

Scientific Evidence Supports Customary and Backup (Conditional) Spanking by Parents: Update of Larzelere and Baumrind (2010)¹ and Fuller (2009)²
Robert E. Larzelere and Jason Fuller
October 25, 2019

I. Two major professional societies recently made official statements opposing all spanking. At the same time, two new articles documented the inadequacies of the scientific evidence underlying those statements.

A. The two official statements that oppose all spanking relied primarily on Gershoff & Grogan-Kaylor's³ meta-analytic (quantitative) summary of relevant studies.

1. American Academy of Pediatrics (AAP) in December 2018⁴
2. American Psychological Association (AAP) in February 2019⁵
 - a. APA's resolution was supported by a task force that claimed sufficient causal evidence to oppose all spanking.⁶
3. If opposing all spanking was supported by scientific evidence and was therefore best for the welfare of all children, we would support those position statements.
4. One of us (Larzelere) has been doing research on physical discipline for 40 years to:
 - a. find more-effective alternate tactics to spanking, and
 - b. study whether spanking *causes* any adverse outcomes it is correlated with

B. Two articles published in November and December 2018 documented the inadequacy of the scientific evidence used to oppose all spanking.

1. A new meta-analytic summary of the most relevant studies showed that the evidence against customary spanking disappears when more-appropriate analytical methods are used.⁷
 - a. The new meta-analysis⁸ only included studies that (1) measured the outcome after the spanking (which 55% of Gershoff & Grogan-Kaylor's studies failed to do⁹), (2) adjusted statistically for preexisting child differences on the outcome variable (in contrast to Gershoff & Grogan-Kaylor, who used "only bivariate comparisons or correlations"¹⁰), and (3) investigated only *customary* or *open-handed* spanking.
 - b. The new meta-analysis replicated the results of two previous meta-analyses
 - i. The new meta-analysis replicates the correlation-based results from Gershoff & Grogan-Kaylor, whose effect sizes were based mostly on cross-sectional correlations (55%) and whose strongest evidence against spanking came from retrospective (21%) and longitudinal (21%) correlations;¹¹ but such correlations are especially misleading to make causal conclusions about corrective actions.
 1. Cross-sectional correlations cannot tell whether spanking occurred before or after the child outcome.
 2. Cross-sectional correlations make even the most effective corrective actions look harmful, because the problems being corrected ("outcomes") are measured during the same time period the corrective action was used. The following corrective actions would thus look harmful according to cross-sectional correlations:
 - a. Medical treatments, because patients are sicker when treated compared to other people in the community
 - b. Psychotherapy, because clients have worse symptoms when treated than other people in the community
 - c. Home furnaces, because home temperatures are cooler when they run (e.g., 61°F in my Omaha home) than when they do not run (mean 67°F throughout the year)

- d. Sprinkler systems in public buildings, because there is more water and fire damage when they turn on than in other buildings where sprinkler systems do not turn on.
3. Cross-sectional correlations are never used as scientific evidence against corrective actions by professionals, so why were they 55% of the evidence against spanking in the meta-analysis relied on by AAP & APA?
 - a. A well-intentioned, preexisting bias against spanking seems to be the most plausible explanation
 - ii. The new meta-analysis¹² also replicated Ferguson's (2013)¹³ "trivial" average effect of spanking from the same data. These results are preferred to the correlation-based meta-analysis relied on by the AAP and APA, because causally relevant conclusions should (1) require studies to measure the child outcome after the spanking and (2) adjust for preexisting child differences on the child outcome variable.
 - iii. In addition, the new meta-analysis¹⁴ showed that an alternative method to adjust for preexisting child differences resulted in the trivial effects of spanking being beneficial, not harmful.
 1. Recent statistical books and articles have shown that the previous way to adjust for preexisting differences is biased.¹⁵ One of those articles¹⁶ illustrates a corrected analysis, which results in beneficial outcomes of spanking, consistent with the new meta-analysis's beneficial trivial effects.¹⁷
 2. At a minimum, the new meta-analysis indicates that the average causal effect of customary spanking is so close to zero that it requires a bias either for or against it to get significant effects in either direction (whether harmful or helpful).
2. The second late-2018 publication showed that the strongest causal evidence against customary spanking (i.e., the trivial adverse-average effect, above) fits a pattern of evidence typical of all corrective actions, whether by parents or professionals.¹⁸
 - a. The following corrective actions by parents showed trivial, harmful-looking effects similar to spanking (mean $\beta = .08$), when analyzed the same way:¹⁹
 - i. Other disciplinary responses
 1. Privilege removal (mean $\beta = .09$)²⁰
 2. Grounding ($\beta = .12$)²¹
 3. Send to room (mean $\beta = .08$)²²
 - ii. Corrective actions for other problems
 1. Helping with homework (mean harmful $\beta = .13$)²³
 2. Warning adolescents of the risks of premature sex ($\beta = .03$)²⁴
 3. Talking to adolescents about the dangers of smoking ($\beta = .13$)²⁵
 4. Forbidding associating with deviant peers ($\beta = .11$)²⁶
 - b. The following corrective actions by professionals showed similar trivial or larger harmful-looking effects, when analyzed the same way:²⁷
 - i. Psychotherapy for children (mean $\beta = .15$)²⁸
 - ii. *Ritalin* ($\beta = .06$)²⁹
 - iii. Medications for depression ($\beta = .07$)³⁰
 - iv. Therapy for depression ($\beta = .06$)³¹
 - v. Medications for physical-health problems ($\beta = .08$)³²
 - vi. Emergency-room treatment or hospitalization ($\beta = .05$)³³
 - vii. Out-of-home placements for youths (mean $\beta = .05$)³⁴

1. Note that the causal evidence for out-of-home placements looks harmful, just like the strongest causal evidence against customary spanking.
2. This study is one of the three least-biased evaluations of outcomes of out-of-home care for maltreated children, according to a meta-analysis with rigorous criteria.³⁵
3. Similar pattern of results as for customary spanking
 - a. Analyses of residualized change produces a trivial but statistically significant harmful effect from out-of-home placement on internalizing problems, but that harmful effect disappears in analyses of simple change.³⁶
 - b. The only meta-analysis³⁷ and studies³⁸ that analyzed both residualized change and simple change for customary spanking obtained the same pattern of disappearing adverse effects as was found for out-of-home placements.
- viii. Child-care subsidies³⁹
- ix. Job-training programs for the unemployed⁴⁰
- c. Why is this evidence (along with weaker unadjusted correlations) sufficient to oppose all spanking, when identical analyses make almost all corrective actions seem just as harmful?

II. Most published meta-analyses of child outcomes of physical punishment do not oppose all spanking

A. Only Gershoff's meta-analyses have been used to oppose all spanking

1. Gershoff & Grogan-Kaylor (2016),⁴¹ updating Gershoff (2002)⁴²
 - a. Primary authority on which the APA⁴³ and AAP⁴⁴ rely to oppose all spanking. *See* critiques of both position statements and rejoinders.⁴⁵
 - b. As indicated above, all their meta-analytic evidence against spanking was based on unadjusted correlations, mostly cross-sectional (61 of 111 effect sizes; 55%) plus some retrospective ($n = 23$; 21%) and some longitudinal ($n = 23$; 21%) correlations.⁴⁶

B. None of the other four published meta-analyses opposed all spanking. Here are quotes from their abstracts:

1. Paolucci & Violato (2004)
 "The results of the present meta-analysis suggest that exposure to corporal punishment does not substantially increase the risk to youth of developing affective, cognitive, or behavioral pathologies."⁴⁷
2. Larzelere & Kuhn (2005)
 "The results indicated that effect sizes significantly favored *conditional* spanking over 10 of 13 alternative disciplinary tactics for reducing child noncompliance or antisocial behavior. *Customary* physical punishment yielded effect sizes equal to alternative tactics, except for one large study favoring physical punishment. Only *overly severe* or *predominant* use of physical punishment compared unfavorably with alternative disciplinary tactics" (emphasis in original).⁴⁸
3. Ferguson (2013)
 "[T]he impact of spanking and [corporal punishment] on the negative outcomes evaluated here (externalizing, internalizing behavior and low cognitive performance) are minimal."⁴⁹
4. Larzelere, Gunnoe, & Ferguson (2018)
 "Significant results indicated a small risk or a small benefit of spanking *depending on the adjustment method*" (emphasis in original).⁵⁰

C. Are the Gershoff meta-analyses better than the other meta-analyses? How do they compare to the other meta-analyses in going beyond correlational evidence to approximate causal evidence from predominantly correlational studies?

1. Unlike Gershoff's meta-analyses, two other meta-analyses used only longitudinal studies to ensure that spanking preceded the child outcome, emphasizing ones that adjusted statistically for preexisting differences to the child outcome
 - a. Ferguson (2013),⁵¹ Larzelere et al. (2018)⁵²
2. Another meta-analysis is the only one to compare child outcomes of physical discipline with the outcomes of alternate disciplinary tactics in studies that use the same statistical analyses on the same sample. This strategy approximates valid causal evidence to the extent that the selection bias due to persistent defiance affects both disciplinary responses similarly and, therefore, cancels each other out in direct comparisons of the outcomes of the two disciplinary responses:
 - a. Larzelere & Kuhn (2005)⁵³
3. The other meta-analyses made no attempt to improve causal evidence beyond simple unadjusted correlations
 - a. Gershoff (2002)⁵⁴: cross-sectional correlations (59%), retrospective (25%), longitudinal correlations (13%), randomized (3% with no adverse results)
 - b. Gershoff & Grogan-Kaylor (2016)⁵⁵: cross-sectional correlations (55%), retrospective (21%), longitudinal correlations (21%), randomized (4% with no adverse results)
 - c. Paolucci & Violato (2004):⁵⁶ cross-sectional (39%), retrospective (22%), longitudinal (27%), other (12%)

D. Why then did the APA and AAP position statements rely on Gershoff & Grogan-Kaylor's meta-analysis more than the other meta-analyses? It is likely that the decision makers merely preferred its conclusion rather than the conclusions of the other three available meta-analyses (plus the fourth meta-analysis from 2018). Evidence:

1. The AAP prominently cited an advocacy group, the Global Initiative to End Corporal Punishment,⁵⁷
2. The APA Task Force on Physical Punishment of Children was formed in 2015 with the explicit goal to "disseminate information about its dangers,"⁵⁸ and several members were already disseminating that viewpoint to the media even before they started reviewing the literature.⁵⁹
3. Robert Larzelere contacted the co-chairs of the APA Task Force on April 10, 2015 and Junly 17, 2015, requesting them to add "some members whose relevant research supports a more balanced perspective" (mentioning Larzelere, Gunnoe, & Ferguson specifically) than seemed to be represented among the members of the Task Force. Cindy Miller-Perrin replied on April 13, 2015 that "we will seek external reviews and input on any reports resulting from the task force, including input and reviews from a diverse group of scholars with diverse points of view (e.g., additional individuals who have published specifically in the area of physical punishment of children)." To my knowledge, the Task Force never sought information from published researchers who supported a more balanced perspective, at least none of the three or four recommended to the Task Force.
4. The two Gershoff meta-analyses are cited more (2564 & 370 citations) than all the other meta-analyses combined (269, 206, 197, and 3 citations in Google Scholar, 10/8/19). The Task Force

was aware of the other meta-analyses from 2004 to 2016, because they cited most or all of them.⁶⁰ But they based their conclusion only on Gershoff et al. (2018).

III. Even though Gershoff et al. (2018)⁶¹ claimed that the causal evidence is sufficient to oppose all spanking, the pattern of evidence they cite would qualify back-up spanking for the 2nd strongest evidence of beneficial effectiveness (“probably efficacious”), according to the criteria for empirical support used by the APA’s clinical psychology divisions.⁶²

A. The only evidence they cite that would count as relevant by the APA’s clinical divisions are the four randomized comparisons of the traditional two-swat spank enforcement for timeout with two alternate enforcements and a control condition (letting the child decide when to end timeout).⁶³

1. Best practices in meta-analyses⁶⁴ require distinguishing comparisons with a control condition from comparisons with another treatment considered effective, rather than combining those comparisons as Gershoff & Grogan-Kaylor did.
2. The traditional spank enforcement for timeout was more effective than a control condition (letting the child decide when to end timeout) in two randomized studies, meeting the clinical criteria for a *probably efficacious* treatment.
 - a. Bean and Roberts (1981) found that the spank enforcement was significantly more effective than the child-determined release at improving compliance and achieving the criterion for compliance with significantly fewer timeouts required ($p < .01$).⁶⁵ In contrast to Gershoff & Grogan-Kaylor’s⁶⁶ claim that the effect sizes were insignificant after controlling for baseline compliance, the baseline compliances for the two groups were 23.4% and 23.3%, respectively.
 - b. Roberts and Powers (1990) found that children never required more than 6 timeouts to achieve the compliance criterion when timeout was enforced with spanking, but 56% of children in the control condition (child-determined release) required more than six timeouts (and, to achieve treatment success, required a switch to spanking as the timeout enforcement). This was a significant difference ($p < .05$).⁶⁷
 - c. Combining both studies, the spank backup was more effective than the control condition by $d = .74, p < .05$.⁶⁸
3. A brief room isolation was the only timeout enforcement that was equally effective to the traditional spank enforcement, approximating a 2d criterion for *probably efficacious*.
 - a. Three randomized studies⁶⁹ showed that a 1-minute forced room isolation was the only timeout enforcement as effective as the spank enforcement, which had been the preferred enforcement⁷⁰ in the two most-effective treatments for oppositional-defiance disorder in young children.⁷¹
 - b. A literature review of variations in training parents to use timeout concluded from the above studies that spanking and the brief room isolation (the “barrier” method) were the two most effective enforcements for timeout.⁷²
 - c. This falls short of the *probably efficacious* criterion only because neither enforcement met the top criteria (well-supported) and the sample sizes were too small. Nonetheless, being tied as the most effective enforcement for timeout is better evidence for the use of spanking than against the use of spanking.
4. Larzelere and Baumrind (2010)⁷³ summarized why these four randomized studies show that spanking, as a timeout enforcement with defiant 2- to 6-year-olds, is effective and should be retained as a disciplinary option for parents.

B. All other evidence cited against spanking by Gershoff et al. (2018) are irrelevant under the criteria for empirical support of clinical treatments, because they were not randomized comparisons.

C. Most other evidence against spanking is typical of corrective actions, whether implemented by parents or professionals, as summarized above.

D. Gershoff et al. (2018) cite four other studies in their section “Ruling out Spurious Factors.” The following critiques the adequacy of the causal evidence from those studies:

1. Beauchaine et al. (2005)
 - a. Slap/spank/hit was 1 of 6 items on their Harsh Parenting measure;⁷⁴
 - b. Initial Harsh Parenting predicted significantly *greater* reductions in Externalizing problems (path b, pp. 383–4), confirmed by Beauchaine via email.⁷⁵
2. Breitenstein et al. (2012)
 - a. By the 1-year follow up, the treatment group improved more than the controls on some outcomes and reduced their corporal punishment more (e.g., hitting with object); but the study did not look at whether either one depended on the other.⁷⁶
3. Grogan-Kaylor (2004) fixed-effects regression
 - a. If the family was above their own average in spanking during the past week, they tended to be above their own average in antisocial behavior during the past 3 months. Thus, 92% (12/13) of the temporal sequence was the wrong direction.⁷⁷
4. Gershoff et al. (2018) propensity-score matching
 - a. Propensity-score matching depends on including all relevant confounding variables. Steiner et al. (2010)⁷⁸ showed that one of the two most-important variables to control are those that are integral to the process by which parents select spanking (such as the child’s oppositional defiance at home). Instead, Gershoff et al. controlled for many demographic variables,⁷⁹ which failed to reduce bias according to Steiner et al. Steiner et al. also showed that propensity-score methods give the same results as statistically controlled longitudinal analyses, consistent with Gershoff et al.’s results.

E. There is evidence of positive outcomes of the most appropriate ways to use spanking, in contrast to Gershoff et al.’s (2019) title.⁸⁰

1. Backup (“conditional”) spanking is associated with less noncompliance or aggression than 10 of 13 disciplinary tactics to which it has been directly compared.⁸¹
 - a. Backup spanking is non-abusive (e.g., two open-handed swats to the buttocks⁸² when the administrator is not out of control due to anger⁸³)
 - b. Used to backup milder disciplinary tactics, such as timeout⁸⁴ or reasoning.⁸⁵
 - c. In response to defiance,⁸⁶ not milder types of noncompliance.
2. Children who have been spanked have better outcomes than never-spanked children, if spanking is phased out by age 9 or 11, especially in subcultures that view spanking as normative.
 - a. Adolescents with age-delimited spanking (i.e., stopped by age 11) had higher academic rank and more volunteer work and optimism than never-spanked adolescents.⁸⁷
 - b. Children who were spanked near age 3 but not near age 9 had less antisocial behavior at age 9 than never-spanked children among conservative Protestant families, and they were no different in antisocial behavior from other families.⁸⁸
 - c. Backup spanking enhances cooperation with milder tactics, such as timeout and reasoning, so that spanking can be phased out.

3. Austrian and German parents who thought (incorrectly) that mild spanking was still legal were more likely to use mild spanking and (therefore?) *less* likely to resort to severe physical punishment.⁸⁹
 - a. Apparently, the availability of mild spanking helps conclude the most-frustrating discipline episodes before parental frustration escalates toward eruption with verbal or physical abuse.
 - b. This is similar to Baumrind’s observation about overly permissive parents, who admitted more often than other parents to “explosive attacks of rage in which they inflicted more pain or injury upon the child than they had intended . . . because they felt that they could neither control the child’s behavior nor tolerate its effect upon themselves.”⁹⁰

F. Although leading anti-spanking advocates denigrate spanking for having no demonstrated beneficial effects,⁹¹ they have not shown beneficial effects of any other disciplinary response to noncompliance that could replace spanking.

1. Anti-spanking researchers have investigated 14 disciplinary tactics with the correlational methods used to oppose spanking, but they have yet to find one that is significantly related to reductions in behavior problems.⁹²
2. We suspect that is because the correlational methods they use to oppose spanking make all corrective disciplinary responses to noncompliance look more harmful than they are.⁹³

For Presentation on “Is Spanking a Victim of Scholar Advocacy Bias?” at the Conference of the Association of Family and Conciliation Courts, Nov. 1, 2019, Pittsburgh. Robert.larzelere@okstate.edu law.fuller@gmail.com

REFERENCES

-
- ¹ R. E. Larzelere & D. Baumrind, *Are spanking injunctions scientifically supported?*, 73 LAW AND CONTEMPORARY PROBLEMS 57 (2010).
 - ² Jason Fuller, *The science and statistics behind spanking suggest that laws allowing corporal punishment are in the best interests of the child*, 42 AKRON LAW REVIEW 243 (2009).
 - ³ E. T. Gershoff & A. Grogan-Kaylor, *Spanking and child outcomes: Old controversies and new meta-analyses*, 30 JOURNAL OF FAMILY PSYCHOLOGY 453 (2016).
 - ⁴ R. D. Sege et al., *Effective discipline to raise healthy children*, 142 PEDIATRICS 6 (2018).
 - ⁵ American Psychological Association, *Resolution on physical discipline of children by parents*(2019), available at <https://www.apa.org/about/policy/physical-discipline.pdf>.
 - ⁶ E. T. Gershoff et al., *The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers*, 73 AMERICAN PSYCHOLOGIST 626 (2018).
 - ⁷ R. E. Larzelere, Larzelere, M. L. Gunnoe, & C. J. Ferguson, *Improving causal inferences in meta-analyses of longitudinal studies: Spanking as an illustration*, 89 CHILD DEVELOPMENT 2038 (2018).
 - ⁸ *Id.* at 2042.
 - ⁹ Gershoff & Grogan-Kaylor, *supra* note 3, at 463 (61 of 111 effect sizes were cross-sectional).
 - ¹⁰ *Id.* at 456.
 - ¹¹ *Id.* at 463.
 - ¹² Larzelere et al., *supra* note 7 at 2044.
 - ¹³ C. J. Ferguson, *Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies*, 33 CLINICAL PSYCHOLOGY REVIEW 196, 196 (2013).
 - ¹⁴ Larzelere et al., *supra* note 7 at 2044.
 - ¹⁵ E. L. Hamaker et al., *A critique of the cross-lagged panel model*, 20 PSYCHOLOGICAL METHODS 102, 102, 111-113 (2015); D. Berry & M. T. Willoughby, *On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse*, 88 CHILD DEVELOPMENT 1186, 1186, 1202 (2017).
 - ¹⁶ Berry & Willoughby, *supra* note 15 at 1186, 1202, Online Appendix S2 at 4-8.
 - ¹⁷ Larzelere et al., *supra* note 7 at 2044.
 - ¹⁸ R. E. Larzelere et al., *Longitudinal biases against corrective actions*, 6 ARCHIVES OF SCIENTIFIC PSYCHOLOGY 243, 243-5 (2018).
 - ¹⁹ *Id.* at 244. Hereinafter, β refers to a standardized regression coefficient. This can be interpreted, like a correlation, as the strength and direction of the association between two variables (e.g., a corrective action and a child outcome). The difference is that β adjusts statistically for other variables (in this case initial scores on the outcome variable).

²⁰ R. E. Larzelere et al., *Do nonphysical punishments reduce antisocial behavior more than spanking? A comparison using the strongest previous causal evidence against spanking*, 10 (10) BMC PEDIATRICS 1, 8 (2010); R. E. Larzelere et al., *Differences in causal estimates from longitudinal analyses of residualized versus simple gain scores: Contrasting controls for selection and regression artifacts*, 34 INTERNATIONAL JOURNAL OF BEHAVIORAL DEVELOPMENT 180, 185 (2010).

²¹ Larzelere et al., BMC PEDIATRICS, *supra* note 20 at 8.

²² *Id.* at 8; Larzelere et al., INTERNATIONAL JOURNAL, *supra* note 20 at 185.

²³ G. Silinskas et al., *The developmental dynamics of children's academic performance and mothers' homework-related affect and practices*, 51 DEVELOPMENTAL PSYCHOLOGY 419, 419, 426-7 (2015); C. Carmichael & A. MacDonald, *Parental influences on primary school children's mathematics achievement: insights from the Longitudinal Study of Australian Children*, 44 EDUCATION 197, 197, 206 (2016).

²⁴ D. P. Deptula et al., *How can parents make a difference? Longitudinal associations with adolescent sexual behavior*, 24 JOURNAL OF FAMILY PSYCHOLOGY 731, 731, 736 (2010).

²⁵ R. N. H. de Leeuw et al., *Do interactions between personality and social-environmental factors explain smoking development in adolescence*, 24 JOURNAL OF FAMILY PSYCHOLOGY 68, 73-4 ("Frequency of communication" coefficient predicting Slope (gain) in smoking in tbl 2) (2010).

²⁶ L. Keijsers et al., *Forbidden friends as forbidden fruit: Parental supervision of friendships, contact with deviant peers, and adolescent delinquency*, 83 CHILD DEVELOPMENT 651, 659-660, fig. 3 and accompanying text (2012).

²⁷ Larzelere et al., ARCHIVES, *supra* note 18 at 245.

²⁸ Larzelere et al., INTERNATIONAL JOURNAL, *supra* note 20 at 185; Larzelere et al., BMC PEDIATRICS, *supra* note 20 at 8.

²⁹ Larzelere et al., INTERNATIONAL JOURNAL, *supra* note 20 at 185.

³⁰ R. E. Larzelere et al., *Trying to overcome selection bias in longitudinal analyses of corrective actions by professionals* 1, fig. 2 (paper presented at the conference of the Society for Research in Child Development, Austin, TX) (2017); H. Lin, *Revealing and resolving contradictory ways to reduce selection bias to enhance the validity of causal inferences from non-randomized longitudinal data*, 1, 77-79 (2018) (dissertation, Oklahoma State University).

³¹ Larzelere et al., *supra* note 30 at fig. 2; Lin, *supra* note 30 at 77-79.

³² Larzelere et al., *supra* note 30 at fig. 1.

³³ *Id.*

³⁴ L. M. Berger et al., *Estimating the 'impact' of out-of-home placement on child well-being: Approaching the problem of selection bias*, 80 CHILD DEVELOPMENT 1856, 1868-9 (2009). Model 4 for Internalizing and Externalizing in Table 3 and corresponding text summary, Models 5, 6, & 7 and their text summary show that the apparent effect reverses when predicting simple change scores, similar to analyses of simple change scores for spanking in meta-analysis by Larzelere et al., *supra* note 30 at 2044.

³⁵ Miriam J. Maclean et al., *Out-of-home care versus in-home care for children who have been maltreated: A systematic review of health and wellbeing outcomes*, 25 CHILD ABUSE REVIEW 251, 254, 261 (2016).

³⁶ Berger et al., *supra* note 34 at 1868. Hereinafter, *residualized change* refers to a statistical estimate of the association of a cause (e.g., out-of-care or spanking) on a subsequent child outcome (e.g., antisocial aggression) after removing what can be predicted in subsequent outcome scores from initial scores on that same variable and predicting only what is left (residual) after that removal, as in analysis of covariance (ANCOVA).

³⁷ Larzelere et al., *supra* note 7 at 2044.

³⁸ Larzelere et al., INTERNATIONAL JOURNAL, *supra* note 20 at 182-185; Larzelere et al., BMC PEDIATRICS, *supra* note 20 at 8-9; Berry & Willoughby, *supra* note 15 at Supplemental Material, 4-8.

³⁹ C. M. Herbst & E. Tekin, *Child care subsidies, maternal health, and child-parent interactions: Evidence from three nationally representative datasets*, 23 HEALTH ECONOMICS 894, 904, tbl. III (2014) (right-hand column).

⁴⁰ S. M. Director, *Underadjustment bias in the evaluation of manpower training*, 3 EVALUATION QUARTERLY 190, 194-6 (1979).

⁴¹ Gershoff & Grogan-Kaylor, *supra* note 3 at 453; Larzelere et al., ARCHIVES, *supra* note 18 at 244.

⁴² E. T. Gershoff, *Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review*, 128 PSYCHOLOGICAL BULLETIN 539 (2002).

⁴³ Gershoff et al., AMERICAN PSYCHOLOGIST, *supra* note 6 at 628-9.

⁴⁴ Sege et al., PEDIATRICS, *supra* note 4 at 3.

⁴⁵ R. E. Larzelere et al., *Should pediatricians base their parenting advice on advocacy or science (Comment)*, 142 (6) PEDIATRICS (2018) 1, 1; R. E. Larzelere et al., *The insufficiency of the evidence used to categorically oppose spanking and its implications for families and psychological science: Comment on Gershoff et al. (2018)*, 74 AMERICAN PSYCHOLOGIST 497, 497-499 (2019); R. P. Rohner & T. Melendez-Rhodes, *Perceived parental acceptance--rejection mediate or moderates the relation between corporal punishment and psychological adjustment: Comment on Gershoff et al. (2018)*, 500, 500-502; E. T. Gershoff et al., *There is still no evidence that physical punishment is effective or beneficial: Reply to Larzelere, Gunnoe, Ferguson, and Roberts (2019) and Rohner and Melendez-Rhodes (2019)*, 503, 503-505; R. D. Sege & B. S. Siegel, *AAP Authors of "Effective discipline to raise healthy children" response to "Should pediatricians base their parenting advice on advocacy or science?" (comment)*, 142 (6) PEDIATRICS 1 (2018).

⁴⁶ Gershoff & Grogan-Kaylor, *supra* note 3 at 456 (“bivariate associations such as standardized differences of means or correlations were selected over adjusted coefficients”); *id.* at tbl. 4 at 464.

⁴⁷ E. O. Paolucci & C. Violato, *A meta-analysis of the published research on the affective, cognitive, and behavioral effects of corporal punishment*, 138 JOURNAL OF PSYCHOLOGY 197, 197 (2004).

⁴⁸ R. E. Larzelere & B. R. Kuhn, *Comparing child outcomes of physical punishment and alternative disciplinary tactics: A meta-analysis*, 8 CLINICAL CHILD AND FAMILY PSYCHOLOGY REVIEW 1, 1 (2005).

⁴⁹ Ferguson, *supra* note 13 at 196.

⁵⁰ Larzelere et al., *supra* note 7 at 2042.

⁵¹ Ferguson, *supra* note 13 at 198.

⁵² Larzelere et al., *supra* note 7 at 2042.

⁵³ Larzelere & Kuhn, *supra* note 48 at 3 (“If the apparently detrimental child outcome reflect causal effects unique to physical punishment, then the effect sizes of physical punishment should compare poorly to the effect sizes of alternative disciplinary tactics. On the other hand, if detrimental child correlates of physical punishment reflect methodological artifacts, then the effect sizes of alternative disciplinary tactics should appear equally detrimental.”).

⁵⁴ Gershoff, PSYCHOLOGICAL BULLETIN, *supra* note 42 at 545-547.

⁵⁵ Gershoff & Grogan-Kaylor, *supra* note 3 at 456 (“bivariate associations such as standardized differences of means or correlations were selected over adjusted coefficients from multivariate models”). See frequency of each study design. *Id.* at 463.

⁵⁶ Paolucci & Violato, *supra* note 47 at 211.

⁵⁷ Sege et al., PEDIATRICS, *supra* note 4 at 2.

⁵⁸ C. Miller-Perrin, *News from the Interdivisional Task Force on Physical Punishment for Children*, 38 THE ADVOCATE: NEWSLETTER OF DIVISION 37 OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION 9, 9 (2015). APA’s Task Force on Physical Punishment was apparently designed to reach a preordained conclusion: “Recently Division 37, along with Division 7 (Developmental Psychology), established the joint Task Force on Physical Punishment of Children. The ultimate purpose of the task force is to review the scientific literature on the effects of physical discipline on children and to disseminate information about its dangers, in an effort to both reduce parents’ use of physical discipline and increase alternative forms of parenting. . . . In addition, several members of the task force published op-eds {and other media information} to discuss the negative outcomes associated with physical punishment of children,” including Elizabeth Gershoff, George Holden, and Alan Kazdin.

⁵⁹ The only two Task Force members known to have published original research on physical discipline were known advocates of spanking bans: G. W. Holden, *Global Summit on Ending Corporal Punishment and Promoting Positive Discipline* (http://scholar.smu.edu/global_summit/) (2011). E. T. Gershoff & S. H. Bitensky, *The case against corporal punishment of children - Converging evidence from social science research and international human rights law and implications for US public policy*, 13 PSYCHOLOGY PUBLIC POLICY AND LAW 231 (2007).

⁶⁰ Gershoff et al., AMERICAN PSYCHOLOGIST, *supra* note 6 at 628; Sege et al., PEDIATRICS, *supra* note 4 at 8-9.

⁶¹ Gershoff et al., AMERICAN PSYCHOLOGIST, *supra* note 6 at 626-632, 635.

⁶² M. A. Southam-Gerow & M. J. Prinstein, *Evidence Base Updates: The Evolution of the Evaluation of Psychological Treatments for Children and Adolescents*, 43 JOURNAL OF CLINICAL CHILD AND ADOLESCENT PSYCHOLOGY 1 (2014).

⁶³ *Id.* at 2 (“Evidence criteria” for Level 2: Probably Efficacious Treatments: “There must be at least two good experiments showing the treatment is a superior (statistically significantly so) to a wait-list control group” or “One (or more) experiments meeting the Well-Established Treatment level except . . . {the experiments did} not involve independent investigatory teams.” The latter level includes randomized evidence that a treatment is “equivalent (or not statistically different) to an already well-established treatment” in tbl. 1).

⁶⁴ D. O’Connor et al., *Defining the review question and developing criteria for including studies*, in COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS 83, 86 (J. P. T. Higgins & S. Green eds., 2008); METHODS GROUP OF THE CAMPBELL COLLABORATION, *METHODOLOGICAL EXPECTATIONS OF CAMPBELL COLLABORATION INTERVENTION REVIEWS: CONDUCT STANDARDS* 1, 4 (Campbell Collaboration. 2016) (Item #7, 16; Item #50, 21; Item #63).

⁶⁵ A. W. Bean & M. W. Roberts, *The effect of time-out release contingencies on changes in child noncompliance*, 9 JOURNAL OF ABNORMAL CHILD PSYCHOLOGY 95, 102-103 (1981).

⁶⁶ Gershoff et al., AMERICAN PSYCHOLOGIST, *supra* note 6 at 631; Gershoff & Grogan-Kaylor, *supra* note 3 at 464.

⁶⁷ M. W. Roberts & S. W. Powers, *Adjusting chair timeout enforcement procedures for oppositional children*, 21 BEHAVIOR THERAPY 257, 263 (1990).

⁶⁸ Larzelere & Kuhn, *supra* note 48 at 20.

⁶⁹ Roberts & Powers, *supra* note 67 at 257; M. W. Roberts, *Enforcing chair timeouts with room timeouts*, 12 BEHAVIOR MODIFICATION 353 (1988); D. E. Day & M. W. Roberts, *An analysis of the physical punishment component of a parent training program*, 11 JOURNAL OF ABNORMAL CHILD PSYCHOLOGY 141 (1983).

⁷⁰ T. L. HEMBREE-KIGIN & C. B. MCNEIL, *PARENT-CHILD INTERACTION THERAPY* 94-95 (Plenum Press. 1995); R. L. FOREHAND & R. J. MCMAHON, *HELPING THE NONCOMPLIANT CHILD* 79-80 (Guilford 1981) (“While we are basically opposed to physical punishment, we have found a mild spanking to be the most feasible backup for the child leaving the timeout chair”). Other similar treatments used the spank enforcement: R. A. BARKLEY, *DEFIANT CHILDREN: A CLINICIAN'S MANUAL FOR PARENT*

TRAINING 117-118 (Guilford. 1987); E. R. CHRISTOPHERSEN, *LITTLE PEOPLE: GUIDELINES FOR COMMON SENSE CHILD REARING* 45 (Westport 3rd ed. 1988).

⁷¹ S. M. Eyberg et al., *Evidence-based psychosocial treatments for children and adolescents with disruptive behavior*, 37 *JOURNAL OF CLINICAL CHILD AND ADOLESCENT PSYCHOLOGY* 215, 226-230 (2008). Of the 4 most effective treatments for children under the age of 7, two had used the spank enforcement: Forehand & McMahon's and Hembree-Kigin & McNeil's. The other two were Incredible Years and the Triple P Positive Parenting Program, which were also based on behavioral parent training, but never used the spank enforcement.

⁷² G. E. Everett et al., *Time-out with parents: A descriptive analysis of 30 years of research*, 33 *EDUCATION & TREATMENT OF CHILDREN* 235, 247 (2010).

⁷³ Larzelere & Baumrind, *supra* note 1 at 71-72.

⁷⁴ Theodore P. Beauchaine et al., *Mediators, Moderators, and Predictors of 1-Year Outcomes Among Children Treated for Early-Onset Conduct Problems: A Latent Growth Curve Analysis*, 73 *JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY* 371, 374 n. 2 (2005) (physical punishment was only 1 of 6 Harsh Parenting items: slap, spank, hit; physical force; lead, drag to corner; 'bad' Time Out; push away; argue, fight).

⁷⁵ *Id.* at 379-80, 383-4. Note that what they call a *mediating effect* (path a in fig. 4, panel c) is the association between simultaneous change in Harsh Parenting and change in Externalizing (slope to slope), which cannot tell which changed first.

⁷⁶ S. M. Breitenstein et al., *The Chicago Parent Program: Comparing 1-year outcomes for African American and Latino parents of young children*, 35 *RESEARCH IN NURSING & HEALTH* 475,482-4 (2012).

⁷⁷ A. Grogan-Kaylor, *The effect of corporal punishment on antisocial behavior in children*, 28 *SOCIAL WORK RESEARCH* 153, 155-157 (2004). Note that the equation at 157 has the same indicator of time (t) for the outcome (antisocial behavior) as for the predictor (spanking), so the prediction is between question responses in the same interview, from spanking in the past week to antisocial behavior in the past 3 months, which is the time period specified in the following handbook: P. C. BAKER et al., *NLSY CHILD HANDBOOK: A GUIDE TO THE 1986-1990 NATIONAL LONGITUDINAL SURVEY OF YOUTH: CHILD DATA* 105 (Center for Human Resource Research, Ohio State University Rev. ed. 1993).

⁷⁸ P. M. Steiner et al., *The importance of covariate selection in controlling for selection bias in observational studies*, 15 *PSYCHOLOGICAL METHODS* 250, 260-2, 265-6 (2010). (e.g., "three different strategies for selecting covariates are possible. First, if all the covariates that influence the selection process are measured and correctly modeled . . . The second best strategy consists in theoretically and empirically grounded investigations of the selection mechanism . . . the third is least desirable"). The point is that variables important in the process by which parents select spanking are essential to get unbiased causal estimates, and Gershoff et al. failed to include any such variable in the covariates used for propensity-score matching.

⁷⁹ E. T. Gershoff et al., *Strengthening causal estimates for links between spanking and children's externalizing behavior problems*, 29 *PSYCHOLOGICAL SCIENCE* 110, 114-7 (2018).

⁸⁰ Gershoff et al., *AMERICAN PSYCHOLOGIST*, *supra* note 6 at 503.

⁸¹ Larzelere & Kuhn, *supra* note 48 at 1.

⁸² Roberts & Powers, *supra* note 67 at 260.

⁸³ M. A. Straus & V. E. Mouradian, *Impulsive corporal punishment by mothers and antisocial behavior and impulsiveness of children*, 16 *BEHAVIORAL SCIENCES AND THE LAW* 353, 357 (1998).

⁸⁴ Roberts & Powers, *supra* note 67 at 260.

⁸⁵ R. E. Larzelere et al., *Punishment enhances reasoning's effectiveness as a disciplinary response to toddlers*, 60 *JOURNAL OF MARRIAGE AND THE FAMILY* 388, 394-399 (1998).

⁸⁶ K. L. Ritchie, *Maternal behaviors and cognitions during discipline episodes: A comparison of power bouts and single acts of noncompliance*, 35 *DEVELOPMENTAL PSYCHOLOGY* 580, 587 (1999); Larzelere & Kuhn, *supra* note 48 at 7, 34; R. E. Larzelere & B. R. Kuhn, *Immediate effectiveness of disciplinary tactics by type of noncompliance (A re-analysis of data from Ritchie, 1999)* 1, 2, 4 (Poster presented at the conference of the American Psychological Association, Washington, DC) (2005).

⁸⁷ M. L. Gunnoe, *Associations between parenting style, physical discipline, and adjustment in adolescents' reports*, 112 *PSYCHOLOGICAL REPORTS: DISABILITY & TRAUMA* 933, 957-959 (2013).

⁸⁸ C. G. Ellison et al., *Does conservative Protestantism moderate the association between corporal punishment and child outcomes?*, 73 *JOURNAL OF MARRIAGE AND FAMILY* 946, 946, 954-956 (2011).

⁸⁹ K. D. Bussmann et al., *Effects of banning corporal punishment in Europe: A five-nation comparison*, in *GLOBAL PATHWAYS TO ABOLISH PHYSICAL PUNISHMENT: REALIZING CHILDREN'S RIGHTS* 299, 316-317 (J. E. Durrant & A. B. Smith eds., 2011).

⁹⁰ D. Baumrind, *The development of instrumental competence through socialization*, in *MINNESOTA SYMPOSIA ON CHILD PSYCHOLOGY* 3, 35 (A. D. Pick ed. 1973).

⁹¹ Gershoff et al., *AMERICAN PSYCHOLOGIST*, *supra* note 6 at 503.

⁹² E. T. Gershoff et al., *Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness*, 81 *CHILD DEVELOPMENT* 487, 493-495 (2010); K. G. Van Leeuwen et al., *Assessing dimensions of parental discipline*, 34 *JOURNAL OF PSYCHOPATHOLOGY AND BEHAVIORAL ASSESSMENT* 216, 227-228 (2012); R. E. Larzelere & R. B. Cox, Jr., *Making valid causal inferences about corrective actions by parents from longitudinal data*, 5 *JOURNAL OF FAMILY THEORY & REVIEW* 282, 283 (2013).

⁹³ Larzelere et al., ARCHIVES, *supra* note 18 at 244.