Journal of Family Psychology
2016, Vol. 30, No. 4, 453–469

0893-3200/16/$12.00 http://dx.doi.org/10.1037/fam0000191

Spanking and Child Outcomes: Old Controversies and New Meta-Analyses

Elizabeth T. Gershoff
University of Texas at Austin

Andrew Grogan-Kaylor
University of Michigan

Whether spanking is helpful or harmful to children continues to be the source of considerable debate among both researchers and the public. This article addresses 2 persistent issues, namely whether effect sizes for spanking are distinct from those for physical abuse, and whether effect sizes for spanking are robust to study design differences. Meta-analyses focused specifically on spanking were conducted on a total of 111 unique effect sizes representing 160,927 children. Thirteen of 17 mean effect sizes were significantly different from zero and all indicated a link between spanking and increased risk for detrimental child outcomes. Effect sizes did not substantially differ between spanking and physical abuse or by study design characteristics.

Keywords: spanking, physical punishment, discipline, meta-analysis

Around the world, most children (80%) are spanked or otherwise physically punished by their parents (UNICEF, 2014). The question of whether parents should spank their children to correct misbehaviors sits at a nexus of arguments from ethical, religious, and human rights perspectives both in the U.S. and around the world (Gershoff, 2013). Several hundred studies have been conducted on the associations between parents' use of spanking or physical punishment and children's behavioral, emotional, cognitive, and physical outcomes, making spanking one of the most studied aspects of parenting. What has been learned from these hundreds of studies? Several efforts have been made to synthesize this large body of research, first in narrative form (Becker, 1964; Larzelere, 1996; Steinmetz, 1979; Straus, 2001) and later through meta-analyses (Ferguson, 2013; Gershoff, 2002; Larzelere & Kuhn, 2005; Paolucci & Violato, 2004). Each of these four meta-analyses included a different set of articles and came to varied conclusions, namely that physical punishment is largely ineffective and harmful (Gershoff, 2002), that physical punishment is effective under certain circumstances (Larzelere & Kuhn, 2005), and that physical punishment is linked with children's cognitive, emotional, and behavioral problems but only modestly (Ferguson, 2013; Paolucci & Violato, 2004). These competing conclusions have left both social science researchers and the public at large confused about what outcomes can and cannot be attributed to spanking.

As this body of work on spanking and physical punishment has accumulated, several nagging questions about the quality, consistency, and generalizability of the research have persisted. Two primary concerns that have been raised about past meta-analyses are that spanking has been confounded with potentially abusive parenting behaviors in some studies and that spanking has only been linked with detrimental outcomes in methodologically weak studies (Baumrind, Larzelere, & Cowan, 2002; Ferguson, 2013; Larzelere & Kuhn, 2005). The goal of the current article is to address these two concerns with a new set of meta-analyses using the most recent research studies to date. Because the social science theories regarding why spanking might be linked with child outcomes have been summarized extensively elsewhere (Donnelly & Straus, 2005; Gershoff, 2002), we will not repeat them here and instead will focus in this paper on key questions about the research conducted to date.

The terms "corporal punishment," "physical punishment," and "spanking" are largely synonymous in American culture. The majority of the studies discussed in our literature review use the term physical punishment which we define as *noninjurious, open-handed hitting with the intention of modifying child behavior*. In our meta-analyses, however, we focused on the most common form of physical punishment which is known in the U.S. as spanking, and which we define as *hitting a child on their buttocks or extremities using an open hand*.

Previous Meta-Analyses of Physical Punishment and Spanking

The question of whether parents' use of spanking or physical punishment is linked with children's outcomes has been addressed in four published meta-analyses in the last 15 years. The first and most widely cited of the meta-analyses was by Gershoff (2002). This review included 88 studies used in separate meta-analyses of the associations between parents' use of physical punishment and 11 child outcomes, four of which were measured in adulthood. Physical punishment was defined as

This article was published Online First April 7, 2016.

Elizabeth T. Gershoff, Department of Human Development and Family Sciences, University of Texas at Austin; Andrew Grogan-Kaylor, School of Social Work, University of Michigan.

We thank our research assistants: Megan Gilster, Jacqueline Hoagland, and Julie Ma.

Correspondence concerning this article should be addressed to Elizabeth T. Gershoff, Department of Human Development and Family Sciences, University of Texas at Austin, 108 E. Dean Keeton St., Stop A2702, Austin, TX 78712. E-mail: liz.gershoff@austin.utexas.edu

"the use of physical force with the intention of causing a child to experience pain but not injury for the purposes of correction or control of the child's behavior" (per Straus, 2001, p. 4) and excluded any methods that would "knowingly cause severe injury to the child" (Gershoff, 2002, p. 543). All 11 meta-analyses were significant and all but one indicated an undesirable association. Specifically, physical punishment was associated with more immediate compliance ($d = 1.13$) but was also associated with lower levels of moral internalization ($d = -.33$), quality of the parent–child relationship ($d = -.58$), and mental health in childhood ($d = -.49$) and adulthood ($d = -.09$), as well as with higher levels of aggression in childhood ($d = .36$) and adulthood ($d = .57$), antisocial behavior in childhood ($d = .42$) and adulthood ($d = .42$), risk of being a victim of physical abuse ($d = .69$), and risk of abusing own child or spouse as an adult ($d = .13$).

The second meta-analytic article on the outcomes associated with physical punishment included 70 studies in three meta-analyses (Paolucci & Violato, 2004). Physical punishment was defined as "a form of nonabusive or customary physical punishment by a parent or adult serving as a parent" (Paolucci & Violato, 2004, p. 208). The outcomes were grouped into very broad and heterogeneous categories of negative outcomes: "affective outcomes" included mental health problems and low self-esteem; "cognitive outcomes" encompassed a wide range of outcomes including academic impairment, suicidality, and attitudes about spanking; and "behavioral outcomes" included disobedience, behavior problems, child abuse, spouse abuse, and hyperactivity. Higher scores on any of these outcome measures indicated negative outcomes. The weighted mean effect sizes were $d = 0.20$ for affective outcomes, $d = 0.06$ for cognitive outcomes, and $d = 0.21$ for behavioral outcomes, each of which was statistically significant. The conclusion afforded by these meta-analyses is that physical punishment was associated significantly, albeit modestly, with more affective, cognitive, and behavioral problems in children, broadly defined.

The third meta-analytic article (Larzelere & Kuhn, 2005) was distinct from the previous two in that each of the effect sizes was based on differences between an effect size for physical punishment and an effect size for another disciplinary method. Using 26 studies, separate meta-analyses were conducted by comparison group rather than by outcome type. Studies' measures of physical punishment were categorized into four types: conditional spanking ("physical punishment that was used primarily to back-up milder disciplinary tactics"), customary physical punishment ("typical parental usage"), overly severe physical punishment ("measures that gave extra points for severity of physical punishment"), and predominant use of physical punishment ("predominant disciplinary tactics . . . or proportional usage") (Larzelere & Kuhn, 2005, p. 17). When the main effects were examined, predominant and overly severe categories of physical punishment were found to be associated with more detrimental outcomes overall, ds $= -.21$ and $-.22$, respectively, whereas the customary and conditional categories of physical punishment were associated with small levels of beneficial outcomes, ds $= .06$ and $.05$, respectively. When these physical punishment categories were compared with other forms of discipline, conditional spanking was found to be associated with lower levels of noncompliance and antisocial behavior than disciplinary alternatives. Customary physical punishment was found to predict more detrimental outcomes when children's initial levels of child misbehavior were statistically controlled, $d = -.19$, but was generally not significantly different from other disciplinary tactics, including reasoning, taking away privileges, and time out, in the strength or direction of its associations with child outcomes. The severe and predominant categories of physical punishment were consistently associated with detrimental outcomes, such as less compliance, lower conscience, lower positive behavior, and higher antisocial behavior (Larzelere & Kuhn, 2005). The authors concluded that, in general, physical punishment was no worse than other disciplinary techniques. This is of course also to say that physical punishment was no better than other disciplinary techniques in promoting beneficial outcomes for children.

The fourth meta-analysis article by Ferguson (2013) focused solely on longitudinal studies and on the outcomes of externalizing behavior problems, internalizing behavior problems, and cognitive performance. The meta-analyses were conducted using 45 studies and calculated separate effect sizes for spanking and for corporal punishment, which was defined as "a wider range of more serious acts, including pushing, shoving, hitting with an object, or striking the face, yet generally falling short of physically injurious or life-threatening acts of violence" (Ferguson, 2013, p. 199). The bivariate effect sizes for spanking and corporal punishment (cp) were significantly different from zero across all three outcomes: externalizing, $d_{cp} = .18$ and $d_{spanking} = .14$; internalizing, $d_{cp} = .21$ and $d_{spanking} = .12$; and cognitive performance $d_{cp} = -.18$ and $d_{spanking} = -.09$. A secondary set of meta-analyses was conducted for studies that reported effect sizes controlling for children's previous behavior; there were not sufficient numbers of studies for all possible comparisons, but reported effect sizes for externalizing behavior problems were $d_{cp} = .08$ and $d_{spanking} = .07$, for internalizing was $d_{spanking} = .10$, and for cognitive performance was $d_{cp} = -.11$, all statistically significant at $p < .05$. The effect sizes for spanking were smaller than for corporal punishment, and the effect sizes for longitudinal associations controlling for the child's previous behavior were smaller than basic longitudinal associations, yet all were significantly different from zero and all indicated detrimental outcomes associated with spanking or corporal punishment.

Taken together, these meta-analyses provide evidence that physical punishment is associated with negative child outcomes, particularly when the outcomes are divided into finer-grained categories (Ferguson, 2013; Gershoff, 2002) rather than when they are grouped into broad categories (Paolucci & Violato, 2004), and that harsher methods of physical punishment are more strongly associated with negative child outcomes than ordinary spanking (Ferguson, 2013; Larzelere & Kuhn, 2005).

Remaining Concerns About the Research on Spanking and Child Outcomes

The meta-analyses in the present study were conducted in order to address two persistent questions about the research to date in order to clarify what is known about the potential impacts of parents' use of physical punishment on children.

Spanking Has Been Confounded With Harsher Forms of Physical Punishment

The main criticism of the Gershoff (2002) meta-analysis has been that it included harsh and potentially injurious behaviors, such as hitting children with objects, in its definition of physical punishment (Baumrind et al., 2002; Benjet & Kazdin, 2003; although note that this criticism applies to the Paolucci & Violato, 2004 meta-analysis as well). This broad definition of physical punishment included parent behaviors that most professionals and most parents would agree were abusive and that may be linked with negative outcomes while spanking is not (Kazdin & Benjet, 2003). Baumrind, Larzelere, and Cowan (2002) reanalyzed the data from Gershoff (2002), separating out what they deemed harsh or potentially abusive forms of physical punishment. They reported that the effect size for the studies using less severe physical punishment was significantly smaller than the effect size for harsh physical punishment ($d_{less\ severe} = .30$ vs. $d_{more\ severe} = .46$, $\chi^2[1, n = 12{,}244] = 74.50$, $p < .001$). They concluded that only severe methods of physical punishment are harmful. However, both effect sizes are significant and positive, indicating that both are associated with more undesirable child outcomes.

To help resolve this debate, our first research question was thus, are past findings that physical punishment is associated with detrimental child outcomes driven by the inclusion of harsh or abusive methods, or is spanking on its own associated with these detrimental outcomes? We addressed this question using two strategies. First, we focused on "studies of parents' behaviors labeled as "spanking" (see definition above) or as" synonymous terms for the same behavior (e.g., "smacking," "slapping," and "hitting"). This definition therefore excluded the use of objects, the use of methods that have a reasonable expectation of causing harm or injury (e.g., beating, burning, choking, whipping), and the use of methods that are gratuitous expressions of parent displeasure without a clear disciplinary component (e.g., pulling hair, shaking, shoving). By restricting our operationalization of physical punishment in this way, we were able to determine the extent to which ordinary spanking is linked with child outcomes.

Our second strategy was to examine the ways in which the strength and direction of the associations between spanking and child outcomes compare with the strength and direction of the associations between clearly abusive methods and child outcomes. We identified studies that assessed the same individuals for exposure to both ordinary spanking and to harsher methods in order to isolate the associations of one from the other. A comparison of studies of spanking to studies of abuse would not be helpful in this regard, because there could be many selection factors that distinguish the individuals reporting spanking from those reporting harsher methods. Some have argued that parents who use harsh or abusive methods are fundamentally different from parents who use only spanking (Baumrind et al., 2002) while some past research has found that genetic factors in the child elicit corporal punishment but not physical abuse (Jaffee et al., 2004). By focusing on studies that assessed the extent to which individuals experience both spanking and abuse, we compared the unique association of spanking with child outcomes to the unique association of abusive behaviors with child outcomes for the same samples of children.

Spanking Has Only Been Linked With Negative Child Outcomes in Cross-Sectional or Methodologically Weak Studies

The primary standard for determining causal relations among variables has been the randomized controlled experiment because potentially confounding selection factors that might distinguish naturally occurring groups (e.g., spankers and nonspankers) are eliminated through randomization (Shadish, Cook, & Campbell, 2001). However, parents' use of spanking is not easily or ethically studied through an experimental design, as children cannot be randomly assigned to parents with varying predispositions to spank, nor can parents typically be randomly assigned to spank or not spank. There are a small handful of experimental studies that examine whether children comply more in a laboratory setting when mothers use spanking (Bean & Roberts, 1981; Day & Roberts, 1983; Roberts, 1988; Roberts & Powers, 1990); we include these studies in the meta-analyses and discuss them more below. There also have been a few efforts to evaluate the effects of interventions designed to reduce spanking (e.g., Beauchaine, Webster-Stratton, & Reid, 2005), but these studies require a sample of parents who are willing to not spank and thus may be fundamentally different from most spankers in the population. The circumstances of experimentally manipulated spanking thus are likely to be unusual, leading to concern that experiments with parental spanking may suffer from a lack of external validity.

The next strongest approach to studying spanking are studies which examine whether it predicts changes in child outcomes over time. Such prospective longitudinal designs meet one of the key criteria for establishing causality, namely temporal precedence of the spanking independent variable (Shadish et al., 2001). Longitudinal effect sizes of the bivariate links between spanking and later child outcomes do not rule out the potential for a child elicitation effect; however, so few studies report a coefficient that controls only for initial child behavior (and not for a range of other covariates) that we are unable to meta-analyze them. Thus, while not a perfect solution, longitudinal bivariate coefficients are decidedly stronger methodologically than within-time coefficients.

Our second research question was thus: Are associations between spanking and child outcomes only found in methodologically weak studies? In order to address this question, we conducted moderator analyses that examined whether the direction and significance of the mean weighted effect sizes were similar across longitudinal, experimental, and cross-sectional studies. We also examined whether effect sizes varied according to five other dimensions of study design: measure of spanking, time period in which spanking was administered, index of spanking, whether the study assessed the associations of spanking with outcomes within a single group, or employed comparisons between two or more groups, and independence of raters of spanking and outcome. Using these dimensions of study quality as moderators allowed us to examine whether spanking is only associated with child outcomes in some types of studies and not others, a finding which would undermine the generalizability of spanking research.

The Present Study

Given the pervasive use of spanking around the world, and in light of concerns raised about spanking by professional organiza-

tions (American Academy of Pediatrics, 2012) and intergovernmental and human rights organizations (Committee on the Rights of the Child, 2006), there is a need for definitive conclusions about the potential consequences of spanking for children. The purpose of the current study was to conduct a new set of meta-analyses to address the two unresolved debates described above and to do so while incorporating an additional 13 years of literature since the first meta-analysis was published (Gershoff, 2002). The present study is distinguished from the previous meta-analyses by focusing exclusively on parents' use of spanking, by including only peer-reviewed journal articles, by using random effects meta-analyses, and by incorporating several dozen new studies not included in previous meta-analyses.

Method

Identification of Potential Studies for Inclusion

The studies for the present meta-analyses were identified from two main sources. The primary source for studies was a comprehensive literature review of articles listed in four academic abstracting databases (ERIC, Medline, PsycInfo, and Sociological Abstracts) that had been published before June 1, 2014. Each database was searched using six terms for physical punishment, namely "spank*," "corporal punishment," "physical punishment," "physical discipline," "harsh punishment," and "harsh discipline." In addition, all of the studies used in the previously published meta-analyses (Ferguson, 2013; Gershoff, 2002; Larzelere & Kuhn, 2005; Paolucci & Violato, 2004) were considered for inclusion. These two methods yielded a total of 1,574 unique articles to be considered for inclusion in the current meta-analyses.

Coding of Studies for Inclusion or Exclusion

Coding of studies involved a two-step process. In the initial step, the titles, abstracts, or full text of the 1,574 studies identified through the sources above were subjected to an initial screening. Studies were excluded at this stage if they were not relevant to or usable in the meta-analyses; examples of studies excluded at this stage were literature reviews, studies of beliefs about rather than use of spanking, and studies that were not available in English. This initial screening process eliminated 1,016 studies and retained 558 potential studies.

In the second step of coding, each of these 558 potential studies was coded independently by each of the authors. Any disagreements in coding were resolved through follow-up discussion. Studies were coded as to whether they met several criteria: (a) the study was published in a peer-reviewed journal; all book chapters, unpublished dissertations, and unpublished conference papers were excluded, even if they had been included in any of the previously published meta-analyses; (b) the study included a measure of parents' use of customary, noninjurious spanking (or slapping or hitting) that was intended to be a correction of a child's misbehavior. The terms "spank" or "smack" were used alone or in combination with other general terms (e.g., slap) in 63% of studies. The remaining studies measured corporal punishment as "physical punishment" or "physical discipline" (19%), "corporal punishment" (10%), and "slap or hit" (8%); (c) the study reported a bivariate association between parents' spanking and the child outcome of interest; and (d) The study included appropriate statistics for calculating effect sizes. The reasons for exclusion of all 1,499 studies are listed the Appendix. Only 75 studies met all four criteria and were retained for the meta-analyses.

Inclusion and Exclusion of Studies From Past Meta-Analyses

All of the 162 unique studies used in the four previously published meta-analyses were considered for inclusion, but only 36 met all of our criteria. Of the 88 studies in Gershoff (2002), 23 were included in the present study. Paolucci and Violato (2004) analyzed 70 studies; 16 were included here. Of the 26 studies in Larzelere and Kuhn (2005), 11 were included. Ferguson (2013) analyzed 45 studies; of these, 11 were included in the current meta-analyses. Reasons for study exclusion are available from the first author. Thus, 39 of the 75 studies included in the current meta-analyses (52%) have not been included in previous meta-analyses.

Coding of Effect Sizes

All study-level effect sizes were calculated independently by each of the authors; for all effect sizes, agreement was achieved to at least the third decimal place. When discrepancies occurred in effect size calculations, the discrepancy was discussed, and then each author independently recalculated the effect size. This process was repeated, if necessary, until consensus was achieved. Study-level effect sizes were transformed into standardized mean difference effect sizes to allow combination across effect sizes using Cohen's formula for d (Cohen, 1988; Sterne, 2009)

$$Cohen's\ d = \frac{mean_{treatment} - mean_{comparison}}{sd_{pooled}}$$

where sd_{pooled} was calculated as

$$sd_{pooled} = \sqrt{\frac{((n_1 - 1) * sd_1^2) + ((n_2 - 1) * sd_2^2)}{n_1 + n_2 - 2}}$$

Calculation of Cohen's d was straightforward when an article reported the sample size, mean and standard deviation of a group exposed to spanking and one that had never been spanked. For articles that did not report effects as group comparisons, we utilized formulas found in Borenstein, Hedges, Higgins, and Rothstein (2009) and Johnson (1993) to convert quantitative measures of association such as correlations and differences of proportions to Cohen's d effect sizes. For each study, we also calculated the standard error of the estimate of Cohen's d utilizing formulas given in Sterne (2009).

Selection or Aggregation of Single Effect Sizes From Studies

Because meta-analyses are focused on simple effects, only bivariate comparisons or correlations can be used (Borenstein et al., 2009); thus, bivariate associations such as standardized differences of means or correlations were selected over adjusted coefficients from multivariate models. When both longitudinal and cross-sectional results were available, the appropriate longitudinal effect sizes were use in the meta-analyses in order to obtain the most

methodologically robust effect size. If a study reported multiple effect sizes for the same outcome, such as when bivariate associations were reported for subgroups but not the whole sample, the weighted average of these subgroup effect sizes was used as the effect size for that study for that outcome. We allowed studies that reported effect sizes for more than one of our target outcomes to contribute to each appropriate meta-analyses; however, each study (or dataset, in the case of multiple articles from one dataset) was permitted to contribute only one effect size to each analysis for a specific outcome, so that a single individual was only counted once in any given meta-analysis for a specific outcome.

Coding of Study-Level Moderators

Seven study characteristics were coded for each study to be used in moderator analyses: (a) study design (experimental, longitudinal, cross-sectional, or retrospective); (b) measure of spanking (observation, parent report, child report, child retrospective, or both parent and child reports); (c) index of spanking (when used [either observed or in an experiment], frequency, frequency and severity, ever in time period, or ever in life); (d) independence of the raters of spanking and the child outcome (same rater or different raters); (e) time period in which spanking was administered (observed, last week, last month, last year, ever, hypothetical, specific time period, or not specified); (f) the country in which the study was conducted (U.S. or other than U.S.); and (g) the age range of children at the time of spanking (less than 2-years-old, 2- to 5-years-old, 6- to 10-years-old, and 11- to 15-years-old). The authors independently coded these characteristics for each study. Any discrepancies were resolved through discussion.

Meta-Analytic Procedure

Once all study effect sizes had been converted to the metric of Cohen's *d*, effect sizes were combined in a meta-analysis. Each study was entered into the model, weighted by its precision ($1/se_d$), and combined into a weighted average of effect sizes for the respective outcome domain. The meta-analyses reported in this paper utilized the random effects model (Borenstein et al., 2009; DerSimonian & Laird, 1986) using the Stata command *metan* (Bradburn, Deeks, & Altman, 2009). The random effects model for meta-analysis does not assume that there is a single underlying effect size of the studies being analyzed and rather allows effect sizes to differ across studies to account for the fact that study samples differ by characteristics such as age, gender, race, ethnicity and nationality. The random effects model calculates the mean effect size, an estimate of statistical significance, and a measure of the heterogeneity of effect sizes in terms of their variation around the estimated mean effect size. We conducted a separate meta-analysis for each child outcome as well as an overall meta-analysis for all of the studies together.

Results

Main Meta-Analyses

A total of 111 unique effect sizes were derived from data representing 204,410 child measurement occasions; these studies included data from a total of 160,927 unique children. The study-level effect sizes, confidence intervals, and sample sizes are listed in Table 1. For between-subjects designs, the subsample sizes for the subgroup that were spanked and the subgroup that was not spanked are presented, whereas for within-subjects designs a single sample size is presented. As a means of graphically representing the effect sizes, this table also includes bar graphs of the effect sizes and their corresponding confidence intervals both for the individual studies and for the random effects mean effect size for each outcome category. For the purposes of comparison and aggregation across meta-analyses, all of the study-level effect sizes were coded so that larger positive values corresponded to more detrimental child outcomes. This meant that for studies in which the outcome variable was a beneficial outcome (e.g., conscience), the effect sizes were recoded so that higher values reflected adverse outcomes rather than beneficial outcomes (e.g., low conscience).

As the effect sizes and bar graphs in Table 1 indicate, the findings across studies were highly consistent. Of the 111 individual effect sizes, 102 were in the direction of a detrimental outcome with 78 of these statistically significant. In contrast, nine of the effect sizes were in the direction of a beneficial outcome but only one (Tennant, Detels, & Clark, 1975) was statistically different from zero. Thus, among the 79 statistically significant effect sizes, 99% indicated an association between spanking and a detrimental child outcome.

Table 2 summarizes the mean weighted effect sizes and confidence intervals for each outcome along with a *Z* test for significant difference from zero and an I^2 statistic that estimates the amount of variation in the mean weight effect size that was attributable to underlying study heterogeneity. Spanking was significantly associated with 13 of the 17 outcomes examined. In each case, spanking was associated with a greater likelihood of detrimental child outcomes. In childhood, parental use of spanking was associated with low moral internalization, aggression, antisocial behavior, externalizing behavior problems, internalizing behavior problems, mental health problems, negative parent–child relationships, impaired cognitive ability, low self-esteem, and risk of physical abuse from parents. In adulthood, prior experiences of parental use of spanking were significantly associated with adult antisocial behavior, adult mental health problems, and with positive attitudes about spanking. The remaining four meta-analyses were not significantly different from zero. The 13 statistically significant mean effect sizes ranged in size from .15 to .64. The overall mean weighted effect size across all of the 111 study-level effect sizes was $d = .33$, with a 95% confidence interval of .29 to .38; this mean effect was statistically different from zero, $Z = 14.84$, $p < .001$.

Moderator Analyses Comparing Spanking With Physical Abuse

To address the concern that the findings of negative outcomes associated with spanking in past research were a result of the confounding of spanking with overly harsh or potentially abusive methods, we identified seven studies that reported bivariate associations for both spanking and physical abuse. The latter was defined variously as "hitting with fist or object, beating up, kicking, or biting" (Bugental, Martorell, & Barraza, 2003), "beaten to

Table 1
Study-Level Effect Sizes for Spanking by Child Outcome

Individual studies by outcome	Spank *n*	No spank *n*	*d*	95% Confidence interval	
Immediate defiance	120	30	.14	−.19	.47
Bean and Roberts (1981)	8	8	−.74	−1.76	.28
Day and Roberts (1983)	4	4	.36	−1.04	1.77
Minton, Kagan, and Levine (1971)	90		.34	−.09	.76
Roberts (1988)	9	9	−.08	−1.01	.84
Roberts and Powers (1990)	9	9	.10	−.82	1.03
Low moral internalization	745	84	.38	.11	.65
Burton, Maccoby, and Allinsmith (1961)	77		.63	.16	1.10
Grinder (1962)	73	66	.19	−.14	.53
Kandel (1990)	222		.47	.20	.74
Olson, Ceballo, and Park (2002)	50		.14	−.42	.70
Oyserman et al. (2005)	164		−.18	−.49	.13
Power and Chapieski (1986)	7	11	1.18	.15	2.22
Regev, Gueron-Sela, and Atzaba-Poria (2012)	145		.70	.35	1.05
Zahn-Waxler, Radke-Yarrow, and King (1979)	7	7	.63	−.45	1.71
Child aggression	4,534	1,069	.37	.13	.61
Berlin et al. (2009)	2,573		.14	.06	.22
Gershoff et al. (2010)	292		.24	.01	.47
Gunnoe and Mariner (1997)	1,112		.30	.18	.42
Kandel (1990)	222		.84	.55	1.12
Pagani et al. (2004)	106	1,069	.90	.70	1.10
Sears (1961)	160		−.14	−.45	.17
Westbrook et al. (2013)	69		.28	−.20	.76
Child antisocial behavior	5,725	1,412	.39	.24	.53
Boutwell, Franklin, Barnes, and Beaver (2011)	1,600		.52	.42	.62
Flynn (1999)	108	153	.29	.04	.54
Gunnoe and Mariner (1997)	1,112		.39	.27	.51
Jackson, Preston, and Franke (2010)	89		.72	.28	1.17
Kahn and Fua (1995)	25	51	.90	.40	1.39
Kohrt et al. (2004)	99		.62	.21	1.03
Oyserman et al. (2005)	164		.00	−.31	.31
Slade and Wissow (2004)	758	1,208	.12	.03	.21
Straus, Sugarman, and Giles-Sims (1997)	1,770		.41	.31	.50
Child externalizing behavior problems	25,988	1,086	.41	.32	.50
Bakoula et al. (2009)	225	1,086	.49	.34	.63
Barnes, Boutwell, Beaver, and Gibson (2013)	1,650		.39	.29	.49
Choe, Olson, and Sameroff (2013)	241		.36	.10	.61
Eisenberg, Chang, Ma, and Huang (2009)	615		.20	.04	.36
Gershoff et al. (2012)	11,044		.30	.27	.34
Hesketh et al. (2011)	2,200		.20	.12	.29

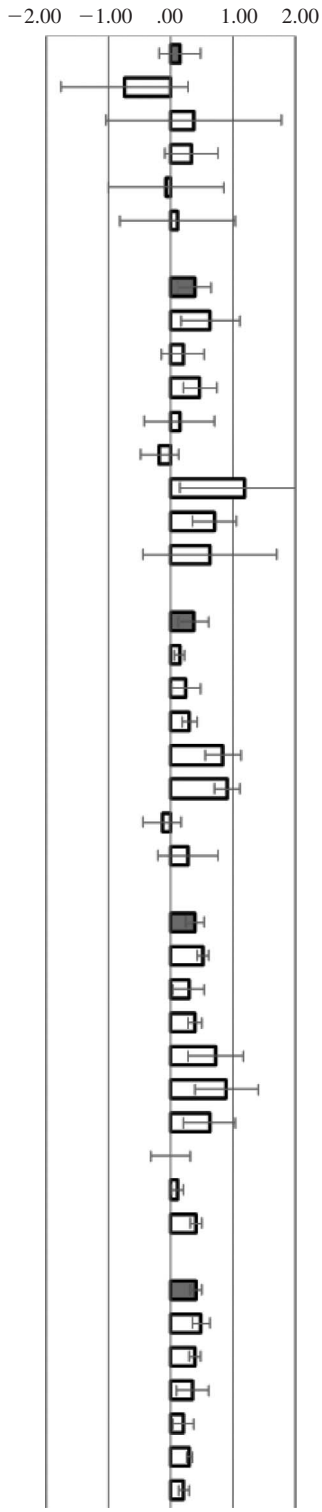
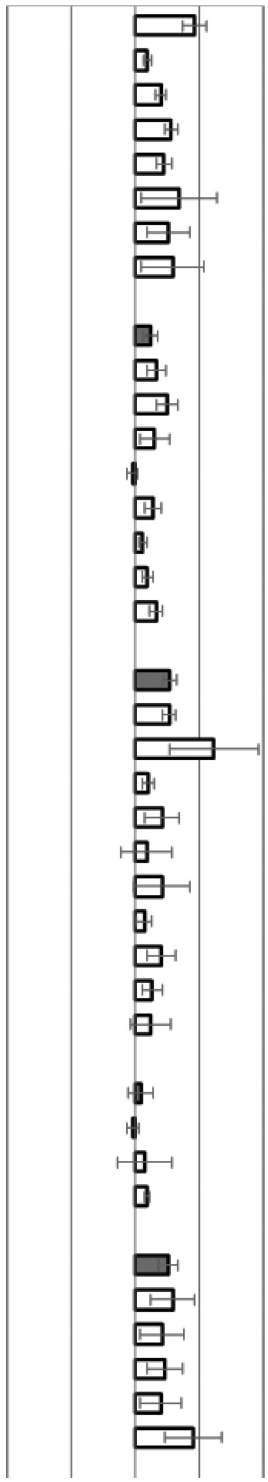


Table 1 (*continued*)

Individual studies by outcome	Spank *n*	No spank *n*	*d*	95% Confidence interval		Beneficial outcomes / Detrimental outcomes
						−2.00 −1.00 .00 1.00 2.00
Lansford et al. (2012)	585		.93	.75	1.10	
Maguire-Jack, Gromoske, and Berger (2012)	3,870		.19	.13	.25	
McKee et al. (2007)	2,582		.40	.32	.48	
McLeod and Shanahan (1993)	1,733		.56	.46	.66	
Mulvaney and Mebert (2007)	979		.45	.32	.58	
Olson, Ceballo, and Park (2002)	50		.68	.08	1.27	
Regev, Gueron-Sela, and Atzaba-Poria (2012)	145		.52	.18	.86	
Westbrook et al. (2013)	69		.58	.09	1.08	
Child internalizing behavior problems	12,413	3,486	.24	.13	.35	
Bakoula et al. (2009)	225	1,086	.34	.19	.48	
Eisenberg, Chang, Ma, and Huang (2009)	587		.49	.33	.66	
Gershoff et al. (2010)	292		.30	.07	.53	
Hesketh et al. (2011)	2,200		−.04	−.12	.04	
Lau et al. (2010)	924	2,400	.28	.15	.41	
Maguire-Jack, Gromoske, and Berger (2012)	3,870		.11	.05	.18	
McKee et al. (2007)	2,582		.19	.11	.27	
McLeod and Shanahan (1993)	1,733		.32	.23	.42	
Child mental health problems	5,122	1,313	.53	.42	.64	
Buehler and Gerard (2002)	1,401		.53	.42	.64	
Bugental, Martorell, and Barraza (2003)	44		1.23	.53	1.93	
Christie-Mizell, Pryor, and Grossman (2008)	1,852		.20	.11	.30	
Kandel (1990)	222		.42	.15	.69	
Kohrt et al. (2004)	99		.18	−.22	.58	
Lau et al. (2003)	22	469	.42	−.01	.85	
Li et al. (2001)	378	844	.14	.02	.26	
Lynam et al. (2009)	338		.41	.19	.63	
McLoyd, Kaplan, Hardaway, and Wood (2007)	606		.26	.10	.42	
Sears (1961)	160		.23	−.08	.55	
Child alcohol or substance abuse	6,621	90,359	.09	−.11	.29	
Alati et al. (2010)	2,784	645	−.04	−.12	.05	
Lau et al. (2003)	22	469	.15	−.28	.58	
Lau et al. (2005)	3,815	89,245	.19	.16	.22	
Negative parent–child relationship	755	0	.51	.36	.66	
Coyl, Roggman, and Newland (2002)	148		.58	.25	.92	
Joubert (1991)	134		.42	.07	.76	
Kandel (1990)	222		.46	.19	.73	
Larzelere, Klein, Schumm, and Alibrando (1989)	157		.40	.08	.72	
Palmer and Hollin (2001)	94		.90	.45	1.34	



(*table continues*)

Table 1 (*continued*)

Individual studies by outcome	Spank *n*	No spank *n*	*d*	95% Confidence interval	
Impaired cognitive ability	8,358	11	.17	.01	.32
Berlin et al. (2009)	2,573		.16	.08	.24
Gest, Freeman, Domitrovich, and Welsh (2004)	76		.17	−.28	.62
Lynam et al. (2009)	338		.14	−.07	.35
Maguire-Jack, Gromoske, and Berger (2012)	3,870		.00	−.06	.07
Oyserman et al. (2005)	164		−.18	−.49	.13
Parkinson, Wallis, Prince, and Harvey (1982)	20		1.22	.16	2.27
Power and Chapieski (1986)	7	11	1.71	.59	2.83
Straus and Paschall (2009)	1,310		.34	.23	.44
Low self-esteem	766	990	.15	.04	.26
Joubert (1991)	134		.12	−.22	.46
Lau et al. (2003)	22	469	.00	−.43	.43
Talillieu and Brownridge (2013)	610	521	.17	.05	.28
Low self-regulation	2,525	0	.30	−.07	.67
Boutwell, Franklin, Barnes, and Beaver (2011)	1,600		.61	.50	.71
Eisenberg, Chang, Ma, and Huang (2009)	587		.06	−.10	.22
Lynam et al. (2009)	338		.22	.01	.44
Victim of physical abuse	3,334	996	.64	.39	1.74
Bugental, Martorell, and Barraza (2003)	44		1.06	.39	1.74
Foshee et al. (2005)	1,146		.49	.38	.61
Frias-Armenta (2002)	102	48	.44	.09	.78
Gagné et al. (2007)	731		1.35	1.18	1.53
Hemenway, Solnick, and Carter (1994)	633	127	.25	.06	.44
Herzberger, Potts, and Dillon (1981)	24		1.00	.08	1.91
Trickett and Kuczynski (1986)	8	32	.31	−.46	1.09
Zolotor et al. (2008)	646	789	.38	.28	.49
Adult antisocial behavior	985	4,206	.36	.06	.65
Fergusson, Boden, and Horwood (2008)	341	2,504	.45	.33	.56
Lynch et al. (2006)	576	1,640	.10	.00	.19
McCord (1991)	68	62	.60	.25	.96
Adult mental health problems	1,855	4,707	.24	.09	.40
Fergusson, Boden, and Horwood (2008)	341	2,504	.21	.09	.32
Joubert (1992)	169		−.03	−.33	.27
Lynch et al. (2006)	576	1,640	.06	−.04	.15
Medina et al. (2001)	46		1.09	.43	1.76
Miller-Perrin, Perrin, and Kocur (2009)	41		.04	−.58	.66
Nettelbladt, Svenson, and Serin (1996)	27	42	.64	.15	1.14
Schweitzer, Zafar, Pavlicova, and Fallon (2011)	45		1.12	.44	1.80
Talillieu and Brownridge (2013)	610	521	.19	.08	.31

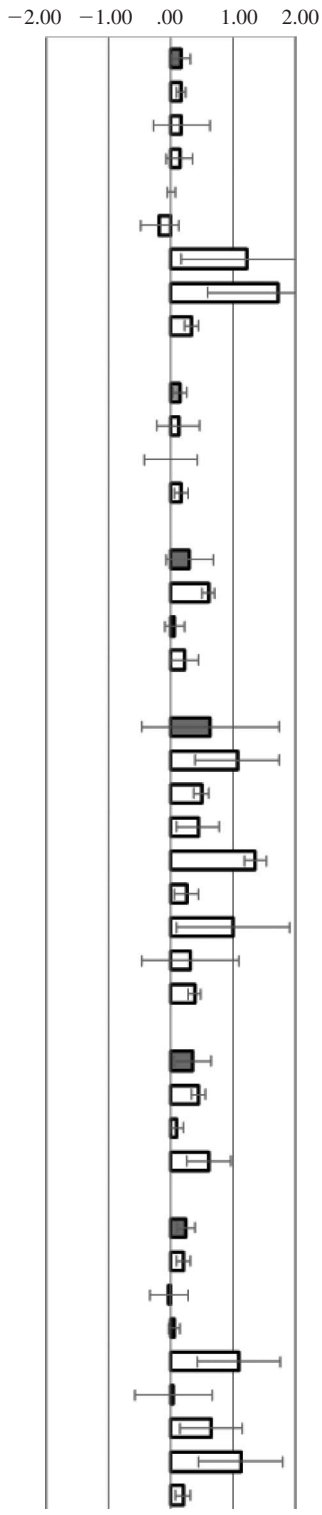


Table 1 (*continued*)

Individual studies by outcome	Spank *n*	No spank *n*	*d*	95% Confidence interval		Beneficial outcomes / Detrimental outcomes
						−2.00 −1.00 .00 1.00 2.00
Adult alcohol or substance abuse	2,596	4,796	.13	−.08	.35	
Baer and Corrado (1974)	93	107	.41	.13	.69	
Fergusson, Boden, and Horwood (2008)	341	2,504	.29	.18	.40	
Lynch et al. (2006)	576	1,640	.05	−.05	.14	
Tennant, Detels, and Clark (1975)	1,586	545	−.13	−.23	−.04	
Adult support for physical punishment	1,016	177	.38	.15	.61	
Deb and Adak (2006)	34	66	.88	.45	1.31	
Durrant (1993)	463	63	.39	.13	.66	
Graziano et al. (1992) U.S. sample	95		.45	.04	.87	
Graziano et al. (1992) India sample	160		.20	−.11	.51	
Hemenway, Solnick, and Carter (1994)	264	48	.12	−.18	.43	
Overall	89,638	114,772	.33	.29	.38	

Note. "Spank *n*" refers to the subsample in between-subjects designs that reported spanking, or to the entire sample in within-subjects designs. "No spank *n*" refers to the subsample in between-subjects designs that did not report spanking.

injury" (Lau, Chan, Lam, Choi, & Lai, 2003), "been injured from a beating" (Lau et al., 2005), "frequent or severe physical punishment" (Fergusson, Boden, & Horwood, 2008), "use of a weapon, punching, or kicking" (Lynch et al., 2006), "severe physical assault" (Miller-Perrin, Perrin, & Kocur, 2009), and "physical abuse leading to bruising" (Schweitzer, Zavar, Pavlicova, & Fallon, 2011). Each of these studies employed a within-subjects design; in each case, the same respondent (either a parent or the adult child recalling the behavior) reported both how often the parent used spanking and, in a separate question, how often the parent used abusive methods of discipline. Two of the studies contributed more than one effect size, yielding a total of 10 pairs of effect sizes for spanking and physical abuse. The effect sizes are presented in Table 3. In three cases, the effect size for spanking was larger than that for physical abuse. The weighted mean effect size for spanking was $d = .25$, while for physical abuse it was $d = .38$. Both were significantly different from zero and both were positive in sign, indicating that both spanking and physical abuse were associated with greater levels of detrimental child outcomes. The magnitude of the mean effect

Table 2
Summary of Spanking Meta-Analyses by Outcome

Detrimental child outcome	*K*	Spank *n*	No Spank *n*	*d*	95% CI		*Z*	I^2
Immediate defiance	5	120	30	.14	−.19	.47	.85	.80%
Low moral internalization	8	745	84	**.38**	.11	.65	2.76***	67.40%
Child aggression	7	4,534	1,069	**.37**	.13	.61	3.07***	91.40%
Child antisocial behavior	9	5,725	1,086	**.39**	.24	.53	5.28***	84.50%
Child externalizing behavior problems	14	25,988	1,086	**.41**	.32	.50	9.19***	88.40%
Child internalizing behavior problems	8	12,413	3,486	**.24**	.13	.35	4.36***	88.50%
Child mental health problems	10	5,122	1,313	**.53**	.42	.64	5.17***	76.00%
Child alcohol or substance abuse	3	6,621	90,359	.09	−.11	.29	.90	91.30%
Negative parent–child relationship	5	755	0	**.51**	.36	.66	6.76***	.00%
Impaired cognitive ability	8	8,358	11	**.17**	.01	.32	2.13*	84.30%
Low self-esteem	3	766	990	**.15**	.04	.26	2.76**	.00%
Low self-regulation	3	2,525	0	.30	−.07	.67	1.58	94.30%
Victim of physical abuse	8	3,334	996	**.64**	.39	1.74	4.07***	93.30%
Adult antisocial behavior	3	985	4,206	**.36**	.06	.65	2.35**	92.00%
Adult mental health problems	8	1,855	4,707	**.24**	.09	.40	3.05**	73.20%
Adult alcohol or substance abuse	4	2,596	4,796	.13	−.08	.35	1.21	91.90%
Adult support for physical punishment	5	1,016	177	**.38**	.15	.61	3.28***	55.50%
Overall effect size	111	89,638	114,722	**.33**	.29	.38	14.84***	88.80%

Note. *K* = number of effect sizes in the meta-analysis; *d* = mean weighted effect size; *Z* = significance test that d differs from zero; *I2* = the variation in the mean effect size attributable to heterogeneity. Bolded effect sizes are significantly different from zero.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
Effect Sizes for Studies That Reported Effect Sizes Separately for Spanking and Physical Abuse

Study	Outcome	Predictor	*d*	95% Confidence interval		Beneficial outcomes / Detrimental outcomes
						−1 0 1 2
Lau et al. (2003)	child externalizing behavior problems	spanking	.15	−.30	.60	
		physical abuse	.65	.19	1.10	
Lau et al. (2005)	child externalizing behavior problems	spanking	.19	.16	.22	
		physical abuse	.33	.30	.37	
Bugental, Martorell, and Barraza (2003)	child mental health problems	spanking	1.23	.77	1.69	
		physical abuse	.40	−.21	1.01	
Lau et al. (2003)	child mental health problems	spanking	.42	−.03	.87	
		physical abuse	.62	.17	1.07	
Lau et al. (2003)	child low self-esteem	spanking	.00	−.45	.45	
		physical abuse	.37	−.08	.82	
Fergusson et al. (2008)	adult antisocial behavior	spanking	.45	.33	.57	
		physical abuse	.25	.13	.37	
Lynch et al. (2006)	adult antisocial behavior	spanking	.10	.00	.20	
		physical abuse	.51	.41	.62	
Fergusson et al. (2008)	adult mental health problems	spanking	.21	.09	.33	
		physical abuse	.55	.43	.66	
Miller-Perrin, Perrin, and Kocur (2009)	adult mental health problems	spanking	.04	−.55	.63	
		physical abuse	.58	−.01	1.17	
Schweitzer, Zafar, Pavlicova, and Fallon (2011)	adult mental health problems	spanking	1.12	.38	1.86	
		physical abuse	.96	.23	1.70	
Overall		spanking	.25	.22	.27	
		physical abuse	.38	.29	.41	

size for spanking was 65% of the magnitude of the mean effect size for physical abuse.

Moderator Analyses by Study Characteristics

We examined whether study-level effect sizes varied across seven study-level characteristics using meta-regression to calculate average effect sizes by study subgroup (Borenstein et al., 2009; Harbord & Higgins, 2009). The results from these moderator analyses are presented in Table 4. All of the comparisons were nonsignificant, indicating that the effect sizes did not vary by study characteristic. The finding that the average effect size for longitudinal studies was the same as that for cross-sectional studies ($\beta = -.07$, *ns*) is important in light of the criticism that previous meta-analyses were overly influenced by cross-sectional studies (Baumrind et al., 2002); for the studies examine here, no evidence was found that the magnitude or direction of effect sizes was smaller in longitudinal than cross-sectional studies. Average effect sizes also did not significantly vary based on how spanking was measured, how it was indexed, whether the raters of spanking and outcome were independent, the time period over which spanking was measured, the country of the study, or the age group of the children studied.

Tests for Publication Bias

One potential threat to the validity of meta-analyses is what is referred to as publication bias, or commonly "the file drawer effect:" Namely, how likely is it that there are many studies with contradictory findings that were not published that would under-

Table 4
Effect Sizes Moderated by Study Characteristics

Moderator	*N* in each subcategory	β for difference from referent	Significant differences among subgroups
Study design (referent = cross-sectional)	61		
Longitudinal	23	−.07	
Retrospective	23	−.03	
Experimental	4	−.50	
Measure (referent = parent report)	65		
Observation	6	−.10	
Child report	11	.03	
Child retrospective	26	−.03	
Both parent and child	3	−.10	
Index of spanking (referent = frequency)	79		
Frequency and severity	8	−.11	
Ever in life	15	−.06	
When used	5	−.29	
Ever in time period	4	−.18	
Raters of spanking and outcome (referent = same rater)	42		
Independent raters	67	−.04	
Time period (referent = not specified)	51		
Observation	6	−.10	
Last week	22	.01	
Last month	3	−.26	
Last year	8	.16	
Ever	8	−.17	
Hypothetical	1	−.40	
Specific time period	12	−.04	
Country (referent = U.S.)	77		
Other than U.S.	34	.04	
Age range of children at time of spanking (referent = 2- to 5-years-old)	36		
Less than 2-years-old	15	.17	
6- to 10-years-old	30	.05	
11- to 15-years-old	27	.01	

Note. βs in third column represent difference from the β for the referent category. None were significantly different from the β for the referent category.

mine the conclusions of the meta-analysis? For each meta-analysis, we conducted the publication bias test developed by Egger, Davey Smith, Schneider, and Minder (1997) and implemented in Stata (Harbord, Harris, & Sterne, 2009; Steichen, 1998). None of the tests were significant, indicating that the risk of publication bias for any of our mean effect sizes is small.

Discussion

The goal of this article was to address two major concerns about past meta-analyses of the association of parents' use of spanking and a range of child outcomes. We will discuss each in turn, but begin with a summary of the overall findings from this new set of meta-analyses.

Spanking Is Associated With Higher Risk for Detrimental Outcomes

Thirteen of the 17 child outcomes examined were found to be significantly associated with parents' use of spanking. Among the outcomes in childhood, spanking was associated with more aggression, more antisocial behavior, more externalizing problems, more internalizing problems, more mental health problems, and more negative relationships with parents. Spanking was also significantly associated with lower moral internalization, lower cognitive ability, and lower self-esteem. The largest effect size was for physical abuse; the more children are spanked, the greater the risk that they will be physically abused by their parents.

Three of the four adult outcomes were significantly associated with a history of spanking from parents: adult antisocial behavior, adult mental health problems, and adult support for physical punishment. While these findings suggest that there may be lasting impacts of spanking that reach into adulthood, they are only suggestive, as adults who engage in antisocial behavior or who are experiencing mental health problems may focus on negative memories of their childhoods and report more spanking than they actually received. The finding that a history of received spanking is linked with more support for spanking of children as an adult may be an example of intergenerational transmission of spanking, or it may be an example of adults selectively remembering their past as a way of rationalizing their current beliefs. Only one of the 20 effect sizes for outcomes in adulthood was from a prospective longitudinal study (McCord, 1991). More longitudinal studies are needed to confirm the direction of effect.

An important observation about the meta-analyses is that the individual studies are highly consistent: 71% of all of the effect sizes, and 99% of the significant effect sizes, indicated a significant association between parental spanking and detrimental child outcomes. The only study that found a significant association with

a beneficial outcome, Tennant, Detels, and Clark (1975), had a unique sample (U.S. Army soldiers stationed in West Germany in 1971 and 1972, most of whom were White [77%]), and found that soldiers who recalled being spanked were less likely to report using amphetamines or opiates. While this is clearly a beneficial outcome, the uniqueness of the sample limits the generalizability of this finding and may explain why this study is an outlier, as it was in the Larzelere and Kuhn (2005) meta-analyses.

Three outcomes in childhood were not significantly associated with spanking in the meta-analyses: immediate defiance, child alcohol or substance abuse, and low-self regulation. The failure to reach significance for immediate defiance appears to result from the small *n* (150), while for child alcohol or substance abuse and low self-regulation the cause appears to be heterogeneity in effect sizes. The finding that spanking was not linked with immediate defiance was unexpected given the opposite findings in the Gershoff (2002) meta-analyses. The disparity arose because we coded the effect sizes from the three experimental studies of compliance (Bean & Roberts, 1981; Day & Roberts, 1983; Roberts & Powers, 1990) differently. Unlike Gershoff (2002) in which effect sizes for each study were calculated by subtracting the rate of compliance among children in the spanking condition from the rate of compliance among children in the comparison condition, we calculated the within condition difference in pre- and postintervention compliance rates for the spank and no-spank groups and then subtracted these two difference scores from each other. Because there were baseline differences between the treatment and control groups in each study, our effect sizes thus captured the extent to which spanking was associated with decreases in immediate defiance over baseline. From these five studies, it appears that children are as likely to defy their parents when they spank as comply with them, but future research will be needed to substantiate this conclusion.

Taken together, these meta-analyses support the conclusion that parents' use of spanking is associated with detrimental child outcomes. As most of the included studies were correlational or retrospective (72%), causal links between spanking and child outcomes cannot be established by these meta-analyses. That said, given that a correlational association is a necessary condition for a causal relationship (Shadish et al., 2001), we can conclude that the data are consistent with a conclusion that spanking is associated with undesirable outcomes.

Spanking Alone Is Associated With Detrimental Outcomes, but in Similar Ways to Physical Abuse

Our first research question was whether spanking would be associated with detrimental child outcomes when studies relying on harsh and potentially abusive methods were removed. The answer to this question is: Yes, it is. As noted above, all of the mean effect sizes indicated that even when a restricted definition of spanking is used, spanking is associated with detrimental child outcomes. The mean effect size across all studies, $d = .33$, was smaller than the overall mean effect size reported by Gershoff (2002), $d = .40$, but still statistically significant.

In order to better compare the findings for spanking with those for abuse, we identified seven studies that reported effect sizes for spanking and for physical abuse for the same child outcomes and conducted a meta-analyses of this set of studies. Spanking in these studies was significantly associated with detrimental outcomes with an effect size $d = .25$, while the mean effect size for physical abuse for these studies was significant and $d = .38$ (see Table 3). However, the mean effect size for all studies from Table 2, $d = .33$, is closer to the mean effect size for physical abuse in Table 3 than it is to the mean effect size for spanking from these 10 select effect sizes, indicating that spanking and physical abuse have relations with child outcomes that are similar in magnitude and identical in direction.

That spanking and physical abuse may have similar associations with child outcomes is consistent with previous literature. Both behaviors involve parents intentionally hitting (and hurting) children, albeit to different degrees (Gershoff, 2013), and most instances of substantiated physical abuse (75%), like all instances of spanking, begin as responses to children's misbehavior (Durrant et al., 2006). In addition, many researchers have argued that spanking and physical abuse are on a continuum of violence against children, and that spanking can escalate into physical abuse (Straus, 2001), an argument supported by our finding that spanking was significantly associated with physical abuse (Table 2; $d = .64$). Clearly not all parents who spank their children also administer more severe punishment; as with all of the meta-analyses presented, the association only indicates that milder and more severe corporal punishment are linked, and that the former may increase the risk that children will also be physically abused.

Spanking Effect Sizes Are Similar Across Study Characteristics

A major concern raised about the spanking literature in general and previous meta-analyses in particular is that their reliance on cross-sectional designs may mask what are truly child-elicitation effects (Baumrind et al., 2002; Ferguson, 2013; Larzelere et al., 2004). In other words, associations between spanking and problematic behavior may reflect the fact that difficult children elicit more spanking from parents, not that spanking causes the problematic behavior in the first place (Baumrind et al., 2002; Larzelere et al., 2004). Longitudinal or experimental designs are needed to isolate the direction of effect, and several were available for inclusion in the meta-regression moderation analyses. While it was indeed true that the majority of studies (70%) were cross-sectional or retrospective in nature, the effect sizes for the longitudinal and experimental studies were not significantly different from the effect sizes for the cross-sectional studies (see Table 4). This finding indicates that methodologically stronger studies did not find significantly smaller effect sizes than methodologically weaker studies, lending more confidence to the findings from the main meta-analyses that include both. The mean effect size for spanking also did not vary by any of the other six study characteristic moderators. The association between spanking and detrimental child outcomes did not depend on how spanking was assessed, who reported the spanking, the country where the study was conducted, or what age children were the focus of the study. Across all categories, methodologically stronger study designs identified the same risk for negative outcomes as did weaker study designs, suggesting that the associations between spanking and child outcomes are robust to study design.

We were surprised that none of the moderators was significant given that most of the I^2 values in Table 2 indicate high levels of

heterogeneity. There is little guidance in the literature about how to interpret significant I^2 values when paired with significant and consistent mean effect sizes. We suspect that the I^2 tests are picking up on expected variability in our independent variable. Unlike clinical trials, for which the I^2 was developed, in which treatment is systematized, it is nearly impossible to manipulate the amount or frequency of exposure to spanking and thus every participant's experience with our independent variable is different. Given that 13 of the 17 mean effect sizes were significantly different from zero and that nearly three quarters (71%) of the studies yielded effect sizes in the direction of detrimental outcomes, including nearly all of the significant effect sizes, we suspect that the I^2 is picking up on heterogeneity in spanking itself rather than in its associations with child outcomes.

Limitations

The primary limitation of these meta-analyses is their inability to causally link spanking with child outcomes. This is problematic because there is selection bias in who gets spanked—children with more behavior problems elicit more discipline generally and spanking in particular (Larzelere, Kuhn, & Johnson, 2004). Cross-sectional designs do not allow the temporal ordering of spanking and child outcomes that could help rule out the selection bias explanation. As noted above, randomized experiments of spanking are difficult if not ethically impossible to conduct, and thus this shortcoming of the literature will be difficult to correct through future studies.

The main viable strategy for doing so is through the use of analytic methods which increase our confidence that the causal direction is as hypothesized. Whenever such strategies have been employed, they have confirmed that spanking is associated with detriments to children. A series of cross-lagged studies (Berlin et al., 2009; Gershoff, Lansford, Sexton, Davis-Kean, & Sameroff, 2012; McLeod, Kruttschnitt, & Dornfeld, 1994; Sheehan & Watson, 2008) has demonstrated that spanking predicts changes in children's behavior, over and above their initial levels and the child effect of early problem behavior on later spanking.

Another statistical method that has been employed to strengthen conclusions is fixed effects regressions, which control for time-invariant unobserved characteristics that may account for observed relationships between spanking and child outcomes, such as children's initial levels of problem behavior. Using fixed effects models with the National Longitudinal Study of Youth (NLSY), Grogan-Kaylor (2005) found that increases in parents' use of spanking predicted increases in children's externalizing behaviors over time.

A third method is to establish spanking as a significant mediator of treatment effects on children for interventions that include a focus on reducing parents' use of spanking. In one example, an evaluation of the Incredible Years intervention for young children with behavior problems (Beauchaine et al., 2005) found that treatment effects on a reduction in conduct problems were significantly mediated through a reduction in parents' use of spanking. Similarly, analysis of data from a national randomized controlled trial of the federal Head Start program for low income children found that parents in the program significantly reduced their spanking, which was in turn linked with decreases in child aggression over time (Gershoff, Ansari, Purtell, & Sexton, 2016). More studies capitalizing on experimental designs are needed.

By looking at change over time and by accounting for potential alternative explanations through statistical methods or by capitalizing on data from experimental designs, such studies support the conclusion that there is a significant association of parents' use of spanking with later child outcomes, over and above children's initial behavior and child elicitation effects. None of these studies using advanced statistical methods found evidence that the pathway is entirely one of selection or child elicitation, or that spanking predicts improvements in children's behavior over time, as critics of the literature on spanking have contended (Baumrind et al., 2002; Ferguson, 2013; Kazdin & Benjet, 2003; Larzelere et al., 2004). Rather, these studies with strong designs provide more, not less, support for a potential causal link between spanking and detrimental child outcomes.

Conclusion

Spanking children to correct misbehavior is a widespread practice, yet one shrouded in debate about its effectiveness and even its appropriateness. The meta-analyses presented here found no evidence that spanking is associated with improved child behavior and rather found spanking to be associated with increased risk of 13 detrimental outcomes. These analyses did not find any support for the contentions that spanking is only associated with detrimental outcomes when it is combined with abusive methods or that spanking is only associated with such outcomes in methodologically weak studies. Across study designs, countries, and age groups, spanking has been linked with detrimental outcomes for children, a fact supported by several key methodologically strong studies that isolate the ability of spanking to predict child outcomes over time. Although the magnitude of the observed associations may be small, when extrapolated to the population in which 80% of children are being spanked, such small effects can translate into large societal impacts. Parents who use spanking, practitioners who recommend it, and policymakers who allow it might reconsider doing so given that there is no evidence that spanking does any good for children and all evidence points to the risk of it doing harm.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

*Alati, R., Maloney, E., Hutchinson, D. M., Najman, J. M., Mattick, R. P., Bor, W., & Williams, G. M. (2010). Do maternal parenting practices predict problematic patterns of adolescent alcohol consumption? *Addiction, 105,* 872–880. http://dx.doi.org/10.1111/j.1360-0443.2009.02891.x

American Academy of Pediatrics. (2012). AAP publications reaffirmed and retired. *Pediatrics, 130,* e467–e468. http://dx.doi.org/10.1542/peds.2012-1359

*Baer, D. J., & Corrado, J. J. (1974). Heroin addict relationships with parents during childhood and early adolescent years. *Journal of Genetic Psychology, 124,* 99–103.

*Bakoula, C., Kolaitis, G., Veltsista, A., Gika, A., & Chrousos, G. P. (2009). Parental stress affects the emotions and behavior of children up to adolescence: A Greek prospective, longitudinal study. *Stress, 12,* 486–498. http://dx.doi.org/10.3109/10253890802645041

*Barnes, J. C., Boutwell, B. B., Beaver, K. M., & Gibson, C. L. (2013). Analyzing the origins of childhood externalizing behavior problems. *Developmental Psychology, 49,* 2272–2284. http://dx.doi.org/10.1037/a0032061

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

*Bean, A. W., & Roberts, M. W. (1981). The effect of time-out release contingencies on changes in child noncompliance. *Journal of Abnormal Child Psychology, 9,* 95–105. http://dx.doi.org/10.1007/BF00917860

Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73,* 371–388.

Becker, W. C. (1964). Consequences of different models of parental discipline. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1, pp. 169–208). New York, NY: Sage.

Benjet, C., & Kazdin, A. E. (2003). Spanking children: The controversies, findings, and new directions. *Clinical Psychology Review, 23,* 197–224. http://dx.doi.org/10.1016/S0272-7358(02)00206-4

*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. http://dx.doi.org/10.1111/j.1467-8624.2009.01341.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley. http://dx.doi.org/10.1002/9780470743386

*Boutwell, B. B., Franklin, C. A., Barnes, J. C., & Beaver, K. M. (2011). Physical punishment and childhood aggression: The role of gender and gene-environment interplay. *Aggressive Behavior, 37,* 559–568.

Bradburn, M. J., Deeks, J. J., & Altman, D. G. (2009). Metan—A command for meta-analysis in Stata. In J. A. C. Sterne (Ed.), *Meta-analysis in Stata: An updated collection from the Stata journal* (pp. 3–28). College Station, TX: Stata Press.

*Buehler, C., & Gerard, J. M. (2002). Marital conflict, ineffective parenting, and children's and adolescents' maladjustment. *Journal of Marriage and the Family, 64,* 78–92.

*Bugental, D. B., Martorell, G. A., & Barraza, V. (2003). The hormonal costs of subtle forms of infant maltreatment. *Hormones and Behavior, 43,* 237–244. http://dx.doi.org/10.1016/S0018-506X(02)00008-9

*Burton, R. V., Maccoby, E. E., & Allinsmith, W. (1961). Antecedents of resistance to temptation in four-year-old children. *Child Development, 32,* 689–710.

*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. http://dx.doi.org/10.1037/a0032054

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

*Christie-Mizell, C. A., Pryor, E. M., & Grossman, E. R. B. (2008). Child depressive symptoms, spanking and emotional support: Differences between African American and European American youth. *Family Relations, 57,* 335–350.

Committee on the Rights of the Child. (2006). *General comment No. 8 (2006): The right of the child to protection from corporal punishment and or cruel or degrading forms of punishment (articles 1, 28(2), and 37, inter alia), 42nd Sess., U. N. Doc. CRC/C/GC/8*. Retrieved from http://www.ohchr.org/english/bodies/crc/docs/co/CRC.C.GC.8.pdf

*Coyl, D., Roggman, L., & Newland, L. (2002). Stress, maternal depression, and negative mother-infant interactions in relation to infant attachment. *Infant Mental Health Journal, 23,* 145–163.

*Day, D. E., & Roberts, M. W. (1983). An analysis of the physical punishment component of a parent training program. *Journal of Abnormal Child Psychology, 11,* 141–152. http://dx.doi.org/10.1007/BF00912184

*Deb, S., & Adak, M. (2006). Corporal punishment of children: Attitude, practice and perception of parents. *Social Science International, 22,* 3–13.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188. http://dx.doi.org/10.1016/0197-2456(86)90046-2

Donnelly, M., & Straus, M. A. (Eds.). (2005). *Corporal punishment of children in theoretical perspective*. New Haven, CT: Yale University Press. http://dx.doi.org/10.12987/yale/9780300085471.001.0001

*Durrant, J. E. (1994). Sparing the rod: Manitobans' attitudes toward the abolition of physical discipline and implications for policy change. *Canada's Mental Health, 41,* 2–6.

Durrant, J., Trocmé, N., Fallon, B., Milne, C., Black, T., & Knoke, D. (2006). *Punitive violence against children in Canada. CECW Information Sheet #41E*. Toronto, Canada: University of Toronto, Faculty of Social Work. Retrieved from www.cecw-cepb.ca/DocsEng/PunitiveViolence41E.pdf

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

*Eisenberg, N., Chang, L., Ma, Y., & Huang, X. (2009). Relations of parenting style to Chinese children's effortful control, ego resilience, and maladjustment. *Development and Psychopathology, 21,* 455–477. http://dx.doi.org/10.1017/S095457940900025X

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.

*Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2008). Exposure to childhood sexual and physical abuse and adjustment in early adulthood. *Child Abuse & Neglect: The International Journal, 32,* 607–619. http://dx.doi.org/10.1016/j.chiabu.2006.12.018

*Flynn, C. P. (1999). Exploring the link between corporal punishment and children's cruelty to animals. *Journal of Marriage and the Family, 61,* 971–981.

*Foshee, V. A., Ennett, S. T., Bauman, K. E., Benefield, T., & Suchindran, C. (2005). The association between family violence and adolescent dating violence onset: Does it vary by race, socioeconomic status, and family structure? *Journal of Early Adolescence, 25,* 317–344.

*Frias-Armenta, M. (2002). Long-term effects of child punishment on Mexican women: A structural model. *Child Abuse & Neglect, 26,* 371–386.

*Gagné, M. H., Tourigny, M., Joly, J., & Pouliot-Lapointe, J. (2007). Predictors of adult attitudes toward corporal punishment of children. *Journal of Interpersonal Violence, 22,* 1285–1304.

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.

Gershoff, E. T. (2013). Spanking and child development: We know enough now to stop hitting our children. *Child Development Perspectives, 7,* 133–137. http://dx.doi.org/10.1111/cdep.12038

Gershoff, E. T., Ansari, A., Purtell, K. M., & Sexton, H. R. (2016). Changes in parents' spanking and reading as mechanisms for Head Start impacts on children. *Journal of Family Psychology*. Advance online publication. http://dx.doi.org/10.1037/fam0000172

*Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., & Dodge, K. A. (2010). Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness. *Child Development, 81,* 487–502. http://dx.doi.org/10.1111/j.1467-8624.2009.01409.x

*Gest, S. D., Freeman, N. R., Domitrovich, C. E., & Welsh, J. A. (2004). Shared book reading and children's language comprehension skills: The

moderating role of parental discipline practices. *Early Childhood Research Quarterly, 19,* 319–336.

*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. http://dx.doi.org/10.1111/j.1467-8624.2011.01732.x

*Graziano, A. M., Lindquist, C. M., Kunce, L. J., & Munjal, K. (1992). Physical punishment in childhood and current attitudes: An exploratory comparison of college students in the United States and India. *Journal of Interpersonal Violence, 7,* 147–155.

*Grinder, R. E. (1962). Parental child-rearing practices, conscience, and resistance to temptation of sixth-grade children. *Child Development, 33,* 803–820.

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

*Gunnoe, M. L., & Mariner, C. L. (1997). Toward a developmental–contextual model of the effects of parental spanking on children's aggression. *Archives of Pediatric and Adolescent Medicine, 151,* 768–775.

Harbord, R. M., Harris, R. J., & Sterne, A. C. (2009). Updated tests for small-study effects in meta-analyses. In J. A. C. Sterne (Ed.), *Meta-analysis in Stata: An updated collection from the Stata journal* (pp. 138–150). College Station, TX: Stata Press.

Harbord, R. M., & Higgins, J. P. T. (2009). Meta-regression in Stata. In J. A. C. Sterne (Ed.), *Meta-analysis in Stata: An updated collection from the Stata journal* (pp. 70–96). College Station, TX: Stata Press.

*Hemenway, D., Solnick, S., & Carter, J. (1994). Child-rearing violence. *Child Abuse & Neglect, 18,* 1011–1020.

*Herzberger, S. D., Potts, D. A., & Dillon, M. (1981). Abusive and nonabusive parental treatment from the child's perspective. *Journal of Consulting and Clinical Psychology, 49,* 81–90.

*Hesketh, T., Zheng, Y., Jun, Y. X., Xing, Z. W., Dong, Z. X., & Lu, L. (2011). Behaviour problems in Chinese primary school children. *Social Psychiatry and Psychiatric Epidemiology, 46,* 733–741.

*Jackson, A. P., & Preston, K. S. J., & Franke, T. M. (2010). Single parenting and child behavior problems in kindergarten. *Race and Social Problems, 2,* 50–58.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Polo-Tomas, M., Price, T. S., & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology, 40,* 1047–1058. http://dx.doi.org/10.1037/0012-1649.40.6.1047

Johnson, B. T. (1993). *DSTAT 1.10: Software for the meta-analytic review of research literatures*. Hillsdale, NJ: Erlbaum.

*Joubert, C. E. (1991). Self-esteem and social desirability in relation to college students' retrospective perceptions of parental fairness and disciplinary practices. *Psychological Reports, 69,* 115–120.

*Joubert, C. E. (1992). Antecedents of narcissism and psychological reactance as indicated by college students' retrospective reports of their parents' behaviors. *Psychological Reports, 70,* 1111–1115.

*Kahn, M. W., & Fua, C. (1995). Children of South Sea island immigrants to Australia: Factors associated with adjustment problems. *International Journal of Social Psychiatry, 41,* 55–73.

*Kandel, D. B. (1990). Parenting styles, drug use, and children's adjustment in families of young adults. *Journal of Marriage and the Family, 52,* 183–196.

Kazdin, A. E., & Benjet, C. (2003). Spanking children: Evidence and issues. *Current Directions in Psychological Science, 12,* 99–103. http://dx.doi.org/10.1111/1467-8721.01239

*Kohrt, H. E., Kohrt, B. A., Waldman, I., Saltzman, K., & Carrion, V. G. (2004). An ecological-transactional model of significant risk factors for child psychopathology in Outer Mongolia. *Child Psychiatry and Human Development, 35,* 163–181.

*Lansford, J. E., Wager, L. B., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2012). Forms of spanking and children's externalizing behaviors. *Family Relations, 61,* 224–236. http://dx.doi.org/10.1111/j.1741-3729.2011.00700.x

Larzelere, R. E. (1996). A review of the outcomes of parental use of nonabusive or customary physical punishment. *Pediatrics, 98,* 824–828.

*Larzelere, R. E., Klein, M., Schumm, W. R., & Alibrando, S. A., Jr. (1989). Relations of spanking and other parenting characteristics to self-esteem and perceived fairness of parental discipline. *Psychological Reports, 64,* 1140–1142.

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,* 1–37. http://dx.doi.org/10.1007/s10567-005-2340-z

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

*Lau, J. T., Chan, K. K., Lam, P. K., Choi, P. Y., & Lai, K. Y. (2003). Psychological correlates of physical abuse in Hong Kong Chinese adolescents. *Child Abuse & Neglect, 27,* 63–75. http://dx.doi.org/10.1016/S0145-2134(02)00507-0

*Lau, J. T., Kim, J. H., Tsui, H. Y., Cheung, A., Lau, M., & Yu, A. (2005). The relationship between physical maltreatment and substance use among adolescents: A survey of 95,788 adolescents in Hong Kong. *The Journal of Adolescent Health, 37,* 110–119. http://dx.doi.org/10.1016/j.jadohealth.2004.08.005

*Lau, J. T. F., Yu, X., Zhang, J., Mak, W. W. S., Choi, K. C., Lui, W. W. S., . . . Chan, E. Y. Y. (2010). Psychological distress among adolescents in Chengdu, Sichuan at 1 month after the 2008 Sichuan earthquake. *Journal of Urban Health, 87,* 504–523. http://dx.doi.org/10.1007/s11524-010-9447-3

*Li, Y., Shi, A., Wan, Y., Hotta, M., & Ushijima, H. (2001). Child behavior problems: Prevalence and correlates in rural minority areas of china. *Pediatrics International, 43,* 651–661.

*Lynam, D. R., Miller, D. J., Vachon, D., Loeber, R., & Stouthamer-Loeber, M. (2009). Psychopathy in adolescence predicts official reports of offending in adulthood. *Youth Violence and Juvenile Justice, 7,* 189–207. http://dx.doi.org/10.1177/1541204009333797

*Lynch, S. K., Turkheimer, E., D'Onofrio, B. M., Mendle, J., Emery, R. E., Slutske, W. S., & Martin, N. G. (2006). A genetically informed study of the association between harsh punishment and offspring behavioral problems. *Journal of Family Psychology, 20,* 190–198. http://dx.doi.org/10.1037/0893-3200.20.2.190

*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. http://dx.doi.org/10.1111/j.1467-8624.2012.01820.x

*McCord, J. (1991). Questioning the value of punishment. *Social Problems, 38,* 167–179. http://dx.doi.org/10.2307/800527

*McKee, L., Roland, E., Coffelt, N., Olson, A. L., Forehand, R., Massari, C., . . . Zens, M. S. (2007). Harsh discipline and child problem behaviors: The roles of positive parenting and gender. *Journal of Family Violence, 22,* 187–196.

McLeod, J. D., Kruttschnitt, C., & Dornfeld, M. (1994). Does parenting explain the effects of structural conditions on children's antisocial behavior? A comparison of Blacks and Whites. *Social Forces, 73,* 575–604. http://dx.doi.org/10.1093/sf/73.2.575

*McLeod, J. D., & Shanahan, M. J. (1993). Poverty, parenting, and children's mental health. *American Sociological Review, 58,* 351–366.

*McLoyd, V. C., Kaplan, R., Hardaway, C. R., & Wood, D. (2007). Does endorsement of physical discipline matter? Assessing moderating influ-

ences on the maternal and child psychological correlates of physical discipline in African American families. *Journal of Family Psychology, 21,* 165–175.

*Medina, A. M., Mejia, V. Y., Schell, A. M., Dawson, M. E., & Margolin, G. (2001). Startle reactivity and PTSD symptoms in a community sample of women. *Psychiatry Research, 101,* 157–169.

*Miller-Perrin, C. L., Perrin, R. D., & Kocur, J. L. (2009). Parental physical and psychological aggression: Psychological symptoms in young adults. *Child Abuse & Neglect, 33,* 1–11. http://dx.doi.org/10.1016/j.chiabu.2008.12.001

*Minton, C., Kagan, J., & Levine, J. A. (1971). Maternal control and obedience in the two-year-old. *Child Development, 42,* 1873–1894.

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

*Nettelbladt, P., Svenson, C., & Serin, U. (1996). Background factors in patients with schizoaffective disorder as compared with patients with diabetes and healthy individuals. *European Archives of Psychiatry and Neurosciences, 246,* 213–218.

*Olson, S. L., Ceballo, R., & Park, C. (2002). Early problem behavior among children from low-income, mother-headed families: A multiple risk perspective. *Journal of Clinical Child & Adolescent Psychology, 31,* 419–430.

*Oyserman, D., Bybee, D., Mowbray, C., & Hart-Johnson, T. (2005). When mothers have serious mental health problems: Parenting as a proximal mediator. *Journal of Adolescence, 28,* 443–463.

*Pagani, L. S., Tremblay, R. E., Nagin, D., Zoccolillo, M., Vitaro, F., & McDuff, P. (2004). Risk factor models for adolescent verbal and physical aggression toward mothers. *International Journal of Behavioral Development, 28,* 528–537.

*Palmer, E. J., & Hollin, C. R. (2001). Sociomoral reasoning, perceptions of parenting and self-reported delinquency in adolescents. *Applied Cognitive Psychology, 15,* 85–100.

Paolucci, E. O., & Violato, C. (2004). A meta-analysis of the published research on the affective, cognitive, and behavioral effects of corporal punishment. *The Journal of Psychology, 138,* 197–221. http://dx.doi.org/10.3200/JRLP.138.3.197-222

*Parkinson, C. E., Wallis, S. M., Prince, J., & Harvey, D. (1982). Research note: Rating the home environment of school-age children: A comparison with general cognitive index and school progress. *Journal of Child Psychology and Psychiatry, 23,* 329–333.

*Power, T. G., & Chapieski, M. L. (1986). Childrearing and impulse control in toddlers: A naturalistic observation. *Developmental Psychology, 22,* 271–275.

*Regev, R., Gueron-Sela, N., & Atzaba-Poria, N. (2012). The adjustment of ethnic minority and majority children living in Israel: Does parental use of corporal punishment act as a mediator? *Infant and Child Development, 21,* 34–51.

*Roberts, M. W. (1988). Enforcing chair timeouts with room timeouts. *Behavior Modification, 12,* 353–370. http://dx.doi.org/10.1177/01454455880123003

*Roberts, M. W., & Powers, S. W. (1990). Adjusting chair timeout enforcement procedures for oppositional children. *Behavior Therapy, 21,* 257–271. http://dx.doi.org/10.1016/S0005-7894(05)80329-6

*Schweitzer, P. J., Zafar, U., Pavlicova, M., & Fallon, B. A. (2011). Long-term follow-up of hypochondriasis after selective serotonin reuptake inhibitor treatment. *Journal of Clinical Psychopharmacology, 31,* 365–368. http://dx.doi.org/10.1097/JCP.0b013e31821896c3

*Sears, R. R. (1961). Relation of early socialization experiences to aggression in middle childhood. *Journal of Abnormal and Social Psychology, 63,* 466–492.

*Slade, E. P., & Wissow, L. S. (2004). Spanking in early childhood and later behavior problems: A prospective study of infants and young toddlers. *Pediatrics, 113,* 1321–1330.

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

Sheehan, M. J., & Watson, M. W. (2008). Reciprocal influences between maternal discipline techniques and aggression in children and adolescents. *Aggressive Behavior, 34,* 245–255. http://dx.doi.org/10.1002/ab.20241

Steichen, T. J. (1998). sbe19: Test for publication bias in meta-analysis. *Stata Technical Bulletin, 41,* 9–15.

Steinmetz, S. K. (1979). Disciplinary techniques and their relationship to aggressiveness, dependency, and conscience. In W. R. Burr, R. Hill, F. I. Nye, & I. L. Reiss (Eds.), *Contemporary theories about the family: Vol. 1. Research based theories* (pp. 405–438). New York, NY: Free Press.

Sterne, J. A. C. (Ed.). (2009). *Meta-analysis in Stata: An updated collection from the Stata journal*. College Station, TX: Stata Press.

Straus, M. A. (2001). *Beating the devil out of them: Corporal punishment in American families* (2nd ed.). Piscataway, NJ: Transaction Publishers.

*Straus, M. A., & Paschall, M. J. (2009). Corporal punishment by mothers and development of children's cognitive ability: A longitudinal study of two nationally representative age cohorts. *Journal of Aggression, Maltreatment, & Trauma, 18,* 459–483. http://dx.doi.org/10.1080/10926770903035168

*Straus, M. A., Sugarman, D. B., & Giles-Sims, J. (1997). Spanking by parents and subsequent antisocial behavior of children. *Archives of Pediatric and Adolescent Medicine, 151,* 761–767.

*Talillieu, T. L., & Brownridge, D. A. (2013). Aggressive parental discipline experienced in childhood and internalizing problems in early adulthood. *Journal of Family Violence, 28,* 445–458.

*Tennant, F. S., Jr., Detels, R., & Clark, V. (1975). Some childhood antecedents of drug and alcohol abuse. *American Journal of Epidemiology, 102,* 377–385.

*Trickett, P. K., & Kuczynski, L. (1986). Children's misbehaviors and parental discipline strategies in abusive and nonabusive families. *Developmental Psychology, 22,* 115–123.

UNICEF. (2014). *Hidden in plain sight: A statistical analysis of violence against children*. New York, NY: UNICEF.

*Westbrook, T. R., Harden, B. J., Holmes, A. K., Meisch, A. D., & Whittaker, J. V. (2013). Physical discipline use and child behavior problems in low-income, high-risk African American families. *Early Education and Development, 24,* 923–945.

*Zahn-Waxler, C., Radke-Yarrow, M., & King, R. A. (1979). Child rearing and children's prosocial initiations toward victims of distress. *Child Development, 50,* 319–330.

*Zolotor, A. J., Theodore, A. D., Chang, J. J., Berkoff, M. C., & Runyan, D. K. (2008). Speak softly-and forget the stick corporal punishment and child physical abuse. *American Journal of Preventive Medicine, 35,* 364–369.

Appendix

Number and Percent of Studies Excluded from the Meta-Analyses by Exclusion Code

Reason for exclusion from meta-analyses	Number of studies excluded	Percent
Spanking not linked with child outcomes (e.g., prevalence only).	238	16
Not an empirical article (e.g., a literature review).	221	15
Definition of physical punishment included harsh methods of physical punishment beyond spanking, slapping, or hitting.	194	13
Spanking was not measured in the study.	171	11
Study was an unpublished dissertation.	104	7
Article was not relevant.	85	6
Attitudes toward, and not use of, physical punishment was assessed.	82	5
Study was of physical punishment in schools or other institutions.	73	5
Study did not include a bivariate association between spanking and the child outcome.	61	4
Study was of an intervention to reduce physical punishment.	47	3
Available statistics were unclear, insufficient, or inappropriate for the meta-analyses.	46	3
Spanking was combined with yelling or some form of psychological aggression.	44	3
Study was not available in English.	32	2
Spanking was combined with other types of discipline.	30	2
Study was published as a book chapter or conference presentation.	23	2
Study used same dataset as another study in the meta-analysis.	23	2
Dependent variable did not fit into other outcome categories.	11	1
Spanking was of animals, not children.	5	<1
Article was unavailable through interlibrary loan.	3	<1
Spanking measure included threats of spanking.	3	<1
Physical punishment measure was nontraditional (i.e., aversive noise; washing mouth out with soap).	2	<1
Study involved a special population of children (chromosomal abnormality).	1	<1
Total number of excluded studies	1,499	100%

Received November 10, 2015
Revision received February 16, 2016
Accepted February 16, 2016 ■