Improving Clinical Trial Analyses Through Meta-Analysis: Estimating Heterogeneity in Meta-Analyses for Binary Outcomes

Jake Baldauf, Nolan R. Cole, Dr. Brinley Zabriskie
BYU Department of Statistics

Abstract

A meta-analysis is a statistical procedure that combines data from multiple studies. This is particularly useful in clinical research when several studies produce conflicting results and investigators must determine a medication's true efficacy. Determining the extent to which these studies differ from one another due to variation in treatment administration or patient populations is key to identifying this efficacy. When this variability is greater than researchers could expect from chance alone, statistical heterogeneity is present. Estimating the amount of heterogeneity present is key to advancing science and protecting patient health. We investigate the performance of heterogeneity variance estimators for binary data in conjunction with continuity corrections. We discover that while specific methods of estimating heterogeneity with certain continuity corrections outperform others, all methods considered consistently underperform. Understanding these methods' performance will allow clinical researchers to better estimate heterogeneity, thereby improving their ability to find new, safe treatments.

Example Meta-Analysis

Multiple myeloma is a cancer with no proven treatment or cure. To find such a treatment, suppose researchers have treated multiple myeloma patients with stem cells, with 10 clinical trials being conducted (See Table 1). To protect patient health, researchers recorded the number of adverse events, such as death, that occurred within each trial. The odds (Events / Non-Events) that a patient receiving stem cell treatment experiences an adverse event compared to the odds of a control patient is an odds ratio (Treatment Odds / Control Odds), a common measure in clinical trials with binary outcomes. Some studies concluded that stem cell treated patients were at lower risk for an adverse event than their control counterparts (odds ratio < 1). Other studies concluded the opposite (Odds ratio > 1). To determine if there truly is an increased risk associated with the stem cell treatment, a meta-analysis can be performed which takes a weighted average of the individual study odds ratios to produce one overall odds ratio. However, if the adverse events are rare, there may be zero events in either the treatment arm, control arm, or both. A study where this occurs is known as a zero study.

Table 1

	Stem Cell Treatment		Control		
Study	Event	Non-Event	Event	Non-Event	Odds Ratio
1	0	23	1	37	0
2	0	10	0	28	NA
3	1	96	3	69	0.24
4	2	83	13	60	0.11
5	3	71	0	41	NA
6	13	98	5	35	1.07
7	0	15	0	83	NA
8	7	61	0	89	NA
9	3	46	8	68	0.55
10	0	63	0	79	NA

These zero studies result in undefined odds ratios and prevents the accurate estimation of heterogeneity. This in turn prevents researchers from safely estimating the risk associated with a treatment. One method of dealing with zero studies is the continuity correction.

Methods

We designed a simulation study based on Zhang, et. al (2020). We use various heterogeneity variance estimators and continuity corrections. The heterogeneity estimators and continuity corrections we included in our study are:

Heterogeneity Estimators
- DerSimonian-Laird
- Restricted Maximum Likelihood
- Paule-Mandel
- Improved Paul-Mandel
- Bootstrap DerSimonian-Laird

Continuity Corrections
- Constant: add 0.5 to each zero cell
- Zero exclusion: exclude zero studies
- Increase all: if one zero exists, add constant to all event cells
- Reciprocal: add (constant / number of patients) to each zero cell
- Empirical: Uses empirical odds ratio that minimizes bias

We simulated data with multiple parameters. We set the true log odds ratio to be 1 and the event rate in the control arm to be 0.03. The level of heterogeneity variance was varied (0, 0.5, 1.0). Each meta-analysis consisted of 10 studies, and each study contained between 50 and 200 individuals. We simulated unbalanced data sets, meaning both the control and treatment arms did not contain an equal number of subjects. Varying these parameters allowed us to simulate realistic data to analyze which pair of continuity correction and heterogeneity estimator minimized the pooled log odds ratio bias.
We ran our simulation using R (version 4.0.3) and the meta library (version 4.16-1). We simulated 100 replicates for each possible combination of all parameters mentioned above.

Figure 1

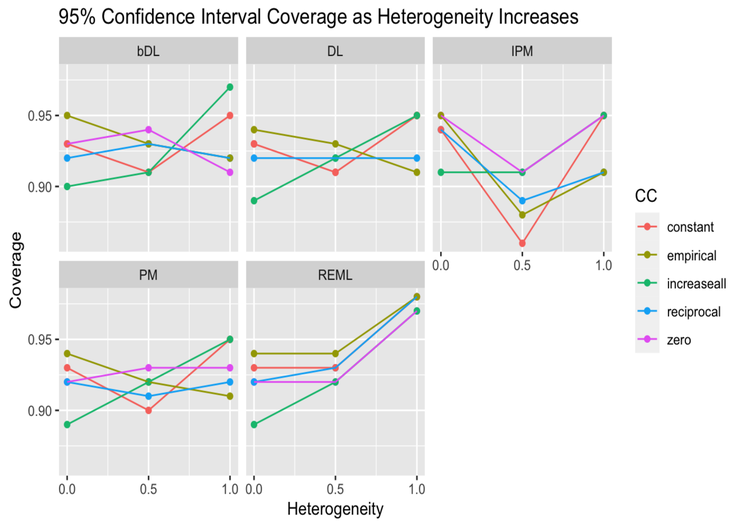


Figure 2

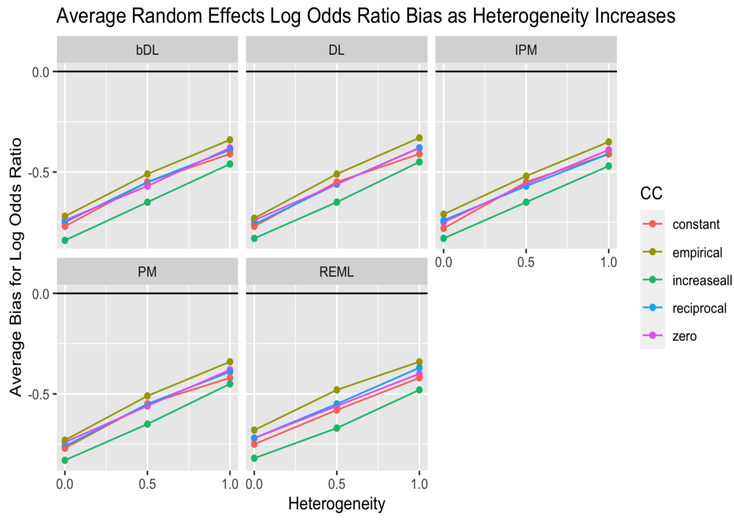
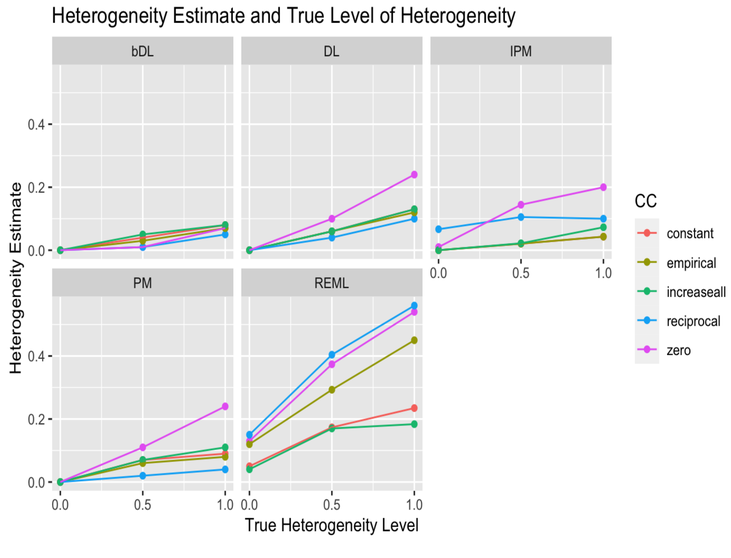


Figure 3



Results

Figure 1 shows the average 95% confidence interval coverage of the log odds ratio for different levels of heterogeneity for each estimator and continuity correction. Each estimator performed decently well, except the Improved Paule-Mandel method when moderate levels of heterogeneity were present. However, it was comparable to the other estimators when there was no or high levels of heterogeneity. Each continuity correction performed differently for each estimator. The zero exclusion continuity correction performed the best for the Improved Paule-Mandel estimator while the empirical continuity correction performed the best for the restricted maximum likelihood estimator. The boostrap DerSimonian-Laid and Improved Paule-Mandel estimators had the highest coverage of the log odds ratio when there was no heterogeneity present. The bootstrap DerSimonian-Laird estimator, with the use of the zero exclusion continuity correction, performed the best with moderate levels of heterogeneity. The restricted maximum likelihood estimator, with the use of the empirical continuity correction or the reciprocal continuity correction, performed the best with high levels of heterogeneity.

Figure 2 illustrates the average bias of the log odds ratio as heterogeneity increases. The empirical continuity correction consistently performed the best at each level of heterogeneity for all estimators. The increase all continuity correction consistently resulted in the most bias for the pooled log odds ratio. When no heterogeneity was present, estimators perform very poorly in general. A trend that holds for all estimator-continuity correction combinations was that bias decreases as heterogeneity variance increases. All other continuity correction methods performed comparably.

Figure 3 depicts the bias of heterogeneity variance estimates. All estimators severely underestimated heterogeneity variance at moderate and high levels of heterogeneity. When no heterogeneity was present, most estimators performed well. The bootstrap DerSimonian-Laird method consistently performed the worst of all estimators tested, regardless of the implemented continuity correction. The restricted maximum likelihood estimator resulted in the highest heterogeneity estimate at high and moderate levels of heterogeneity when using the reciprocal and zero exclusion continuity corrections. It also overestimated the heterogeneity present at no levels of heterogeneity. However, in real practice, it is very unlikely that no heterogeneity will be present. Therefore, we recommend the restricted maximum likelihood estimator based on these results. Overall, the zero exclusion continuity correction resulted in the highest estimates of heterogeneity for almost all estimators observed in the simulation. It is interesting to note there does not appear to be a continuity correction method that performs well for all three metrics.

Conclusion

While meta-analysis is a powerful tool for clinical investigators to determine the true risk of an adverse event, it is not without its own risks. We have demonstrated via simulation the extreme effect that the choice of continuity correction and heterogeneity estimator has in random-effects meta-analysis. For patients with diseases like multiple myeloma, the development of treatments will rely heavily on multiple clinical trials and meta-analysis. Thus, it is of utmost importance to develop methods to effectively estimate heterogeneity and process zero event studies. In the future, it will be beneficial to expand our simulation study by (1) simulating both large and small study sizes, (2) varying the number of studies in the meta-analysis, (3) using balanced versus unbalanced study arms, and (4) varying the pooling method. Currently, we are also deriving a new permutation-based approach to estimate heterogeneity.

Acknowledgements & References

We would like to acknowledge the financial and academic support provided by the Brigham Young University Department of Statistics and College of Physical & Mathematical Sciences.

Zhang, C., Chen, M., and Wang, X. (2020). Statistical methods for quantifying between-study heterogeneity in meta-analysis with focus on rare binary events. Statistics and Its Interface, 13:449–464
Sweeting, M., Sutton, A., and Lambert, P. (2004). What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data.Statistics in Medicine, 23(9):1351–1375.