MATEMATIKA, 2011, Volume 27, Number 2, 121–128

Estimating the Bias in Meta Analysis Estimates for Continuous Data With Non-Random Missing Study Variance

Nik Ruzni Nik Idris
Kulliyyah of Science
International Islamic University Malaysia
25200 Kuantan Pahang, Malaysia
e-mail: ruzni@iium.edu.my

Abstract This paper examines, analytically, the biases introduced in the meta analysis estimates when the study-level variances are missing with non-random missing mechanism (MNAR). Two common approaches in handling this problem is considered, namely, the missing variances are imputed, and the studies with missing study-variances are omitted from the analysis. The results suggest the variance will be underestimated if the magnitude of the study-variances that are missing are mostly larger implying false impression of precision. On the other hand, if the missing variances are mostly smaller, the variance of the effect size will be overestimated.

Keywords meta analysis; variance estimates; not missing at random; imputation;

2010 Mathematics Subject Classification 62P10; 62P12; 62P15; 62P25

1 Introduction

A common drawback with meta analysis data is when the variability measures, particularly the standard deviations or the variances, are not reported, or "missing" in the published report of the individual studies. This is particularly common in earlier publications. Missing study level variance is a serious problem in meta analysis and there are a variety of methods for dealing with the issue. One of the common methods is through indirect approach in which missing values are replaced by a form of imputation [1].

There are a variety of reasons why the outcomes measures are omitted from publications, ranging from journal space constraint to the values which are deemed not interesting or not statistically significant [2]. When the study variances are not reported, it is normal practice in meta analysis to assume that they are missing completely at random (MCAR), implying that recorded observed variances are random sample of the population of the variances from all studies [3]. However, it is possible that some studies do not report the variances because the values are large. Smaller studies, for instance, are more likely not to report the variances compared to those from larger studies. If this is the case then the variances are considered to be missing not at random (MNAR). Studies on the estimates based on random effect model [4] suggested that imputation was a good way of recovering the missing information and increasing the precision of the overall effect and the corresponding variance if the individual study variances are missing under the MCAR mechanism.

This paper examines, analytically, the effects of mean imputation on the overall effect size and the corresponding variance when the individual study variances are not missing at random (NMAR). The estimates are based upon the Fixed Effect model. While the Random Effect model appears to be a more popular choice of model, the Fixed Effect model is often used when the assumptions associated with the Random Effect models seem unrealistic. For instance, the assumptions of normally distributed between-study errors poses problems in both its validity, and in our ability to check that validity for meta analyses based on small number of studies. We considered a scenario where the mechanism of missing study-variances are non-random (MNAR), namely, the "missing" study-variances are larger than or smaller than those that are reported. The main aim is to study the biases introduced in the overall estimates. We look at the case where the missing variances are imputed using the mean imputation and when the studies with missing study-variances are omitted from the analysis.

2 Method

The main investigation is through analytical derivation of the overall effects estimate and the corresponding variance based on (1) complete data, where all studies are assumed to report the variances (2) incomplete data where the studies with missing study-variances are excluded from analysis, and (3) complete imputed data, where missing study-variances are imputed using the mean imputation. The observed and expected biases were derived for estimates based on (2) and (3) against those based on (1).

2.1 Non-Random Mechanism of Missing Study Variances (NMAR)

Assume that there are N studies, each with complete treatment effect size and variance information. To create the non-random missing variances scenario, the variances are split into 2 groups. Let a number x of these N studies do not report the variances information, and we assume that these are the 'missing' variances. Thus, the missing group comprised of σ_x^2 and the available group comprised of σ_{N-x} , with common within study variances in each group. That is, we let σ_i^2 take the following values

$$\sigma_i^2 = \begin{cases} \sigma_x^2 & \text{for } i = 1,2,\ldots,x, \\ \sigma_{N-x}^2 & \text{for } i = x{+}1, \ldots,\text{N} \end{cases} \tag{1}$$

2.2 The Fixed Effects Meta Analysis Model

The estimate of study-specific treatment effects using the Fixed Effect model is given by

$$y_i = \theta + \epsilon_i$$

is the estimate of treatment effect in study i, is the overall true treatment effect and is the random error for study i= 1,2,..., N. We assume that the random error, ϵ_i to be normally distributed with mean 0 and variance σ_i^2. Then the overall fixed effect estimate based on N studies is the weighted average given by

$$\hat{\theta}_{all} = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}$$

where w_i is the inverse of the study specific-variance. The variance of the effect estimate is given by

$$V(\hat{\theta}_{all}) = \frac{1}{\sum_1^N w_i}$$

3 Results

3.1 Incomplete Data

3.1.1 Bias in the Overall Effect Size $B(\hat{\theta}_{omit})$

As the study variances are split into two groups, the overall fixed effect estimate based on all studies is

$$\hat{\theta}_{all} = \frac{\sum_{i=1}^{x} w_i y_i + \sum_{i=x+1}^{N} w_i y_i}{\sum_{i=1}^{x} w_i + \sum_{i=x+1}^{N} w_i}$$

Suppose that σ_i^2 are unknown and replaced by s^2 , the weights w_i , based on the assumption on σ_i^2 given by Equation (1)

$$w_i = \begin{cases} \frac{n_i}{s_x^2} & \text{for } i = 1,2,\ldots,x \\ \frac{n_i}{s_{N-x}^2} & \text{for } i = x+1,\ \ldots,N \end{cases}$$

The estimate based on incomplete data, denoted $\hat{\theta}_{omit}$, in this case is

$$\hat{\theta}_{omit} = \frac{\sum_{i=x+1}^{N} w_i y_i}{\sum_{i=x+1}^{N} w_i}$$

Letting $\sum_{i=1}^{x} n_i = N_x$,$\sum_{i=1}^{N} n_i = N_N$, the observed bias in the effect size estimate is

$$\begin{aligned} B(\hat{\theta}_{omit}) &= \hat{\theta}_{all} - \hat{\theta}_{omit} \\ &= \frac{(N_x)\left[\frac{\sum_1^x n_i y_i}{N_x} - \frac{\sum_{x+1}^N n_i y_i}{N_N - N_x}\right]}{s_x^2\left(\sum_1^x \frac{n_i}{s_x^2} + B\right)} \end{aligned}$$

where $B = \frac{N_N - N_x}{s_{N-x}^2}$, suggesting that there will be some observed bias in the incomplete data which depends on the weighted mean of the missing and available data. However it can be shown that the expected bias in this case is zero, $E[B(\hat{\theta}_{omit})] = E(\hat{\theta}_{all}) - E(\hat{\theta}_{omit}) = 0$

3.1.2 Bias in the Variance of the Overall Effect Size, $B[V(\hat{\theta}_{omit})]$

The variance of the estimates based on complete data can be written

$$V(\hat{\theta}_{all}) = \frac{1}{\sum_1^x n_i/s_x^2 + \sum_{x+1}^N n_i/s_{N-x}^2}$$

and those based on incomplete data

$$V(\hat{\theta}_{omit}) = \frac{1}{\sum_{x+1}^{N} n_i / s_{N-x}^2} = \frac{s_{N-x}^2}{\sum_{x+1}^{N} n_i}$$

The observed bias is

$$\begin{aligned} B[V(\hat{\theta}_{omit})] &= V(\hat{\theta}_{all}) - V(\hat{\theta}_{omit}) \\ &= \frac{s_{N-x}^2}{\sum_1^x n_i \frac{s_{N-x}^2}{s_x^2} + \sum_{x+1}^N n_i} - \frac{s_{N-x}^2}{\sum_{x+1}^N n_i} \end{aligned}$$

To look at the relationship between the available and missing variances, let $r_1 = \frac{s_{N-x}^2}{s_x^2}$ and $n_i = n$. The bias may be written

$$= \frac{s_{N-x}^2}{n(N-x)} \left[\frac{(N-x)}{r_1 x + (N-x)} - 1 \right]$$

Here we note that if $r_1 > 1$ i.e. studies with smaller variances are missing,$s_x^2 < s_{N-x}^2$, the bias is negative, implying that the variance of the overall effect estimate will be overestimated. Similarly, the bias will also take negative value, for the case of $s_x^2 > s_{N-x}^2$, i.e when studies with larger variances are missing.

Derivation of the expected bias may be carried out using Taylor series approximation. The approximation series will have to be considered twice as there is additional ratio term,s_{N-x}^2/s_x^2. Additionally, further approximation is required in order to get the general interpretation of the results. Therefore a numerical approach is more feasible.

3.2 Complete Imputed Data

3.2.1 Bias in the Overall Effect Size $B(\hat{\theta}_{mean})$

The missing variances are imputed using the weighted average of the available variances, s_a^2 ,

$$s_a^2 = \frac{\sum_{x+1}^N (n_i - 1) s_{N-x}^2}{\sum_{x+1}^N (n_i - 1)} = s_{N-x}^2$$

The overall effect estimate based on the imputed data is

$$\hat{\theta}_{mean} = \frac{\sum_{i=1}^x \frac{n_i}{s_{N-x}^2} y_i + \sum_{i=x+1}^N \frac{n_i}{s_{N-x}^2} y_i}{\frac{N_x}{s_{N-x}^2} + \frac{N_N - N_x}{s_{N-x}^2}}$$

Letting

$$
\begin{aligned}
A &= \sum_{i=x+1}^{N} \frac{n_i}{s^2_{N-x}} y_i \\
B &= \frac{N_N - N_x}{s^2_{N-x}} \\
C &= \sum_{i=1}^{x} n_i y_i,
\end{aligned}
$$

the observed bias in the mean imputed data is

$$
B(\hat{\theta}_{mean}) = \frac{(AN_x - CB)\left[\frac{1}{s^2_{N-x}} - \frac{1}{s^2_x}\right]}{\left[\frac{N_x}{s^2_x} + B\right]\left[\frac{N_x}{s^2_{N-x}} + B\right]}
$$

Consider the terms $(AN_x - CB)$ of the above equation which involves y_i ,

$$
\begin{aligned}
AN_x - CB &= \left(\sum_{x+1}^{N} \frac{n_i y_i}{\sigma^2_{N-x}}\right) N_x - \left(\sum_{1}^{x} \frac{n_i y_i}{\sigma^2_{N-x}}\right) N_N - N_x \\
&= \frac{N_x(N_N - N_x)}{\sigma^2_{N-x}} \left[\frac{1}{(N_N - N_x)} \sum_{x+1}^{N} n_i y_i - \frac{1}{N_x} \sum_{1}^{x} n_i y_i\right]
\end{aligned}
$$

which implies that the observed bias depends on the difference between the weighted means of the available data and the data which is missing, and the differences in the variances. The magnitude of the bias here is also less than those when the studies are omitted.

The expected bias, $E[B(\hat{\theta}_{mean})]$ is also zero as the term

$$
E\left[\frac{1}{(N_N - N_x)} \sum_{x+1}^{N} n_i y_i - \frac{1}{N_x} \sum_{1}^{x} n_i y_i\right]
$$

tends to zero.

3.2.2 Bias in the Variance of the Overall Effect Size, $B[V(\hat{\theta}_{mean})]$

The variance of the effect estimate based on complete imputed data is

$$
\begin{aligned}
V(\hat{\theta}_{mean}) &= \frac{1}{\sum_1^x n_i/s^2_{N-x} + \sum_{x+1}^N n_i/s^2_{N-x}} \\
&= \frac{s^2_{N-x}}{\sum_1^N n_i}
\end{aligned}
$$

Giving observed bias in variance of the effect estimate of

$$\begin{aligned} B[V(\hat{\theta}_{mean})] &= V(\hat{\theta}_{all}) - V(\hat{\theta}_{mean}) \\ &= \frac{s^2_{N-x}}{\sum_1^N n_i}\left[\frac{\sum_1^N n_i}{\sum_1^x n_i\left(s^2_{N-x}/s^2_x\right) + \sum_{x+1}^N n_i} - 1\right] \end{aligned}$$

Again if we let $r_2 = \frac{s^2_{N-x}}{s^2_x}$, the bias can now be written

$$= \frac{s^2_{N-x}}{\sum_1^N n_i}\left[\frac{\sum_1^N n_i}{\sum_1^x n_i r_2 + \sum_{x+1}^N n_i} - 1\right]$$

suggesting that when $r_2 = 1$ that is $s^2_x = s^2_{N-x}$ then there is no bias. If $r_2 > 1$, that is when the studies with smaller variances are missing, $s^2_x < s^2_{N-x}$, the bias takes negative value implying that the estimate of variance of the effect estimate is overestimated. When $r_2 < 1$, i.e. studies with larger variance are missing, $s^2_x > s^2_{N-x}$,the bias is positive, implying that the variance of the effect estimate is underestimated.

3.3 Conclusion

We investigated, under the NMAR missing mechanism, the effects of imputation of missing study variances on the estimates of overall effects size and its corresponding variances. Two common approaches in handling the missing variances were considered, namely, when the studies with missing variances were excluded from the analysis, and when the missing variances were imputed using the available variances from the data.

The results suggest that, in both approaches, although there was observed bias, the estimate of overall effect size is expected to be unbiased. These results are consistent with those from earlier empirical studies [5, 6] which were based on MCAR missing mechanism. However, it was suggested that studies which do not report the variances may not be a random selection of all studies so there is the possibility of a biased point estimate of the treatment effect using just the studies with adequate variance information.

Under MNAR, the missing study variances, however, does have greater effect on the estimate of the variance of the overall effect size estimate. We investigated the relation between the magnitude of the missing variances, available variances and their effect on the estimate of the variance of the effect size. We let the study variances take two different values, one for the set of missing data, and another for the set of available data. Although this is an extreme case where all the variances that are missing are assumed to be either greater than or smaller than those that are available, it allows investigation on the biases when the missing studies are more or less variable than the studies with available variances.

Generally, exclusion of studies with missing variances will result in overestimation in the estimate of the variance of the overall effect size, thus making the overall effect to be less visible. The results hold irrespective of the magnitude of the missing variances relative to the available variances. This is expected because with exclusion of the studies with missing variances, the sum of the inverse variances, $\sum w_i$ in available studies, would be smaller. This would result in higher overall variances of the estimates. As the ratio of the two variances tends to zero however, bias gets smaller.

On the other hand, if the missing study variances are imputed using the mean imputation, the estimate of the variance of the overall effect depends on the magnitude of the variances that are missing relative to those that are available. If the within-study variances that are missing are mostly larger, the estimate of the variances of the overall effect will be under-estimated. This is because the missing variances are being replaced by those with smaller values, which carry more weights, resulting in smaller variance of the estimate than expected. So mean imputation gives false impression of precision as the estimated variance of the overall effect is too small. This is different from earlier results [4]based on MCAR missing mechanism where most form of imputation, including the mean imputation, produces practically unbiased estimates of the overall effect size and its corresponding variances. For the MNAR missing scenario, the general conclusion from the same study which is based on random effect model, suggest that mean imputation of study variances may resulted in overestimation of the variance of the overall effect estimate. We showed analytically that if the Fixed Effect model is used, the results seem to indicate otherwise. This is because the Fixed Effect model does not take into account the between-study variability. It was suggested that imputation has the effect of overestimating the between-study variances in the Random effect model, which will thus increase the estimate of the variance of the overall effect estimates.

In this paper, assumptions were made to allow simplification of the derivations and for ease of interpretation and analysis. For instance, we assumed that the missing study variances and the available variances take common values, namely s_x^2 and s_{N-x}^2 , respectively. If no assumption is made on the magnitude of s_i^2 , the information regarding the bias can only be obtained in terms of the sums of inverse within study variances and practical interpretation of this is difficult. We also looked at only one technique of non-parametric imputation, i.e. the weighted mean imputation. This technique is quite popular and commonly used in filling in the missing variances in many research areas as it is easy to execute. However, it is known to have the effect of reducing the variability for variable as a constant value is substituted for all the missing instances [7].

In practice, it is impossible to determine whether the mechanism of the missing variances occur completely at random or not for a particular data set. However, the author believes that the results presented here could serve as a cautionary note. The analysts are advised to consider the possibility of non-random missing as under the wrong assumption of MCAR imputation of missing study variances may potentially lead to biased estimate of overall variance of the effect size.

Acknowledgemnts

This work was partly supported by the IIUM Research Endowment Fund (EDW B0802-106).

References

[1] Wiebe, N., Vandermeer, B., Platt, R. W., Klassen, T. P., Moher, D. and Barrowman, N. J. A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology*. 2006. 59: 342–353.

[2] Sutton, A. J. and Pigott, T. D. *Publication Bias in Meta Analysis -prevention, Assessment and Adjustment* New York: John Wiley. 2005.

[3] Little, J. A. and Rubin, D. B. *Statistical Analysis with Missing Data*. London: John Wiley & Sons. 1987.

[4] Idris, N. R. N. and Robertson, C. The effects of imputing the missing standard deviations on the standard error of the meta analysis estimates *Communication in Statistics - Simulation and Computation* . 2009. 38(3): 513–526

[5] Furukawa, T. A., Barbui, C., Cipriani, A., Brambilla, P. and Watanabe, N. Imputing missing standard deviations in meta analyses can provide accurate results. *Journal of Clinical Epidemiology*. 2006. 59: 7–10.

[6] Thiessen, P. H., Barrowman, N. and Garg, A. X. Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *Journal of Clinical Epidemiology*. 2007. 60(3): 228–240.

[7] Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. and Song, F. *Methods for Meta-analysis in Medical Research*. New York : John Wiley. 2000. 199–204.