JOURNAL
OF THE CANADIAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY / DE L'ACADÉMIE CANADIENNE DE PSYCHIATRIE DE L'ENFANT ET DE L'ADOLESCENT

INVITED COMMENTARY

An update on the scientific evidence *for* and *against* the legal banning of disciplinary spanking

Marjorie Lindner Gunnoe PhD[1]; Robert E. Larzelere, PhD[2]; Christopher J. Ferguson PhD[3]; Ronald B. Cox, Jr. PhD[2]

Abstract

In 2004, the Canadian Supreme Court upheld the right of Canadian parents to use open-handed swats to a child's buttocks to correct child misbehavior, but only as developmentally appropriate (i.e., between the ages of 2 and 12). Some social scientists believe that the Canadian Supreme Court did not go far enough. These researchers support *total bans* on physical discipline. Other social scientists support Canada's existing *age-specific legislation*. This commentary provides a concise overview of physical discipline research since 2004, emphasizing the methodological rigor of the research used to argue *for* and *against* total spanking bans. Advocates of total bans primarily cite reviews based on bivariate correlations and non-randomized methods known to be inherently biased against disciplinary actions (i.e., methods known to make *all* disciplinary responses to defiance *appear* harmful). In contrast, those who support Canada's existing legislation have systematically compared methods known to be inherently biased in opposite directions (i.e., harmful- and beneficial-looking), to demonstrate that the true average effect size of customary spanking on child outcomes is likely very near zero. These researchers also emphasize four randomized clinical trials in which spanking increased compliance in defiant preschoolers. Other issues discussed in this commentary are: the developmental trajectories for children who do not learn to comply with parental directives while they are young; and children's risk of assault in countries with and without total bans. We conclude that the most rigorous empirical studies and available crime statistics validate the appropriateness of Canada's existing legislation on disciplinary spanking.

Key Words: Spanking, physical discipline, corporal punishment, legal ban, Canada, Sweden

[1]Calvin University, Grand Rapids, Michigan, USA
[2]Oklahoma State University, Stillwater, Oklahoma, USA
[3]Stetson University, DeLand, Florida, USA
Corresponding E-Mail: marjorie.gunnoe@calvin.edu
Submitted: January 19, 2024; Accepted: November 20, 2024

A primary goal of parenting is to ensure that children become well-adjusted, responsible adults. Most parents and professionals recognize that good parenting requires some discipline and thus welcome guidance about the most appropriate ways to administer discipline. Toward this end, all three levels of the Canadian courts have wrestled with extensive social scientific and legal evidence *for* and *against* disciplinary spanking.

In 2004, the Canadian Supreme Court agreed with two lower-court decisions to retain the right of parents to use reasonable force to correct their children (Section 43), while limiting legal forms of reasonable force (1). The court saw value in permitting parents to use open-handed swats to a child's buttocks to correct child misbehavior, but only as developmentally appropriate (i.e., between the ages of 2 and 12).

This commentary provides a concise overview of physical discipline research since 2004. In Part I, we raise some general methodological considerations that consumers of this research should attend to. We do this to set up Part II, in which we review the specific studies most frequently cited to argue *for* and *against* spanking bans. In Part III, we raise additional considerations for parents and professionals concerned about the well-being of Canada's children. We conclude by lauding the appropriateness of Canada's current approach to the legislation of disciplinary spanking.

Part I: Methodological Considerations

Spanking is a corrective action, that is, an action taken to correct (or ameliorate) a problem. Examples of corrective actions include: chemotherapy (to correct cancer), methylphenidate (to correct ADHD symptomology), and spanking and other disciplinary responses (to correct child misbehavior). When evaluating studies of corrective actions, it is imperative to know: how the corrective action is being defined, the presenting problems in the population receiving the corrective action, the dosage regimen, and whether the corrective action works alone or in the context of a comprehensive treatment package. Also critical, is an understanding of whether the evidence obtained permits causal inference.

Definition, age, regimen/context

How researchers define spanking typically reflects their predisposing views on spanking. Researchers who consider any parental use of physical force to be abusive sometimes fail to distinguish spanking from actions that *all* discipline researchers view as abusive (e.g., kick, beat up). When defining "physical discipline" or "physical punishment" for review articles, these researchers often cast a wide inclusion net, lumping spanking together with many forms of severe and abusive punishment. Specific examples of these problematically-broad definitions have been identified by Baumrind et al. (2) and Larzelere et al. (3). Broad definitions can make non-harmful types of physical discipline appear harmful due to the conflation of spanking with severe and abusive physical punishment.

In contrast, researchers who believe that spanking can be administered in a non-abusive way distinguish "customary spanking" from overly harsh/abusive discipline, and sometimes make even finer discriminations. For example, some use the term "conditional spanking" (4) to refer to spanking used to back-up milder techniques (e.g., for refusal to cooperate with time-out).

Researchers opposed to spanking have also tended to amalgamate children aged 0-17 in their analyses (5,6). Lumping children with infants and teens (for whom physical discipline is not developmentally appropriate) elevates group means for undesirable outcomes and reduces the means for desirable ones. Physical discipline may appear harmful for the full group, even if it was neutral or beneficial for children in a specific age range.

In contrast, researchers who strive to distinguish spanking from abuse are more likely to limit their investigations to children aged 18 months through 11 years. As such, research conducted by this group is more *directly* applicable to the current legal debate in Canada over non-abusive spanking with children ages 2-12.

Four categories of evidence

Predisposing views also impact the types of evidence that researchers create, cite, and emphasize. Four broad categories of evidence are distinguished here (to set up our presentation of the *specific* studies most cited in the current spanking debate).

The first category is **correlational** evidence. Because corrective actions are typically applied only *when there is a problem to correct*, both cross-sectional and longitudinal studies will virtually always yield a positive correlation between the use of a corrective action and the severity of the problem it is intended to correct. (Methylphenidate use will always be positively correlated with ADHD symptomatology; discipline use will always be positively correlated with frequency of misbehavior.) As with corrective medical actions, no causal inferences can be drawn from correlational studies of spanking and child outcomes.

The second category is **baseline-partialled longitudinal** (BPL) evidence (aka *controlled longitudinal* (3)) evidence). When predicting the future severity of a problem, BPL studies statistically control for (i.e., partial out) the severity of the problem at baseline. BPL studies are often used to make causal inferences in the social sciences, even though most BPL analyses are known to reduce -- but not eliminate -- inherent statistical biases.

The most common type of BPL study used to assess spanking relies on ANCOVA-type statistics. ANCOVA-type longitudinal analyses have been shown to be biased *against* corrective actions (i.e., tend to make these actions appear harmful: 7-9). ANCOVA analyses of corrective actions yield a positive association between baseline correction and subsequent severity because people with more problems at baseline typically have more problems at follow-up, compared to those who had no need for the corrective action in the first place. ANCOVA only partially reduces the selection confound. This apparent harm (i.e., positive association) has been documented not just for spanking, but also for nonphysical disciplinary responses, psychotherapy, methylphenidate (9-11) and out-of-home placements (12).

Another type of BPL study relies on *change-score* (aka "slope") *analyses* wherein the corrective action is used to predict change in the severity of problems from baseline to follow-up (i.e., the slope of the line from baseline to follow-up). Change score analyses are typically biased *in favor* of corrective actions. Because a corrective action is usually applied when a problem is more severe (e.g., a child is particularly defiant), there may be natural regression toward the mean prior to follow-up. Statistical methodologists have been exploring ways to overcome the biases particular to these two BPL methods (discussed later).

The third category of evidence is **randomized experiments.** Randomized experiments are the standard in medical research but occur rarely in child discipline research. We know of only four randomized experiments of spanking, conducted when clinical researchers were searching for alternatives to the traditional use of spanking to enforce compliance with a chair time-out. In these studies conducted by Roberts and his colleagues (13-16), clinically-referred 2- to 6-year-olds who left chair time-out prematurely were prescribed "two swats" versus an alternative "back-up" procedure, and then returned to the chair. Unfortunately, these studies are now quite dated (1981-1990) and an anti-punishment bias among many contemporary parenting researchers will likely preclude their replication.

The fourth category of evidence is a **literature review.** Some reviews are *narrative* (box-score). Authors of narrative reviews simply count the number of studies documenting harmful- or beneficial-looking results. Other reviews are *meta-analytic*. Authors of meta-analyses (MAs) calculate and provide average effect sizes. Regardless of the type, a review is only as good as the individual studies on which it is based. Causal inference cannot be drawn from reviews of bivariate correlations; reviews

of ANCOVAs or change-scores are subject to the biases inherent in these two baseline-adjustment methods. Consequently, reviews of randomized trials are strongly preferred in medical research (17).

Part II: Studies Used to Argue *For* and *Against* Spanking Bans

Studies cited in Canada's 2004 Supreme Court decision

The 2004 court decision emphasized two reviews and two BPL studies. The first review was a narrative review from a scientific consensus conference on corporal punishment (18). The second review was a preliminary draft of Gershoff's (5) MA based exclusively on bivariate correlations obtained from studies of children 0-17 (*mean* ages 1-16). The individual studies were the first two BPL studies of physical discipline (19, 20). Like the authors of the two reviews, the authors of the two BPL studies came to *some* shared and *some* contradictory conclusions.

Studies emphasized to promote spanking bans

Since the 2004 Supreme Court ruling, researchers who are unconditionally-opposed to spanking have tended to emphasize three types of studies.

1. Meta-analyses of bivariate correlation coefficients. Some of the most frequently cited studies of physical discipline are Gershoff's initial MA (5) and an updated version co-authored with Grogan-Kaylor (21). The update is an improvement over the initial version, in that 24% of the studies in the update were longitudinal, and many studies lumping spanking with severe corporal punishment were omitted. But correlations are still correlations, which the researchers acknowledge "do not rule out the potential for a child elicitation effect" (p. 455). Thus, "causal links between spanking and child outcomes cannot be established by these meta-analyses" (21, p. 464).

Despite these acknowledgements, the *American Academy of Pediatrics* (AAP) and the *American Psychological Association* (APA) repeatedly cited these unadjusted correlations in their empirical rationale for officially opposing all spanking in 2018 (22) and 2019 (23, 24). In response to various critiques of this evidence (25, 26), the APA Task Force cited a lack of evidence of any beneficial outcome of spanking (27).

2. Baseline-partialled longitudinal studies (mostly ANCOVA-type). A more recent narrative review (6) went beyond bivariate correlations by tallying the number of BPL studies that reported harmful- and beneficial-looking associations of baseline physical punishment with subsequent child outcome. Children had *mean* ages of 6 months to 14 years when punished. Because the vast majority of these studies employed ANCOVA-type statistics (known to be biased against corrective actions), it is not surprising that 69% of the studies reported harmful-looking effects. Only 1% found more beneficial- than harmful-looking outcomes; 30% had mixed significant and non-significant outcomes. The researchers presented this tally as "compelling evidence that physical punishment is harmful to children's development and wellbeing," asserting that all countries should therefore join UN efforts "by prohibiting physical punishment in all forms and all settings" (6, p. 361).

3. Attempts to correct known bias. Attempts to overcome the selection bias inherent in ANCOVA-type analyses of spanking have included the use of: propensity score methods (28), entropy balancing (29), and fixed-effects regression (30). Unfortunately, systematic comparisons have revealed that propensity score methods do not necessarily overcome the selection bias (31, 32). For example, Lin (33) found that both prescription medications for mothers and therapy for mothers still *looked* harmful (i.e., were positively associated with their own subsequent depression severity) according to propensity-score matching. Entropy balancing can be understood as a flexible propensity score method (32). Fixed-effects regression can be less biased than ANCOVA-type analyses when researchers attend carefully to temporal sequence, but many datasets do not lend themselves to the hypotheses researchers want to test. In one application of fixed-effects regression to physical discipline (30), a spanking-ban advocate demonstrated that above average spanking for a specific child in the *past week* was associated with above-average antisocial behavior during the *past three months* (i.e., temporal direction was reversed, providing evidence for child elicitation of spanking rather than a spanking effect).

Despite recent advances in approximating unbiased causal effects in longitudinal studies, eliminating selection bias remains a difficult challenge (32). Studies *have* demonstrated the importance of partialling out the most relevant covariates (31), identifying treatment assignment mechanisms, and designing studies to focus on similar comparison groups (34) (e.g., equivalent presenting problems, such as oppositional defiance). Meta-analyses of BPL studies suggest that the most important covariates to partial when conducting spanking research are baseline child adjustment and child age (3, 35, 36).

Summary of the evidence used to support spanking bans. The most cited empirical evidence used to promote spanking bans are reviews of children aged 0 to 17 years that are either correlational or based on statistical methods known to be biased against corrective actions. Efforts to eliminate these biases have not been convincing.

Studies emphasized to support Canada's current age-specific approach

Researchers who support Canada's existing age-delimited approach to spanking (i.e., only between the ages of 2 and 12) tend to focus on four lines of evidence.

1. Meta-analyses of baseline-partialled longitudinal studies. Three MAs of physical discipline have gone beyond correlations by adjusting for baseline differences on child outcomes (3, 35, 36). These MAs also distinguished customary spanking from more severe physical discipline. Two were limited to children aged 18 months through 11 years (3, 36).

In all three of these MAs, the researchers emphasized the *trivial* effect sizes obtained from BPL studies of customary spanking. Although the majority of qualifying studies in each MA (45-79%) did report a harmful-looking effect, spanking explained *less than* 1% of the variance in child outcome (0.64% for externalizing problems; 0.16% for internalizing, cognitive achievement, and social competence in the most recent MA (3)). Effect sizes this small can easily be explained as *residual confounding* (37, p. 99) due to researchers' failure to include all relevant covariates.

The researchers also addressed the biases inherent in the usual ANCOVA analyses and in change-score analyses. The two most recent MAs computed effect sizes for the *two* (oppositely-biased) methods. As expected, average ANCOVA results suggested *harmful*-looking effects of spanking, and average change-score results suggested *beneficial*-looking effects of spanking. In other words, all "effects" were consistent with each method's known bias and too small to overcome the opposite bias in the other method.

Per Angrist and Pischke's (38) assertion that these two methods likely bracket the true effect, the researchers concluded that the average effects of customary spanking on child outcomes are very near zero. If these results occurred for a medical intervention, researchers and physicians would next attempt to evaluate the effectiveness of *various implementations* of the corrective action (as a function of presenting problems, dosage regimen, other patient characteristics, etc.). This would be consistent with the likelihood of genuinely heterogeneous effects demonstrated by Imbens & Xu (32). Unfortunately, such nuances are largely ignored in the research on parental discipline.

2. Comparisons of spanking with alternative discipline tactics. Another way to evaluate spanking is to compare it with its alternatives -- investigated with the *same* methods in the *same* families. The only MA comparing effect sizes for various types of physical discipline with alternative disciplinary tactics (4) indicated that children *benefitted* the most from *conditional* spanking (i.e., spanking used only to back up milder techniques). For increasing compliance and reducing antisocial behavior, conditional spanking outperformed 10 of 13 alternatives. Not surprisingly, children were *harmed* the most when physical discipline was overly severe or was parents' predominant response to children's misbehavior. Most studies in this MA had investigated *customary* spanking (i.e., "spanking" as defined by the parent), which performed in the middle. Child outcomes associated with customary spanking did not differ from those associated with other disciplinary responses (e.g., privilege removal, reasoning), except in one retrospective study favoring customary spanking for drug abuse eight or more years later (39).

Subsequent to this MA, there have been three more BPL studies that compared spanking with alternative disciplinary techniques. Two utilized longitudinal data from the US (10, 40). The third (11) is the only known BPL study of physical punishment in Canada. Table 1 is an updated synthesis of the results of these three studies with other published mean effect sizes, demonstrating the pattern of biases typical of all corrective actions, c.f. (41).

Table 1. Standardized Associations of Three Corrective Actions with Externalizing-Type Problems by Type of Evidence[a]

Type of Evidence	Spanking	Nonphysical Punishments	Psychotherapy
Non-randomized studies Cross-sectional correlations	.20***[b]	.17***[d]	.10**[d]
Longitudinal correlations	.16***[b]	.19***[e]	.16***[e]
ANCOVA/beta-type controls for initial differences	.07***[b]	.05*[e]	.12*[e]
Within-person changes (slope)	-.04*[b]	-.05*[e]	.00[e]
Randomized trials Corrective action vs. control group	-.35*[c]	-.63*[f]	-.24***[g]

[a] All coefficients are standardized. For all cell elements, positive coefficients indicate harmful-looking associations (e.g., spanking is correlated with more externalizing-type problems); negative coefficients indicate beneficial-looking associations.

[b] Mean effect sizes (*r*, b) from Larzelere, Gunnoe, & Ferguson (36, p. 2044, Table 1) for externalizing. These values are similar to correlations reported by Gershoff & Grogan-Kaylor (21) and to the ANCOVA-type coefficients reported by Ferguson (35).

[c] Difference between the effect size *d* for conditional physical punishment vs. the control condition (child release from timeout) for compliance reported by Larzelere & Kuhn (4, p. 20, top half of Table IV; *d* = -.74 is equivalent to *r* = -.35).

[d] Unweighted average of effect sizes for three nonphysical tactics and effect size for therapy with the antisocial outcome in Larzelere, Cox, & Smith (10, p. 6, Table 2).

[e] Nonphysical punishment is the unweighted average of three effects from three studies by Larzelere and colleagues (each study contributing one effect). The first effect is for one item assessing two tactics (11, p. 185, Table 1, antisocial). The second is the average of three tactics (10, p. 8, Table 5, continuous antisocial). The third is the average of three tactics (40, p. 530 text & 531, Table 1, externalizing). Therapy is the unweighted average of three effect sizes from the same three publications/same DVs.

[f] Mean effect size for time-out only for conduct problems reported by Larzelere et al. (55, p. 300; *d* = -1.62, equivalent to *r* = -.63).

[g] Mean effect size reported by Weisz et al. for conduct problems (56, p. 227; *g* = -.50, equivalent to *r* = -.24).

*$p < .05$. ** $p < .01$. ***$p < .001$.

To create Table 1, we calculated *five* mean effect sizes (cross-sectional correlations, longitudinal correlations, ANCOVA betas, change-score slope coefficients, and treatment effects from randomized trials) for *three* types of corrective actions (spanking, nonphysical discipline techniques, and psychotherapy) for externalizing problems (the most frequent outcome in studies of parental discipline). As seen in the first three rows of the tabled data, correlations and ANCOVA betas make *all* three corrective actions appear harmful, although ANCOVA reduces the bias somewhat, relative to correlations. In contrast, the fourth and fifth rows show that slope coefficients from change-score analyses and randomized trials make all three corrective actions appear beneficial.

3. Attempts to correct known bias. In addition to measurement issues, ANCOVA-type analyses are biased by between-child differences that do not change during the study period (7,8). Recommendations for overcoming this bias include slope-type change-score innovations (e.g., random-intercepts cross-lagged panel model). To date, these recommended methods have produced either non-significant associations between spanking and child outcomes or have *reversed* the apparent effects of spanking on externalizing problems from harmful- to beneficial-looking (7, pp. 4-6 of online Supplemental Material; 42).

4. Randomized trials. The *only* method permitting researchers to confidently infer causation is a randomized experiment. This makes the four trials by Roberts and his colleagues (13-16) very important – and a focus of debate in the spanking literature. Promoters of spanking bans have emphasized the small sample sizes and attempted to explain away the benefits of spanking documented in this series of studies (21, 24). Defenders of Canada's current approach acknowledge the small sample sizes but emphasize the large beneficial effect of this spanking on children's compliance to both parental commands and time-out (3).

Compliance to time-out is critical because the effectiveness of *behavioral parent training* (BPT; the most empirically-supported non-medication treatment for oppositional defiant disorder and ADHD) depends on time-out cooperation. Cooperation with time-out permits parents to quickly phase-out or reduce the need for disciplinary spanking over time (a goal shared by all spanking researchers).

Summary of the evidence used to support Canada's current approach. Researchers who support Canada's existing age-specific approach to spanking emphasize randomized trials and effect sizes obtained in MAs that go beyond unadjusted correlations. They tend to limit their studies to children (excluding infants and teens) and strive for transparency concerning biases inherent in various statistical approaches. Moreover, they have investigated nonphysical discipline techniques to make recommendations for replacing spanking but have not found any that is more effective than customary spanking.

Part III: Additional Considerations

In this section, we address two concerns likely raised with our review of the most emphasized studies. We also *speculate* on some possible reasons for differences in child abuse and delinquency statistics across countries with and without spanking bans.

1. If customary spanking and nonphysical punishments "work the same way," can't we just ban spanking?

Canadian parents currently have the option to use both physical and nonphysical punishment with children ages 2-12. We believe they should retain both options for two reasons.

First, equivalent mean effects do not mean that different types of discipline work equally well for all children (just as different ADHD medications with equivalent mean effects do not work equally well for all children with ADHD). The fact that *both* spanking and a brief barrier-enforced room isolation increased compliance with chair time-out in Roberts' randomized studies (13-16) has often been cited as evidence that parents do not "need" spanking. In the medical field, when two corrective actions are both demonstrated to be effective, both are kept available as treatment options. Having options in medicine is critical because individual clients often respond better to one type of treatment than another. Similarly, in Roberts' lab, some children responded better to one back-up condition than the other (16).

Second, corrective actions sometimes work better in combination. Most children who intially refused to cooperate in Roberts' studies developed the self-control to cooperate *when* the parental directive to remain in chair time-out was consistently backed-up with a more forceful consequence (3, 43). The combination of a less forceful corrective action (time-out) with a more forceful one (spanking or brief room isolation) permitted the phasing out of the more forceful action (16). Consistent with these randomized results, two non-randomized studies found that *never*-spanked children in certain parenting contexts appeared less well-adjusted than those whose spanking had been phased out by ages 9 or 11 (44, 45).

2. Do spanking bans protect children?

Spanking bans are often promoted as a way to improve children's lives by reducing their risk of assault. Whether they actually do this is unclear.

Spanking bans do seem to be related to *parental reports* of disciplinary tactics. In Bussmann et al.'s comparison of five European countries (three with spanking bans and two without), parents reported less use of customary spanking in countries where spanking was banned (46). This was particularly true for Sweden, the first country to ban spanking (in 1979) and the country with the most extensive publicity campaigns to

promote awareness of the ban. Patterns for *more severe* physical discipline were somewhat less consistent. Seven years after their spanking ban, more German parents reported "beating with an object" (5.2%) than French parents (4.5%, no ban) and more "severe beating" (9.0%) than Spanish parents (4.4%, no ban). (Statistical significance was not reported in the original article).

While parental reports across countries suggest that Sweden's ban may be protecting Swedish children, Swedish *crime statistics* suggest a more complicated story. Within-Sweden, across-time comparisons suggest that several types of assault increased markedly for the first cohort of children raised after the ban, and then again for the second cohort. Sweden not only posted increased allegations of child abuse (which may or may not (47) be attributable to changes in definition, social awareness, and reporting) but also increased allegations of violence perpetrated by minors. From 1981 to 2010, Sweden saw a 22-fold increase in alleged cases of child abuse and a 24-fold increase in alleged criminal assaults by minors against children ages 7-14 (47).

We do *not* presume that Sweden's ban *caused* these apparent increases. Non-experimental data does not permit causal inference, and these increases are much too large to be reasonably attributed to any one cause. A thorough analysis would need to consider many factors including budgets for child protection and forensic services, the general economy, changes in family structure, media exposure, immigration patterns, secularization, etc. The ban *may* be a *partial* contributor to the dramatic increase in Swedish assaults on children. At minimum, the ban failed to take Sweden in the direction that Swedish legislators intended.

In contrast, Canada's 2004 legislation was followed by a *decrease* in child abuse. Trocmé et al. (48) reported that *substantiated* physical child abuse throughout Canada (excluding Quebec) decreased from 5.31 cases per 1000 children in 2003 to 3.19 cases per 1000 in 2008. Within Quebec, Clement et al. (49) documented a significant decrease in "minor physical violence" and a nonsignificant decrease in "severe physical violence" from 2004 to 2012. Again, we do not presume that Canada's 2004 legislation *caused* this decrease. Indeed, rates were already falling in the period from 1999 to 2004. Our point is simply that Canada's *age-specific* ban on physical discipline was followed by a decrease in substantiated cases of child abuse, whereas Sweden's *total* ban was followed by a large increase in alleged child abuse cases.

If these two bans did increase risk for Swedish children and decrease risk for Canadian children, how might we explain this difference? One possible explanation is that Canadian parents have retained the authority necessary to help young children grow into responsible adolescents and adults, whereas the authority of Swedish parents was undermined. Such undermining is a serious concern. In a review of studies from the US and Scandinavia, Patterson and Fisher (50) concluded that most children's disposition to comply with adult directives increases steadily from ages 2-8, and parents' failure to secure compliance about 70% of the time foreshadows serious dysfunction.

One manifestation of family dysfunction is child abuse, which may stem (in part) from parental anger over limited authority/control (51). Baumrind observed that permissive parents were more likely than authoritative parents to admit to "explosive attacks of rage in which they inflicted more pain or injury upon the child than they had intended" (51, p. 35). The possibility that customary spanking helps suppress children's misbehavior before parents' escalating frustration leads to undisputed abuse is supported by Bussmann et al.'s (46) analyses predicting "mild" and "severe" violence in three European countries that have banned spanking. As shown in Table 24.2 of the Bussman report, parents' knowledge that mild corporal punishment was prohibited in Sweden, Austria and Germany was negatively associated with the use of mild corporal punishment (b = -.26, $p < .001$), but *positively* associated with severe physical punishment ($b = .09$, $p < .001$).

Perhaps some parents in the ban-countries followed the law until their children's behavior became unbearable and then temporarily "snapped," surprising themselves with an atypical burst of severe punishment. Atypical bursts may help explain the marked increased in child abuse allegations in the Swedish crime statistics following the 1979 ban. They may also help explain data from a cross-country comparison of hospital admission codes. Gilbert et al. (52) analyzed hospital trends from 1979 to 2005/6 in six industrialized countries/regions. Sweden had a higher rate of hospital admissions than Canada for maltreatment injuries in children ages 1-11, even as it had a lower rate of violent child deaths.

Low parental authority also increases the risk of serious antisocial behavior related to early oppositional defiance. Unquelled "authority conflict" in childhood has been identified as one of the *earliest* (53) and *strongest* (54) predictors of criminal delinquency. The possibility that nonphysical punishment alone was insufficient to deter some Swedish children's early antisocial tendencies may help explain Sweden's marked increases in alleged minor-on-minor assaults following the ban.

Support for Canada's Existing Age-Specific Approach

Since Canada legislated its current approach to physical discipline in 2004, some groups have continued to lobby for a ban on all spanking. As demonstrated in this commentary, the evidence underlying this well-intentioned advocacy is weak.

Methodologically sophisticated MAs of BPL studies suggest that the mean effect of customary spanking on subsequent child adjustment is near zero. More importantly, the only randomized studies of spanking have shown that back-up spanking can be part of an effective treatment protocol for those children at the highest risk of chronic delinquency. Admittedly, this latter evidence is dated. Social scientists should now be attempting to

replicate these randomized studies and to design new ones to develop more effective discipline programs. If exact replication is too controversial, there are methods used in medical research to approximate unbiased causal effects when ethical concerns prohibit full randomization (e.g., randomized consent, encouragement designs, regression discontinuity designs).

Finally, we know of no evidence to suggest that total bans protect children more than Canada's current approach. In contrast to the increased allegations of assaults on Swedish children following Sweden's total ban on physical discipline, Canada's child abuse rates have continued falling since Canada's age-specific ban. We believe this is because Canada currently has one of the best approaches in the world. We hope that Canada will continue to lead with empirically-informed, developmentally appropriate legislation that empowers parents to help children become well-adjusted adults.

Conflicts of Interest

The authors have no financial relationships or other ties to disclose.

References

1. *Canadian Foundation for Children, Youth and the Law v. Canada*. 2004. 1 SCR 76.
2. Baumrind D, Larzelere RE, Cowan PA. Ordinary physical punishment: Is it harmful? Comment on Gershoff (2002). Psychological Bulletin. 2002;128:580-9.
3. Larzelere RE, Gunnoe ML, Pritsker J, Ferguson CJ. Resolving the contradictory conclusions from three reviews of controlled longitudinal studies of physical punishment: A meta-analysis. Marriage & Family Review. 2024; https://doi.org/10.1080/01494929.2024.2392672
4. 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):1-37.
5. Gershoff ET. Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. Psychological Bulletin. 2002;128(4):539-79.
6. Heilmann A, Mehay A, Watt RG, Kelly Y, Durrant JE, van Turnhout J, Gershoff ET. Physical punishment and child outcomes: A narrative review of prospective studies. The Lancet. 2021;398(10297):355-64.
7. Berry D, Willoughby MT. On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. Child Development. 2017;88(4):1186-206.
8. Hamaker EL, Kuiper RM, Grasman RPPP. A critique of the cross-lagged panel model. Psychological Methods. 2015;20(1):102-16.
9. Larzelere RE, Lin H, Payton ME, Washburn IJ. Longitudinal biases against corrective actions. Archives of Scientific Psychology. 2018;6(1):243-50.
10. 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.
11. 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(2):180-9.
12. Berger LM, Bruch SK, Johnson EI, James S, Rubin D. Estimating the 'impact' of out-of-home placement on child well-being: Approaching the problem of selection bias. Child Development. 2009;80:1856-76.
13. Bean AW, Roberts MW. The effect of time-out release contingencies on changes in child noncompliance. Journal of Abnormal Child Psychology. 1981;9(1):95-105.
14. Day DE, Roberts MW. An analysis of the physical punishment component of a parent training program. Journal of Abnormal Child Psychology. 1983;11(1):141-52.
15. Roberts MW. Enforcing chair timeouts with room timeouts. Behavior Modification. 1988;12(3):353-70.
16. Roberts MW, Powers SW. Adjusting chair timeout enforcement procedures for oppositional children. Behavior Therapy. 1990;21(3):257-71.
17. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions. 2nd ed. Hoboken, NJ: Wiley-Blackwell; 2019.
18. Larzelere RE. A review of the outcomes of parental use of nonabusive or customary physical punishment. Pediatrics. 1996;98:824-8.
19. Straus MA, Sugarman DB, Giles-Sims J. Spanking by parents and subsequent antisocial behavior of children. Archives of Pediatrics and Adolescent Medicine. 1997;151(8):761-7.
20. Gunnoe ML, Mariner CL. Toward a developmental-contextual model of the effects of parental spanking on children's aggression. Archives of Pediatrics and Adolescent Medicine. 1997;151(8):768-75.
21. Gershoff ET, Grogan-Kaylor A. Spanking and child outcomes: Old controversies and new meta-analyses. Journal of Family Psychology. 2016;30(4):453-69.
22. 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).
23. American Psychological Association. Resolution on physical discipline of children by parents. 2019. Retrieved from https://www.apa.org/about/policy/physical-discipline.pdf
24. Gershoff ET, Goodman GS, Miller-Perrin CL, Holden GW, Jackson Y, Kazdin AE. The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers. American Psychologist. 2018;73(5):626-38.
25. Larzelere RE, Gunnoe ML, Ferguson CJ, Roberts MW. The insufficiency of the evidence used to categorically oppose spanking and its implications for families and psychological science: Comment on Gershoff et al. (2018). American Psychologist. 2019;74(4):497-9.
26. Rohner RP, Melendez-Rhodes T. Perceived parental acceptance--rejection mediates or moderates the relation between corporal punishment and psychological adjustment: Comment on Gershoff et al. (2018). American Psychologist. 2019;74:500-2.
27. Gershoff ET, Goodman GS, Miller-Perrin C, Holden GW, Jackson Y, Kazdin AE. 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). American Psychologist. 2019;74:503-5.
28. Gershoff ET, Sattler KMP, Ansari A. Strengthening causal estimates for links between spanking and children's externalizing behavior problems. Psychological Science. 2018;29(1):110-20.
29. Kang J. Spanking and children's social competence: Evidence from a US kindergarten cohort study. Child Abuse & Neglect. 2022;132:105817.
30. Grogan-Kaylor A. The effect of corporal punishment on antisocial behavior in children. Social Work Research. 2004;28(3):153-62.
31. 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(3):250-67.
32. Imbens G, Xu Y. LaLonde (1986) after nearly four decades: Lessons learned. arXiv. 2024;2406.00827v2.
33. Lin H. Revealing and resolving contradictory ways to reduce selection bias to enhance the validity of causal inferences from non-randomized longitudinal data [Dissertation]. Stillwater, OK: Oklahoma State University; 2018.
34. Rosenbaum PR. Observation and experiment: An introduction to causal inference. Cambridge, MA: Harvard University Press; 2017.

35. Ferguson CJ. Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies. Clinical Psychology Review. 2013;33(1):196-208.
36. Larzelere RE, Gunnoe ML, Ferguson CJ. Improving causal inferences in meta-analyses of longitudinal studies: Spanking as an illustration. Child Development. 2018;89(6):2038-50.
37. Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern epidemiology. 4th ed: Wolters Kluwer Health; 2021.
38. Angrist JD, Pischke J-S. Mostly harmless econometrics: An empiricist's approach. Princeton, NJ: Princeton University Press; 2009.
39. Tennant FS, Jr., Detels R, Clark V. Some childhood antecedents of drug and alcohol abuse. American Journal of Epidemiology. 1975;102:377-85.
40. Larzelere RE, Knowles SJ, Adkison-Johnson C, Cox RB, Jr., Lin H, Mandara J. Ethnic differences in the effects of five disciplinary tactics on subsequent externalizing behavior problems. Marriage & Family Review. 2023;59(8):523-48.
41. Larzelere RE, Reitman D, Ortiz C, Cox RB, Jr. Parental punishment: Don't throw out the baby with the bathwater. In: Frisby CL, Redding RE, O'Donohue WT, Lilienfeld SO, editors. Ideological and political bias in psychology: Nature, scope, and solutions. Cham, Switzerland: Springer; 2023. p. 561-84.
42. Pritsker J. Spanking and externalizing problems: Examining within-subject associations. Child Development. 2021;92(6):2595-602.
43. Everett GE, Hupp SDA, Olmi DJ. Time-out with parents: A descriptive analysis of 30 years of research. Education & Treatment of Children. 2010;33(2):235-59.
44. Gunnoe ML. Associations between parenting style, physical discipline, and adjustment in adolescents' reports. Psychological Reports: Disability & Trauma. 2013;112(3):933-75.
45. 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(5):946-61.
46. Bussmann KD, Erthal C, Schroth A. Effects of banning corporal punishment in Europe: A five-nation comparison. In: Durrant JE, Smith AB, editors. Global pathways to abolish physical punishment: Realizing children's rights. New York: Routledge 2011. p. 299-322.
47. Larzelere RE, Swindle T, Johnson BR. Swedish trends in criminal assaults against minors since banning spanking, 1981-2010. International Journal of Criminology and Sociology. 2013;2:129-37.
48. Trocmé N, Fallon B, MacLaurin B, Chamberland C, Chabot M, Esposito T. Shifting definitions of emotional maltreatment: An analysis child welfare investigation laws and practices in Canada. Child Abuse & Neglect. 2011;35:831-40.
49. Clement ME, Chamberland C, Bouchard C. Prevalence, co-occurrence and decennial trends of family violence toward children in the general population. Canadian Journal of Public Health. 2015;106(7, Suppl. 2):eS31-eS7.
50. Patterson GR, Fisher PA. Recent developments in our understanding of parenting: Bidirectional effects, causal models, and the search for parsimony. In: Bornstein MH, editor. Handbook of parenting: Vol 5 Practical issues in parenting. 2nd ed. Mahwah, NJ: Erlbaum; 2002. p. 59-88.
51. Baumrind D. The development of instrumental competence through socialization. In: Pick AD, editor. Minnesota Symposia on Child Psychology. 7. Minneapolis: University of Minnesota Press; 1973. p. 3-46.
52. Gilbert R, Fluke J, O'Donnell M, Gonzalez-Izquierdo A, Brownell M, Gulliver P, et al. Child maltreatment: variation in trends and policies in six developed countries. The Lancet. 2012;379(9817):758-72.
53. Kelley, BT, Loeber, R., Keenan, K., & DeLamatre, M. (1997, December). *Developmental pathways in boys' disruptive and delinquent behavior*. Washington, DC: Office of Justice Programs Retrieved from https://ojjdp.ojp.gov/library/publications/developmental-pathways-boys-disruptive-and-delinquent-behavior
54. Timmermans M, van Lier PAC, Koot HM. Pathways of behavior problems from childhood to late adolescence leading to delinquency and underachievement. Journal of Clinical Child & Adolescent Psychology. 2009;38(5):630-8.
55. 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(4):287-319.
56. Weisz JR, Kuppens S, Ng MY, Vaughn-Coaxum RA, Ugueto AM, Eckshtain D, Corteselli KA. 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(2):216-37.