

The Outcomes of Physical Punishment are Typical of *All* Corrective Actions:

A Response to Heilmann et al.'s (2021) Narrative Review

Robert E. Larzelere, Ph.D., Oklahoma State University
Marjorie Lindner Gunnoe, Ph.D., Calvin University
Joshua Pritsker, Purdue University
Christopher J. Ferguson, Ph.D., Stetson University
Carla Adkison-Johnson, Ph.D., Western Michigan University
Jelani Mandara, Ph.D., Matteson, IL
Den A. Trumbull, M.D., Pediatric Healthcare, Montgomery, AL

Published in *PsyArXiv* (2022, Jan. 8) doi: 10.31234/osf.io/5xbnm

Corresponding Author: Robert E. Larzelere, Dept. of Human Development and Family Science, 233 Nancy Randolph Davis Bldg., Oklahoma State University, Stillwater, OK 74078 robert.larzelere@okstate.edu (405) 744-2053

Marjorie Lindner Gunnoe, Dept. of Psychology, Calvin University, 1734 Knollcrest Cir., Grand Rapids, MI 49546 mgunnoe@calvin.edu

Joshua Pritsker, Dept. of Psychological Sciences, 703 Third St., Purdue University, West Lafayette, IN 47907 jpritsk@purdue.edu

Christopher J. Ferguson, Dept. of Psychology, 421 N. Woodland Blvd., Stetson University, DeLand, FL 32723 cjfergus@stetson.edu

Carla Adkison-Johnson, Dept. of Counselor Education & Counseling Psychology, Western Michigan University, 1903 N. Michigan Ave., Kalamazoo, MI 49008 carla.adkison-johnson@wmich.edu

Jelani Mandara, 162 White Oaks Rd., Matteson, IL 60443 jelani.mandara@yahoo.com
(previous: tenured Associate Professor, School of Education and Social Policy, Northwestern University, Evanston, IL)

Den A. Trumbull, Pediatric Healthcare, LLC, 4700 Woodmere Blvd., Montgomery, AL 36106 trumbulld3@gmail.com

The Outcomes of Physical Punishment are Typical of *All* Corrective Actions:

A Response to Heilmann et al.'s (2021) Narrative Review

Abstract

The reputation of psychological science depends on the adequacy of the science underlying its policy recommendations. This commentary raises concerns about the science used by Heilman et al. (2021) in their recent *narrative* (not meta-analytic) review that encourages spanking bans worldwide. By reviewing controlled longitudinal studies, Heilmann et al. provided stronger causal evidence than the two meta-analyses of unadjusted correlations most frequently cited to support spanking bans. However, the two previously published *meta-analyses* of controlled longitudinal studies of spanking do not support spanking bans, due to the trivial size of the average adverse-looking effect of customary spanking in those studies. Moreover, several lines of evidence indicate that this trivial average effect is likely due to inadequate statistical controls rather than an actual adverse causal effect of typical spanking. We need stronger causal evidence for policy recommendations for both the welfare of children and the reputation of psychological science.

Heilmann et al.¹ recently published a narrative review claiming that physical punishment was consistently associated with problematic child outcomes in prospective longitudinal studies even after controlling statistically for initial differences in those problems. (A narrative review is a "box-score" literature review that counts the number of studies reporting beneficial, detrimental, or neutral effects. It differs from a meta-analytic review in that it does not report average effect sizes.)

Heilmann et al.'s narrative review constitutes stronger evidence than the meta-analysis of correlations² that the American Academy of Pediatrics³ and the American Psychological Association⁴ relied on to oppose all physical punishment in 2018 and 2019. Its summaries are also consistent with the results of the two known meta-analyses of controlled longitudinal studies of physical punishment,^{5,6} which documented consistent, but "trivial" adverse-looking effects of physical punishment on externalizing problems (partial $r = .075$, equivalent to $d = .15$, and AOR = 1.32). Heilman et al.'s recommendation that spanking be banned worldwide is *not* consistent with the recommendations offered by the authors of these more rigorous meta-analyses.

The tiny effect sizes documented in meta-analyses of controlled, longitudinal studies of spanking can easily be explained by the effects of confounding variables, which epidemiologists recognize as *residual confounding*⁷ rather than a true causal effect. Residual confounding is the best explanation of the following evidence:

1. The apparent effect of spanking on externalizing problems becomes non-significant when measurement of the primary confound (initial differences) is improved.⁸
2. The ANCOVA-type of statistical control used in most longitudinal studies has been shown to be biased,^{9,10} and the pure within-person difference-score analyses that are recommended instead show non-significant or beneficial-looking average effects of customary spanking.¹⁰⁻¹² These contradictory *apparent* effects derived from different types of statistical analyses are shown in Table 1 (“spanking” column) where we use data from the Larzelere et al. meta-analysis⁶ to compare effect sizes featured in various meta-analyses of physical punishment.^{2,5,6} Externalizing problems actually decreased more from Time 1 to Time 2 in children who were spanked at Time 1 than in children who were not spanked then (in the average longitudinal study, $r = -.04$, $p < .05$).⁶ That decrease appears to be slightly harmful in the usual ANCOVA-type analyses ($\beta = +.07$, $p < .001$) because its results are biased by unchanging differences between the children being compared, differences already apparent at Time 1.^{9,10}
3. The bias in ANCOVA-type controls makes all corrective actions appear harmful, even ones known to be helpful (e.g., timeout, psychotherapies, hospitalizations, prescription medications).¹³ The common denominator seems to be that ANCOVA-type controls reduce, but *fail to eliminate*, the bias inherent in outcomes of any corrective action due to the poor prognosis of the problem being corrected.

This residual confounding also explains why other corrective actions for children’s oppositional defiance look as harmful as spanking in the only direct comparisons in ANCOVA-controlled longitudinal studies, whether implemented by parents (privilege removal, timeout) or professionals (therapy, methylphenidate hcl [*Ritalin*]).^{8,12} It explains why researchers have yet to find an effective replacement for spanking with this type of study.^{8,12} Although Heilmann et al. attempted to address this issue with studies using propensity score matching and fixed-effects

models, these methods do not eliminate the remaining bias. Propensity score methods retain residual confounding and thus replicate the results of ANCOVA-type controls when controlling for the same variables.^{14,15} All known fixed-effects models of spanking to date have focused on *cross-sectional* associations and found that spanking and antisocial behavior tend to both be higher than that family's average at some times and below their own average at other times, but they cannot determine whether the spanking or the antisocial behavior came first. When the near-equivalent of two-occasion fixed-effects models were used to predict within-person changes subsequent to spanking, spanking no longer predicted externalizing problems significantly,¹⁰ and even predicted decreased externalizing problems when limited in use.¹¹

Before opposing a widely-used medical or psychological treatment, most practitioners would require reasonable causal evidence that its *optimal implementation for appropriate presenting problems* is consistently *less effective than an alternative* treatment for those problems. Instead, the linear statistics in a typical longitudinal study of physical punishment mostly compare its over-usage (median: 4+ times weekly)⁶ with no use for a limited time (median: one week)⁶ without specifying the presenting problems or showing that any alternative disciplinary response is more effective.

Other research that distinguishes milder spanking from inappropriate physical punishment has found:

1. *Spanking appears effective as a back-up for timeout for the most defiant children.* Two randomized trials found spanking to be a more effective back-up for timeout than a child-release control condition (mean $d = -.74$, equivalent to $\beta = -.35$, $p < .05$)¹⁶; three trials found that a brief isolation in an ideal safe room was the only back-up as effective as spanking ($d = .20$ & $\beta = .10$, n.s.).¹⁶ The spanking back-up is no longer taught in clinical programs, but that may help explain why clinical treatments for child conduct problems are half as effective now as they were 30 to 50 years ago, when spanking was the primary back-up for timeout.¹⁷
2. *Back-up spanking appears to deter defiance more effectively than most alternatives.* In direct comparisons using clinical and non-clinical samples, back-up spanking (aka "conditional" spanking) resulted in *less* defiance or aggression than 10 of 13 alternative tactics.¹⁶ Physical punishment led to worse outcomes than alternative tactics only when used too severely or as the primary tactic.¹⁶

3. *Adverse-looking effects generally disappear with proper statistical controls for more appropriate types of spanking at appropriate ages.* The few controlled longitudinal studies that have focused explicitly on *open-handed*¹¹ or phased-out¹⁸ spanking in non-clinical samples have found significant beneficial-looking effects. Heilmann et al.'s review misleads by neglecting differential effects by how it is implemented and by the child's age — important discriminations made in some of the more rigorous meta-analyses^{5,6,16} and by parents themselves.

The effects of spanking also vary by the parent-child context. Some studies have found that spanking is not harmful when children perceive it as motivated by love and concern for the child¹⁹ or as part of an optimal authoritative parenting pattern.^{20,21} A recent qualitative study of African American families found that *when* and *if* physical discipline was used, it was rarely a stand-alone technique, but was reserved for blatantly defiant behavior and was contingent on child age and the disciplinary context.²²

Although we appreciate Heilmann et al.'s efforts to understand the effects of spanking on children's wellbeing, we are concerned that current evidence does not begin to justify total injunctions against spanking. Prohibiting appropriate spanking without providing an alternative may do more harm than good. We call for parenting researchers to conduct research that the medical and psychological establishments would consider sufficient to prohibit a widely used professional practice.

References

1. Heilmann A, Mehay A, Watt RG, et al. Physical punishment and child outcomes: A narrative review of prospective studies. *The Lancet* 2021; **398**: 355-364.
2. Gershoff ET, Grogan-Kaylor A. Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology* 2016; **30**: 453-69.
3. Sege RD, Siegel BS, Council on Child Abuse and Neglect, Committee on Psychosocial Aspects of Child and Family Health. Effective discipline to raise healthy children. *Pediatrics* 2018; **142**(6: e20183112).
4. American Psychological Association. Resolution on physical discipline of children by parents. 2019. <https://www.apa.org/about/policy/physical-discipline.pdf>.
5. Ferguson CJ. Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies. *Clinical Psychology Review* 2013; **33**: 196-208.
6. Larzelere RE, Gunnoe ML, Ferguson CJ. Improving causal inferences in meta-analyses of longitudinal studies: Spanking as an illustration. *Child Development* 2018; **89**: 2038-50.
7. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia, PA: Wolter Kluwer; 2008.
8. Larzelere RE, Cox RB, Jr., Smith GL. Do nonphysical punishments reduce antisocial behavior more than spanking? A comparison using the strongest previous causal evidence against spanking. *BMC Pediatrics* 2010; **10**(10): 1-17.
9. Hamaker EL, Kuiper RM, Grasman RPPP. A critique of the cross-lagged panel model. *Psychological Methods* 2015; **20**: 102-16.
10. Berry D, Willoughby MT. On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development* 2017; **88**: 1186-1206.
11. Pritsker J. Spanking and externalizing problems: Examining within-subject associations. *Child Development* 2021; **92**: 2595-2602.
12. Larzelere RE, Ferrer E, Kuhn BR, Danelia K. 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* 2010; **34**: 180-9.
13. Larzelere RE, Lin H, Payton ME, Washburn IJ. Longitudinal biases against corrective actions. *Archives of Scientific Psychology* 2018; **6**: 243-50.
14. Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods* 2010; **15**: 250-67.
15. Gershoff ET, Sattler KMP, Ansari A. Strengthening causal estimates for links between spanking and children's externalizing behavior problems. *Psychological Science* 2018; **29**: 110-20.
16. Larzelere RE, Kuhn BR. Comparing child outcomes of physical punishment and alternative disciplinary tactics: A meta-analysis. *Clinical Child and Family Psychology Review* 2005; **8**: 1-37.
17. Weisz JR, Kuppens S, Ng MY, et al. Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science* 2019; **14**: 216-37.
18. Ellison CG, Musick MA, Holden GW. Does conservative Protestantism moderate the association between corporal punishment and child outcomes? *Journal of Marriage and Family* 2011; **73**: 946-61.
19. Larzelere RE, Klein M, Schumm WR, Alibrando SA, Jr. Relations of spanking and other parenting characteristics to self-esteem and perceived fairness of parental discipline. *Psychological Reports* 1989; **64**: 1140-2.
20. Baumrind D, Larzelere RE, Owens EB. Effects of preschool parents' power assertive patterns and practices on adolescent development. *Parenting: Science and Practice* 2010; **10**: 157-201.
21. Gunnoe ML. Associations between parenting style, physical discipline, and adjustment in adolescents' reports. *Psychological Reports: Disability & Trauma* 2013; **112**: 933-75.
22. Adkison-Johnson C. Child discipline in African American families: Culturally responsive policies. Lanham, MD: Lexington Books; 2021.
23. Gershoff ET. Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin* 2002; **128**: 539-79.
24. Larzelere RE, Gunnoe ML, Roberts MW, Lin H, Ferguson CJ. Causal evidence for *exclusively positive parenting* and for timeout: Rejoinder to Holden, Grogan-Kaylor, Durrant, and Gershoff (2017). *Marriage & Family Review* 2020; **56**: 287-319.

Table 1

Associations of Four Corrective Actions with Externalizing Behavior Problems and Hyperactivity with Standardized Regression Coefficients by Type of Evidence^a

Type of Evidence	Spanking	Nonphysical Punishment	Therapy for the Child	<i>Ritalin</i>
Cross-sectional correlations	.20*** ^b	.20** ^c	.12** ^c	.11** ^d
Longitudinal correlations	.16*** ^b	.18** ^c	.13** ^c	.13** ^d
ANCOVA-type control for initial differences	.07*** ^b	.08 ^{^c}	.14 ^{^c}	.06* ^d
Subsequent within-person changes	-.04* ^b	-.02 ^{^c}	.01 ^c	.02 ^d
Randomized vs. control	-.35* ^e	-.64** ^f	-.18** ^g	--

^aPositive standardized regression coefficients (β 's) indicate harmful-looking associations, whereas negative β 's indicate beneficial-looking associations. $\beta = r$ in first 2 rows.

^bFrom Larzelere, Gunnoe, & Ferguson's meta-analysis (p. 2044),⁶ similar to mean effect sizes from other meta-analyses of correlations^{2,23} and of partial correlations for ANCOVA-type controls (Ferguson, pp. 201-02).⁵

^cMean effect sizes from the two known statistically controlled longitudinal analyses of physical punishment that also investigated equivalent effect sizes for alternative corrective actions (nonphysical punishments, therapy, and methylphenidate hcl [*Ritalin*] on subsequent externalizing behavior problems and hyperactivity (Larzelere, Ferrer et al., 2010,¹² pp. 182-85), and on subsequent antisocial behavior (Larzelere, Cox, & Smith, 2010,⁸ pp. 5-8). Both studies controlled for an initial Time-1 measure of each outcome).

^dMean effect sizes from Larzelere, Ferrer, et al. (2010),¹² pp. 182-185.

^eTwo relevant studies cited in Larzelere, Gunnoe, & Ferguson's meta-analysis;⁶ with their mean effect size estimated from Larzelere & Kuhn's meta-analysis.¹⁶

^fFrom meta-analysis of timeout.²⁴

^gCurrent effect size from Weisz et al.'s¹⁷ meta-analysis of randomized trials of psychosocial treatments for conduct problems from the 1960s through 2017 (down from $\beta = -.36$ in the 1960s and 1970s, when spanking was the preferred back-up for timeout).

* $p < .05$ (for at least 2 of the 3 coefficients averaged from the two studies in last three columns).

** $p < .01$ for all coefficients that were averaged in the last three columns. *** $p < .001$.

[^]Only one of the three standardized coefficients averaged in the last 3 columns was significant at $p < .05$