Author's response to reviews

Title: Do Nonphysical Punishments Also Increase Antisocial Behavior? A Replication of the Strongest Causal Evidence Against Customary Spanking

Authors:

Robert E Larzelere (Robert.Larzelere@okstate.edu)
Ronald B Cox Jr (r.cox@okstate.edu)
Gail L Smith (smithg@girlsandboystown.org)

Version: 3 Date: 11 September 2009

Author's response to reviews: see over
Responses to Reviewers’ Comments on the First Draft  
Robert E. Larzelere and Ronald B. Cox, Jr.

We thank both reviewers for considering our manuscript as important for the field and suggesting several ways to improve it and to circumscribe its major conclusions. The paper has two major inter-related purposes, to find alternative disciplinary tactics that do a better job of reducing antisocial behavior than spanking and to test whether the effects might be artifactual due to residual confounding. The comments focused only the latter purpose and its implications for the causal evidence against customary spanking. This suggests that we emphasized the latter purpose too much in the original submission. To correct this imbalance, we have emphasized the first purpose more in the revised manuscript. If we would have found that some of the nonphysical punishments significantly reduced antisocial behavior in these analyses, we would not have needed to emphasize an explanation for the similarity of these results, such as residual confounding.

The following considers the comments of Natalie Pafitis, the action editor, and then those by Drs. Capaldi and Jolly.

Responses to Comments by Dr. Natalie Pafitis

1. Ensure that your conclusions are in line with the data and tone down any claims made in reference to the original article.

We have tried to do a better job of making appropriate conclusions without making statements that go beyond our data. We are not sure where we needed to tone down our claims about the original article. We now introduce the article by Straus et al. (1997) as a seminal study that was the first to show the causally strongest evidence against customary spanking. We agree with your suggestion that he and his colleagues should have an opportunity to respond to our study. We also deleted a paragraph about their Figure, which seems to show a larger effect size than any of the statistically controlled longitudinal studies by others or by our 10 close replications of their data. That information is now only in footnote c to Table 4.

Responses to Comments by Dr. Deborah Capaldi

1. The study seems to be a re-analysis rather than a replication.  
Replications usually involve a different data set, or at least different waves of the same data set. This study involves the same two waves of the same data set in the Straus et al. study.

Straus et al. (1997) featured the cohort with the strongest longitudinal effect sizes out of 5 cohorts that they considered. If we replicated their results using another data set or another cohort, then we would most likely be selecting a data set with a weaker effect of spanking on subsequent antisocial behavior. Accordingly, most replications of similar net-effects regression (i.e., predicting subsequent behavior problems from spanking frequency, controlling statistically for Wave-1 scores on the same behavior problems) obtained smaller effect sizes that usually do not achieve the same consistently significant effects for all subsamples that were obtained by
Straus et al. (1997). Because neither Straus et al. nor any of these replications have conducted similar analyses for alternative disciplinary tactics that parents could use instead of spanking, we wanted to compare the effects of alternative tactics against a strong test of the detrimental effects of spanking frequency, not a watered-down test, which would be likely for replications in other cohorts. Additionally, a major objective was to address the methodological issues present in the Straus et al. (1997) study that need to be improved in future research on this important topic. Therefore we felt it important to duplicate Straus et al. as closely as possible.

2. I draw rather different conclusions from the current study than the authors - mostly that maternal reports of discipline, particularly as assessed in this study, are weak. It has been observed many times in the literature that parents are often poor reporters of discipline practices - especially of harsh discipline practices. Also, the survey only asked about maternal discipline and did not include paternal behavior. Each type of discipline was assessed by just one question covering discipline only for the past week. The reliability of single items is generally poor; thus, the measures of discipline were weak. This seems to be the main conclusion to draw, along with the fact that the findings of the Straus et al. study did not hold up on the modified reanalysis.

We agree with the reviewer’s observation that 1) relying solely on maternal report; 2) using only one question to measure parental use of a disciplinary tactic; and 3) extrapolating data from one week as representative of ongoing practice has its limitations. We retain these limitations, however, because (1) they were used in all the studies with the strongest causal evidence against customary spanking, (2) it is all that is available in this widely used NLSY data set, (3) and multiple items would not allow us to compare the outcomes of specific alternative disciplinary tactics with spanking. We added a section acknowledging these limitations as a possible reason for non-significant effects after improving the measure of initial differences in antisocial behavior or externalizing behavior problems. We also tried to clarify the major purpose of this study, which is to replicate the strongest causal evidence against nonabusive spanking by repeating the same analyses on the same sample to determine (1) what alternative disciplinary tactics produce beneficial or less detrimental effects using the same procedures and (2) whether that causal evidence can be accounted for by residual confounding.

3. Another major issue that was not clear why from a sample of 7,725 there were only 807 families in the Straus study and 785 in the current study. It is not clear from the description on pages 9-10 how many were excluded due to planned subsample design features, such as randomly selecting one child from each family, and how many were lost due to attrition between the time points, missing data, and so forth. This is an important issue in interpreting the findings as if there was large data loss from nonplanned factors that could affect the findings - e.g., more abusive parents may have more missing data.

We have now included a more complete and accurate description of the number of women in the original study and the subset who had children who qualified with valid data for the original study, and therefore our replication.

4. The most major issue with the study is the overall approach and conclusions
regarding spanking. There are moral reasons for opposing physical punishment - it is very different from other disciplines in that it involves hitting. Parents who hit are likely to be at risk for using more severe hitting when they are angry. There may be other very negative consequences to spanking than antisocial behavior - e.g., fear and anxiety, damage to the parent-child relationship. Although spanking is mentioned as an effective back up, it is not clear that the evidence is strong for that. The implication seems to be that children will always comply once spanked. That is far from the case - children can become very upset and defiant to physical discipline - and what should the parent conclude is the back up to that? To hit harder? Also note that the current study only involves children who were 6-9 years of age at the first wave. Thus effects on children of other ages are not known.

The fact is that physical discipline is not necessary as there are effective alternatives. Statements such as regarding Sweden's smacking ban and the large increase in criminal assaults by minors are hard to interpret. Usually such increases involve changes in crime definitions, arrest policies, police coverage, and so forth. If the smacking ban was replaced by no discipline, that is obviously a problem. It does not mean that smacking should be reinstated.

We have attempted to make our point more clear in this revision. We agree that there are moral consideration and other empirical findings that go beyond the scope of the current manuscript whose purpose is not to critique the morality of corporal punishment, but to address and critique strongest type of evidence that customary spanking causes adverse child outcomes. That being said, the current situation is that some countries have imposed absolute bans on all spanking, including its mildest forms and others are considering whether or not to do so. If the case against mild spanking is based only on moral issues, then the field needs to consider when moral arguments are sufficiently convincing to warrant imposing one group’s values upon entire societies, including important subgroups who hold differing values concerning nonabusive spanking. When do we tolerate pro-choice positions regarding parental discipline and when do other moral issues become more important than parental liberty to raise their own children as they deem best? Certainly child abuse is an example where the welfare of the child takes priority over parental liberty, but when does it become inappropriate to impose one group’s moral views about parental discipline on all parents? In the current manuscript we do not attempt to address any of these questions, leaving them for a different venue. However, we do agree with the reviewer that there are multiple considerations involved in determining what disciplinary tactics most benefit a child. We explicitly cite the relevance of moral considerations in our new concluding paragraph.

In regards to the reviewers’ other comments that suggest that there are other reasons that spanking may be counterproductive, we agree that there is certainly more to consider than the replication of one study before making decisions on this important issue. We would agree that recommendations and practices should be based on relevant scientific empirical evidence as well as relevant moral considerations. Again, however, we believe that the empirical issues raised by the reviewer go beyond the scope and the intent of the current manuscript. To address all of these issues adequately would require a much longer and less-focused article.
One important point of our paper is that we need to investigate the effects of alternative disciplinary enforcements using similar methods to be able to compare their effectiveness to corporal punishment. This is necessary to move the field forward and to inform parents about what to use instead of spanking. Except for overly severe and predominant use of physical punishment, our meta-analysis could locate no study showing that spanking resulted in more detrimental effects than alternative disciplinary tactics. This leads to our conclusion that more research with more discriminating causal evidence is needed to resolve the issue.

Although the purpose of the manuscript does not allow for a full discussion of the points the reviewer raises, we offer the following in the spirit of scientific debate.

*Parents who hit more are likely to be at risk for using more severe hitting when they are angry.* The broader issue is how to prevent escalations in parental frustrations and anger within disciplinary incidents. This is a very important hypothesis, but we know of only correlational data on it. For example, Gershoff’s (2002) meta-analysis includes 10 studies that show an overall positive association between corporal punishment and child abuse, but 9 of those studies use cross-sectional correlations and the 10th study uses correlations between the same time periods, based on retrospective recall. Behavioral parent training trains parents to use mild nonphysical punishments, such as time out, consistently and contingently. If parents can use that successfully, then this provides a tool to prevent escalations of anger to levels that increase the risk of abuse. With young clinically defiant children, however, Mark Roberts showed that time out must be enforced with a more forceful tactic before many of these children will cooperate with time out. Although a 2-swat spank was used for that purpose by most behavioral parent training programs for 2- to 6-year-olds for 25 years, those programs now generally use the forceful brief room isolation, which was the only alternative demonstrated to be equally effective for that purpose in Roberts’s studies. Wherever spanking is prohibited, it is important to identify effective enforcement tactics for time out, so that defiant young children will cooperate with time out, thereby preventing escalations toward levels of frustration and anger that put the child at risk for abuse. Otherwise, parents are at risk of concluding that no type of punishment is appropriate and therefore may be at risk for the permissive type of coercive process found in Norway by experts such as Marion Forgatch and Patterson and Fisher (2002, p. 74). Our review of the literature in 1999 found that Sweden’s first spanking reduced mild spanking more than more problematic types of physical punishment (e.g., physically punishing adolescents or spanking when upset enough), which might account for the fact that, if anything, physical child abuse had increased in the 15 years after the spanking ban and/or occurred at higher rates than in the USA (e.g., 3% of Swedish parents reporting on an anonymous survey that they had beaten up their child in the year following the spanking ban, compared to substantially lower rates in the USA: Gelles & Edfeldt, 1986). The point is that the best way to prevent escalating frustration during discipline episodes is to teach parents how to use mild disciplinary enforcements effectively, but that requires more forceful back-ups for some young children.

*There may be other negative consequences to spanking.* Other outcome variables are beyond the scope of our study. Evidence about mental health (e.g., anxiety) and parent-child relationship quality in Gershoff’s (2002) meta-analysis is based entirely on correlations, from cross-sectional (91% of child-outcome studies for mental health, 50% for parent-child relationships) or retrospective studies (100% of adult-outcome studies and the remaining child-outcome studies).
Evidence for spanking as effective back-up may be weak. The four studies by Roberts and his colleagues are the only randomized clinical trials of spanking, limited to its traditional use in behavioral parent training to enforcement of compliance with time out in clinically defiant 2- to 6-year-old children. On any other topic, it would be inconsistent to call the results of four replicated randomized clinical trials using observational data weak while basing objections on other literature that is predominantly, if not entirely, correlational from cross-sectional and retrospective verbal reports.

What if children respond defiantly to physical discipline? Persistent defiance by children is a crucial problem regardless of the disciplinary tactic is used, especially for last-resort tactics after preferred milder disciplinary tactics fail to work to the parent’s satisfaction. One study in Quebec found that as spanking falls out of favor, parents use more verbal hostility (Clement & Chamberland, 2007), which has been shown to have more detrimental child effects than even physical child abuse in some studies (Bremner, Vermetten, & Mazure, 2000; Teicher, Samson, Polcari, & McGreenery, 2006; Vissing, Straus, Gelles, & Harrop, 1991). On this issue, we need research that is capable of discriminating between effective vs. counterproductive disciplinary enforcements, including tactics parents can use when milder tactics do not work to their satisfaction. In general, parents need more disciplinary options, along with training and education so that they can use the mildest disciplinary tactic skillfully, thus minimizing the need to resort to more forceful disciplinary tactics. Our re-analysis of Ritchie’s (1999) data indicated that the most effective tactics at putting an immediate stop to defiance in 3-year-olds were, in order, spanking, time out, and physical power assertion. Of these, time out would be preferred whenever possible, but the others become the best options if a young child defiantly refused to cooperate with time out.

This study is only on 6- to 9-year-olds. Age of the child is an important variable, that distinguished between mostly effective vs. counterproductive use of spanking according to a review of the literature by Larzelere (2000). Straus et al. (1997) found weaker evidence against spanking for the 2 younger cohorts they examined. However, data on alternative tactics are available in the NLSY data only for children 6 years of age and older. Personally we think that skillful parents should have phased out their last-resort tactics by that age range. Therefore, we expect that it is more likely to find that age-appropriate nonphysical punishments are more effective than spanking for 6- to 9-year-olds than for younger children. So our particular test compares child outcomes of spanking vs. alternative tactics for ages where spanking should be more detrimental than at younger ages (e.g., 2 to 6 years old).

Physical discipline is not necessary because there are effective alternatives. Medical practice is enhanced by having multiple options for prescription drugs, so that a drug can be matched to the particular person and their responsiveness to other drugs. For similar reasons, parents need more disciplinary options, other things being equal. Different disciplinary options are each better suited for distinct situations for the same child and their effectiveness also varies between children. Our meta-analysis showed that nonabusive spanking was associated with better child outcomes than 10 of 13 alternative tactics it has been compared to when it was used to respond to defiant refusal to cooperate with milder disciplinary tactics, such as time out, in 2- to 6-year-olds. Only a brief forced room isolation was equally effective as spanking in those situations in more
than one study, in this case for enforcing compliance with time out. Roberts’s fourth randomized clinical trial (Roberts & Powers, 1990) showed common-sense advantages of having multiple options for parents to enforce time out. The spank back-up and the forceful room isolation each worked for parents when the other one was slow to be effective. Second, the study gave parents their preference for a back-up to use at home, and the majority selected the spank back-up. If the parents followed the psychologist’s recommendation to limit spanking to enforcing time out, their data showed that the majority of parents decreased their use of spanking to zero spans per week by the third week after training. In short, there are advantages for parents to have multiple options, especially for last-resort tactics, so that they can get mild disciplinary tactics to work and thereby phase out the more forceful tactics that may sometimes be necessary to get young clinically defiant children to cooperate with time out.

*Trends in criminal assaults by minors against minors are difficult to interpret.* We acknowledge that those increasing trends might be due to changes in what gets reported to police. At some point, we should see *some* decreases in objective statistics if spanking bans are successful. Ten years ago, my colleagues and I called for more objective, unbiased evaluations of spanking bans. We know of no such evaluations since then. Although the available evidence can be debated from both perspectives, there is no convincing evidence that either child abuse or societal violence has been decreased due to spanking bans. Part of the reason might be that studies have failed to investigate what alternative tactics are at least as effective as spanking in the situations where parents traditionally have used mild spanking. Based on our review of the available literature in 1999, we concluded that spanking bans primarily suppress the kind of mild spanking that can stop escalations toward abuse. This would help account for evidence from multiple types of data sources that show, if anything, increases in physical child abuse and criminal assaults since the first spanking ban in Sweden in 1979. This is also consistent with Marion Forgatch’s informed personal opinion that spanking bans can be iatrogenic, especially when they get over-generalized against all forms of nonphysical punishments as well as physical punishments (see Patterson & Fisher, 2002, p. 74). We are not saying necessarily that spanking should be reinstated, but parents need to know what alternatives are associated with better child outcomes in studies replicating the strongest causal evidence against nonabusive spanking.

5. In the Introduction, residual confounding should be explained more clearly.

We thank Deborah and have the revised the comments on residual confounding in the introduction to increase clarity.

*Responses to Dr. Damien Jolley’s comments:*

1. Problems involving residual confounding. Thank you for agreeing with one of our major points, that fallible statistical controls for crucial confounds are not improved when the measures of those confounds are weaker in validity and reliability.

2. Wild over-interpretaions of the results of the re-analysis. You certainly have a valid point that the estimated effect sizes (betas contrasting zero use of the corrective disciplinary actions vs. using each of them one or more time in the previous week) remain about the same after improved statistical controls, even though they are no longer significant, even marginally so
We checked our data and found it puzzling that the effect sizes are that similar, even though the $F$ value for spanking (for example) dropped to $F(3, 745) = 0.36, p = .79$, when controlling for the 16-item Externalizing Problems scale. We have therefore modified our no-effect conclusions to call them non-significant effects, accordingly. For the same reason, we do not consider the effect sizes for Remove Privileges as necessarily less detrimental than those for spanking when using the original 3-level control for antisocial behavior, even though one is statistically significant and the other is not. Nonetheless, the effect size for spanking is then identical to the mean effect size for the 3 alternative disciplinary tactics and less than half of the effect size for psychotherapy. Given that none of those effects are even marginally significant (i.e., $p > .10$), we do not want to make too much of comparison of the effect sizes for psychotherapy vs. disciplinary punishments. After all, by definition, non-significant associations could plausibly be due to random fluctuations around a population value of $b = 0$. Moreover, the effect size is likely to be inflated somewhat, because Straus et al. (1997) featured the cohort with the strongest effect size out of five cohorts in their original study, which we have replicated as closely as possible. Features of the sample that increased the effect size for spanking may also increase the effect size for alternative tactics and for psychotherapy.

3. **SEM evidence that the effect of corporal punishment on antisocial behavior vanishes is uninterpretable in the last row, featuring gain scores.** We apologize for failing to make this type of analysis more clear, and have described it more thoroughly in the revision. We did not predict Wave-2 outcomes controlling statistically for gain scores, as Dr. Jolley understood us to do. Instead, the outcome variable was changed from a residualized change score (Wave-2 antisocial controlling for Wave-1 antisocial) to a simple change score (Wave-2 antisocial minus Wave-1 antisocial). Although an important conference decided that residualized change scores were preferable to simple change scores in 1970, several leading statistical methodologists have argued since then that simple change scores are often preferable (Willett, Rogosa, Allison). An in-press article on another longitudinal data set showed that all significant outcomes of all corrective actions varied by whether the analyses predicted residualized change scores or simple change scores (Larzelere, Ferrer, Kuhn, & Danelia, In press). Those corrective actions included disciplinary corrective actions by parents, psychotherapy, and Ritalin. From analyses of residualized change scores, all significant effects of those corrective actions on subsequent antisocial behavior or hyperactivity indicated detrimental effects. In contrast, in analyses of simple change scores, all significant effects of those corrective actions indicated beneficial reductions in those same behavior problems, using identical data. These results are consistent with previous research that has shown that residualized change scores can be biased against corrective actions, as shown by Lord’s (1967) famous paradox and by the supposedly detrimental effects of Head Start found in 1969, which were re-analyzed by Campbell and his associates in several publications throughout the subsequent 2 or 3 decades. The results in the last row of Table 4 are consistent with those contrasting biases, although none of the last-row coefficients are significantly in the opposite direction.

**Appropriateness for a reply by Straus et al.** We wholeheartedly agree that Straus and his colleagues should be given an opportunity to respond to these analyses that replicate their important analyses for alternative disciplinary tactics that parents could use instead. When policies are being based on absolute conclusions opposing a disciplinary tactic used by the vast majority of parents in most cultures for generations, we think that the evidence needs to be
debated as fairly and openly as possible. If nonabusive spanking is invariably so adverse to children that anti-spanking values should be imposed on all parents by policy and by changes in criminal laws (as in New Zealand in 2007), then it should be easy for evidence against nonabusive spanking to stand up against appropriate kinds of scientific scrutiny. Further, all parents who choose not to use spanking or are mandated to avoid spanking need to know what alternative disciplinary enforcements are at least as effective as spanking.

Robert E. Larzelere, Ph.D. and Ronald B. Cox, Jr.
(405) 744-2053  (405) 744-2800 (dept. fax)
E-Mail: Robert.Larzelere@okstate.edu

References

Bremner, J. D., Vermetten, E., & Mazure, C. M. (2000). Development and preliminary psychometric properties of an instrument for the measurement of childhood trauma: The Early Trauma Inventory. Depression and Anxiety, 12, 1-12.


