It is commonly hypothesized in violent media research that children should be more susceptible to the effects of violent media exposure than adults. Children are generally considered to be more impressionable: they lack experience and sophistication for evaluating media, they are still developing schema for how the world works and how to behave, and their self-control is not yet fully developed.

Curiously, research has yet to observe larger effects for children than for adults. In the very thorough and highly-cited meta-analysis by Anderson and colleagues (2010), the authors report no significant moderation of the effect size by participant age group (children, adults). [INSERT DIRECT QUOTE]

In re-analyzing this data set, my colleagues and I found that the studies gathered by Anderson et al. may be badly contaminated by some form of selection bias. That is, studies that found statistical significance seemed more likely to find publication in journals and more likely to be rated as good methodology by the meta-analysts. Adjustment for this selection bias suggested that the true effect may be smaller than it seems.

In the discussion, we suggested that this selection bias may be the reason that research has not found stronger effects for children than for adults. It seems to me that, after you filter for statistical significance, the observed effect size is heavily dependent on the sample size. Large samples can reach statistical significance at small observed effect sizes, whereas small samples must observe a very large effect to reach statistical significance. This is called a “small-study effect.” We suggested that contamination by small-study effects may conceal differences in the true effect size across populations.

This problem is not limited to violent media research, of course. I have long found it perplexing that all psychological therapies appear equally effective [citation needed]. Therapy trials seem at high risk of publication bias, as a trial that reports a significantly effective therapy is much more publishable than a trial that does not. Additionally, such trials are expensive to conduct and do not have particularly large samples. Therefore, the efficacy of therapies may also be overestimated, and differences in efficacy may be obscured by small-study effects.

Although I made this argument in the paper, and it seems reasonable enough, I never conducted any simulations to test this hypothesis. In this post, I present simulations of meta-analyses on a variety of true effect sizes under different degrees of publication bias. The results indicate that publication bias causes not only an overestimation of the true effect size, but indeed a loss of power to detect differences between effects. These simulations reveal how publication bias may make it difficult to distinguish individual differences in susceptibility to violent media, differential efficacies of different psychotherapies, or other potential moderators in meta-analysis.

It is a common goal of meta-analysis to provide not only an overall average effect size, but also to test for moderators that cause the effect size to become larger or smaller. For example, researchers who study the effects of violent media would like to know who is most at risk for adverse effects. Researchers who study psychotherapy would like to recommend a particular therapy as being most helpful.

However, meta-analysis does not often generate these insights. For example, research has not found that violent-media effects are larger for children than for adults (Anderson et al. 2010). Similarly, it is often claimed that all therapies are roughly equally effective (citation needed).

It seems to me that publication bias may obscure such patterns of moderation. Publication bias introduces a “small-study effect” in which the observed effect size is highly dependent on the sample size. Large-sample studies can reach statistical significance with smaller effect sizes. Small-sample studies can only reach statistical significance by reporting enormous effect sizes. The observed effect sizes gathered in meta-analysis, therefore, may be more a function of the sample size than they are a function of theoretically-important moderators such as age group or treatment type.

In this simulation, I compare the statistical power of meta-analysis to detect moderators when there is, or when there is not, publication bias.

Method

Simulations cover 4 scenarios in a 2 (Effects: large or medium) × 2 (Pub bias: absent or present) design.

When effect sizes were large, the true effects were δ = 0 in the first population, δ = 0.3 in the second population, and δ = 0.6 in the third population. When effect sizes were medium, the true effects were δ = 0 in the first population, δ = 0.2 in the second population, and δ = 0.4 in the third population. Thus, each scenario represents one group with no effect, a group with a medium-small effect, and a group with an effect twice as large.

When studies were simulated without publication bias, twenty studies were conducted on each population, and all were reported. When studies were simulated with publication bias, studies were simulated, then published and/or file-drawered until at least 70% of the published effects were statistically significant. When results were not statistically significant and were file-drawered, further studies were simulated until 20 statistically significant results were obtained. This keeps the number of studies *k* constant at 20, which prevents confounding the influence of publication bias with the influence of fewer observed studies.

For each condition, I report the observed effect size for each group, the statistical power of the test for moderators, and the statistical power of the Egger test for publication bias. I simulated 500 meta-analyses within each condition in order to obtain stable estimates.

Results

Large effects. In the absence of publication bias, the meta-analysis has good power to detect the differences in effect size between groups. In 92% of the metas, the difference between δ = 0 and δ = 0.3 was detected. In 100% of the metas, the difference between δ = 0 and δ = 0.6 was detected. Only in 4.2% of cases was the δ = 0 group mistaken as having a significant effect.

Type I error rate for the Egger test for publication bias was observed as 11%, larger than it ought to be. Effect sizes within each group were accurately estimated (in the long run) as δ = 0, 0.3, and 0.6.

With publication bias, power was poorer. 100% of the metas mistook the δ = 0 group as having a significant effect. Only 15% of the metas were able to tell the difference between δ = 0 and δ = 0.3. Still, 91% of meta-analyses were able to tell the difference between δ = 0 and δ = 0.6. The Egger test detected significant publication bias in 76% of cases. Effect sizes within each group were overestimated: d = .45, .58, and .73 instead of 0, 0.3, and 0.6.

Medium effects. In the absence of publication bias, the meta-analysis still had pretty good power to detect the difference in effect size between groups. 60% of metas detected the difference between d = 0 and d = 0.2. 99% detected the difference between d = 0 and d = 0.4. The Type I error rate in the delta = 0 group was 5.6%. Egger’s test demonstrated a Type I error rate of 7%. In the long run, effect sizes within each group were accurately recovered as d = 0, 0.2, and 0.4.

With publication bias, power was much poorer. As above, 100% of meta-analyses made a Type I error, mistaking the delta = 0 group as reflecting a significant effect. Only 2.2% of the meta-analyses were able to detect the difference between d = 0 and d = 0.2, and only 35% were able to detect the difference between d = 0 and d = 0.4. Egger’s test detected significant publication bias in 77% of cases. Every meta-analysis misreported the d = 0 group as showing a significant effect. Effect sizes within each group were overestimated: d = .46, .53, and .62 instead of d = 0, 0.2, and 0.4.

Conclusion:

Publication bias can have a deleterious effect on statistical power for moderators in meta-analysis. Publication bias leads to the exaggeration of effect sizes, and this exaggeration is largest when effect sizes are small or zero. Obvious differences such as that between d = 0 and d = 0.6 may retain decent power, but power will fall dramatically for more modest differences such as that between d = 0 and d = 0.4.

One possibility is to employ a hybrid meta-analytic model that tests for moderation while simultaneously adjusting for publication bias. For example, Hedges & Vevea’s selection model can specify effects of moderators and of significance thresholds simultaneously. Similarly, it has been suggested that one can use meta-regression models like PET-PEESE to simultaneously estimate the effect of moderators and small-study effects. I will try these in a later blog post, but it does not seem promising.