Reviewer’s report

**Title:** Is the Strongest Causal Evidence Against Customary Parental Spanking Due to Residual Confounding? A Replication of the Straus et al. (1997) Landmark Study with Alternative Disciplinary Tactics and Psychotherapy

**Version:** 2  **Date:** 15 June 2009

**Reviewer:** Damien Jolley

**Reviewer’s report:**

This manuscript attempts to show that evidence against parental corporal punishment (CP) from a 1997 paper is due to residual confounding, rather than being causal.

To establish this, the authors have gone to the same source of data as the original work, obtained a sample of about the same size (consisting the same families and of mostly the same children), and replicated the results using the same confounding variables to estimate an association between CP and antisocial behaviour (ASB). To demonstrate the existence of residual confounding, the authors use a series of progressively finer measures for the main confounder, and show an attenuated effect between CP and ASB.

Two problems exist with the analysis reported in Straus et al. (1997). The first is always a problem in longitudinal research: while the outcome (ASB in 1990) is measured after the key predictor variable (CP in 1988), the most influential confounding variable is the concurrent measure of outcome (ASB in 1988). The second is less common: the outcome (ASB in 1990) is measured and analysed as a continuous variable from a standardized 6-item composite index, the confounding variable (ASB in 1988) is measured and included in ANOVA as a three-level categorical variable (“low, middle or high quartile” [sic]).

It is interesting that Straus et al. (1997) argue on the first point that “our test of the effect of CP on ASB in 1990 will be extremely conservative”, citing the standard textbook of MV analysis in psychology, Tabachnik & Fidell (1996). The whole point of the present manuscript is precisely the opposite. I have generally found Tabachnik & Fidell to be an unreliable source of advice for multivariable analysis.

Certainly, most analysts nowadays would choose to include the concurrent level of the outcome variable as a confounder in a multiple regression analysis of change over time. However, the authors of the manuscript under review are correct to point out that the use of a coarse, trichotomous aggregate measure at baseline will leave open the possibility of residual confounding.

Larzelere and co-authors remedy this in their replicated analysis by including a continuous measure of ASB in 1988 as their covariate (Table 4). However, they proceed then to wildly over-interpret the results of this re-analysis, making a
typical Type II error. Their results clearly show an attenuation from a standardized effect of beta=0.10 in the “confounded” analysis to beta=0.09 in the corrected analysis using the continuous measure of ASB in 1988.

The fact that the former is significant (P < 0.05) while the latter is not is irrelevant. These authors have committed the error of assuming that the non-significant effect is no effect. Their statements that “all effect sizes tended to decrease in magnitude and change from apparently detrimental to neutral effects” (p 14), and “When initial differences in levels of antisocial behaviour were measured more comprehensively, spanking frequency no longer predicted higher antisocial behaviour two years later” (p 17) are simply untrue.

The authors also claim the results of structural equation modelling (SEM) as further evidence that the effect of CP on ASB vanishes, specifically that the effect size is negative in the last row of their Table 4. But controlling for “gain in latent externalising” is equivalent to the inclusion of the outcome measure as a predictor in this model – effect sizes of other predictor variables in this model are uninterpretable.

On reflection, my advice to the editors of BMC Pediatrics is that these authors are best advised to discuss the potential methodological limitations and alleged errors of Straus et al (1997) with the editors of the original journal (Arch Pediatr Adolesc Med). Once these have been aired and Straus et al have been offered the right of reply, then the results of extension analyses – appropriately interpreted – might be considered for publication.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.