Clinical Psychology Review 33 (2013) 196–208

Contents lists available at SciVerse ScienceDirect

Clinical Psychology Review

Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies☆

Christopher J. Ferguson *

Department of Psychology and Communication, Texas A&M International University, 5201 University Blvd., Laredo, TX, 78041, USA

HIGHLIGHTS

- Presents a meta-analysis of longitudinal studies of negative effects of spanking.
- Spanking had a small but non-trivial negative effect on cognitive performance.
- Effects of spanking were largely trivial for other behavior problems.
- Spanking has not only few benefits, but also fewer negative consequences than often assumed.

ARTICLE INFO

Article history:
Received 29 June 2011
Received in revised 1 November 2012
Accepted 22 November 2012
Available online 3 December 2012

Keywords:
Spanking
Corporal punishment
Child development
Aggression
Cognition
Internalizing symptoms

ABSTRACT

Social scientists continue to debate the impact of spanking and corporal punishment (CP) on negative child outcomes including externalizing and internalizing behavior problems and cognitive performance. Previous meta-analytic reviews have mixed long- and short-term studies and relied on bivariate *r*, which may inflate effect sizes. The current meta-analysis focused on longitudinal studies, and compared effects using bivariate *r* and better controlled partial *r* coefficients controlling for time-1 outcome variables. Consistent with previous findings, results based on bivariate *r* found small but non-trivial long-term relationships between spanking/CP use and negative outcomes. Spanking and CP correlated .14 and .18 respectively with externalizing problems, .12 and .21 with internalizing problems and −.09 and −.18 with cognitive performance. However, when better controlled partial *r* coefficients (*pr*) were examined, results were statistically significant but trivial (at or below *pr* = .10) for externalizing (.07 for spanking, .08 for CP) and internalizing behaviors (.10 for spanking, insufficient studies for CP) and near the threshold of trivial for cognitive performance (−.11 for CP, insufficient studies for spanking). It is concluded that the impact of spanking and CP on the negative outcomes evaluated here (externalizing, internalizing behaviors and low cognitive performance) are minimal. It is advised that psychologists take a more nuanced approach in discussing the effects of spanking/CP with the general public, consistent with the size as well as the significance of their longitudinal associations with adverse outcomes.

Contents

☆ The author would like to thank Robert E. Larzelere and Ronald B. Cox, Jr. for the comments on an early draft of this manuscript.
* Tel.: +1 956 326 2636.
E-mail address: CJFerguson1111@aol.com.

0272-7358/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cpr.2012.11.002

1. Introduction

Spanking, usually defined as a mild open-handed strike to the buttocks or extremities (Friedman & Schonberg, 1996; McLoyd & Smith, 2002), and corporal punishment, which also includes more severe use of physical punishments, such as striking the face, hitting with an object or shaking or pushing a child, have been issues for considerable debate in social science and in the general public. The American Academy of Pediatrics has counseled against the use of spanking as a disciplinary strategy, citing potential negative child outcomes such as increased aggressiveness and potential physical harm to the child (American Academy of Pediatrics, 1998). Sweden was the first country to ban spanking, eventually leading the way for a total of 32 countries that do not allow the use of corporal punishment in the home (GITEACPOC, 2012).

Several prominent family violence scholars have been passionate in advocating against the use of spanking. Writing for the general public, psychologist Alan Kazdin (2008) claimed that spanking is linked with a host of negative outcomes ranging from aggression, poor academic performance and depression in childhood to "poor physical-health outcomes (cancer, heart disease, chronic respiratory disease)" in adulthood. Others such as sociologist Murray Straus (2008) have argued that corporal punishment is one of the key originating variables related to a wide range of violence related outcomes.

Despite calls against spanking and corporal punishment, these disciplinary practices remain in wide use, particularly within the United States where one recent study suggested that 65% of 3-year-olds had been spanked in the previous month (Taylor, Lee, Guterman, & Rice, 2010). Yet, concerns have been expressed that causal links between spanking and negative outcomes may have been exaggerated (Baumrind, Larzelere, & Cowan, 2002; Morris & Gibson, 2011), with problems in measurement and proper statistical controls inflating estimates of harm. Thus considerable debate remains regarding the impact of spanking and corporal punishment on long-term outcomes. The current study seeks to address some of the gaps in the literature by conducting a meta-analytic review of longitudinal studies of spanking and corporal punishment (CP).

2. The debate on spanking and corporal punishment (CP)

As noted in the first lines of this paper spanking and CP are not synonymous. Spanking generally is used to refer to relatively mild physical punishment using an open hand on the buttocks or extremities. Corporal punishment generally is used to refer to a broader class of behaviors. Spanking may be included within CP but CP generally also includes hitting with an object such as a switch, shaking, pushing, slapping the face, etc. Nonetheless it is further necessary to clarify that CP does not include highly injurious child abuse such as causing serious lacerations or broken bones. Therefore any conclusions about spanking and CP should not be extended to more serious forms of child abuse.

Although debates about spanking are not new (the American Psychological Association passed a resolution condemning corporal punishment in schools as far back as 1975) a useful starting point to understanding recent debates probably begins with Gershoff's (2002a) meta-analytic review of CP studies. As a technical note, Gershoff reported her results using the effect size index "d" although in most cases this was calculated from correlation (r) values. Effect sizes r and d are readily convertible from one to another, and as might be expected, randomized experiments on spanking/CP are very few. As such I use the effect size "r" consistently through the manuscript, converting from d where necessary for ease of communication. Gershoff linked CP with increased aggression ($r = .18$, i.e., $d = .36$) and decreased mental health ($r = -.24$) in childhood. Longer term effects on aggression appeared to remain consistent ($r = .27$), although deleterious effects on mental health declined long-term ($r = -.05$). However, these longer term effects are based on a combination of retrospective (89%) and longitudinal (11%) studies, and only 13% of all the effect sizes in the Gershoff meta-analysis were longitudinal in nature. A later meta-analysis by Paolucci and Violato (2004) found lower effect sizes ($r = .10$ for externalizing problems; $r = .10$ for internalizing problems and $r = .03$ for cognitive problems). Both meta-analyses concluded that CP could have small but significant deleterious effects on child outcomes. It is worth noting, however, that all the adverse effect sizes in Gershoff (2002a) and probably in Paolucci and Violato (2004) were based on bivariate r correlations, presumably to maintain homogeneity between the effect size estimates across all types of studies.

Several scholars have remained skeptical of claims of causal harm due to spanking and CP, however (Baumrind et al., 2002; Gunnoe, Hetherington, & Reiss, 2006; Larzelere, 2008). For instance a further meta-analysis by Larzelere and Kuhn (2005) found that negative effects for spanking differ very little from other disciplinary strategies. Specifically, although overly severe CP was related to more negative outcomes than disciplinary alternatives, conditional spanking had better outcomes than 10 of 13 non-physical discipline alternatives such as ignoring or privilege removal. Concerns with Gershoff's analyses and many of the studies which underlie her analysis include:

1) Conflation of spanking with more severe forms of corporal punishment. It has been contended that measures, have not carefully distinguished between various types of physical punishment, particularly in earlier studies. Conflating severe forms of CP with spanking may result in inflated effect sizes.
2) The temporal order of spanking and negative outcomes is not well documented. From cross-sectional correlational studies, it is not possible to determine whether spanking and CP lead to negative outcomes, or whether children with greater problem behaviors are more likely to be spanked. One way of establishing the temporal order is through the use of longitudinal designs. If time-1 spanking/CP to be found to predict time-2 outcomes the argument that spanking/CP comes first in the temporal order is strengthened. Although both Gershoff (2002a) and Paolucci and Violato (2004) included longitudinal studies in their analyses, they consisted of a minority of their studies and their effect sizes were not well distinguished from cross-sectional or retrospective designs. In Gershoff, 13% of the reported effect sizes were from longitudinal studies, with 21.8% of the studies included in Paolucci and Violato being of longitudinal design.
3) Controlling third variables. As Baumrind et al. (2002) point out, effect sizes based on bivariate correlations (as those in Gershoff, 2002a, appear to be) run the risk of inflating effect size estimate due to failure to control for other relevant variables. The use of

well-controlled designs is generally considered state of the art in violence research (Savage & Yancey, 2008). For meta-analyses to rely on bivariate *r* presents an issue in which the resultant effect size estimates may ultimately fail to represent the well-controlled studies from which they are culled. However, this presents something of a conundrum for meta-analysis. Although standardized regression coefficients (β) can, in a sense, be interpreted similarly to bivariate *r* (Ferguson, 2009), the difference in level of control makes it difficult to directly compare un-adjusted *r*s to covariate-controlled betas, and adds heterogeneity to meta-analyses. In addition, covariate-adjusted βs will have additional heterogeneity due to differences in the covariates across studies. Yet including only bivariate *r*s can inflate effect sizes and lead to mistaken conclusions about the causal role of a particular risk factor. Controlling for time 1 outcomes scores is an essential control variable for longitudinal studies.

4) Overreliance on retrospective recall. Many of the long-term studies used in both Gershoff (2002a) and Paolucci and Violato (2004) were retrospective rather than prospective (17% and 22% respectively). Retrospective reports of past CP may ultimately say more about the respondent's current psychological status rather than accurate memories of disciplinary acts, which took place years or decades earlier (Baumrind et al., 2002). Thus their utility is limited.
5) Shared method variance. In many of the studies included in the two meta-analyses discussed, respondents for both the spanking/CP variables and child outcomes were the same individual. This shared method variance can result in inflated correlations between variables (Kenny & Kashy, 1992) and thus inflate effect size estimates.

From the debates between Gershoff (2002a) and Baumrind et al. (2002) in particular, it becomes apparent that short-term correlational designs (whether cross-sectional or retrospective) are not adequate for establishing an argument in favor of a causal link between spanking/CP and long-term harm. Thus, increased emphasis must be placed on the use of longitudinal studies of spanking/CP. Such studies are able to place spanking/CP and negative outcomes in a temporal sequence, which would provide stronger evidence in support of the potential for harm, particularly when experimental studies are impractical or unethical.

3. Longitudinal studies of spanking and corporal punishment

In the decade since Gershoff (2002a), a plethora of longitudinal studies have been published, examining links between spanking and CP and a number of negative outcomes, although aggression remains a primary focus of the majority of studies. A large number of these longitudinal studies actually come from a single dataset, the National Longitudinal Survey of Youth (NLSY; Baker, Keck, Mott, & Quinlan, 1993). The NLSY followed thousands of children over many years and considered numerous behavioral data points. As such, the NLSY has been a rich source of data on child behavioral health. As one cautionary note, when a single dataset has such a dominant position in the field, there is always the possibility that the peculiarities of this dataset may have undue influence on the field at large. That having been said, not all studies published using the NLSY agree on whether CP results in long-term harm or not. Many studies using this dataset have concluded that CP is linked with long-term harm to children (e.g., Christie-Mizell, Pryor, & Grossman, 2008; Eamon, 2001; Grogan-Kaylor, 2004; Straus & Paschall, 2009) although effect sizes were generally small. However, a reanalysis by Larzelere, Cox, and Smith (2010) found little difference in the effect for spanking as compared to using grounding or even using psychotherapy for behavioral problems. Larzelere, Ferrer, Kuhn, and Daniela (2010) also found that using improved statistical controls for third variables reduced the spanking effect to non-significance.

Longitudinal datasets independent of the NLSY have likewise tended to produce relatively small, and not always consistent effects (Berlin et al., 2009; Larzelere, Ferrer et al., 2010; Lau, Litrownik, Newton, Black, & Everson, 2006; Morris & Gibson, 2011; Stacks, Oshio, Gerard, & Roe, 2009). Nonetheless, in reviewing these longitudinal studies, it appears that the majority of authors of longitudinal studies conclude that spanking and CP may have at least small long-term effects on negative outcome in childhood and early adulthood.

Nonetheless the methodological circumstances under which longitudinal studies of spanking and CP find negative effects remain unclear. Many of the questions raised by Baumrind et al. (2002) in the wake of the Gershoff (2002a) analysis regarding methodological issues that may inflate effect size estimates remain unaddressed. Further, no meta-analysis has specifically focused on longitudinal studies. Both Gershoff (2002a) and Paolucci and Violato (2004) mixed longitudinal studies in with cross-sectional and retrospective designs. Arguably, longitudinal studies provide the best non-experimental evidence for or against the belief in causal spanking/CP effects given the ability of such studies to delineate the temporal sequence. Thus, examining these studies more closely will be of particular value.

The current meta-analysis has several aims:

1.) To consider the long-term influence of spanking and CP on externalizing and internalizing symptoms, and cognitive performance.
2.) To examine how the influences of spanking/CP contrast to other disciplinary strategies such as negative verbalizations, arbitrary and inconsistent discipline, and disciplinary strategies generally identified as positive such as positive encouragement, supervision and redirection, and rewards for "good" behavior (American Academy of Pediatrics, 1998). It is worth noting that positive parenting and spanking may be used in differing disciplinary scenarios, an issue not yet addressed in the literature.
3.) The current meta-analysis will also examine moderators that may explain outcome differences between studies. Among moderators considered will be source independence (Baumrind et al., 2002), the length of the longitudinal period and age of the children considered and potential differences between the NLSY and other longitudinal designs. It had been hoped that racial differences in longitudinal designs could be examined, but too few studies parceled out effects across racial groups for a reliable meta-analysis.

4. Method

4.1. Selection of studies

Identification of relevant studies involved a search of the PsycINFO, Medline, Dissertation Abstracts and Digital Dissertations databases using the search term "spanking OR corporal punishment OR physical punishment." In addition, recent reviews of the CP/spanking literature were examined for papers that may have been missed in the literature search. Included studies had to meet the following criteria:

1) Each study had to present a longitudinal design.
2) Each study had to measure the influence of spanking/CP on at least one of the outcomes noted earlier (externalizing or internalizing behavior problems or cognitive performance).
3) Studies which focused exclusively on severe child abuse (i.e., causing serious injuries, breaking bones, actions resulting in arrest or removal of child custody, sexual abuse) were excluded.
4) Each study had to present statistical outcomes or data that could be meaningfully converted into effect size "*r*".

This search netted 45 studies, of which 6 were doctoral dissertations. Given the enormous time invested in longitudinal studies, finding a relative minority of doctoral dissertations with longitudinal designs is not surprising. One dissertation that was located (Buemi, 2009) appeared to overlap in data with a published study (Christie-Mizell et al., 2008) and thus was discarded in favor of the published study. Non-indexed unpublished studies were not included out of concern

that a search for non-indexed studies might introduce selection bias (Ferguson & Brannick, 2012), given it may be more likely to receive unpublished studies from certain groups of authors (for instance those already well-published in a particular field) relative to others (see Egger & Smith, 1998). The 45 studies in the current analysis provided 111 separate effect size estimates linking the disciplinary and outcome variables of interest. As these involved separate outcomes analyzed separately here, the independence of effect size estimates in the meta-analysis was not compromised. The list of studies is presented in Appendix A.

5. Operationalization of relevant terms

One issue that has been raised as a confound in spanking/CP research is the difficulty in distinguishing spanking from CP and from more severe abuse (Baumrind et al., 2002). For instance, as noted earlier, spanking is typically defined as open handed swats to the buttocks or extremities. However, few studies restrict their measure of spanking to that definition. Many studies simply ask parents how often they "spank" their child, while leaving spanking undefined, for instance. Given that individual studies are not always clear on this issue, it represents a serious challenge for meta-analyses. Furthermore, studies that satisfy a broader definition of spanking (e.g., reported frequency) seldom exclude those who also use more severe CP or physical abuse. When such a study does not also assess and remove variance due to more severe forms of CP and abuse, it remains possible that the parents who respond affirmatively to spanking questions may also be those prone to more serious CP or physical abuse of children. Some studies, such as those using the Conflict Tactics Scale, could potentially remove cases in which respondents acknowledge more serious abuse, and compare remaining spanking parents versus parents who do not. However, few studies employ such techniques.

That said, the theoretical definition of spanking for the current study was open-hand swats to the buttocks or extremities, but most qualifying studies used reported occurrence or frequency of spanking as defined by the participants. Corporal punishment was operationalized as consisting of a wider range of generally 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. Again, it has been amply argued (e.g., Baumrind et al., 2002; Larzelere & Kuhn, 2005) that the operationalization of spanking and CP across studies has generally been weak, and clear delineations among open-handed spanking, other spanking, CP, and injurious abuse are absent. This remains a serious limitation of this research field and one which is difficult to fix in meta-analysis. For the current analysis there is little recourse but to take the individual studies at face value, although it must be emphasized that this cautionary note should be recalled when interpreting the results.

Negative verbalizations were operationalized as including measures of parental yelling, screaming or insulting of children. Example items from various measures include "I have called my child lazy or dumb or some other word like that" or "I shouted, yelled or screamed at my child." Arbitrary discipline included measures of inconsistency in parental discipline, and a general tendency toward inconsistent application of harsh disciplines for minor infractions. Example items from various measures include "You threaten to punish your child, but then do not actually punish him or her" and "How often do you yell at your child after you've had a bad day?" Positive discipline was broadly operationalized to include parental acts that focused on responding to misbehavior with supervision, redirection and reasoning such as giving rationales for behaving appropriately. Example items from various measures include "I explained why something was wrong" or "I gave my child something else to do instead of what she or he was doing wrong." It is once again important to note that the operationalizations here are, to some extent, dependent upon the definitions and measurements used in individual studies. It would have been ideal, for instance, to distinguish positive disciplinary strategies that were proactive as opposed to responsive to misbehavior, although clarity on this issue was not available in the present set of studies.

Outcome measures were operationalized in the following ways. Externalizing problems referred to behavior problems related to aggression, rule breaking, antisocial and oppositional behaviors. Internalizing problems referred to issues related primarily to depression, anxiety and stress. Cognitive performance was operationalized as measures of intellectual capacity, aptitude or achievement.

6. Effect size estimates

One issue to arise from the Gershoff (2002a) meta-analysis was that of proper estimates of effect size (Baumrind et al., 2002). Gershoff based her effect sizes only on bivariate *r*s, except for her only beneficial outcome (immediate compliance). This practice is, of course, fairly standard for meta-analyses (Hunter & Schmidt, 2004). The differences in variance explained by bivariate *r*s and better controlled effect size estimates such as betas or partial *pr*s makes the aggregation of effect size estimates across studies difficult. However, as noted by Baumrind et al. (2002), reliance on bivariate *r* raises the potential of upwardly inflated effect size estimates from selection bias due to child effects. Given that well-controlled multivariate analyses are considered the gold standard in aggression research (Savage & Yancey, 2008), for meta-analyses to rely solely on bivariate *r* leads to increased risks of misleading causal conclusions coming from these analyses. This is particularly the case when many longitudinal studies take great care to employ careful statistical controls. For a meta-analysis to remain rooted to bivariate *r*, it would be theoretically possible for every single longitudinal study to conclude that any correlation between spanking/CP was reduced to non-significance once other factors were controlled, yet for a meta-analysis of these studies to conclude spanking/CP had a significant effect. In this circumstance, reliance on the bivariate *r*, when examining well-controlled multivariate longitudinal studies in meta-analysis is problematic. As just one example of this concern, Gunnoe and Mariner (1997, p. 772) reported bivariate correlations of .15 and .19 between spanking and externalizing outcomes, which were included as such in the Gershoff (2002a) meta-analysis (as *d*s of .30 and .39 in her Table 3, p. 545). Yet these correlations were greatly reduced if not eliminated or reversed once statistical controls were employed, leading Gunnoe and Mariner to conclude (p. 768) "For most children, claims that spanking teaches aggression seem unfounded."

If reliance on bivariate *r* is problematic, the solution is unclear. Some authors have suggested formulas for imputing bivariate *r* from betas (Peterson & Brown, 2005), providing estimates of the missing bivariate *r*. However, calculating r from beta simply returns us to the problem of the potential undesirability of basing causal inferences on un-adjusted bivariate *r*. Several authors have suggested that betas indeed can be used as effect size estimates in meta-analyses. As Rosenthal and DiMatteo (2001) note betas can be used as effect size estimates, with the cautionary note to recall that betas are better controlled than rs. Other authors have echoed this basic view (Farley, Lehmann, & Sawyer, 1995; Raju, Fralicx, & Steinhaus, 1986). Thus, it is concluded that bivariate *r* is potentially undesirable for the present analysis as reliance on bivariate *r* would not advance understanding beyond the criticisms raised by Baumrind et al. (2002) of the Gershoff meta-analysis. In her response to Baumrind et al. (2002), Gershoff (2002b) noted the issue of aggregating betas and bivariate *r*s together in a meta-analysis, but also acknowledged that examination of "third variables" could be important to study. Gershoff noted that relatively few studies in her meta-analysis controlled for third variables. Gershoff (p. 606) stated "It is unfortunate that the state of research on corporal punishment does not support the examination of important third variables, and I sincerely hope that future meta-analyses of parental corporal punishment will have sufficient data on third variables to include them either as control or moderator variables." As such there appears to be general agreement that control of third variables is of great importance in understanding long-term effects of

spanking/CP. Arguably theoretically relevant confounds that are most likely to minimize bias in the predictor/outcome relationship should take greatest priority (Steiner, Cook, Shadish, & Clark, 2010). In the case of spanking/CP, child characteristics that might theoretically lead to greater parental spanking/CP (i.e., child temperament or initial aggressiveness or behavioral problems) are a particularly important variable to control. However, as Gershoff (2002b) rightly notes, this approach interjects heterogeneity into the meta-analysis. This is particularly true where the control variables used across studies varies considerably. As such the use of betas may be too great a threat to the homogeneity assumption of meta-analysis.

In the current study, a compromise position was employed, examining partial *r* values using consistent controls across studies, specifically controlling for time 1 outcome variables. Controlling for the time 1 outcome was the most commonly employed statistical control, used in most (but not all) longitudinal designs. Thus, using the partial *r* controlling for the time 1 outcome, represents a better controlled effect size estimate, yet is homogeneous across longitudinal studies. Authors of studies that did not provide enough data to calculate partial *r* were contacted to provide the necessary data. It is argued that this approach is the best "middle path" between the pitfalls of reliance on bivariate *r* and the heterogeneity introduced by using betas. Sufficient data for calculating the partial *r* was available for 69 of the 124 (56%) individual effect size estimates, thus allowing a subset analysis for these studies.

However, meta-analytic results using bivariate *r* will also be presented, so that the results can be compared. Not all longitudinal studies included tables of bivariate correlations between variables. When these were missing, the authors of individual studies were contacted to supply bivariate *r* values. In several cases either the data were no longer available, or authors did not respond to the request for additional data. In such cases the formula for imputing missing bivariate *r* from standardized regression coefficients provided by Peterson and Brown (2005) was employed. However, in some cases data were not available to calculate the bivariate *r* or the partial *r* or they were not available from the authors.

7. Moderator variables

Several variables were considered important to examine as moderators. The issue of source independence has been raised as a potential moderator, with the potential that data points obtained from the same source may inflate effect size estimates (Baumrind et al., 2002). As such, source independence will be considered. The length of the longitudinal period and age of the children at time 1 will be considered, to examine whether the effects of spanking/CP are age dependent, and whether they remain stable over time. Finally whether a particular longitudinal study used data from the NLSY versus other data will be examined as a moderator. There is no reason to believe that the NLSY is a superior or inferior data set, but when a single data set dominates the research literature (even with different specific samples and methodological designs) the potential exists that this dataset may have undue influence on the field at large. When a particular data set or research lab is particularly productive it is worth considering this as a moderator (Starr & Davila, 2008).

8. Statistical and publication bias analyses

The Comprehensive Meta-Analysis (CMA) software program was used to fit both random and fixed effects models. Hunter and Schmidt (2004) argue that random effects models are appropriate when population parameters may vary across studies, as is likely here. As such, only random effects will be reported.

As for publication bias, it has been known for many years (e.g., Rosenthal, 1979) that the selective publication of statistically significant reports can bias research fields and meta-analyses, which draw from them. Doctoral dissertations were included to help offset potential publication bias processes. However, the position is taken here that it is naïve to assume that the mere inclusion of a set of unpublished studies represents a panacea for publication bias when the true scope of unpublished studies is unknown and essentially unknowable (most of them being non-indexed and, thus, difficult to retrieve in a non-biased manner). As such, the issue of publication bias will be carefully considered in the current analysis.

Testing for publication bias can be undertaken using a variety of techniques. However, some controversy has arisen that individual publication bias detection techniques may be prone to Type I error under varying circumstances (Macaskill, Walter, & Irwig, 2001). One way of addressing this concern is to use several tests of publication bias as suggested by Ferguson and Brannick (2012). Given that their individual weaknesses differ, combining them in order to make decisions about publication bias reduces the potential for type I error. Therefore, the following publication bias techniques and decision criteria will be employed, consistent with recent recommendations (Ferguson & Brannick, 2012):

1.) Orwin's FSN number (Orwin, 1983) is lower than the number of studies (*k*), using the criterion of effect size $r = .10$.
2.) Either the rank order correlation (Begg & Mazumdar, 1994) or Egger's regression (Egger, Davey-Smith, Schneider, & Minder, 1997) demonstrates significant results.
3.) Trim and Fill (Duvall & Tweedie, 2000) results are significant for bias
4.) A careful examination of included studies does not reveal consistent methodological differences between smaller and larger studies (i.e., small study effects).

9. Results

Individual study results are presented in Appendix B. Note that a minimum number of 3 studies in any given category of analysis were required for reporting of meta-analytic results in this article. Where potential results are not reported, this indicates an absence of this minimal number of studies. All reported effect sizes are means weighted by a function of sample size.

Aggregate results for the three main outcomes (externalizing and internalizing behaviors and cognitive performance) are presented in Table 1. Table 1 presents these values for bivariate *r* coefficients. Table 2 presents the same data using partial *pr* coefficients controlling for time 1 measures of the outcome variable. All mean effect sizes reached the threshold for statistical significance, $p < .05$, unless specifically indicated otherwise. As can be seen, results for bivariate *r*s were larger than for partial *pr* coefficients. Effects for bivariate *r*s were generally small to moderate but non-trivial (i.e., greater than Cohen's, 1992, small effect size of $r = .10$). On average, spanking correlated .14 with externalizing, .12 with internalizing, and -.09 with cognitive outcomes. Corporal punishment (CP) correlated .18 with externalizing, .21 with internalizing, and −.18 with cognitive outcomes. Harsh verbal punishment correlated .14 with externalizing and .14 with internalizing problems, whereas arbitrary discipline correlated .18 with externalizing outcomes. Positive parenting was not significantly related to externalizing behaviors ($r = -.03$), and there were too few studies to examine other outcomes ($k < 3$).

For the more conservative partial *pr* coefficients, which controlled for pre-existing difference on the outcomes, the effect sizes for spanking and corporal punishment on externalizing problems and cognitive performance were significant, but smaller and often trivial, defined herein as at or lower than $pr = .10$ (Cohen, 1992; Ferguson, 2009). Trivial effect sizes may often have limited practical significance. Effect sizes at or lower than $pr = .10$ are considered trivial herein (Cohen, 1992). The mean partial *pr* correlations (or βs, which are nearly equivalent for small effect sizes) for spanking were .07 with externalizing and .10 with internalizing problems. Corporal punishment (CP) had mean partial *pr* correlations of .08 with externalizing problems and -.11 with

Table 1
Meta-analytic results for main analysis including publication bias analysis, bivariate *r* values.

Outcome	Discipline	k	r_+	95% C.I.	OFSN	RCT	RT	Outlier?	SSE?	Bias?
Externalizing	CP	32	.18*	(.14, .22)	26	$p>.05$	$p>.05$	No	No	No
	Spanking	19	.14*	(.10, .18)	6	$p>.05$	$p>.05$	No	No	No
	Neg verbal	8	.14*	(.07, .22)	4	$p>.05$	$p>.05$	No	No	No
	Arbitrary	4	.18*	(.07, .29)	4	$p>.05$	$p>.05$	No	No	No
	Positive	20	−.03	(−.08, .02)	0	$p>.05$	$p>.05$	No	No	No
Internalizing	CP	5	.21*	(.06, .35)	2	$p=.04$	$p=.04$	No	No	Yes (r_+ =.08)*
	Spanking	4	.12*	(.08, .16)	1	$p>.05$	$p>.05$	No	No	No
	Neg Verbal	3	.14*	(.01, .27)	2	$p>.05$	$p>.05$	No	No	No
Cognitive performance	CP	4	−.18*	(−.13, −.22)	3	$p>.05$	$p>.05$	No	No	No
	Spanking	4	−.09*	(−.06, −.12)	0	$p>.05$	$p>.05$	No	No	No

Discipline = discipline strategy; k = number of independent observations; r_+ = pooled partial correlation coefficient; C.I. = confidence intervals; OFSN = Orwin's Fail-safe N; RCT = significance of Begg & Mazumdar's rank correlation test; RT = significance of Egger's Regression; SSE = small study effect; Neg verbal = Negative verbalizations; arbitrary = arbitrary and inconsistent discipline; Positive = positive discipline; study clusters with less than 3 studies were not analyzed; trim and fill analyses were non-significant in all cases, with the exception of CP on Internalizing behaviors where the adjusted value is presented in parentheses.
* $p<05$.

cognitive performance. Harsh verbal punishment had mean *pr*s of.08 with externalizing and arbitrary discipline had a mean *pr* of .08 with externalizing problems. The differences in *pr*s for externalizing and other outcomes might be explained by the fact that time-1 externalizing problems control both for pre-existing outcome differences and for the selection process by which parents resort to disciplinary punishments, e.g., due to child oppositionalism. In contrast, time 1 scores on the other outcomes control for their pre-existing differences but not for the process for selecting disciplinary punishments. Steiner et al. (2010) showed that both types of controls are necessary to get unbiased causal effects. Because preexisting externalizing problems could provoke greater spanking/CP among parents, it could be argued that time 1 externalizing problems is an important control variable for all outcomes, not only for time 2 externalizing problems. Too few studies provided sufficient data to examine this, but it seems plausible that controlling for time 1 externalizing problems for all negative outcomes could further reduce *pr* coefficients to the trivial range. In any case, the current results do suggest that controlling for time 1 outcomes reduces the relationship between spanking/CP and negative outcomes.

10. Publication bias analyses

As indicated in Tables 1 and 2, several procedures were employed to test for publication bias in spanking/CP research. Results generally indicated that publication bias was not an issue with the current body of data. In the few cases in which the tandem procedure indicated that publication bias may have inflated effect size estimates, this appears to be due to random variation in relatively small groups of studies, rather than systematic publication bias in the research field. Adjustments for Trim and Fill were generally negligible. The only exception was for the bivariate results for CP influences on internalizing disorders wherein the effect size was adjusted to $r=.08$ to account for potentially missing "file drawer" studies. Because the longitudinal studies tended to have large N and thus small confidence intervals, they may be less susceptible to typical publication bias effects, particularly biases more typical of small samples.

11. Moderator analyses

For the effect sizes reported in Tables 1 and 2, tests of heterogeneity were significant in most cases (i.e., significant *Q* values with I^2 values ranging between .66 for CP and .88 for spanking), indicating potential for moderators to account for non-random variation across studies. For *r* effect size, statistically significant *Q* values ranged from 20.53 through 227.76. For *pr* effect size, statistically significant *Q* values ranged from 9.73 through 50.11. Non-significant exceptions for *r* values were for arbitrary discipline on externalizing behaviors ($Q=6.67$, $p=.08$), negative verbalization on internalizing symptoms ($Q=1.80$, $p=.41$), spanking on internalizing symptoms ($Q=3.55$, $p=.31$), spanking on cognitive performance ($Q=7.04$, $p=.07$) and CP on cognitive performance ($Q=3.87$, $p=.28$). Non-significant exceptions for *pr* values were for arbitrary discipline on externalizing symptoms ($Q=.079$, $p=.85$), CP on cognitive performance ($Q=5.83$, $p=.12$) and spanking on internalizing symptoms ($Q=6.02$, $p=.11$). Significant heterogeneity results may be the product of precise effect size estimates from large sample sizes. Given that the majority of studies dealt with externalizing behaviors as their outcome, the moderator analyses were limited to these effect sizes. Results for moderator analyses are presented separately for spanking, CP and positive discipline on externalizing disorders in Table 3. These moderator tests were conducted on the more conservative *pr* estimates.

Table 2
Meta-analytic results for main analysis including publication bias analysis, partial *pr* controlling for time 1 outcome values.

Outcome	Discipline	k	pr_+	95% C.I.	OFSN	RCT	RT	Outlier?	SSE?	Bias?
Externalizing	CP	18	.08*	(.04, .12)	0	$p>.05$	$p>.05$	No	No	No
	Spanking	8	.07*	(.02, .12)	0	$p>.05$	$p>.05$	No	No	No
	Neg verbal	5	.08*	(.03, .14)	0	$p>.05$	$p>.05$	No	No	No
	Arbitrary	4	.08*	(.01, .16)	0	$p>.05$	$p>.05$	No	No	No
	Positive	8	−.04	(−.10, .03)	0	$p>.05$	$p>.05$	No	No	No
Internalizing	Spanking	4	.10*	(.04, .15)	0	$p>.05$	$p>.05$	No	No	No
Cognitive performance	CP	4	−.11*	(−.05, −.17)	1	$p>.05$	$p>.05$	No	No	No

Discipline = discipline strategy; k = number of independent observations; pr_+ = pooled partial correlation coefficient; C.I. = confidence intervals; OFSN = Orwin's fail-safe N; RCT = significance of Begg & Mazumdar's rank correlation test; RT = significance of Egger's Regression; SSE = small study effect; Neg verbal = negative verbalizations; arbitrary = arbitrary and inconsistent discipline; positive = positive discipline; study clusters with less than 3 studies were not analyzed; trim and fill analyses were non-significant in all cases.
* $p<.05$.

In regard to basic statistics for the moderator variables, the average study length was 6.5 years ($SD = 6.7$). The average age of the child participants at time 1 was 6.4 years ($SD = 4.08$). Regarding individual effect size estimates, 73 effect sizes (66%) were source independent, and 21 (19%) were from the NLSY study.

Outcomes for categorical moderators related to source independence or NLSY data were generally non-significant. Particularly in relation to the issue of source independence, this may be because *pr* values were generally and uniformly low. Given the trivial to small effect sizes, the power of moderator analyses related to source independence may have been low.

Continuous moderator variables were analyzed using meta-regression techniques. Age at time 1 significantly moderated the effect sizes of both spanking and positive discipline on externalizing problems. For both moderators, effect sizes were higher for older than for younger children; spanking became more adverse, whereas positive discipline became more beneficial with age. Study length did not impact effect size estimates for any outcome.

The moderating effect of age was further addressed by examining *pr* effect sizes for age categories divided near the mean age (6.4) from the studies. Given that few studies examined participants under age 4 or over 14, this led to age blocks below age 7, those aged 7–11, and those above the age of 11. Effect sizes by age blocks are presented in Table 4. Although age was not found to be a significant moderator in meta-regression for CP, meta-regression can be underpowered and age differences for CP effect sizes are further analyzed here. For spanking, effects on externalizing symptoms were generally trivial in younger children (mean $pr = .06$, n.s.), but non-trivial in children seven and older ($pr = .12$, $p < .05$). Similarly for CP, effects were trivial for children under 11 ($pr = .07$, $p < .05$), but non-trivial for children older than 11 ($pr = .12$, $p < 05$). Results are also presented for positive discipline which had non-significant effect sizes for both younger and older children. These results suggest that for both spanking and CP, effects worsen for older children, although with the caveat that moderator results for age on CP were non-significant in the meta-regression.

12. Discussion

The current study examined the relationship between spanking and CP on externalizing and internalizing problems and on cognitive ability in longitudinal analyses. Previous reports (Baumrind et al., 2002; Gershoff, 2002a, 2002b) appear to agree that reliance on bivariate *r* correlations in analyses may result in inflated effect size estimates. Indeed one of the strengths across the majority of longitudinal designs examined here was the sophisticated effort to control for other relevant variables that might potentially explain a link between spanking/CP and later negative outcomes. Thus, both effects for bivariate *r* and for better-controlled partial *pr* coefficients controlling for time 1 differences on outcome variables were presented.

Consistent with Gershoff (2002a, 2002b) and Paolucci and Violato (2004), small to moderate but non-trivial (i.e., $r > .10$) bivariate correlational relationships were found between spanking/CP and negative outcomes including externalizing and internalizing symptoms and lower cognitive performance. Bivariate outcomes were about midway between those seen in Paolucci and Violato (2004) and Gershoff's (2002a, 2002b) previous meta-analyses, which seemed to be based only on bivariate correlations. Similar effects were found for harsh verbal punishment and arbitrary discipline, but not for positive disciplinary strategies which were not significantly correlated with externalizing symptom outcomes, the only outcome with a sufficient number of studies.

Table 3
Moderator analyses for spanking, CP and positive discipline on externalizing behaviors.

Moderator	Spanking	CP	Positive discipline
Source independence (effect size)			
Independent	.07	.10	−.04
Non-independent	.06	.07	.00
NLSY data (effect size)			
No	.05	N/A	-.05
Yes	.08	.06	
Age, time 1 (meta-regression)	5.47*	0.74	5.31*
Study length (meta-regression)	0.91	2.78	2.67

Effect size = meta-analytic partial *r* effect size reported for subgroups of studies; Meta-regression statistics are Q statistics for model with statistical significance denoted.
* $p < .05$.

Table 4
Differences in spanking and cp effects on externalizing problems across age categories.

Parental discipline	Age category	k	*pr*	95% CI
Spanking	Under 7	5	.06	−.02, .14
	7–11	3	.12*	.04, .19
	Older than 11	0	N/A	N/A
Corporal punishment	Under 7	7	.07*	.02, .12
	7–11	6	.07*	.02, .11
	Older than 11	5	.12*	.00, .24
Positive discipline	Under 7	4	−.03	−.14, .09
	7–11	3	−.07	−.18, .05
	Older than 11	1	N/A	N/A

k = number of studies; *pr* = partial *r* effect size; 95% CI = 95% confidence interval.
* $p < .05$.

Perhaps more critically, however, controlling for pre-existing differences on the outcomes with partial *pr* coefficients reduced the effect sizes between spanking/CP and negative outcomes, although only the effect size of spanking with externalizing problems for 4- to 7-year-olds was reduced to non-significance. All effect sizes for externalizing behaviors were reduced to the trivial range ($pr < .10$), whereas the effect size between spanking and internalizing behavior was at the threshold for triviality ($pr = .10$). Only for cognitive performance did *pr* effect size estimates for CP remain barely above the trivial threshold, an outcome warranting more research.

It is not clear why CP effects on cognitive performance appear to be slightly stronger, even in better controlled studies, compared to other outcomes of spanking and CP. It may be that CP detracts from a nurturing environment under which learning most ideally occurs. It is also possible that externalizing problems may predict both the use of parental CP and lower cognitive performance, although few studies have examined this possibility. It should be noted that the only previous meta-analysis to include this outcome found a trivial effect size of $r = -.03$ between corporal punishment and cognitive performance (Paolucci & Violato, 2004). Thus more research is needed to parse the relationship between CP and cognitive performance.

Overall it is concluded that, when sophisticated well-controlled longitudinal designs are employed, results indicate a trivial to very small significant relationship between spanking and negative outcomes. Nonetheless, it is worth emphasizing that spanking and CP do appear to be significantly associated with small increases in negative outcomes, although these correlations may not be as substantial as sometimes implied in public discussions or scholarly comments on the topic.

The intent is not to disparage scholars or professional groups who have made these well-intentioned claims and arguments. Instead the issue is the comparative lack of guidance for how scholars should use statistically significant but trivial effect sizes. Three issues stand out: trivial effects are more easily accounted for by remaining confounds. Second, the closer the true average causal effect is to zero, the larger the subset of spanking that may actually *reduce* outcomes such as externalizing problems. The third issue is how these trivial but significant effects should be communicated to scholars and the general public.

On the first issue, effect sizes could be reduced even further if other confounding variables were controlled for. Although longitudinal data cannot provide causal evidence as conclusively as randomized designs, their results can approximate unbiased causal evidence more closely by controlling for potential confounds. Controlling only for time 1 outcomes

may leave residual confounds which inflate effect sizes, sometimes referred to as the under-adjustment bias (Campbell & Boruch, 1975). This possibility is demonstrated by studies providing stronger causal evidence. For example, some research suggests that links between spanking/CP (but not more serious abuse) may be accounted for by genetic effects (Jaffee et al., 2004). Stronger causal evidence from propensity score matching (Morris & Gibson, 2011) suggests that causal links between spanking and adverse outcomes may be negligible, except for the most compliant children. Research by Larzelere, Ferrer et al. (2010) further suggests that small but significant links between spanking and negative outcomes become non-significant when more careful controls or gain scores are employed. Researcher hypothesis bias inflation can also account for small effects, particularly in "hot" or politically and ideologically charged research fields (Ioannidis, 2005), although large sample sizes reduce this bias due to their small confidence intervals. In short, it remains a reasonable hypothesis that the small and/or trivial effects found in the present meta-analysis may nonetheless be inflated by remaining methodological confounds.

If biases and residual confounding can account for the remaining trivial effect sizes, this increases the possibility that a subset of spanking may actually decrease problematic outcomes such as externalizing problems. The effect size estimates the *average* causal effect, not the distribution of those effects (Heckman, 2005). If the true average causal effect size is close to zero, then the distribution of causal effects would be split evenly into slightly beneficial and slightly harmful implementations of spanking. Recent research by Lansford, Wager, Bates, Pettit, and Dodge (2012) suggests that parceling out mild from harsh and frequent CP is important, as negative outcomes become even more trivial for mild as opposed to harsh or frequent CP (their mean *pr* was .015). Given that few studies clearly differentiate optimally defined spanking (e.g., one or two open-handed swats for defiance) from less appropriate spanking, CP, or even abuse, it remains possible that the effectiveness of an optimal type of spanking may be obscured by contamination with overlapping variance from CP and physical abuse. On the other hand, it is entirely possible that future research may provide stronger support for zero tolerance of all disciplinary spanking, but the field would benefit from a constructive debate about whether the existing evidence is strong enough to support such an absolute stance currently. Unfortunately the present crop of longitudinal studies had insufficient information to examine these issues with moderator analyses.

13. The interpretation of very small effects

It is worth putting the effect sizes found in this meta-analysis into some perspective. Most of the effect sizes for the controlled *pr* outcomes were less than $r = .10$, the traditional trivial effect size identified by Cohen (1992). This means that disciplinary practices generally accounted for less than 1% of the variance in the outcome variable. For instance, the influence of spanking on later externalizing problems was equivalent to $pr = .07$, or put in terms of shared variance, spanking explains 0.49% of additional variance in externalizing problems after controlling for pre-existing externalizing problems. Bivariate effects were larger, as expected, although the largest of these, for CP on internalizing problems was equivalent to $r = .21$ or 4.4% of the variance explained. When effect sizes are small but "statistically significant" they can be difficult to properly interpret. Perhaps because Cohen (1992) and others (e.g., Ferguson, 2009) have been reluctant to make rigid guidelines some well-respected scholars have argued that even the smallest effect sizes are "practically significant." For instance, several scholars (e.g., Bem & Honorton, 1994; Rosnow & Rosenthal, 2003) have drawn favorable comparisons between small effect sizes in psychological and medical research, where effect sizes for the Salk Vaccine for polio or the Physician's Aspirin Study are near $r = .00$ ($r = .011$ and .03, respectively). Their point is that tiny psychological effects might also be practically important.

However, the statistics upon which such statements are made have been criticized numerous times (Block & Crain, 2007; Crow, 1991; Ferguson, 2009; Hsu, 2004). These criticisms are complex, but one aspect is the large sample size in studies of infrequently occurring problems. For instance, in reference to the Salk vaccine for polio, Ferguson (2009) noted that even if the Salk Vaccine perfectly prevented *every single case* of polio, the calculated effect size using the flawed statistics (phi or Cramer's V) would be a mere $r = .017$. This can be compared to the odds ratio from the initial study, which was a substantial 3.48. Thus the medical effects in these comparisons are being artificially truncated by badly skewed dichotomous outcome variables, which leads to extremely small maximum *r* values that do not reflect the actual impact of these medical treatments. Thus, this line of reasoning is questionable.

14. The practical element: what should psychologists be saying?

The current best evidence suggests that spanking or CP has a small or trivial but statistically significant impact on negative outcomes, at least for externalizing and internalizing symptoms and cognitive performance. The only exception was for spanking children under the age of 7 which was non-significant for externalizing symptoms. Although it is certain that debate on the effects of spanking and CP will continue, it is recommended that scholars remain cautious in nuancing their communications in line with the magnitude of effects and the potential that accounting for additional confounds may result in even more trivial effect sizes. This is not to say that scholars are remiss in communicating concerns about negative outcomes, just that these should be kept in moderation given the trivial to small effect sizes. On the other hand, there was no evidence from the current meta-analysis to indicate that spanking or CP held any particular advantages. There appears, from the current data, to be no reason to believe that spanking/CP holds any benefits related to the current outcomes, in comparison to other forms of discipline. Although effect sizes were all quite small, positive parenting practices had a non-significant relationship with externalizing behaviors and could be considered as "least negative." As such it would be within reason to argue that, although the negative effects of spanking and CP on internalizing/externalizing behaviors and cognitive performance may be small to minimal, spanking and CP conveys no particular advantages for these outcomes at least in the typical ways they are used in available studies.

The small effects seen here for spanking and CP should not be generalized to more severe forms of child abuse. Indeed there are sound reasons to suggest that physical and sexual abuse that causes physical injury, sexual violation or psychopathological symptoms (i.e., post traumatic stress disorder) may be very different in long-term outcomes from spanking and milder forms of CP which were the focus of the current study.

15. Limitations and directions for future research

As with any study, the current meta-analysis is not without limitations, and these should be noted. First, although the use of partial *pr* coefficients to control for time 1 outcome variables represents a more careful scientific approach, not all studies provide sufficient data for the calculation of partial *pr* coefficients (or did not respond affirmatively to requests for additional data). Therefore meta-analyses based on partial *pr* represent a subset of the total longitudinal studies. Second, a meta-analysis is only as good as the studies upon which it relies. The current set of longitudinal studies are reasonably consistent in using well-validated outcome measures, thus avoiding some of the pitfalls of other aggression-related fields (see Ferguson & Kilburn, 2009, for a discussion of measurement problems in media violence research, for instance). However, longitudinal studies are less consistent in employing controls for time 1 child characteristics that may influence the use of spanking/CP. Particularly in the spanking literature, using these control variables decreased resultant effect sizes, highlighting the importance of these controls (Morris & Gibson, 2011). It appears that better controlled studies produce the weakest effects, an issue which should be

kept in mind when interpreting this literature. Third, dissertations using longitudinal data were relatively uncommon, resulting in few included herein. Although publication bias did not appear to be a major issue for the current set of studies, it remains possible that having a broader array of unpublished studies could have influenced results. Next, although the current study focused on spanking and milder forms of CP, it is not always possible to easily divide spanking from CP and CP from more serious physical abuse. Were it possible to more clearly delineate these behaviors from each other, the results of the study may be somewhat different. The current study did take care not to include studies of severe physical abuse and to delineate spanking from CP, but I acknowledge inevitable errors at the margins. Likely, however, such errors would inflate, not deflate effect size estimates. Next, it should be noted that the current study focused only on 3 outcomes, albeit those most often represented in the literature, and cannot be generalized to other potential outcomes, whether positive or negative.

As for future directions, it may be of value to further differentiate spanking and mild CP from more serious forms of abuse in future studies. It may be that harsher forms of violence toward children have much greater adverse effects than milder forms. For instance Jaffee et al. (2004) found that relationships between CP and aggression, but not more serious maltreatment and aggression, could be explained by genetic effects. Similar findings were reported by Lynch et al. (2006). More sophisticated statistical designs such as the use of propensity score matching also suggest that small correlations found between spanking or mild CP and negative outcomes may be spurious, except for the most compliant children (Morris & Gibson, 2011). This topic needs more research designs that can yield stronger causal evidence. More research that compares the child outcomes of spanking and CP to positive parenting practices and other alternatives (e.g., time out) would also be of great value. This research remains in short supply.

It is important to note that the outcomes of spanking, CP and other types of parental punishment are very complex and take place within a broader cultural, family and child–parent attachment relationship than is possible to examine in a single meta-analysis. Indeed too much of the debate on spanking and CP assumes that those practices are either invariably bad or invariably acceptable for children and may miss important nuances about their use. In an atmosphere in which the discussion on a topic such as this is politically charged it may be difficult to address spanking/CP in this more nuanced and complex manner, but our understanding of this phenomenon would benefit from such an approach. For instance the degree of influence of spanking and CP may relate to the context in which the discipline is rendered, including the warmth and general positivity of the family environment. Although some studies have begun to examine such relationships (e.g., Baumrind, Larzelere, & Owens, 2010), these are in relatively short supply. The current results lend some support for this notion, particularly regarding age, as noted for both spanking and CP, effects were more pronounced for older children than for younger. For younger children, the influence of spanking on externalizing problems did not differ statistically from zero. For older children, they were small but significant. Thus it may be mistaken for scholars to consider spanking as either invariably harmful or always acceptable, but its effects may vary by the contexts, age, and manner in which spanking is used.

Research would be particularly welcome addressing these contextual issues regarding spanking. For instance, do the effects of spanking change given the cultural context? Does cultural approval of spanking influence the potential for spanking to have harmful effects on children's outcomes? It would be valuable for future research to consider the degree to which a parent's emotional state during a spanking incident may influence outcomes. For instance does spanking delivered in a calm, pragmatic manner differ from spanking delivered by an angry, frustrated parent? Similarly it may be worth exploring whether parental debriefing and explaining of the incident and consequences may mitigate potential negative effects of spanking. There would be considerable value in examining cultural and family factors as potential moderator variables. At present too few studies provided effect sizes for such analyses in meta-analysis, aside from racial and ethnic groups. As such, it would be prudent for future research to consider other potential cultural moderators of spanking/CP effects. These issues have yet to be carefully explored and may help parse out situations in which spanking is more or less negative. It may also be valuable for more studies to control for time 1 externalizing problems, even when predicting other negative outcomes. Differences in pre-existing problem behaviors among children also deserve attention as potential moderators. Although some studies considered this, mainly as a control issue, too few longitudinal studies presented independent effect sizes for high, moderate and low levels of pre-existing problems to examine moderator influences in meta-analysis. It would be helpful for researchers, in the future, to examine children with differing levels of pre-existing problem behaviors independently as this issue could have important clinical implications, presenting independent effect size estimates for children with differing levels of pre-existing problems.

16. Concluding statements

Results from the current study indicate a trivial to small, but generally significant relationship between the use of spanking and CP and long-term negative outcomes. It is recommended that social scientists take a more conservative approach when discussing the effects of spanking and CP to the general public than has sometimes been the case. That is to say, scholars should take greater care not to exaggerate the magnitude and conclusiveness of the negative consequences of spanking/CP to the general public, particularly when their statements may generalize beyond the evidence. This does not mean that scholars should endorse spanking; there may be reasonable arguments to suggest that spanking confers no particular benefits and thus might easily be replaced with alternative discipline strategies. However, over-generalizing from the data might easily backfire, decreasing the credibility of scholarly statements on parenting research overall. It is hoped that the current study will be a positive contribution to the scholarly debates on spanking and CP effects.

Appendix A. Studies included in the current meta-analysis

Baumrind, D., & Larzelere, R., & Owens, E. (2010). Effects of preschool parents' power assertive patterns and practices on adolescent development. *Parenting: Science and Practice, 10*, 157–201.

Berlin, L., Ispa, J., Fine, M., Malone, P., Brooks-Gunn, J., Brady-Smith, C., et al. (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. http://dx.doi.org/10.1111/j.1467-8624.2009.01341.x.

Bradley, R., Corwyn, R., Burchinal, M., Pipes McAdoo, H., & García Coll, C. (2001). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development*, *72*(6), 1868–1886. http://dx.doi.org/10.1111/1467-8624.t01-1-00383.

Brendgen, M., Vitaro, F., Tremblay, R., & Wanner, B. (2002). Parent and peer effects on delinquency-related violence and dating violence: A test of two mediational models. *Social Development*, *11*(2), 225–244. http://dx.doi.org/10.1111/1467-9507.00196.

Bugental, D., Martorell, G., & Barraza, V. (2003). The hormonal costs of subtle forms of infant maltreatment. *Hormones and Behavior*, *43*(1), 237–244. http://dx.doi.org/10.1016/S0018-506X(02)00008-9.

Cadmus-Romm, D. (2004). Emotional abuse and neglect in adolescence as predictors of psychosocial adjustment in young adulthood. *Dissertation Abstracts International.*

Christie-Mizell, C., Pryor, E., & Grossman, E. (2008). Child depressive symptoms, spanking, and emotional support: Differences between African American and European American youth. *Family*

Relations: An Interdisciplinary Journal of Applied Family Studies, *57*(3), 335–350. http://dx.doi.org/10.1111/j.1741-3729.2008.00504.x.

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*, 32(6), 1065–1072. http://dx.doi.org/10.1037/0012-1649.32.6.1065.

Eamon, M. (2001). Poverty, parenting, peer and neighborhood influences on young adolescent antisocial behavior. *Journal of Social Service Research*, *28*(1), 1–23. http://dx.doi.org/10.1300/J079v28n01_01.

Eron, L. D. (1982). Parent–child interaction, television violence, and aggression of children. *American Psychologist*, 37(2), 197–211. http://dx.doi.org/10.1037/0003-066X.37.2.197.

Eron, L. D., Huesmann, L., & Zelli, A. (1991). The role of parental variables in the learning of aggression. In D. J. Pepler, K. H. Rubin, D. J. Pepler, K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 169–188). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Fine, S., Trentacosta, C., Izard, C., Mostow, A., & Campbell, J. (2004). Anger Perception, Caregivers' Use of Physical Discipline, and Aggression in Children at Risk. *Social Development*, *13*(2), 213–228. http://dx.doi.org/10.1111/j.1467-9507.2004.000264.x.

Foshee, V., Ennett, S., Bauman, K., 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? *The Journal of Early Adolescence*, *25*(3), 317–344. http://dx.doi.org/10.1177/0272431605277307.

Grogan-Kaylor, A. (2005). Corporal Punishment and the Growth Trajectory of Children's Antisocial Behavior. *Child Maltreatment*, *10*(3), 283–292. http://dx.doi.org/10.1177/1077559505277803.

Gunnoe, M., & Mariner, C. (1997). Toward a developmental-contextual model of the effects of parental spanking on children's aggression. *Archives of Pediatrics & Adolescent Medicine*, 151(8), 768–775.

Johannesson, I. (1974). Aggressive behavior among school children related to maternal practices in early childhood. In J. DeWit & W. W. Hartup (Eds.), *Determinants and origins of aggressive behavior* (pp. 347–425). The Hague, the Netherlands: Mouton.

Keiley, M., Howe, T., Dodge, K., Bates, J., & Pettit, G. (2001). The timing of child physical maltreatment: A cross-domain growth analysis of impact on adolescent externalizing and interalizing problems. *Development and Psychopathology*, *13*(4), 891–912.

Kessenich, A. (2006). The impact of parenting practices and early childhood curricula on children's academic achievement and social competence. *Dissertation Abstracts International.*

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158*, 848–856.

Lahey, B., Van Hulle, C., Keenan, K., Rathouz, P., D'Onofrio, B., Rodgers, J., et al. (2008). Temperament and parenting during the first year of life predict future child conduct problems. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology*, *36*(8), 1139–1158. http://dx.doi.org/10.1007/s10802-008-9247-3.

Lansford, J., Criss, M., Pettit, G., Dodge, K., & Bates, J. (2003). Friendship quality, peer group affiliation, and peer antisocial behavior as moderators of the link between negative parenting and adolescent externalizing behavior. *Journal of Research on Adolescence*, *13*(2), 161–184.

Lansford, J., Deater-Deckard, K., Dodge, K., Bates, J., & Pettit, G. (2004). Ethnic differences in the link between physical discipline and later adolescent externalizing behaviors. *Journal of Child Psychology and Psychiatry*, *45*(4), 801–812. http://dx.doi.org/10.1111/j.1469-7610.2004.00273.x.

Larzelere, R., Ferrer, E., Kuhn, B., & Daniela, K. (2010). Differences in causal estimates from longitudinal analyses of residualized versus simple gain scores: Contrasting controls for selection and regression artifacts. *International Journal of Behavioral Development, 34*(2), 180–189.

Larzelere, R., Cox, R., & Smith, G. (2010). Do nonphysical punishments reduce antisocial behavior more than spanking? a comparison using the strongest previous causal evidence against spanking. *BMC Pediatrics, 10*(10),

Lau, A., Litrownik, A., Newton, R., Black, M., & Everson, M. (2006). Factors affecting the link between physical discipline and child externalizing problems in Black and White families. *Journal of Community Psychology*, *34*(1), 89–103. http://dx.doi.org/10.1002/jcop.20085.

Lengua, L. (2008). Anxiousness, frustration, and effortful control as moderators of the relation between parenting and adjustment in middle-childhood. *Social Development*, *17*(3), 554–577. http://dx.doi.org/10.1111/j.1467-9507.2007.00438.x.

Lynam, D., Miller, D., Vachon, D., Loeber, R., & Stouthamer-Loeber, M. (2009). Psychopathy in adolescence predicts official reports of offending in adulthood. *Youth Violence and Juvenile Justice*, *7*(3), 189–207. http://dx.doi.org/10.1177/1541204009333797.

Magee, T. (2004). A longitudinal exploration of the bi-directional relationship between parenting practices and child aggression and the potential moderating effect of race on that relationship. *Dissertation Abstracts International.*

McCord, J. (1988). Parental behavior in the cycle of aggression. *Psychiatry: Journal for the Study of Interpersonal Processes*, 51(1), 14–23.

Morris, S. Z., & Gibson, C. L. (2011). Corporal punishment's influence on children's aggressive and delinquent behavior. *Criminal Justice and Behavior*, 38(8), 818–839. http://dx.doi.org/10.1177/0093854811406070.

Mulvaney, M., & Mebert, C. (2007). Parental corporal punishment predicts behavior problems in early childhood. *Journal of Family Psychology*, *21*(3), 389–397. http://dx.doi.org/10.1037/0893-3200.21.3.389.

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*(4), 443–463. http://dx.doi.org/10.1016/j.adolescence.2004.11.004.

Pardini, D., Lochman, J., & Powell, N. (2007). The development of callous-unemotional traits and antisocial behavior in children: Are there shared and/or unique predictors? *Journal of Clinical Child and Adolescent Psychology*, *36*(3), 319–333.

Pardini, D., Fite, P., & Burke, J. (2008). Bidirectional associations between parenting practices and conduct problems in boys from childhood to adolescence: The moderating effect of age and African-American ethnicity. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology*, *36*(5), 647–662. http://dx.doi.org/10.1007/s10802-007-9162-z.

Pardini, D. (2002). The interaction between childhood temperament and parenting practices in predicting adolescent substance abuse. Unpublished doctoral dissertation.

Pettit, G. S., Bates, J. E., & Dodge, K. A. (1997). Supportive parenting, ecological context, and children's adjustment: A seven-year longitudinal study. *Child Development*, 68(5), 908–923.

Schmitz, M. (2003). Influences of Race and Family Environment on Child Hyperactivity and Antisocial Behavior. *Journal of Marriage and Family*, *65*(4), 835–849. http://dx.doi.org/10.1111/j.1741-3737.2003.00835.x.

Sears, R. R. (1961). Relation of early socialization experiences to aggression in middle childhood. *The Journal of Abnormal and Social Psychology*, 63(3), 466–492. http://dx.doi.org/10.1037/h0041249.

Simons, R. L., Lin, K., & Gordon, L. C. (1998). Socialization in the family of origin and male dating violence: A prospective study. *Journal of Marriage & the Family*, 60(2), 467–478. http://dx.doi.org/10.2307/353862.

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. *Infant and Child Development*, *18*(2), 178–194. http://dx.doi.org/10.1002/icd.596.

Stattin, H., Janson, H., Klackenberg-Larsson, I., & Magnusson, D. (1995). Corporal punishment in everyday life: An intergenerational perspective. In J. McCord, J. McCord (Eds.), *Coercion and punishment in long-term perspectives* (pp. 315–347). New York, NY US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527906.019.

Strassberg, Z., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Spanking in the home and children's subsequent aggression toward kindergarten peers. *Development and Psychopathology*, 6(3), 445–461. http://dx.doi.org/10.1017/S0954579400006040.

Straus, M., & Paschall, M. (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*(5), 459–483. http://dx.doi.org/10.1080/10926770903035168.

Straus, M., Sugarman, D., & Giles-Sims, J. (1997). Spanking by parents and subsequent antisocial behavior of children. *Archives of Pediatrics & Adolescent Medicine*, 151(8), 761–767.

Wager, L. (2009). Racial differences in the relationship between child externalizing and corporal punishment: The role of other discipline strategies. *Dissertation Abstracts International*.

Yu, X. (2005). The relation of family influences and personal inhibition to childhood aggression in China and in the Unites States: A cross-cultural comparison. *Dissertation Abstracts International*.

Appendix B. Studies characteristics from the current meta-analysis

Study Name	Discipline type	Outcome	*r* (bivar)	Partial *r*	Source ind.	Age at time 1	NLSY	Length	Ethnicity
Baumrind (2010)	Arbitrary	Cognitive	−0.23	−0.1	Yes	4.5	No	10.6	Mixed
N = 87	Arbitrary	Extern	0.09	0.02	Yes	4.5	No	10.6	Mixed
	Arbitrary	Intern	0.37	0.25	Yes	4.5	No	10.6	Mixed
	CP	Cognitive	−0.19	−0.18	Yes	4.5	No	10.6	Mixed
	CP	Extern	0.02	−0.01	Yes	4.5	No	10.6	Mixed
	CP	Intern	0.23	0.18	Yes	4.5	No	10.6	Mixed
	Spank	Cognitive	−0.09	−0.03	Yes	4.5	No	10.6	Mixed
	Spank	Extern	−0.02	−0.11	Yes	4.5	No	10.6	Mixed
	Spank	Intern	0.15	0.17	Yes	4.5	No	10.6	Mixed
	Verbal	Cognitive	−0.22	−0.14	Yes	4.5	No	10.6	Mixed
	Verbal	Extern	0.02	−0.05	Yes	4.5	No	10.6	Mixed
	Verbal	Intern	0.23	0.3	Yes	4.5	No	10.6	Mixed
Berlin (2009)	Spank	Cognitive	−0.08	−0.08	Yes	1	No	2	Mixed
N = 2573	Spank	Extern	0.07	0.009	No	1	No	2	Mixed
	Verbal	Cognitive	−0.08	−0.06	Yes	1	No	2	Mixed
	Verbal	Extern	0.09	0.12	No	1	No	2	Mixed
Bradley (2001)	Spank	Cognitive	−0.07	N/A	No	0	No	13	Mixed
N = 5310	Spank	Extern	0.22	N/A	No	0	No	13	Mixed
Brendgen (2002)	CP	Extern	0.25	0.24	No	12	No	5	Cau
N = 336									
Christie-Mitzel (2008)	Spank	Intern	0.08	0.05	No	14	Yes	2	Cau
N = 1852	Spank	Intern	0.12	0.08	No	14	Yes	2	Af
Bugental (2003)	CP	Intern	0.53	N/A	No	1.5	No	1	Mixed
N = 44	Verbal	Intern	−0.02	N/A	No	1.5	No	1	Mixed
Eamon (2001)	Positive	Extern	−0.16	0.02	No	11	Yes	2	Mixed
N = 963	Spank	Extern	0.18	0.04	No	11	Yes	2	Mixed
Fine (2004)	CP	Extern	0.22	0.17	Yes	4.9	No	6	Mixed
N = 87									
Foshee (2005)	CP	Extern	0.12	0.13	No	13.8	No	1.5	Af
N = 1218	CP	Extern	−0.03	−0.04	No	13.8	No	1.5	Cau
Grogan (2005)	Positive	Extern	−0.03	N/A	No	8.5	Yes	10	Cau
N = 6912	Spank	Extern	0.17	N/A	No	8.5	Yes	10	Cau
	Positive	Extern	−0.03	N/A	No	8.5	Yes	10	Af
	Spank	Extern	0.17	N/A	No	8.5	Yes	10	Af
	Positive	Extern	−0.03	N/A	No	8.5	Yes	10	Hisp
	Spank	Extern	0.17	N/A	No	8.5	Yes	10	Hisp
Yu (2005)	CP	Extern	0.17	N/A	Yes	9.5	No	3.5	Cau
N = 246	CP	Extern	0.09	N/A	Yes	9.5	No	3.5	Asian
	Verbal	Extern	0.28	N/A	Yes	9.5	No	3.5	Cau
	Verbal	Extern	0.02	N/A	Yes	9.5	No	3.5	Asian
	Positive	Extern	0.05	N/A	Yes	9.5	No	3.5	Cau
	Positive	Extern	0.01	N/A	Yes	9.5	No	3.5	Asian
	Spank	Extern	0.2	N/A	Yes	9.5	No	3.5	Cau
	Spank	Extern	0.07	N/A	Yes	9.5	No	3.5	Asian
	Positive	Extern	0.14	N/A	Yes	9.5	No	3.5	Cau
	Positive	Extern	0.1	N/A	Yes	9.5	No	3.5	Asian
Keiley (2001)	CP	Extern	0.25	N/A	Yes	5	No	9	Mixed
N = 578	CP	Intern	0.05	N/A	Yes	5	No	9	Mixed
Kessenich (2006)	Positive	Cognitive	0.04	N/A	Yes	5	No	4	Mixed
N = 13698	Positive	Extern	−0.08	N/A	Yes	5	No	4	Mixed
	Spank	Cognitive	−0.11	N/A	Yes	5	No	4	Mixed
	Spank	Extern	0.08	N/A	Yes	5	No	4	Mixed
Lahey (2008)	CP	Extern	0.19	N/A	No	0.5	Yes	8.5	Mixed
N = 1863									
Lansford (2003)	CP	Extern	0.3	0.23	Yes	12	No	1.5	Mixed
N = 362									

Appendix B (*continued*)

Study Name	Discipline type	Outcome	*r* (bivar)	Partial *r*	Source ind.	Age at time 1	NLSY	Length	Ethnicity
Lansford (2004)	CP	Extern	0.17	0.13	Yes	5	No	11	Cau
N = 453	CP	Extern	−0.05	−0.1	Yes	5	No	11	Af
Larzelere (2010)	CP	Extern	0.21	0.07	No	5	No	6	Mixed
N = 1464	Verbal	Extern	0.22	0.06	No	5	No	6	Mixed
Larzelere/Cox (2010)	Spank	Extern	0.27	0.16	No	7.5	Yes	2	Mixed
N = 785									
Lau (2006)	CP	Extern	0.26	N/A	No	4	No	4	Mixed
N = 442									
Lengua (2008)	Arbitrary	Extern	0.31	0.11	Yes	9.5	No	1	Mixed
N = 188	Arbitrary	Intern	0.18	−0.02	Yes	9.5	No	1	Mixed
	CP	Extern	0.44	0.11	Yes	9.5	No	1	Mixed
	CP	Intern	0.3	0.01	Yes	9.5	No	1	Mixed
Lyman (2009)	Arbitrary	Extern	0.1	0.07	Yes	13	No	9	Mixed
N = 338	CP	Extern	0.12	0.06	Yes	13	No	9	Mixed
	Positive	Extern	0.045	0.01	Yes	13	No	9	Mixed
Magee (2002)	CP	Extern	0.23	0.17	Yes	9.5	No	4	Mixed
N = 126									
Mulvaney (2007)	Positive	Extern	−0.2	−0.14	No	1.26	No	5	Mixed
N = 979	Spank	Extern	0.2	0.16	No	1.26	No	5	Mixed
	Positive	Intern	−0.23	−0.18	No	1.26	No	5	Mixed
	Spank	Intern	0.16	0.15	No	1.26	No	5	Mixed
Oyserman (2005)	CP	Extern	0	N/A	Yes	14.5	No	5.5	Mixed
N = 164	Verbal	Extern	0.18	N/A	Yes	14.5	No	5.5	Mixed
Pardini (2007)	Arbitrary	Extern	0.22	0.13	Yes	10.66	No	1	Mixed
N = 120	CP	Extern	0.27	0.07	Yes	10.66	No	1	Mixed
	Positive	Extern	0	−0.13	Yes	10.66	No	1	Mixed
Pardini (2008)	CP	Extern	0.14	0.09	Yes	7.7	No	10	Mixed
N = 1517									
Pardini (2002)	CP	Intern	0.05	N/A	Yes	10.35	No	5	Mixed
N = 243									
Cadmus-Romm (2004)	Verbal	Intern	0.12	N/A	No	12	No	9	Mixed
N = 87									
Schmidtz (2003)	Positive	Extern	−0.07	N/A	No	7	Yes	6	Af
N = 2307	Spank	Extern	0.22	N/A	No	7	Yes	6	Af
	Positive	Extern	0.17	N/A	No	7	Yes	6	Cau
	Spank	Extern	−0.01	N/A	No	7	Yes	6	Cau
	Positive	Extern	−0.11	N/A	No	7	Yes	6	Hisp
	Spank	Extern	−0.09	N/A	No	7	Yes	6	Hisp
Stacks (2009)	Spank	Extern	0.05	0	Yes	1.2	No	2	Mixed
N = 2792									
Wager (2009)	Positive	Extern	0.27	0.1	Yes	5	No	3	Mixed
N = 574	Spank	Extern	0.34	0.17	Yes	5	No	3	Mixed
	Verbal	Extern	0.25	0.14	Yes	5	No	3	Mixed
Straus (2009)	CP	Cognitive	−0.12	−0.17	Yes	3	Yes	4	Mixed
N = 1510	CP	Cognitive	−0.21	−0.09	Yes	7	Yes	4	Mixed
Deater Decard	CP	Extern	0.31	N/A	Yes	3	No	4	Cau
N = 460	CP	Extern	−0.07	N/A	Yes	3	No	4	Af
Eron (1982)	CP	Extern	0.21	N/A	Yes	7	No	2	Mixed
N = 505									
Eron (1991)	CP	Extern	0.12	N/A	Yes	8	No	22	Mixed
N = 535									
Gunnoe (1997)	Spank	Extern	0.17	0.05	Yes	7.5	No	5	Mixed
N = 1112	Positive	Extern	−0.13	−0.12	Yes	7.5	No	5	Mixed
Johannesson (1974)	CP	Extern	0.02	N/A	Yes	0.1	No	13	Cau
N = 212									
McCord (1988)	CP	Extern	0.28	N/A	Yes	10.5	No	36.5	Mixed
N = 130									
Morris (2011)	CP	Extern	0.21	0	No	9	No	2.5	Mixed
N = 1346									
Pettit (1997)	CP	Extern	0.17	0.11	Yes	4	No	7	Mixed
N = 585	CP	Cognitive	−0.2	−0.05	Yes	4	No	7	Mixed
	Positive	Extern	−0.17	−0.02	Yes	4	No	7	Mixed
	Positive	Cognitive	0.27	0.13	Yes	4	No	7	Mixed
Sears (1961)	CP	Extern	−0.07	−0.06	Yes	5	No	6	Mixed
N = 160	Positive	Extern	−0.05	−0.04	Yes	5	No	6	Mixed
	Verbal	Extern	−0.06	−0.05	Yes	5	No	6	Mixed
Simons (1998)	CP	Extern	0.19	N/A	Yes	12	No	5	Mixed
N = 113	Positive	Extern	−0.19	N/A	Yes	12	No	5	Mixed
Stattin (1995)	CP	Extern	0.42	N/A	No	1	No	25	Mixed
N = 212									
Strassberg (1994)	Spank	Extern	0.12	N/A	Yes	5	No	0.5	Mixed
N = 273									
Straus (1997)	CP	Extern	0.2	0.07	No	7.5	Yes	2	Mixed
N = 807									

Note: CP = corporal punishment; Cau = Caucasian, Af = African American; Hisp = Hispanic.

References

American Academy of Pediatrics (1998). Guidance for effective discipline. *Pediatrics*, *101*(4), 723–728.

Baker, P. C., Keck, C. K., Mott, F. L., & Quinlan, S. V. (1993). *NLSY child handbook: A guide to the 1986–1990 National Longitudinal Survey of Youth Child Data* (Revised ed.). Columbus: Ohio State University, Center for Human Resource Research.

Baumrind, D., Larzelere, R., & Cowan, P. (2002). Ordinary physical punishment: Is it harmful? Comment on Gershoff (2002). *Psychological Bulletin*, *128*(4), 580–589, http://dx.doi.org/10.1037/0033-2909.128.4.580.

Baumrind, D., Larzelere, R., & Owens, E. (2010). Effects of preschool parents' power assertive patterns and practices on adolescent development. *Parenting: Science and Practice*, *10*, 157–201.

Begg, C., & Mazumdar, G. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088–1101.

Bem, D., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*, 4–18.

Berlin, L., Ispa, J., Fine, M., Malone, P., Brooks-Gunn, J., Brady-Smith, C., et al. (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, http://dx.doi.org/10.1111/j.1467-8624.2009.01341.x.

Block, J., & Crain, B. (2007). Omissions and errors in 'Media violence and the American public. *The American Psychologist*, *62*, 252–253.

Buemi, S. (2009). *How race-gender status affects the relationship between spanking and depressive symptoms among children and adolescents*. (Unpublished doctoral dissertation.) Kent State University, Ohio.

Campbell, D. T., & Boruch, R. F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett, & A. A. Lumsdaine (Eds.), *Evaluation and experiment: Some critical issues in assessing social programs* (pp. 195–296). New York: Academic Press.

Christie-Mizell, C., Pryor, E., & Grossman, E. (2008). Child depressive symptoms, spanking, and emotional support: Differences between African American and European American youth. *Family Relations: An Interdisciplinary Journal of Applied Family Studies*, *57*(3), 335–350, http://dx.doi.org/10.1111/j.1741-3729.2008.00504.x.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159, http://dx.doi.org/10.1037/0033-2909.112.1.155.

Crow, E. (1991). Response to Rosenthal's comment, "How are we doing in soft psychology?". *The American Psychologist*, *46*, 1083.

Duvall, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–99.

Eamon, M. (2001). Poverty, parenting, peer and neighborhood influences on young adolescent antisocial behavior. *Journal of Social Service Research*, *28*(1), 1–23, http://dx.doi.org/10.1300/J079v28n01_01.

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.

Egger, M., & Smith, G. (1998). Meta-analysis: Bias in location and selection of studies. *British Medical Journal*, *7124* [Retrieved 10/16/09 from: http://www.bmj.com/archive/7124/7124ed2.htm].

Farley, J. U., Lehmann, D. R., & Sawyer, A. (1995). Empirical marketing generalization using meta-analysis. *Marketing Science*, *14*, G36–G46.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling and implications for the use of meta-analyses. *Psychological Methods*, *17*(1), 120–128.

Ferguson, C. J., & Kilburn, J. (2009). The Public health risks of media violence: A meta-analytic review. *The Journal of Pediatrics*, *154*(5), 759–763.

Friedman, S., & Schonberg, S. K. (1996). Consensus statements. *Pediatrics*, *98*, 853.

Gershoff, E. (2002a). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, *128*(4), 539–579, http://dx.doi.org/10.1037/0033-2909.128.4.539.

Gershoff, E. (2002b). Corporal punishment, physical abuse, and the burden of proof: Reply to Baumrind, Larzelere, and Cowan (2002), Holden (2002), and Parke (2002). *Psychological Bulletin*, *128*(4), 602–611, http://dx.doi.org/10.1037/0033-2909.128.4.602.

GITEACPOC (2012). States with full abolition. Retrieved 5/30/12 from. http://www.endcorporalpunishment.org/pages/progress/prohib_states.html.

Grogan-Kaylor, A. (2004). The effect of corporal punishment on antisocial behavior in children. *Social Work Research*, *28*(3), 153–162.

Gunnoe, M., Hetherington, E., & Reiss, D. (2006). Differential impact of fathers' authoritarian parenting on early adolescent adjustment in conservative protestant versus other families. *Journal of Family Psychology*, *20*(4), 589–596, http://dx.doi.org/10.1037/0893-3200.20.4.589.

Gunnoe, M., & Mariner, C. (1997). Toward a developmental–contextual model of the effects of parental spanking on children's aggression. *Archives of Pediatrics & Adolescent Medicine*, *151*(8), 768–775.

Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology*, *35*, 1–97, http://dx.doi.org/10.1111/j.0081-1750.2006.00164.x.

Hsu, L. M. (2004). Biases of success rate differences shown in binomial effect size displays. *Psychological Bulletin*, *9*(2), 183–197.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124 Retrieved 7/14/09 from: http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124.

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*(6), 1047–1058, http://dx.doi.org/10.1037/0012-1649.40.6.1047.

Kazdin, A. (2008). Spare the rod: Why you shouldn't hit your kids. Slate. Retrieved 9/13/10 from: http://www.slate.com/id/2200450/?from=rss.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.

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: An Interdisciplinary Journal of Applied Family Studies*, *61*(2), 224–236, http://dx.doi.org/10.1111/j.1741-3729.2011.00700.x.

Larzelere, R. (2008). Disciplinary spanking: The scientific evidence. *Journal of Developmental and Behavioral Pediatrics*, *29*(4), 334–335, http://dx.doi.org/10.1097/DBP.0b013e3181829f30.

Larzelere, R., Cox, R., & Smith, G. (2010). Do nonphysical punishments reduce antisocial behavior more than spanking? A comparison using the strongest previous causal evidence against spanking. *BMC Pediatrics*, *10*(10).

Larzelere, R., Ferrer, E., Kuhn, B., & Daniela, K. (2010). Differences in causal estimates from longitudinal analyses of residualized versus simple gain scores: Contrasting controls for selection and regression artifacts. *International Journal of Behavioral Development*, *34*(2), 180–189.

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.

Lau, A., Litrownik, A., Newton, R., Black, M., & Everson, M. (2006). Factors affecting the link between physical discipline and child externalizing problems in Black and White families. *Journal of Community Psychology*, *34*(1), 89–103, http://dx.doi.org/10.1002/jcop. 20085.

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*(2), 190–198, http://dx.doi.org/10.1037/0893-3200.20.2.190.

Macaskill, P., Walter, S., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, *20*, 641–654.

McLoyd, V., & 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*, *64*(1), 40–53, http://dx.doi.org/10.1111/j.1741-3737.2002.00040.x.

Morris, S. Z., & Gibson, C. L. (2011). Corporal punishment's influence on children's aggressive and delinquent behavior. *Criminal Justice and Behavior*, *38*(8), 818–839, http://dx.doi.org/10.1177/0093854811406070.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159, http://dx.doi.org/10.2307/1164923.

Paolucci, E., & Violato, C. (2004). A meta-analysis of the published research on the affective, cognitive, and behavioral effects of corporal punishment. *Journal of Psychology: Interdisciplinary and Applied*, *138*(3), 197–221, http://dx.doi.org/10.3200/JRLP.138.3.197-222.

Peterson, R., & Brown, S. (2005). On the use of beta coefficients in meta-analyses. *The Journal of Applied Psychology*, *90*(1), 175–181.

Raju, N. S., Fralicx, R., & Steinhaus, S. D. (1986). Covariance and regression slope models for studying validity generalization. *Applied Psychological Measurement*, *10*, 195–211.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641, http://dx.doi.org/10.1037/0033-2909.86.3.638.

Rosenthal, R., & DiMatteo, M. (2001). Meta analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59–82.

Rosnow, R., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, *57*, 221–237.

Savage, J., & Yancey, C. (2008). The effects of media violence exposure on criminal aggression: A meta-analysis. *Criminal Justice and Behavior*, *35*, 1123–1136.

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. *Infant and Child Development*, *18*(2), 178–194, http://dx.doi.org/10.1002/icd.596.

Starr, L., & Davila, J. (2008). Excessive reassurance seeking, depression, and interpersonal rejection: A meta-analytic review. *Journal of Abnormal Psychology*, *117*(4), 762–775, http://dx.doi.org/10.1037/a0013866.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 250–267.

Straus, M. (2008). The special issue on prevention of violence ignores the primordial violence. *Journal of Interpersonal Violence*, *23*(9), 1314–1320, http://dx.doi.org/10.1177/0886260508322266.

Straus, M., & Paschall, M. (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*(5), 459–483, http://dx.doi.org/10.1080/10926770903035168.

Taylor, C. A., Lee, S. J., Guterman, N. B., & Rice, J. C. (2010). Use of spanking for 3-year-old children and associated intimate partner aggression or violence. *Pediatrics*, *126*(3), 415–424, http://dx.doi.org/10.1542/peds.2010-0314.