Mitigating Publication Bias Using Bayesian Stacking

Thomas A. Gibson

August 2, 2021

1 Introduction

Results from a meta-analysis may be skewed and unreliable in the presence of publication bias, where the publication or non-publication of a study depends on the statistical significance or magnitude of its results (Rothstein et al., 2006). Statistical methods for publication bias have been designed for sensitivity analysis, testing for the presence/magnitude of publication bias, and calculating bias-corrected parameter estimates. Most methods are either based on the funnel plot or selection models.

Methods based on the funnel plot (Light and Pillemer, 1984) – a scatterplot of effect sizes against their standard errors – use the plot's asymmetry to test or correct for bias. The methods assume that publication bias operates primarily on smaller studies and that larger studies are much less affected. A popular non-parametric test for publication bias (Begg and Mazumdar, 1994) uses Kendall's tau to measure the rank correlation between standardized observed effect sizes and the effect sizes' standard errors. Egger's test (Egger et al., 1997) fits a linear regression of observed standard normal deviates against the observed precision of estimates, with the null hypothesis being that the regression intercept is zero. Other regression methods

(Macaskill et al., 2001; Rücker et al., 2008; Thompson and Sharp, 1999; Peters et al., 2006) are similar and use regression weights or transformations to improve upon Egger's test in the presence of heterogeneity or for dichotomous outcomes (Jin et al., 2015). Lin and Chu (2018) develops a measure for the severity of publication bias based on the skewness of standardized deviates. The trim-and-fill method (Duval and Tweedie, 2000) estimates the number of missing studies and their effect sizes using funnel plot asymmetry and gives an adjusted pooled effect estimate. The authors recommend using it as a sensitivity analysis based on the potential number of missing studies, with general guidelines given in Shi and Lin (2019). The trim-and-fill method is the only funnel plot-based method that offers an adjusted mean estimate, and it is not recommended if there is heterogeneity present (Jin et al., 2015).

A second class of methods is based on *selection models*, first described in Hedges (1984). Some models explicitly model the probability of publication for individual studies as a function of their p-values (Hedges, 1992; Givens et al., 1997; Vevea and Hedges, 1995) or as a function of both the effect size and standard error (Copas, 1999; Copas and Shi, 2000, 2001). Earlier selection models were recommended for bias-corrected effect size estimates, and were later recommended only for sensitivity analyses because of identifiability issues in smaller meta-analyses (Vevea and Woods, 2005; Jin et al., 2015). Sensitivity analyses use a grid representing varying levels of publication bias and estimate the mean effect under each assumed scenario. If results do not change much under sever publication bias they are considered robust, and if results do change under mild publication bias they are considered sensitive to publication bias. Bayesian implementations of the Copas selection model (Mavridis et al., 2013; Bai et al., 2020) have again allowed for estimation of mean effect sizes.

Recent approaches to mitigating publication bias have used Bayesian model averaging (BMA) to consider a set of potential selection functions. Guan and Vandekerckhove (2016) considers four different selection functions, including a no-bias model, an extreme-bias model where results with p-values $p > \alpha$ are never published, a 1-step function where results with $p > \alpha$ are published with some probability $\pi < 1$, and a model inspired by Givens et al. (1997) where the probability of publication decreases exponentially with p. The authors only implement the models in a fixed-effects framework. Maier et al. (2020) considers a set of 12 models, using a $2 \times 2 \times 2$ factorial design with fixed/random effects, a true null/alternative hypothesis, and the presence/absence of publication bias. The authors consider two-step and three-step selection functions based on p-values when publication bias is assumed.

BMA effectively assumes that one of the considered models is the "true" model, which we call the $\mathcal{M} - \textit{closed}$ setting. BMA does not perform as well under the $\mathcal{M} - \textit{complete}$ or $\mathcal{M} - \textit{open}$ settings, where the true data generating mechanism is too complex to implement or to put into a probabilistic framework (Clyde and Iversen, 2013). Multiple issues arise for BMA in these settings, including a) the need to specify prior model probabilities, which makes little sense when we know the true model is not in our list, and b) the model weights from BMA will converge to the 1 for the model "closest" to the true model in terms of Kullback-Leibler divergence, and 0 for all others. Bayesian stacking (Yao et al., 2018, 2021) is a related method that outperforms BMA and solves the above issues by calculating optimal weights using expected log-predictive densities and leave-one-out cross validation.

We propose a method using Bayesian stacking to mitigate publication bias by incorporating multiple different selection models. Copas and Shi (2001) recommends a sensitivity analysis over a grid of possible patterns of publication bias, each of which returns a bias-adjusted mean effect size estimate. A stacked-average estimate can be obtained through a weighted average of estimates with the weights coming from Bayesian stacking. Assumed patterns of publication bias that do not fit the data will be given little weight. We also combine multiple types of models, including an exponential decay model (Givens et al., 1997), step functions (Hedges, 1992; Vevea and Hedges, 1995), and the Copas selection model (**?**).

References

Bai, R., Lin, L., Boland, M. R., and Chen, Y. (2020). A robust bayesian copas selection model for quantifying and correcting publication bias. *arXiv preprint arXiv:2005.02930* .

Begg, C. B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* pages 1088–1101.

Clyde, M. and Iversen, E. S. (2013). Bayesian model averaging in the m-open framework. *Bayesian theory and applications* **14,** 483–498.

Copas, J. (1999). What works?: selectivity models and meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **162,** 95–109.

Copas, J. and Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics* **1,** 247–262.

Copas, J. and Shi, J. Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical methods in medical research* **10,** 251–265.

Duval, S. and Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56,** 455–463.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj* **315,** 629–634.

Givens, G. H., Smith, D., and Tweedie, R. (1997). Publication bias in meta-analysis: a bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science* **12,** 221–250.

Guan, M. and Vandekerckhove, J. (2016). A bayesian approach to mitigation of publication bias. *Psychonomic bulletin & review* **23,** 74–86.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics* **9,** 61–85.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science* **7,** 246–255.

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

Light, R. J. and Pillemer, D. B. (1984). Summing up: the science of reviewing research.

Lin, L. and Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics* **74,** 785–794.

Macaskill, P., Walter, S. D., and Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in medicine* **20,** 641–654.

Maier, M., Bartoš, F., and Wagenmakers, E.-J. (2020). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging.

Mavridis, D., Sutton, A., Cipriani, A., and Salanti, G. (2013). A fully bayesian application of the copas selection model for publication bias extended to network meta-analysis. *Statistics in medicine* **32,** 51–66.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama* **295,** 676–680.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Rücker, G., Schwarzer, G., and Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* **27,** 746–763.

Shi, L. and Lin, L. (2019). The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine* **98,**.

Thompson, S. G. and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine* **18,** 2693–2708.

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

Vevea, J. L. and Woods, C. M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological methods* **10,** 428.

Yao, Y., Pirš, G., Vehtari, A., and Gelman, A. (2021). Bayesian hierarchical stacking. *arXiv preprint arXiv:2101.08954* .

Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis* **13,** 917–1007.