




## RESEARCH METHODS

# Approaches to Assessing and Adjusting for Selective Outcome Reporting in Meta-analysis

Jeffrey L Jackson, MD MPH<sup>1</sup> , Ethan M. Balk, MD MPH<sup>2</sup>, Noorie Hyun, PhD<sup>3</sup>, and Akira Kuriyama, MD DrPH PhD<sup>4</sup>

<sup>1</sup>General Medicine Section, Zablocki VAMC, Milwaukee, WI, USA; <sup>2</sup>Center for Evidence Synthesis in Health, Brown University School of Public Health, Providence, RI, USA; <sup>3</sup>Institute of Health and Equity, Center for Advancing Population Science, Medical College of Wisconsin, Milwaukee, WI, USA; <sup>4</sup>Emergency and Critical Care Center, Kurashiki Central Hospital, Okayama, Japan.

**BACKGROUND:** Selective or non-reporting of study outcomes results in outcome reporting bias.

**OBJECTIVE:** We sought to develop and assess tools for detecting and adjusting for outcome reporting bias.

**DESIGN:** Using data from a previously published systematic review, we abstracted whether outcomes were reported as collected, whether outcomes were statistically significant, and whether statistically significant outcomes were more likely to be reported. We proposed and tested a model to adjust for unreported outcomes and compared our model to three other methods (Copas, Frosi, trim and fill). Our approach assumes that unreported outcomes had a null intervention effect with variance imputed based on the published outcomes. We further compared our approach to these models using simulation, and by varying levels of missing data and study sizes.

**RESULTS:** There were 286 outcomes reported as collected from 47 included trials: 142 (48%) had the data provided and 144 (52%) did not. Reported outcomes were more likely to be statistically significant than those collected but for which data were unreported and for which non-significance was reported (RR, 2.4; 95% CI, 1.9 to 3.0). Our model and the Copas model provided similar decreases in the pooled effect sizes in both the meta-analytic data and simulation studies. The Frosi and trim and fill methods performed poorly.

**LIMITATIONS:** Single intervention of a single disease with only randomized controlled trials; approach may overestimate outcome reporting bias impact.

**CONCLUSION:** There was evidence of selective outcome reporting. Statistically significant outcomes were more likely to be published than non-significant ones. Our simple approach provided a quick estimate of the impact of unreported outcomes on the estimated effect. This approach could be used as a quick assessment of the potential impact of unreported outcomes.

**KEY WORDS:** Meta-analysis; Outcome reporting bias; Statistical adjustment.

*The views in this article should not be construed, in any way, to represent those of the Department of Veterans Affairs.*

Received May 26, 2021

Accepted September 2, 2021

Published online October 19, 2021

J Gen Intern Med 37(5):1247–53

DOI: 10.1007/s11606-021-07135-3

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Rigorously conducted systematic reviews based on high-quality studies provide strong evidence for decision-making by patients, clinicians, and policy makers. However, there are a number of biases that can produce inaccurate or misleading results.<sup>1</sup> These can lead to inappropriate treatment and misguided policy. Two common biases that can distort systematic reviews and meta-analyses are publication bias and selective outcome reporting.<sup>1, 2</sup>

Studies with small or non-significant treatment effects are more likely to go unpublished or to experience publication delay.<sup>3–5</sup> Thus, published articles may report biased, overestimations of benefit. Similarly, outcome reporting bias (ORB) may occur when study authors selectively report outcomes, often based on statistical significance.<sup>2</sup> While publication bias occurs when the entire study is unpublished, selective outcome reporting occurs when trials collect a number of outcomes and, post hoc, publish only selected ones.<sup>6–8</sup> Both biases result in non-significant outcomes being omitted from meta-analyses. Comparisons of published papers to original protocols have found that ORB is common; statistically significant results are more likely to be published, protocol primary outcomes are often changed, and most trials underreport their outcomes.<sup>6–8</sup> Comparisons of data submitted for drug approval and reported in subsequent publications also provide evidence of ORB and biased meta-analyses.<sup>9</sup> There are several tests that assess publication bias (also referred to as small-study effects) including funnel plots,<sup>10, 11</sup> nonparametric tests,<sup>2</sup> and regression analyses.<sup>3, 4, 12</sup> There are also approaches to adjust pooled estimates when publication bias is suspected to be present.<sup>10, 11</sup>

While the Cochrane Risk of Bias tool includes a question to assess risk of ORB, this assessment has been shown to be insensitive.<sup>13</sup> The Outcome Reporting Bias in Trials (ORBIT) study proposed a tabular approach to assess for the presence of

outcome reporting bias.<sup>14</sup> In this method, missing outcomes are stratified as high or low risk of being biased based on reviewer judgment or contacting authors for information.

ORB is essentially a missing data problem. There are three patterns of missingness: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).<sup>15</sup> MCAR occurs if the reason data are missing is unrelated to any of the data values, whether missing or observed; the missing data are a random subset of the full data. For example, an audio-recording that is of insufficient quality to code. MAR occurs if the reason the data are missing is unrelated to the missing values but is related to (conditional on) the observed data. If one can control for the conditional variable, one can get a subset independent of the missing mechanism. An example would be residents missing an in-service examination because they are ill. While one may be able to predict missing the assessment from data on their health, it would not have been related to test scores if the resident had been present. Data are NMAR if there is an association between the value of the missing data and their likelihood of being missing (e.g., in studies on depression incidence, patients who develop depression are less likely to complete follow-up surveys). Compared to MCAR and MAR, it is more challenging to address NMAR because any method requires assumptions that are untestable with the observed data.

The best approach for correcting ORB is to obtain the unreported data from the authors. However, trials may have been conducted decades ago, authors may be difficult to locate, or they may no longer have or may be unwilling to share the data. Most systematic reviews have little success obtaining unreported data from the authors.<sup>16</sup> Thus, statistical methods to adjust for ORB have been proposed; we focus on three approaches.<sup>17–19</sup> The Copas method uses the ORBIT risk classification system to stratify the risk from ORB as high or low. For outcomes with high risk of bias, the Copas method uses a maximum likelihood approach to adjust the pooled outcome, accounting for the relative sample size of the studies with missing outcome data.<sup>20</sup> Copas calculates a maximum bias bound, based on a worst-case scenario for missing data and adds this value to the pooled effect. This effectively moves the estimate closer to the null. The Copas method does not use the other observed outcomes in the adjustment. A web-based tool is available using this method to adjust for ORB, but at present is available only for binary outcomes.<sup>21</sup>

The Frosi method uses multivariate modeling to jointly synthesize multiple correlated outcomes. This approach imputes unreported outcome data from reported data from other studies, based on estimated within-study correlations.<sup>18</sup> Unfortunately, the iterative algorithm may not converge and thus may produce no useable result. This commonly occurs when there are few observations. The method assumes a correlation between different reported outcomes within studies. Therefore, if other outcomes are also selectively reported, the Frosi adjustment will perpetuate ORB. Both the Copas and Frosi

methods are computationally complex and are not provided by meta-analytic packages. Consequently, few systematic reviews explore the impact of potential ORB in reporting their results. A third, simpler, approach is based on the trim and fill method, which was created to adjust for publication bias. It is included in most meta-analytic packages. Trim and fill adjusts for small-study effects, and is based on a nonparametric test of asymmetry of the funnel plots, effect size plotted against study variance.<sup>22</sup>

We had two study purposes: (1) to propose a test to assess for the likelihood of ORB and (2) to propose a new, simple method to assess and adjust pooled outcomes when there is evidence that ORB exists. In addition, we compare this proposed method to previously reported ORB adjustment methods and tested our proposed method and other methods using simulation.

## METHODS

We use data from our review on the efficacy of beta-blockers on migraine headache prophylaxis.<sup>23</sup> We included published, randomized, placebo-controlled clinical trials of at least 4 weeks' duration that evaluated prophylactic treatment of episodic migraines with beta-blockers in adults. Abstracted outcomes included the following: (1) headache frequency, (2) headache days, (3) headache index (which involved combinations of frequency, severity, and duration), (4) severity, (5) duration, (6) analgesic medication use, (7) health-related quality of life, (8) 50% headache reduction, and (9) work absence. We used headache frequency as the primary outcome in our original meta-analysis and in this study because it was the most frequently reported outcome and is the headache outcome recommended by the International Headache Society.<sup>24</sup>

## Assessment for Risk of Bias

We abstracted whether these potential outcomes were reported in manuscripts or protocols as having been collected. We also abstracted whether these outcomes were reported to be statistically significant (including for outcomes without other reported data). We assessed for likelihood of ORB using the ORBIT approach (Appendix Table 1).<sup>20</sup> We pooled data using a random effects maximum likelihood (REML) estimator based on a marginal normal distribution to estimate between-study variances.<sup>25, 26</sup> All analyses were conducted using STATA (v 16.1, College Station, TX).

We categorized each of the nine headache variables as (1) data provided, statistically significant ( $p < 0.05$ ); (2) data provided, statistically non-significant; (3) data not provided, reported to be statistically significant; (4) data not provided, reported to be statistically non-significant; and (5) data and statistical significance not reported. We compared these categories using contingency tables and calculated the relative risk of publishing significant versus non-significant findings. Outcomes that were reported as collected, but for which no further

information was available, were assumed to be statistically non-significant, a common assumption.<sup>17-19</sup> We tested the sensitivity of a formal test for whether data were MCAR using Little's approach, which involves a chi-squared test for the MCAR assumption and accommodates arbitrary missing-value patterns.<sup>27</sup>

## Adjustment for Outcome Reporting Bias

We created and tested a simple model to adjust for the potential effect of unreported outcomes. In this model, we assumed that all unpublished data have a null effect:  $\theta_i = 0$  for continuous data, standardized mean differences (SMD), weighted mean differences (MD), or log (odds ratio [OR]). We assessed the impact of including these missing data on the pooled outcome. To "fill in" missing outcomes, we also needed to estimate the standard deviations (SD) for missing data. We tested two approaches to handle missing SDs for unreported outcomes. First, we marginally fit a linear regression model for each SD to observed data SDs with the independent variables including sample size and effect size for the outcome.<sup>28</sup> Then, we imputed the predicted value obtained from the linear model for any missing SDs.<sup>15</sup> In our second approach, we imputed missing SDs using the multiple imputation (MI) command in STATA using additional variables (e.g., year, country). We estimated the pooled effects including these "filled-in estimates" using a REML model (*meta* command, STATA).

We compared our proposed approach to the Copas,<sup>17</sup> Frosi,<sup>18</sup> and trim and fill<sup>22</sup> methods. The Copas method provides a bound for publication bias,  $|b|$ , and a possible bound of the pooled effect as

$$\hat{\theta}_p \pm |b|, \text{ where } |b| = \frac{n+m}{n} \phi \left\{ \Phi^{-1} \left( \frac{n}{n+m} \right) \right\} \frac{\sum (\sigma_i^2 + \tau^2)^{-1/2}}{\sum (\sigma_i^2 + \tau)^{-1}},$$

where  $n$  is the total number of trials in the meta-analysis,  $m$  is the number of missing studies that are high risk for ORB,  $\tau^2$  is the between-study variance,  $\phi(\cdot)$  is the standard normal density function, and  $\Phi^{-1}(\cdot)$  is the inverse function of the cumulative standard normal distribution.

For the Frosi method, we used the *mvmeta* function in STATA with a REML approach to adjust for unreported outcomes. We used the trim and fill approach in STATA's *metatrim* command.

## Simulation Study

To test the accuracy of the various approaches, we generated data with five outcomes and 56 study arms being compared with placebo using a multivariate normal distribution, with the mean and covariance for outcomes based on the values from our beta-blocker data set. We selected the five most commonly reported outcomes for purposes of simulation (frequency, headache index, duration, severity, 50% improvement). We then randomly deleted observations for each outcome to create a dataset with missing data for each outcome to match the number of missing outcomes in our beta-blocker set. This

produced a dataset of 56 comparisons of interventions with varying numbers of missing observations. We generated NMAR data by randomly deleting a portion of the non-significant results (from 5 to 40%, in increments of 5%). We then tested all approaches (ours, Copas, Frosi, and trim and fill) to adjust for ORB for headache frequency, the most commonly reported outcome. Based on 1000 replications at each level of missing data, we compared the results to the unbiased results from the full simulated dataset without missing observations.

## Sensitivity Analysis

To test the sensitivity of the various methods of adjusting for ORB, we randomly selected a subset of our simulated studies ( $n = 40, 30, 20, 10, 5$  studies) and calculated the unadjusted pooled effect on frequency as well as the adjusted values from the four adjustment approaches.

## RESULTS

We included data from 56 study arms compared with placebo from 47 randomized controlled trials of beta-blockers to prevent episodic migraine headache. Commonly collected outcomes were headache frequency and severity. Among the 286 total reported outcomes reported as having been collected, 142 (48%) provided the data in the manuscript, and the remaining 144 (52%) provided no data. Our assessment of risk of ORB is shown in Appendix Table 2.

### Test for Likelihood of Outcome Reporting Bias Being Present

Among studies reporting data, 86 outcomes were statistically significant and 55 were not. Among the 142 outcomes that were reported as collected, but for which no data were provided, 8 were reported to be significant, 22 were reported to be non-significant, and 112 did not report significance (Table 2). Reported outcomes were more likely to be significant than outcomes that were collected but not reported (RR, 2.4; 95% CI, 1.9 to 2.9). This suggests that the data are not missing at random. The Little test failed to find evidence that this data was not MCAR ( $p = 0.43$ ).

### Efficacy Adjusted for Missing Data

Based on reported data, there was evidence that beta-blockers were effective in preventing migraine headaches for most outcomes (Table 1). Headache frequency was reduced:  $-1.2$  headaches/month (95% CI,  $-1.5$  to  $-0.8$ ). Our approach to adjust for ORB reduced the benefit to  $-1.0$  headaches/month (95% CI,  $-1.3$  to  $-0.7$ ). Our approach adjusted estimated effects toward the null for all outcomes analyzed with reductions ranging from 7 to 88% of the unadjusted values (Table 2). For estimates that used SMDs, the adjustment resulted in effect sizes deemed to be trivial for most

**Table 1 Outcomes Reported by Studies as Collected (With or Without Abstractable Data) or Uncollected**

	Outcome collected, results reported			Outcome collected, results not reported			Outcome not collected
	Statistically significant	Not significant	No comment on significance	Statistically significant	Not significant	No comment on significance	
Headache frequency	22	18	0	1	1	6	8
Headache days	8	5	0	2	1	6	34
50% improvement	10	11	0	0	4	1	30
Headache index	11	12	0	1	3	26	3 (23 studies had variables to calculate the headache index)
Severity	13	9	0	3	5	21	5
Duration	5	12	0	1	3	20	14
Analgesic use	6	4	0	6	7	27	6
HRQOL	1	1	0	0	1	1	52
Work absenteeism	0	1	0	0	0	1	54
All outcomes	75	71	0	14	24	107	209

HRQOL health-related quality of life

outcomes.<sup>29</sup> There was no significant difference in estimates of variance when using either regression or multiple imputation approaches.

Our proposed approach produced values similar to those of the Copas method (Table 2). The Frosi method failed to converge for several outcomes. For headache frequency, the Frosi method resulted in no change in the pooled estimate ( $-1.1$ ; 95% CI,  $-1.4$  to  $-0.8$ ). For all outcomes, the Frosi method increased the pooled benefit by 13 to 215% (Table 3). The trim and fill method failed to identify the presence of biased data for any outcome.

## Simulation Results

Our simulated dataset reduced the frequency of headaches ( $-1.4$  headaches/month; 95% CI,  $-1.8$  to  $-1.0$ ; 40 studies;  $I^2 = 96%$ ). This served as our estimate of the “true” effect of the

intervention. Table 3 summarizes the average of unadjusted and adjusted pooled effects over 1000 replications using varying rates of missing (NMAR) data. As the percentage of missing data increased, the pooled benefit also increased, from a reduction of 1.4 headaches per month (95% CI,  $1.0$ – $1.8$ ) when 5% of outcomes were missing to the method (1.9 fewer headaches (95% CI,  $1.4$ – $2.3$ )) when 40% of the data were missing. Using the simulation data with missing data, both our method and the Copas method showed reductions in pooled estimates that were similar to the true values. The Frosi method yielded similar results when the missing rate was at most 25%. With higher degrees of missing data, the Frosi method resulted in adjustments that suggested greater benefit than the true effect. The trim and fill method failed to detect any evidence of ORB for all scenarios.

**Table 2 Summary Effects, Unadjusted and Adjusted for Missing Outcomes**

Variable	Reported (Missing)	Unadjusted (REML)	Proposed Method	Copas Adjustment	Frosi Adjustment	Trim and Fill
Frequency (Headaches/month)	40 (8)	$-1.2$ ( $-1.5$ to $-0.8$ ), df=39, $I^2=68%$	$-1.0$ ( $-1.3$ to $-0.7$ ), df=47*, $I^2=65%$	$-0.9$ ( $-1.8$ to $-0.5$ )	$-1.1$ ( $-1.4$ to $-0.8$ )	NC
Headache Days	13 (9)	$-0.8$ ( $-1.1$ to $-0.5$ ), df=12, $I^2=0%$	$-0.6$ ( $-0.9$ to $-0.4$ ), df=21, $I^2=0%$	$-0.4$ ( $-0.7$ to $-0.1$ )	$-0.9$ ( $-1.2$ to $-0.5$ )	NC
50% Improvement (OR)	21 (5)	2.5 (1.8–3.4), df=20, $I^2=51%$	2.2 (1.7 to 2.9), df=25, $I^2=42%$	2.0 (1.5 to 2.7)	2.7 (2.1 to 3.7)	NC
Severity (SMD)	23 (29)	$-0.4$ ( $-0.5$ to $-0.2$ ), df=21, $I^2=27%$	$-0.1$ ( $-0.2$ to $-0.06$ ), df=48, $I^2=24%$	$-0.08$ ( $-0.2$ to 0.02)	$-0.3$ ( $-0.5$ to $-0.2$ )	NC
Analgesic medications used (Pills/month)	10 (37)	$-1.8$ ( $-2.8$ to $-0.8$ ), df=9, $I^2=66%$	$-0.05$ ( $-0.2$ to 0.1), df=46, $I^2=0%$	$-1.0$ ( $-1.9$ to 0.07)	No convergence	NC
Headache Index (SMD)	23 (29)	$-0.3$ ( $-0.5$ to $-0.2$ ), df=22, $I^2=22%$	$-0.1$ ( $-0.2$ to $-0.1$ ), df=51, $I^2=17%$	0.0 ( $-0.2$ to 0.2)	$-0.4$ ( $-0.6$ to $-0.2$ )	NC
Duration (Hours/month)	17 (29)	$-1.6$ ( $-2.3$ to $-0.9$ ), df=16, $I^2=0%$	$-0.04$ ( $-0.4$ to 0.3), df=34, $I^2=0%$	$-0.8$ ( $-1.5$ to $-0.1$ )	$-5.4$ ( $-8.8$ to $-2.0$ )	NC
Health-related Quality of Life (SMD)	2 (3)	$-0.2$ ( $-0.6$ to 0.2), df=1, $I^2=N/A$	$-0.08$ ( $-0.24$ to 0.09), df=4, $I^2=0%$	0.1 ( $-0.3$ to 0.5)	No convergence	NC
Work Absence (Days/month)	1 (0)	0.09 ( $-0.3$ to 0.5) Only 1 study	N/A	N/A	N/A	N/A

Abbreviations: N/A not applicable, NC no change, REML random effects maximum likelihood, SMD standard mean difference

\*Degrees of freedom based on imputing for studies reporting outcome was collected but not reported



**Table 3 Impact of Varying Proportion of Missing Data (Simulated Data with 1000 Repetitions)**

Percent dropped	Full data set (no missing data, 40 studies)	Unadjusted	Proposed method	Copas method	Frosi method	Trim and fill
5%	- 1.4 (- 1.8 to - 1.0)	- 1.4 (- 1.8 to - 1.0) 38 studies	- 1.3 (- 1.7 to - 1.0)	- 1.3 (- 1.7 to 0.9)	- 1.4 (- 1.9 to - 1.0)	NC
10%	- 1.4 (- 1.8 to - 1.0)	- 1.5 (- 1.9 to - 1.3) 36 studies	- 1.4 (- 1.8 to - 1.2)	- 1.4 (- 1.8 to - 1.1)	- 1.3 (- 1.6 to - 0.9)	NC
15%	- 1.4 (- 1.8 to - 1.0)	- 1.5 (- 1.9 to - 1.2) 34 studies	- 1.3 (- 1.7 to - 0.9)	- 1.2 (- 1.6 to - 0.8)	- 1.4 (- 1.7 to - 1.0)	NC
20%	- 1.4 (- 1.8 to - 1.0)	- 1.6 (- 2.0 to - 1.2) 32 studies	- 1.3 (- 1.7 to - 0.9)	- 1.2 (- 1.6 to - 0.8)	- 1.8 (- 2.1 to - 1.5)	NC
25%	- 1.4 (- 1.8 to - 1.0)	- 1.6 (- 2.1 to - 1.2) 30 studies	- 1.3 (- 1.7 to - 0.9)	- 1.2 (- 1.7 to - 0.8)	- 1.9 (- 2.2 to - 1.7)	NC
30%	- 1.4 (- 1.8 to - 1.0)	- 1.7 (- 2.1 to - 1.3) 28 studies	- 1.3 (- 1.7 to - 0.9)	- 1.3 (- 1.7 to - 0.8)	- 2.0 (- 2.2 to - 1.8)	NC
35%	- 1.4 (- 1.8 to - 1.0)	- 1.8 (- 2.2 to - 1.4) 26 studies	- 1.3 (- 1.6 to - 0.9)	- 1.3 (- 1.7 to - 0.9)	- 2.1 (- 2.3 to - 1.9)	NC
40%	- 1.4 (- 1.8 to - 1.0)	- 1.9 (- 2.3 to - 1.4) 24 studies	- 1.2 (- 1.6 to - 0.9)	- 1.4 (- 1.8 to - 1.0)	- 1.9 (- 2.0 to - 1.8)	NC

NC no change in estimate

**Sensitivity Analysis**

Varying the number of included studies produced similar adjustments with both our approach and the Copas method. The Frosi method failed to converge for our 5-sample data set. The trim and fill method failed to recognize the presence of ORB in all subsamples (Table 4).

**DISCUSSION**

In this cohort of randomized controlled trials evaluating how well beta-blockers prevent migraine headaches, significant outcomes were twice as likely to be reported than non-significant ones. Stratifying outcomes into contingency tables based on author reports is a simple method to assess the likelihood that ORB occurred. We found that our simple approach to adjusting for unreported outcomes gave results similar to those of the Copas method’s more complex approach. Frosi’s multivariate method overestimated benefits, suggesting that reported outcomes were systematically biased. The trim and fill method performed poorly and should not be used to account for ORB. While none of the adjusted outcomes resulted in significant results becoming statistically

non-significant, many became clinically small or even trivial.<sup>29</sup>

In our simulation, as the unreported rate for non-significant observations increased, the pooled effect estimates became increasingly more biased. Our simple method and the Copas method provided bias corrections that brought the adjusted values close to the true pooled estimate. Frosi’s multivariate method gave similar results when the unreported rate was less than 25%. However, when the unreported rate increased, Frosi’s multivariate adjustment resulted in estimates of benefit that were greater than the “true” values. The trim and fill method performed poorly, identifying no potentially missing studies.

Our method could allow meta-analysts to quickly and simply estimate the potential impact of ORB on treatment effects with similar accuracy as prior complex methods. The simplicity of our method could allow more frequent use in meta-analyses of randomized controlled trials, which in turn could have important clinical implications. It would allow clinicians and policy makers to have a better understanding of the range of likely benefit.

Multivariate approaches to deal with missing data are useful when most data are present and there is a correlation between unreported and reported outcomes. Unfortunately, systematic

**Table 4 Adjusted and Unadjusted Results for Frequency with Progressively Smaller Data Sets**

Number of Studies	Included/Missing high-risk studies	Unadjusted	Proposed Method	Copas Method	Frosi Method	Trim and Fill
40	32/8	-1.2 (-1.5 to -0.8), df=31, I <sup>2</sup> =76%	-0.9 (-1.3 to -0.6), df=39, I <sup>2</sup> =72%	-0.9 (-1.2 to -0.5)	-1.1 (-1.6 to -0.7)	NC
30	24/6	-0.9 (-1.3 to -0.6), df=23, I <sup>2</sup> =70%	-0.8 (-1.1 to -0.5), df=29, I <sup>2</sup> =68%	-0.6 (-0.7 to -1.0)	-0.9 (-1.3 to -0.5)	NC
20	17/3	-0.9 (-1.3 to -0.4), df=16, I <sup>2</sup> =83%	-0.8 (-1.2 to -0.3), df=19, I <sup>2</sup> =81%	-0.6 (-1.6 to -0.2)	-0.8 (-1.5 to -0.1)	NC
10	9/1	-1.3 (-1.8 to -0.9), df=8, I <sup>2</sup> =53%	-1.2 (-1.7 to -0.8), df=9, I <sup>2</sup> =52%	-1.2 (-1.6 to -0.8)	-1.2 (-1.8 to -0.8)	NC
5	3/2	-1.5 (-1.6 to -1.4), df=2, I <sup>2</sup> =0.0%	-1.0 (-1.8 to -0.3), df=4, I <sup>2</sup> =60%	-1.4 (-1.6 to -1.3)	No Convergence	NC

NC No change

reviews often address numerous outcomes, but few studies provide results from all outcomes of interest. When many outcomes are simultaneously missing (a common occurrence), multivariable approaches face convergence issues. Our data suggest that imputing missing data, based on biased datasets, amplifies the bias. This may explain why the multivariate approach (Frosi) yielded progressively more biased findings as the rate of unreported data increased.

There are several important limitations to our study. Our method shares the same limitation as the Copas method. It does not account for correlation among the various outcomes. This study was limited to a single intervention type of a single disorder. Additional evaluations of other sets of studies are necessary to test generalizability. In addition, whether our approach will work for studies incorporating non-randomized trials needs to be investigated. It is possible that unreported outcomes had benefit, but were non-significant. If so, our adjustment would be conservative and would reduce the true treatment effect more than including the unreported outcomes would have. On the other hand, if some of the missing data showed non-significant harm, our approach could underestimate the impact of missing data. By setting all missing effects to zero, we may have underestimated the true heterogeneity. Finally, clinical trial data with unreported outcomes were not obtained to assess the performance of our proposed models. Obtaining unreported outcome data from the original study authors is the best approach for dealing with data that are not missing at random. In practice, though, this is difficult to achieve.

There are possible long-term solutions. Requiring protocol registration at inception could reduce selective reporting.<sup>30</sup> This would require that protocols be more complete and sufficiently detailed than most clinical protocol registration sites are able to enforce.<sup>31, 32</sup> Another approach would be to require researchers to make all study data publicly available. Many researchers are nervous about this requirement, because they may have secondary analyses they hope to perform and would prefer to exhaust their data before posting it. It might be necessary to limit other researcher's access to the data for purposes other than systematic reviews. At a minimum, researchers should annotate that additional data are forthcoming and should be required to "close out" the study, after exhausting their analyses, by posting all data to the registry. A final approach would be to require articles to list all outcomes collected with the effect estimate with confidence interval for all outcomes.<sup>14</sup>

In summary, we found strong evidence that authors selectively report outcomes in this cohort of migraine headaches being treated by beta-blockers. We evaluated a simple method to explore the impact of selective reporting in randomized controlled trials, finding that it does well as long as less than 30% of outcomes are missing. We recommend that meta-analysts consider exploring the impact of ORB on their results as a sensitivity analysis for the possibility that the results are overstated. There are several additional studies that could be

considered. First, it would be interesting to assess whether making a distributional assumption about missing values centered on the null could affect the results, though this would complicate the calculations. In addition, studies need to be done to explore how well this method works for non-randomized trials and in other clinical data sets.

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11606-021-07135-3>.

**Acknowledgements:** To Katherine Sherman for discussion about possible statistical approaches for this manuscript.

**Corresponding Author:** Jeffrey L Jackson, MD MPH; General Medicine Section, Zablocki VAMC, Milwaukee, WI, USA (e-mail: [jjackson@mcw.edu](mailto:jjackson@mcw.edu)).

**Data Availability** All the data for this study is available via Figshare:7870019.

**Declarations:**

**Conflict of Interest:** The authors declare that they do not have a conflict of interest.

## REFERENCES

- Jackson JL, Kuriyama A. From the Editors' Desk: Bias in Systematic Reviews-Let the Reader Beware. *J Gen Intern Med*. 2018;33(2):133-135. Not in File. doi:<https://doi.org/10.1007/s11606-017-4236-2>
- Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*. 2013;8(7):e66844. Not in File. doi:<https://doi.org/10.1371/journal.pone.0066844>
- Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med*. 2006;25(20):3443-3457. Not in File. <https://doi.org/10.1002/sim.2380>
- Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA*. 2006;295(6):676-680. Not in File. <https://doi.org/10.1001/jama.295.6.676>
- Rothstein HR, Sutton AJ, Borenstein M. *Publication bias in meta-analysis. Prevention, assessment and adjustments*. John Wiley & Sons; 2005.
- Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457-2465. Not in File. doi:<https://doi.org/10.1001/jama.291.20.2457>
- >Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials*. :11:52. <https://doi.org/10.1186/1745-6215-11-52>
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 302(9):977-984. Not in File. <https://doi.org/10.1001/jama.2009.1242>
- Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326(7400):1171-1173. Not in File. doi:<https://doi.org/10.1136/bmj.326.7400.1171>
- Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med*. 26(25):4544-4562. Not in File. <https://doi.org/10.1002/sim.2889>
- Rosenthal R. The file drawer problem and tolerance for null results. *Psychological Bulletin* 1979. 86(3):638-641. Not in File.
- Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315(7109):629-634. Not in File.

13. Page MJ, Higgins JP. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Syst Rev*. 2016;5(1):108. Not in File. doi:<https://doi.org/10.1186/s13643-016-0289-2>
14. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365. Not in File.
15. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Wiley.; 1987.
16. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One*. 2009;4(9):e7078. <https://doi.org/10.1371/journal.pone.0007078>
17. Copas J, Dwan K, Kirkham J, Williamson P. A model-based correction for outcome reporting bias in meta-analysis. *Biostatistics*. Apr 2014;15(2):370-83. doi:<https://doi.org/10.1093/biostatistics/kxt046>
18. Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. *J Clin Epidemiol*. 2015;68(5):542-550. Not in File. <https://doi.org/10.1016/j.jclinepi.2014.11.017>
19. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med*. 2012;31(20):2179-2195. Not in File. <https://doi.org/10.1002/sim.5356>
20. Kirkham JJ, Altman DG, Chan AW, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. *Bmj*. 2018;362:k3802. <https://doi.org/10.1136/bmj.k3802>
21. Copas Adjustment Method. Accessed Feb 16 2020. <http://outcome-reporting-bias.org/HOME/Copas>
22. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. Jun 2000;56(2):455-63. doi:<https://doi.org/10.1111/j.0006-341x.2000.00455.x>
23. Jackson JL, Kuriyama A, Kuwatsuka Y, et al. Beta-blockers for the prevention of headache in adults, a systematic review and meta-analysis. *PLoS One*. 2019;14(3):e0212785. doi:<https://doi.org/10.1371/journal.pone.0212785>
24. Silberstein SD, Holland S, Freitag F, Dodick DW, Argoff C, Ashman E. Evidence-based guideline update: pharmacologic treatment for episodic migraine prevention in adults: report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Headache Society. *Neurology*. 2012;78(17):1337-1345. Not in File. <https://doi.org/10.1212/WNL.0b013e3182535d20>
25. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med*. 1996;15(6):619-29. [https://doi.org/10.1002/\(sici\)1097-0258\(19960330\)15:6<619::Aid-sim188>3.0.Co;2-a](https://doi.org/10.1002/(sici)1097-0258(19960330)15:6<619::Aid-sim188>3.0.Co;2-a)
26. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55-79. doi:<https://doi.org/10.1002/jrsm.1164>
27. Little R. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988;1988(83):1198-1202.
28. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. *J Clin Epidemiol*. 1992;45(7):769-773. Not in File.
29. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27(3 Suppl):S178-S189. Not in File.
30. Chalmers I. Underreporting research is scientific misconduct. *JAMA*. 1990;263(10):1405-1408. Not in File.
31. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract*. 2002;8(3):353-359. Not in File.
32. Norris SL, Holmer HK, Fu R, Ogden LA, Viswanathan MS, Abou-Setta AM. Clinical trial registries are of minimal use for identifying selective outcome and analysis reporting. *Res Synth Methods*. 2014;5(3):273-284. Not in File. <https://doi.org/10.1002/jrsm.1113>

**Publisher's Note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.