections cannot be made.

4.2. Implications for policy and practice

This study provides a new analytical or statistical lens to a long-standing debate, and contributes to a large body of research that shows that parental spanking is associated with undesirable increases in children's behavior problems. These associations are robust across a large sample, and notably, in this analysis evidence suggests that associations are largely consistent across the racial and ethnic groups studied. Thus, both policy and practice would do well to developing strategies to prevent parental use of corporal punishment across all racial and ethnic groups of children.

Notably, substantive effect sizes in this study are non-trivial if we divide the β regression coefficient for every additional spanking by the standard deviation of the outcome variable, externalizing behavior at age 5. We see that every additional level of spanking is associated with a 9% of a standard deviation increase in externalizing behavior ($.024 / .28 = .09$). In a follow up analysis, we removed the interaction term of spanking with race or ethnicity for Black families, as this credible interval included zero. The size of the regression coefficient for spanking in this follow up analysis increased to .031, representing an 11% of a standard deviation

increases in externalizing behavior (.031 / .28 = .11). In families where children are spanked a number of times a week, it is therefore not inconceivable that the increase in externalizing behavior associated with corporal punishment could be a third to a half of a standard deviation. The findings here are consonant with the mean effect size found by Gershoff and Grogan-Kaylor in their large meta-analytic review of 50 years of research on corporal punishment (2016a).

4.3. Implications for research

The non-trivial effect sizes of spanking upon children's behavior problems are relevant to research on children and families broadly, as such non-trivial effect sizes indicate that continued research on parental use of physical punishment will prove fruitful and salient in understanding children's behavior problems. These findings also add to more traditional frequentist approaches, suggesting that the effects of spanking are largely consistent across racial and ethnic groups. Lastly, this research suggests that variants of Bayesian regression strategies may prove fruitful in the study of children's behavior problems, and the study of child development more generally. Bayesian methods provide an alternative method of parameter estimation, but are also consistent with a theoretical perspective that attempts to find both similarities and differences between groups of families and children. Bayesian regression methods are becoming increasingly common, and are increasingly available in standard statistical software that is available to researchers (StataCorp, 2017a; StataCorp, 2017b).

5. Conclusion

Using Bayesian methods, this study provides strong evidence that parental use of physical punishment is associated with increases in externalizing behavior problems for children. These increases appear to be consistent between White and Black families whereas the relationship between spanking and externalizing behavior was stronger for Hispanic families than White families. Both advocates and child and family researchers should work to reduce the use of spanking by parents and suggest the use of non-physical parenting practices.

References

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4–18, YSR, and TRF profiles*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Achenbach, T. M. (1992). *Manual for the child behavior Checklist/2–3 and 1992 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Baumrind, D., Larzelere, R. E., & Cowan, P. A. (2002). Ordinary physical punishment: is it harmful? Comment on Gershoff. *Psychological Bulletin*, 128(4), 511–580.
- 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(5), 1403–1420. <https://doi.org/10.1111/j.1467-8624.2009.01341.x>.
- 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. *The American Psychologist*, 55(2), 218–232. <https://doi.org/10.1037/0003-066X.55.2.218>.
- 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. <https://doi.org/10.1207/s15327965pli0803>.
- Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1996). *Physical discipline among African American and European American mothers: Links to children's externalizing behaviors*. *Developmental Psychology*. US: American Psychological Association <https://doi.org/10.1037/0012-1649.32.6.1065>.
- Deater-Deckard, K., Ivy, L., & Petrill, S. A. (2006). Maternal warmth moderates the link between physical punishment and child externalizing problems: A parent-offspring behavior genetic analysis. *Parenting: Science and Practice*. https://doi.org/10.1207/s15327922par0601_3.
- Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*, 22(2), 240–261. <https://doi.org/10.1037/met000065>.
- Figueiredo Filho, D. B., Paranhos, R., da Rocha, E. C., Batista, M., da Silva, J. A., Jr, Santos, D., et al. (2013). When is statistical significance not significant? *Brazilian Political Science Review*, 7(1), 31–55. <https://doi.org/10.1590/S1981-38212013000100002>.
- García Coll, C., Lambert, G., Jenkins, R., McAdoo, H. P., Crnic, K., Wasik, B. H., et al. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development*, 67, 1891–1914. <https://doi.org/10.2307/1131600>.
- Gershoff, E. T., & Font, S. A. (2016). Corporal punishment in U.S. Public schools: Prevalence, disparities in use, and status in state and federal policy. *SRCD Social Policy Report*, 30(1), 1–25.
- Gershoff, E. T., & Grogan-Kaylor, A. (2016a). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology*, 30(4), 453–469. <https://doi.org/10.1037/fam0000191>.
- Gershoff, E. T., & Grogan-Kaylor, A. (2016b). Race as a moderator of associations between spanking and child outcomes. *Family Relations*, 65(3), 490–501. <https://doi.org/10.1111/fare.12205>.
- Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, 128(4), 539–579.
- Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., et al. (2010). Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness. *Child Development*, 81(2), 487–502.
- Gershoff, E. T., Sattler, K. M. P., & Ansari, A. (2018). Strengthening causal estimates for links between spanking and children's externalizing behavior problems. *Psychological Science*, 29(1), 110–120. <https://doi.org/10.1177/0956797617729816>.
- Grogan-Kaylor, A. (2004). The effect of corporal punishment on antisocial behavior in children. *Social Work Research*, 28(3), 153–162. <https://doi.org/10.1093/swr/28.3.153>.
- Hoeve, M., Dubas, J. S., Eichelsheim, V. I., Van Der Laan, P. H., Smeenk, W., & Gerris, J. R. M. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*. <https://doi.org/10.1007/s10802-009-9310-8>.
- Holden, G. W., Grogan-Kaylor, A., Durrant, J. E., & Gershoff, E. T. (2017). Researchers deserve a better critique: Response to Larzelere, Gunnoe, Roberts, and ferguson. *Marriage & Family Review*, 53(5), 465–490. <https://doi.org/10.1080/01494929.2017.1308899>.
- Ioannidis, J. A. (2018). The proposal to lower p value thresholds to .005. *JAMA*, 319(14), 1429–1430. <https://doi.org/10.1001/jama.2018.1536>.
- Kawabata, Y., Alink, L. R. A., Tseng, W. L., van Ijzendoorn, M. H., & Crick, N. R. (2011). Maternal and paternal parenting styles associated with relational aggression in children and adolescents: A conceptual analysis and meta-analytic review. *Developmental Review*, 31(4), 240–278. <https://doi.org/10.1016/j.dr.2011.08.001>.
- Kazdin, A. E., & Benjet, C. (2003). Spanking Children: Evidence and Issues. *Current Directions in Psychological Science*, 12(3), 99–103. <https://doi.org/10.1111/1467-8721.01239>.
- Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press.

- Kruschke, J. (2014). *Doing bayesian data analysis. ICPSR summer workshop* Ann Arbor, MI.
- Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4), 722–752. <https://doi.org/10.1177/1094428112457829>.
- Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., ... Quinn, N. (2005). Physical Discipline and Children's Adjustment: Cultural Normativeness as a Moderator. *Child Development*, 76(6), 1234–1246. <https://doi.org/10.1111/j.1467-8624.2005.00847.x>.
- 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(1), 24–35. <https://doi.org/10.1080/01494929.2016.1145613>.
- Larzelere, R. E., & Kuhn, B. R. (2005). Comparing child outcomes of physical punishment and alternative disciplinary tactics: A meta-analysis. *Clinical Child and Family Psychology Review*, 8(1), 1–37. <https://doi.org/10.1007/s10567-005-2340-z>.
- Lee, S. J., Altschul, I., & Gershoff, E. T. (2013). Does warmth moderate longitudinal associations between maternal spanking and child aggression in early childhood? *Developmental Psychology*, 49(11), 2017–2028. <https://doi.org/10.1037/a0031630>.
- Lee, S. J., Grogan-Kaylor, A., & Berger, L. M. (2014). Parental spanking of 1-year-old children and subsequent child protective services involvement. *Child Abuse and Neglect*, 38(5), <https://doi.org/10.1016/j.chiabu.2014.01.018>.
- Ma, J., & Grogan-Kaylor, A. (2017). Longitudinal associations of neighborhood collective efficacy and maternal corporal punishment with behavior problems in early childhood. *Developmental Psychology*, 53(6), 1027–1041. <https://doi.org/10.1037/dev0000308>.
- Ma, J., & Klein, S. (2018). Does Race/Ethnicity Moderate the Associations between Neighborhood and Parenting Processes on Early Behavior Problems? *Journal of Child and Family Studies*, 27(11), 3717–3729. <https://doi.org/10.1007/s10826-018-1200-7>.
- Ma, J., Grogan-Kaylor, A., & Lee, S. J. (2018). Associations of neighborhood disorganization and maternal spanking with children's aggression: A fixed-effects regression analysis. *Child Abuse & Neglect*, 76, 106–116. <https://doi.org/10.1016/j.chiabu.2017.10.013>.
- Maguire-Jack, K., Gromoske, A. N., & Berger, L. M. (2012). Spanking and child development during the first 5 years of life. *Child Development*, 83(6), 1960–1977. <https://doi.org/10.1111/j.1467-8624.2012.01820.x>.
- McLoyd, V. C., & Smith, J. (2002). Physical discipline and behavior problems in African American, European American, and Hispanic children: Emotional support as a moderator. *Journal of Marriage and Family. US: National Council on Family Relations*. <https://doi.org/10.1111/j.1741-3737.2002.00040.x>.
- Nuzzo, R. (2014). Statistical errors: P values, the “gold standard” of statistical validity, are not as reliable as many scientists assume. *Nature*, 506(7487), 150–152. <https://doi.org/10.1136/bmj.1.6053.66>.
- Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry*, 21(2), 77–99. <https://doi.org/10.1080/10478401003676501>.
- Pinheiro, P. S. (2006). *World report on violence against children* Geneva, Switzerland: United Nations Secretary-General's Study on Violence Against Children.
- Reichman, N. E., Teitler, J. O., Garfinkel, I., & McLanahan, S. S. (2001). Fragile Families: sample and design. *Children and Youth Services Review*, 23(4–5), 303–326. [https://doi.org/10.1016/S0190-7409\(01\)00141-4](https://doi.org/10.1016/S0190-7409(01)00141-4).
- Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “neighborhood effects”: Social processes and new directions in research. *Annual Review of Sociology*, 28, 443–478. <https://doi.org/10.1146/annurev.soc.28.110601.141114>.
- Smith, T. W., Davern, M., Freese, J., & Hout, M. (2017). *General social surveys, 1972-2016. [machine-readable data file]. Principal investigator, Tom W. Smith; Co-principal investigators, Peter V. Marsden and Michael Hout, NORC ed.* Chicago: NORC 1 data file (62,466 logical records) and 1 codebook (3,689 pp).
- Stacks, A., Oshio, T., Gerard, J., & Roe, J. (2009). *The Moderating Effect of Parental Warmth on the Association between Spanking and Child Aggression: A Longitudinal Approach, Vol. 18* <https://doi.org/10.1002/icd.596>.
- StataCorp (2017a). *Stata statistical software: Release 15*. College Station, TX: StataCorp LLC.
- StataCorp (2017b). *Stata 15 stata bayesian analysis reference manual*. College Station, TX: Stata Press.
- Sterne, J. A. C., Smith, G. D., & Cox, D. R. (2001). Sifting the evidence—what's wrong with significance tests? *BMJ*, 322(7280), 226. <https://doi.org/10.1136/bmj.322.7280.226>.
- Straus, M. A., & Donnelly, M. (2005). *Theoretical Approaches to Corporal Punishment. Corporal Punishment in Theoretical Perspective*. Yale University Press <https://doi.org/10.12987/yale/9780300085471.003.0001>.
- UNICEF (2014). *Hidden in plain sign: A statistical analysis of violence against children*. New York, NY.
- United Nations General Assembly (2007). *Rights of the Child, No. 62/141*. United Nations.
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. <https://doi.org/10.1111/cdev.12169>.
- Zolotor, A. J., Robinson, T. W., Runyan, D. K., Barr, R. G., & Murphy, R. A. (2011). The emergence of spanking among a representative sample of children under 2 years of age in north Carolina. *Frontiers in Psychiatry*, 2, 36. <https://doi.org/10.3389/fpsy.2011.00036>.