

## Critique of Gershoff 2017 study:

### Strengthening Causal Estimates for Links between Spanking and Children's Externalizing Behavior Problems<sup>i</sup>

#### Study Deficiencies

This spanking study suffers from the following common deficiencies:

- ✓ - Uses a vague or inaccurate definition for “spanking”
- Combines overly severe physical punishments along with ordinary spanking
- Includes physical punishment of adolescents, which is an inappropriate age
- Fails to control for frequency of spanking usage
- Fails to limit spanking to younger ages, especially under 7 years
- Relies upon statistical correlation rather than true causation
- ✓ - Fails to control for, or define, child behavior prior to use of spanking
- ✓ - Fails to compare spanking with other available correction measures
- Relies upon retrospective interviews, rather than longitudinal observations

In her abstract, Gershoff concedes that *causal* links are difficult to establish without prospective experiments. She admits to a key criticism contained within most of the anti-spanking research, “correlational studies cannot rule out potential selection factors.” Nonetheless, she cites her 2016 review of the scientific literature,<sup>ii</sup> claiming that it confirms a *link* (not causation) between spanking and externalizing behavioral problems. Referring to this study, she admits, “**However, the majority of these studies are correlational, leading to fundamental questions about causality.**” This is a rare confession, and a fact that undermines the evidence most often cited to oppose spanking. Sadly, it is a recurring theme in research used to support opposition to spanking to identify negative *correlations* as though they are proof of *causation*. In fact, all of Gershoff's studies cited in her 2016 work suffer from one or more of the study deficiencies listed above.

This 2017 study provides Gershoff's strongest causal evidence to date. In it she eliminates some of the flaws that rendered her 2016 meta-analyses evidence against spanking unjustified (as it is based solely on inadequate, unadjusted correlations). While this latest study is a much-needed step in the right direction, we need more studies with stronger causal evidence before proclaiming that the vast majority of parents have been harming their children each and every time they spank them, regardless of how appropriately it was used, given the child's age and type of noncompliance at that time (e.g., persistent defiance even when parents try milder disciplinary tactics). Although her measure does not emphasize inappropriately severe physical punishment at inappropriate ages, it does nothing to focus on the most appropriate ways to use spanking, nor does it do anything to tell whether the spanking was appropriate given the disciplinary circumstances (e.g., persistent defiance vs. childish error or an accident that made the parent angry).

The results of this 2017 study are flawed for the following reasons.

- **Spanking technique is not defined.** In this study, parents were simply asked, “How many times, if any, within the past week how you spanked your child?” Regardless of the sophistication of statistics used to analyze this data, without defining *how* and *when* these parents spank, the results will be flawed when looking for the effects of appropriate spanking. Using the term “spanking” in the parent questionnaire, however, is an improvement over many previous research studies where they were asked about physical punishment included abusive practices such as face slapping or punching.
- **Fails to compare with other measures.** There was no comparison of spanking to other correction measures (such as time-out) used by these parents. Therefore, for the children needing correction (those spanked), one can't know whether any other disciplinary measures would have been more successful. In the only study where Gershoff and her colleagues investigated other alternatives, none of

the other 10 disciplinary measures were successful in reducing externalizing behavior problems, despite multiple statistical analyses for each one — an indictment they also level against spanking.

- **Relies upon interviews.** Although prospective in design, this study relies upon parent interviews, not actual observations of their use of spanking.
- **Fails to control for initial behavior.** It does control for teacher reports of externalizing behavior problems at age 5, thereby taking that into account before predicting externalizing behavior problems at school at age 6 and at age 8. However, it does not control for the child's behavior problems AT HOME, which directly influences the frequency and type of disciplinary responses to that misbehavior. In the technique they used to improve causal inferences, called propensity-score adjustments, all the control variables should be measured BEFORE the measure of the causal variable (spanking). This was not done. Instead the control variables were all based on the same interview as the measure of spanking, when the children were in kindergarten. This is a problem, especially for their measure of whether they ever spanked their children before the past week.
- **Behavior at home was not measured.** Similar to another of Gershoff's studies, they control only for the teacher's report of externalizing behavior problems at Time 1 (kindergarten), not for the parents' perspective. The latter is necessary to take into account how oppositional and defiant the child is at home. Furthermore, it looks like the teacher's report of externalizing behavior problems was based on only 5 items, but that is not very clear. Behavior at school often does not mirror a child's behavior at home. Propensity-score methods are only as good as the control variables that are included.
- **Trivial association.** Finally, the results they get are consistent with Ferguson's (2013)<sup>iii</sup> meta-analysis, indicating that the adverse effects of spanking are trivial (i.e., smaller than  $\beta = .10$ ).  $\beta = .10$  is considered a small effect in the social sciences. A trivial effect can be explained away by a trivial amount of confounding that is not yet controlled for adequately. For example, the failure to control for mothers' perceptions of how oppositional and defiant the child is at home could easily account for this trivial effect. If the effect were larger (e.g., like the association between heavy smoking and lung cancer), a confounding variable would have to have an equally large effect to explain away the large association to make it a non-causal risk factor (like staying overnight in a hospital) rather than being a causal risk factor. The magnitude of the association between heavy smoking and lung cancer is more than 20 times as large as the magnitude of the association in this study. So even if the trivial effect in this study represents a true average causal effect (which we doubt), it would indicate that almost half of the children spanked did have a positive outcome in externalizing behavior problems (i.e., better than the average effect), which would make the decision to spank or not a valid option for parents, depending on other factors (the level of their misbehavior, whether milder tactics had been tried, how spanking was implemented, a positive parent-child relationship).

Authors: Den A. Trumbull, MD & Robert E. Larzelere, PhD  
December 2017

## References

---

<sup>i</sup> Gershoff ET, Satler KMP. Links between Spanking and Children's Behavior. *Psychological Science*. November 2017. 1-11.

<https://doi.org/10.1177/0956797617729816>

<sup>ii</sup> Gershoff, E. T., & Grogan-Kaylor, A. (2016). Corporal punishment by parents and its consequences for children: Old controversies and new meta-analyses. *Journal of Family Psychology*. 30, 453–469. doi:10.1037/fam0000191

<sup>iii</sup> Ferguson, CJ. *Clinical Psychology Review*. 2013. 33:196–208. <https://doi.org/10.1080/01494929.2016.1145613>