# Finding Common Ground in Meta-Analysis "Wars" on Violent Video Games

Maya B. Mathur[1,2*] & Tyler J. VanderWeele[1]

[1]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA

[2]Quantitative Sciences Unit, Stanford University, Palo Alto, CA, USA

**Abstract**

Independent meta-analyses on the same topic can sometimes yield seemingly conflicting results. For example, prominent meta-analyses assessing the effects of violent video games on aggressive behavior reached apparently different conclusions, provoking ongoing debate. We suggest that such conflicts are sometimes partly an artifact of reporting practices for meta-analyses that focus only on the pooled point estimate and its "statistical significance". Considering statistics that focus on distributions of effect sizes and that adequately characterize effect heterogeneity can sometimes indicate reasonable consensus between "warring" meta-analyses. We show using novel analyses that this seems to be the case in the video-game literature. Despite seemingly conflicting results for the "statistical significance" of the pooled estimates in different meta-analyses of video game studies, in fact all of the meta-analyses point to the conclusion that, in the vast majority of settings, violent video games do increase aggressive behavior but that these effects are almost always quite small.

## Introduction

Meta-analyses are often intended to improve scientific certainty and settle debates. However, in a recent article, de Vrieze (2018) described instances in which independent meta-analyses on the same topic yield apparently conflicting results, seeming only to exacerbate controversy and uncertainty. He noted that prominent meta-analyses assessing the effects of violent video games on aggressive behavior have been widely interpreted as yielding "opposite" conclusions, which has provoked ongoing heated debate (Anderson et al., 2010; Ferguson, 2015). de Vrieze (2018) described compelling possibilities to help adjudicate between results of such meta-analyses: namely, minimizing researcher degrees of freedom when conducting new meta-analyses, ensuring full analytic reproducibility, and conducting prospective multisite replications of the phenomenon of interest. We agree emphatically with all of these recommendations. In the specific context of the video game debate, others have also provided interesting commentary on potential scientific reasons for apparently conflicting conclusions (e.g., Kepes et al. (2017); Prescott et al. (2018)).

However, we also believe that reporting practices for meta-analyses can sometimes produce an illusion of conflict in the meta-analyses' scientific implications when in fact little conflict exists, as we will illustrate for meta-analyses of violent video games. Meta-analyses are usually reported with nearly exclusive focus on the pooled point estimate and its "statistical significance" and are sometimes treated as conflicting merely because one point estimate attains "statistical significance" while the other does not, even when the point estimates and confidence intervals are quite similar (e.g., Appleton et al. (2010); Bloch & Hannestad (2012)). Even when one focuses on comparing the point estimates rather than the $p$-values, this approach still not fully characterize evidence strength when the effects represent a heterogeneous distribution (Mathur & VanderWeele, 2018).

Some "meta-wars" might be reduced to smaller skirmishes or entirely resolved if investigators were to compare evidence strength between the meta-analyses in a manner that characterizes effect heterogeneity and focuses on the distributions of effect sizes rather than "statistical significance". Specifically, an investigator could select thresholds above which an effect size might be considered of scientifically meaningful size, depending on the context of the effect under investigation. (There is a large, interdisciplinary literature considering how to choose such thresholds, as summarized by Mathur & VanderWeele (2018).) With such a threshold in mind, it is statistically straightforward to estimate the percentage of true effects in each meta-analysis that are stronger than this threshold; we have provided an R package, `MetaUtility`, to do so (Mathur & VanderWeele, 2018). It is also possible to assess the percentage of scientifically meaningful effect sizes in the "unexpected" direction that is opposite in sign from the pooled point estimate. These metrics can help identify if: (1) there are few effects of scientifically meaningful size despite a "statistically significant" pooled point estimate, (2) there are some large effects despite an apparently null point estimate, or (3) strong effects in the direction opposite the pooled estimate also regularly occur (and thus, potential moderators should be examined).

## Methods

For three prominent meta-analyses on violent video games (Anderson et al., 2010; Ferguson, 2015; Prescott et al., 2018), we meta-analytically estimated[1] the percent of effects that surpass standardized effect size thresholds of $q = 0$, 0.10, and 0.20 (where standardized effect sizes were either Fisher-transformed correlations or standardized multiple regression coefficients). The threshold of 0 is the least stringent in that it considers all detrimental effects, regardless of the magnitude. The more stringent thresholds of 0.10 and 0.20 consider only effects of at least modest sizes. Estimating the percentage (and 95% confidence interval) of true effects stronger than these thresholds can be done meta-analytically, an approach that is distinct from simply counting the "significant" $p$-values in the observed sample of studies, as we

[1] Per Mathur & VanderWeele (2018), we used closed-form inference when the estimated percent was $>15\%$ and $<85\%$. Otherwise, we used bias-corrected and accelerated bootstrapping with 1,000 iterates, or percentile bootstrapping if needed to alleviate computational problems.

have described elsewhere (Mathur & VanderWeele, 2018). For each meta-analysis, we first conducted an analysis that reproduced as closely as possible the main results as reported in the respective paper's abstract[2]. We also assessed the percentage of effects below effect sizes of -0.10 or -0.20, which would indicate beneficial, rather than detrimental, effects of violent video games. Additionally, for more direct scientific comparability across analyses, we also conducted a second "controlled" analysis that included only longitudinal studies (Anderson et al., 2010; Prescott et al., 2018) or studies controlling for baseline aggression through statistical adjustment or randomization (Ferguson, 2015). (For one meta-analysis (Prescott et al., 2018), the main and controlled analyses were identical.) Secondarily, we also compared the consistency of these metrics upon correction for publication bias of a form that favors studies with "statistically significant" positive results (Vevea & Hedges, 1995).

## Results

The results (Table 1) suggest considerable common ground between these three meta-analyses, considering both main and controlled analysis specifications. All six analyses suggest that a large majority (point estimate at least 80%, with confidence intervals all bounded above 57%) of effects are above 0, indicating frequent detrimental effects of violent video games, albeit possibly of negligible size (Table 1, column 5). Also, five of the six meta-analyses suggest that very few effects are above 0.20, with the five confidence intervals all bounded below 12% (Table 1, last column). The remaining meta-analysis (Anderson et al. (2010)'s main analysis) suggests that this percentage of effects above 0.20, while not negligible, still represents a minority of effects (29%; 95% CI: [14%, 45%]). The meta-analyses only diverge meaningfully in their estimation of effects above 0.10 (Table 1, column 6), suggesting that the "conflict"

[2] We fit meta-analyses using restricted maximum likelihood with Knapp-Hartung adjusted standard errors (IntHout et al., 2014). In each meta-analysis, some studies appeared to contribute multiple, potentially non-independent point estimates, though all original analyses used standard methods assuming independence. For the meta-analysis with the most apparent clustering, we performed a sensitivity analysis by refitting the model using robust methods (Hedges et al., 2010), yielding nearly identical results. For the others, limitations in available data precluded this sensitivity analysis, but clustering appeared to be minimal and so unlikely to affect results.

between these analyses is limited to their estimation of effects in the narrow range between 0.10 and 0.20. For example, considering the controlled analyses for the two meta-analyses that have been most central in the perceived "war", we estimate for the Anderson et al. (2010) meta-analysis that 100% (CI: [99%, 100%]), 0% (CI: [0%, 69%]), and 0% (CI: [0%, 0%]) of effects surpass the thresholds of 0, 0.10, and 0.20 respectively; similarly, we estimate for the Ferguson (2015) meta-analysis that 80% (CI: [57%, 100%]), 9% (CI: [0%, 29%]), and 0% (CI: [0%, 3%]) of effects surpass these thresholds.

Considering instead the percentage of effects suggesting beneficial, rather than detrimental, effects of violent video games, the three meta-analyses all estimate that no such effects (i.e., 0%) are stronger than (i.e., more beneficial than) an effect size of -0.20 (with confidence intervals all bounded below 3%) or even -0.10 (with confidence intervals all bounded below 21%); these analyses are presented in the Supplement. The sensitivity analysis correcting for publication bias suggested similarly consistent evidence across the three meta-analyses (Supplement).

## Discussion

In practice, we would interpret these various meta-analyses as providing consistent evidence that the effects of violent video games on aggressive behavior are nearly always detrimental in direction but are rarely stronger than a standardized effect size of 0.20. These conclusions are not intended to trivialize important methodological critiques and debates in this literature, for example regarding demand characteristics, expectancy effects, confounding, measurement of aggression, and publication bias in experiments with behavioral outcomes (e.g., Ferguson (2015); Hilgard et al. (2017); Markey (2015)). Our claim is not that our re-analyses resolve these methodological problems, but rather that widespread perceptions of "conflict" among the results of these meta-analyses – even when taken at face value without reconciling their substantial methodological differences – may be partly an artifact of statistical reporting practices in meta-analyses. Indeed, our quantitative findings seem to support a recent task force's suggestion that, heuristically, the "warring" meta-analyses may indicate similar effect sizes (Calvert et al., 2017).

Our findings also in no way undermine de Vrieze (2018)'s and many others' recommendations for designing scientifically robust meta-analyses and for adjudicating between seemingly conflicting results. Corroborating his discussion of analytic reproducibility, we were able to obtain raw data for the three discussed meta-analyses but experienced challenges in attempting to analytically reproduce several published results; these challenges persisted after contact with the authors. Additionally, for one meta-analysis that we initially intended to include because of its historical prominence in the "war" (Ferguson & Kilburn, 2009), contact with the original author indicated that neither the data nor the list of the studies included in the meta-analysis still existed, though with the author's assistance, we were able to obtain data for a subsequent, partly overlapping, meta-analysis (Ferguson, 2015). Ultimately, even in light of potential methodological problems, suboptimal reproducibility, and researcher degrees of freedom as noted by de Vrieze (2018), we believe that these "warring" meta-analyses in fact provide considerable consensus in favor of consistent, but small, detrimental effects of violent video games on aggressive behavior.

## Online Supplement

The Online Supplement is available at https://osf.io/eunz3/.

## Acknowledgements

Anna Prescott et al. provided raw data in their manuscript. Christopher Ferguson and Craig Anderson provided raw data for re-analysis and answered analytic questions. MM and TVW were supported by NIH grant R01 CA222147. The funders had no role in the design, conduct, or reporting of this research.

## Reproducibility

Data from the Prescott meta-analysis were obtained from the published forest plot and are publicly available (https://osf.io/eunz3/). Data from the Ferguson and Anderson meta-analyses cannot be made public at the authors' requests, but will be made available upon request to individuals who have secured permission from the original authors. All code required to reproduce our re-analyses is publicly available (https://osf.io/eunz3/).

## References

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., ... Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, *136*(2), 151.

Appleton, K. M., Rogers, P. J., & Ness, A. R. (2010). Updated systematic review and meta-analysis of the effects of $\eta$-3 long-chain polyunsaturated fatty acids on depressed mood. *The American Journal of Clinical Nutrition*, *91*(3), 757–770.

Bloch, M. H., & Hannestad, J. (2012). Omega-3 fatty acids for the treatment of depression: systematic review and meta-analysis. *Molecular Psychiatry*, *17*(12), 1272.

Calvert, S. L., Appelbaum, M., Dodge, K. A., Graham, S., Nagayama Hall, G. C., Hamby, S., ... Hedges, L. V. (2017). The american psychological association task force assessment of violent video games: Science in the service of public interest. *American Psychologist*, *72*(2), 126.

de Vrieze, J. (2018). The metawars. *Science*, *361*(6408), 1184.

Ferguson, C. J. (2015). Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science*, *10*(5), 646–666.

Ferguson, C. J., & Kilburn, J. (2009). The public health risks of media violence: A meta-analytic review. *The Journal of Pediatrics*, *154*(5), 759–763.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65.

Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al. (2010). *Perspectives on Psychological Science*, *143*(7).

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, *14*(1), 25.

Kepes, S., Bushman, B. J., & Anderson, C. A. (2017). Violent video game effects remain a societal concern: Reply to Hilgard, Engelhardt, and Rouder (2017). *Perspectives on Psychological Science*, *143*(7).

Markey, P. M. (2015). Finding the middle ground in violent video game research: Lessons from ferguson (2015). *Perspectives on Psychological Science*, *10*(5), 667–670.

Mathur, M. B., & VanderWeele, T. J. (2018). New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, 1336âĂŞ1342.

Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences*, *115*(40), 9882–9888.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435.

**Table 1.** Estimates from video-game meta-analyses, including pooled point estimates, estimated heterogeneity (the standard deviation of the true effect distribution), and estimated percent of true effects above three thresholds of scientific importance. Effect sizes are on the standardized difference scale. "Main": main analysis specification reported in the paper's abstract; "controlled": including only longitudinal studies or studies otherwise controlling for baseline aggression. The two specifications were identical for the Prescott et al. (2018) meta-analysis.

| | **Number of estimates** | **Pooled point estimate [95% CI]** | **Heterogeneity estimate [95% CI]** | **% above 0 [95% CI]** | **% above 0.10 [95% CI]** | **% above 0.20 [95% CI]** |
|---|---|---|---|---|---|---|
| **Anderson et al.** | | | | | | |
| Main | 75 | 0.17 [0.15, 0.19] | 0.05 [0.02, 0.07] | 100 [99, 100] | 90 [79, 98] | 29 [14, 45] |
| Controlled | 12 | 0.08 [0.05, 0.10] | 0 [0.00, 0.04] | 100 [98, 100] | 0 [0, 69] | 0 [0, 0] |
| **Ferguson** | | | | | | |
| Main | 166 | 0.05 [0.04, 0.07] | 0.06 [0.04, 0.06] | 83 [76, 90] | 20 [12, 28] | 0 [0, 2] |
| Controlled | 22 | 0.04 [0.01, 0.07] | 0.05 [0.00, 0.07] | 80 [57, 100] | 9 [0, 29] | 0 [0, 3] |
| **Prescott et al.** | | | | | | |
| Main and controlled (the two here are identical) | 25 | 0.10 [0.07, 0.13] | 0.05 [0.01, 0.07] | 98 [75, 100] | 53 [29, 76] | 3 [0, 12] |

# *Supplement*: Finding Common Ground in Meta-Analysis "Wars" on Violent Video Games

## Sensitivity analyses with correction for publication bias

We conducted sensitivity analyses using pooled point estimates and heterogeneity estimates that were corrected for publication bias via a selection model (Vevea & Hedges, 1995). Selection models are a likelihood-based class of methods, and they assume that publication bias selects for "statistically significant" $p$-values (see Jin et al. (2015) and McShane et al. (2016) for reviews). These models specify a parametric form for the true effect distribution as well as for the dependence of a study's publication probability on its $p$-value. Specifically, we used the R package `weightr` to fit a model in which the weight function was a step function such that positive point estimates with a two-tailed $p < 0.05$ may be published with higher probability than negative point estimates or those with $p \geq 0.05$ (Vevea & Hedges, 1995). The model assumes that the true effects are normally distributed prior to the introduction of publication bias. The meta-analytic parameters of interest and the parameters of the weight function can be jointly estimated by maximum likelihood. The results are presented in Table S1 and are qualitatively similar to those in the main text.

## Percentage of effects in the opposite direction

Table S2 presents the estimated percentages of effect sizes below thresholds of -0.10 or -0.20.

**Table S1:** *Publication bias-corrected estimates. "N/A" indicates a model estimation problem (e.g., reflecting too few "nonsignificant" studies).*

| Meta-analysis | No. estimates | Pooled point estimate | Heterogeneity estimate | % above 0 | % above 0.10 | % above 0.20 |
|---|---|---|---|---|---|---|
| **Anderson** | | | | | | |
| Main | 75 | 0.14 [0.11, 0.16] | 0.05 [0.02, 0.07] | 100 [99, 100] | 76 [55, 96] | 10 [7, 47] |
| Controlled | 12 | N/A | N/A | N/A | N/A | N/A |
| **Ferguson** | | | | | | |
| Main | 166 | 0.05 [0.03, 0.07] | 0.06 [0.04, 0.06] | 81 [72, 91] | 18 [9, 27] | 0 [0, 2] |
| Controlled | 22 | N/A | N/A | N/A | N/A | N/A |
| **Prescott** | | | | | | |
| Main and controlled (the two here are identical) | 25 | 0.10 [0.06, 0.15] | 0.05 [0.00, 0.07] | 98 [78, 100] | 53 [19, 88] | 3 [0, 13] |

**Table S2:** *Estimated proportion of effects stronger than thresholds indicating beneficial, rather than detrimental, effects of video games. These estimates do not correct for publication bias. The first three columns reproduce those in in Table 1 of the main text.*

| Meta-analysis | No. estimates | Pooled point estimate | Heterogeneity estimate | % below -0.10 | % below -0.20 |
|---|---|---|---|---|---|
| **Anderson** | | | | | |
| Main | 75 | 0.17 [0.15, 0.19] | 0.05 [0.02, 0.07] | 0 [0, 0] | 0 [0, 0] |
| Controlled | 12 | 0.08 [0.05, 0.10] | < 1e-05 [0.00, 0.04] | 0 [0, 0] | 0 [0, 0] |
| **Ferguson** | | | | | |
| Main | 166 | 0.05 [0.04, 0.07] | 0.06 [0.04, 0.06] | 0 [0, 1] | 0 [0, 0] |
| Controlled | 22 | 0.04 [0.01, 0.07] | 0.05 [0.00, 0.07] | 0 [0, 21] | 0 [0, 3] |
| **Prescott** | | | | | |
| Main and controlled (the two here are identical) | 25 | 0.10 [0.07, 0.13] | 0.05 [0.01, 0.07] | 0 [0, 4] | 0 [0, 0] |

## References

Jin, Z.-C., Zhou, X.-H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, *34*(2), 343–360.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435.