

COMMENT

The Insufficiency of the Evidence Used to Categorically Oppose Spanking and Its Implications for Families and Psychological Science: Comment on Gershoff et al. (2018)

Robert E. Larzelere
Oklahoma State University

Marjorie Lindner Gunnoe
Calvin College

Christopher J. Ferguson
Stetson University

Mark W. Roberts
Idaho State University

Gershoff et al. (2018) recently summarized the scientific evidence against disciplinary spanking, using epidemiological and psychological criteria for causal validity. Unfortunately, the evidence they cited would make most actions to correct serious problems *appear* to be harmful, whether implemented by parents (e.g., timeout) or professionals. The reason is that the type of evidence that Gershoff et al. consider adequate is insufficient for establishing a causal connection between *any* disciplinary response to persistent defiance and problem behaviors in children, whether that response is spanking or an effective alternative to spanking. Before opposing a widespread practice such as spanking, researchers need to document stronger causal evidence against it and identify an alternative demonstrated to be more effective.

Keywords: causal inference, research methodology, physical punishment, parental discipline

Spanking is controversial. According to Gershoff et al. (2018), over 80% of American parents spank their children. Thirty years ago, most empirically supported clinical treatments for oppositional defiance employed spanking to enforce timeout (Everett, Hupp, & Olmi, 2010, p. 246). Many social scientists are opposed to spanking for moral/philosophical reasons, and some believe there is sufficient research evidence to oppose it. Gershoff et al. (2018) asserted that the evidence against spanking “has met the requirements for causal conclusions” (p. 635). To their credit, Gershoff’s team organized their evidence using criteria for valid causal inferences from epidemiology as well as psy-

chology. Nonetheless, this Comment argues that Gershoff et al. overstated the strength of their causal evidence against spanking. It also expresses concern that researchers are opposing a disciplinary technique without adequate evidence for a more effective alternative.

Gershoff et al. placed too much confidence in unadjusted correlations. Instead of reporting the range of effect sizes across the five meta-analyses they cited, they reported only the largest mean effect size ($d = .33$, equivalent to $r = .16$), which was based almost entirely on unadjusted correlations from Gershoff & Grogan-Kaylor (2016: 55% cross-sectional, 21% retrospective, 21% longitudinal). The shortcoming of unadjusted correlations is that corrective actions are almost always positively correlated with the problems they are attempting to correct. Because of this inherent selection bias, most psychologists would *not* claim that psychotherapy ($r = .15$) or nonphysical punishment (timeout and privilege removal: $r = .17$) *cause* subsequent antisocial behavior among children (Larzelere, Ferrer, Kuhn, & Danelia, 2010). It is therefore problematic to rely on unadjusted correlations to claim that spanking causes adverse outcomes—even when these correlations are con-

Robert E. Larzelere, Department of Human Development and Family Science, Oklahoma State University; Marjorie Lindner Gunnoe, Department of Psychology, Calvin College; Christopher J. Ferguson, Department of Psychology, Stetson University; Mark W. Roberts, Department of Psychology, Idaho State University.

Correspondence concerning this article should be addressed to Robert E. Larzelere, Department of Human Development and Family Science, Oklahoma State University, 233 Human Sciences Building, Stillwater, OK 74078. E-mail: robert.larzelere@okstate.edu

sistent across studies and cultures, as Gershoff et al. emphasized. Insufficient criteria do not become sufficient just because they are frequently applied.

To their credit, the Gershoff team did cite 13 studies that controlled for preexisting scores on the child outcome. If their covariates measured all confounding variables perfectly, these effects would be unbiased. Unfortunately, imperfect covariates do not eliminate selection bias; they only reduce it. This failure to eliminate selection bias was evidenced by Larzelere et al. (2010), who demonstrated that various nonphysical punishments and *Ritalin* remained positively associated with the problems they were intended to correct, despite controlling for preexisting child differences longitudinally.

Although Gershoff's et al. (2018) evidence for opposing spanking is insufficient, we recognize that spanking is often overused. As a consequence, some of us have sought to identify effective alternatives to spanking. Unfortunately, acceptance of the causal criteria used by Gershoff et al. (2018) may hinder this identification. Gershoff et al. (2010) investigated physical punishment and 10 alternative disciplinary tactics. Because correlations between corrective actions and the problems they are intended to correct are usually positive, none of those disciplinary tactics predicted significantly lower levels of children's aggression or anxiety (i.e., were negatively correlated with these outcomes). Perhaps that is why Gershoff and Holden are now advocating a "strong positive parenting," void of *all* disciplinary consequences (Holden, Grogan-Kaylor, Durrant, & Gershoff, 2017).

A review of timeout variations concluded that spanking and a brief forced room isolation were the two most effective enforcements for timeout cooperation (Everett et al., 2010, pp. 246–247). Gershoff et al. (2018) cited the equivalent effectiveness of these enforcements as evidence that psychologists should oppose spanking. Instead, this is the kind of evidence that psychologists should prefer for evaluating both spanking and its alternatives. It is based on three randomized clinical trials that compared the most effective use of spanking with alternative procedures that could replace it when needed most; namely, to enforce cooperation with timeout in defiant young children.

Gershoff et al.'s reliance on bivariate associations in an article emphasizing the strength of causal evidence is particularly disconcerting, given their awareness of smaller effects from higher quality evidence. Although there are few randomized studies of spanking, two meta-analyses cited by Gershoff et al. went beyond unadjusted correlations. One showed that longitudinal associations of spanking with externalizing problems become trivial (partial correlation < .10) after controlling for prior externalizing problems (Ferguson, 2013). The other showed that spanking compared unfavorably with alternative disciplinary tactics only when

spanking was used predominantly or severely (Larzelere & Kuhn, 2005).

A unique feature of the Larzelere and Kuhn meta-analysis is that it compared child outcomes of physical punishment with the outcomes of alternative tactics from the same studies. It found that nonabusive spanking was *more* effective than most alternatives when 2- to 6-year-olds refused to cooperate with milder tactics, such as timeout. This *backup spanking* led to less aggression or defiance than 10 of 13 alternatives it was directly compared with. When backup spanking enhances subsequent cooperation with the milder tactic, spanking can then be phased out.

More recently, Larzelere, Gunnoe, and Ferguson (2018) combined the best features of two major meta-analyses. Like Gershoff and Grogan-Kaylor (2016), they employed correlations to maintain consistency across studies. To approximate Ferguson's (2013) less-biased causal estimates, they used additional correlations to adjust for initial differences on the outcome variable. The direction of the resulting trivial effect sizes *varied by the type of adjustment*. Analysis of covariance (ANCOVA) suggested a trivial *detrimental* effect of spanking on children's externalizing problems, whereas linear growth analyses of the same data suggested a trivial *beneficial* effect.

We believe that Gershoff and her colleagues are opposing spanking in good faith. We also recognize that Gershoff et al.'s stance has intuitive appeal to social scientists who oppose spanking on moral grounds and have raised their own children without spanking. But opposing spanking or alternatives with mostly correlational evidence may hurt families more than it helps them. Clinicians would not eschew a widely used clinical intervention without identifying a more effective alternative. Neither should family psychologists. Causal assertions based primarily on correlational evidence may also undermine the integrity of psychological science. We call upon the psychological community to take an appropriately cautious approach when making or accepting blanket statements "against" (or "for") any nonabusive disciplinary tactic based primarily on correlational evidence.

References

- Everett, G. E., Hupp, S. D. A., & Olmi, D. J. (2010). Time-out with parents: A descriptive analysis of 30 years of research. *Education & Treatment of Children, 33*, 235–259. <http://dx.doi.org/10.1353/etc.0.0091>
- Ferguson, C. J. (2013). Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies. *Clinical Psychology Review, 33*, 196–208. <http://dx.doi.org/10.1016/j.cpr.2012.11.002>
- Gershoff, E. T., Goodman, G. S., Miller-Perrin, C. L., Holden, G. W., Jackson, Y., & Kazdin, A. E. (2018). The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers. *American Psychologist, 73*, 626–638. <http://dx.doi.org/10.1037/amp0000327>

- Gershoff, E. T., & Grogan-Kaylor, A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology, 30*, 453–469. <http://dx.doi.org/10.1037/fam0000191>
- Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., & Dodge, K. A. (2010). Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness. *Child Development, 81*, 487–502. <http://dx.doi.org/10.1111/j.1467-8624.2009.01409.x>
- Holden, G. W., Grogan-Kaylor, A., Durrant, J. E., & Gershoff, E. T. (2017). Researchers deserve a better critique: Response to Larzelere, Gunnoe, Roberts, and Ferguson (2017). *Marriage & Family Review, 53*, 465–490. <http://dx.doi.org/10.1080/01494929.2017.1308899>
- Larzelere, R. E., Ferrer, E., Kuhn, B. R., & Danelia, 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*, 180–189. <http://dx.doi.org/10.1177/0165025409351386>
- Larzelere, R. E., Gunnoe, M. L., & Ferguson, C. J. (2018). Improving causal inferences in meta-analyses of longitudinal studies: Spanking as an illustration. *Child Development, 89*, 2038–2050. <http://dx.doi.org/10.1111/cdev.13097>
- Larzelere, R. E., & Kuhn, B. R. (2005). Comparing child outcomes of physical punishment and alternative disciplinary tactics: A meta-analysis. *Clinical Child and Family Psychology Review, 8*, 1–37. <http://dx.doi.org/10.1007/s10567-005-2340-z>

Received September 5, 2018

Revision received February 2, 2019

Accepted February 14, 2019 ■