Poster ID: 1212

Estimating the bias in meta analysis based on fixed effect model for data with missing variability measures

Nik Ruzni Nik Idris

Kulliyyah of Science, International Islamic University Malaysia

Phone: 09 5167400, Fax: 09 516789, E-mail: ruzni@iium.edu.my

abstract

A common drawback with meta analysis is when the variability measures, particularly the variances , are not reported, or "missing" in the individual study. Among the approaches adopted in handling this problem is through exclusion of the studies with missing variances. Alternatively, the missing study-variances could be imputed. This paper examines the analytical implications of these two approaches on the overall effect estimate and the corresponding variances. The bias in these estimates are derived using the Fixed Effect model. The results show that no bias is expected in the estimate of the overall effect using both approaches. Similarly, there is no bias in the variance of the effect estimate when the missing study-variances are imputed and homogeneous study-variances are assumed across the studies. However, if the magnitude of the missing study-variances are mostly larger than those that are reported, imputation leads to under estimation of the variance of the effect estimate. This is a likely case in meta analysis. When studies with missing variances were excluded from analysis, the variances of the effect estimate are overestimated, and the magnitude of the bias in this case is relatively larger when compared to those from complete imputed data.

motivation

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* .

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 . 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 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.

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.

Non-Random mechanism of missing study variances (NMAR)

Assume that there are N studies, each with complete treatment effect size and variance information.
Let a number x of these N studies do not report the variances information, and we assume that these are the 'missing' variances.
Assume that σ_i^2 take the following values

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

The Fixed Effects Meta Analysis Model

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

$$y_i = \theta + \epsilon_i$$

y_i is the estimate of treatment effect in study i, θ is the overall true treatment effect and ϵ_i is the random error for study $i = 1,2,\ldots, N$, assumed to be normally distributed with mean 0 and variance σ_i^2

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

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

result

Incomplete Data - Bias in the overall effect size

The overall fixed effect estimate based on all studies $\longrightarrow$ $\hat\theta_{all} = \frac{\sum_{i=1}^{x} w_i y_i + \sum_{i=x+1}^{N} w_i y_i}{\sum_{i=1}^{x} w_i + \sum_{i=x+1}^{N} w_i}$

The estimate based on incomplete data $\longrightarrow$ $\hat\theta_{omit} = \frac{\sum_{i=x+1}^{N} w_i y_i}{\sum_{i=x+1}^{N} w_i}$

The observed bias in the effect size estimate is $\longrightarrow$

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

where $B = \frac{N_N - N_x}{s_{N-x}^2}$, $\sum_{i=1}^{N} n_i = N_N$ and $\sum_{i=1}^{x} n_i = N_x$

Incomplete Data - Bias in the variance of the overall effect size

The variance of the estimate based on all studies $\longrightarrow$ $V(\hat\theta_{all}) = \frac{1}{\sum_1^x n_i/s_x^2 + \sum_{x+1}^N n_i/s_{N-x}^2}$

The variance of the estimate based on incomplete data $\longrightarrow$ $V(\hat\theta_{omit}) = \frac{1}{\sum_{x+1}^N n_i/s_{N-x}^2} = \frac{s_{N-x}^2}{\sum_{x+1}^N n_i}$

The observed bias in the variance of the effect size estimate is $\longrightarrow$

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

Complete Imputed Data - Bias in the overall effect size

The estimate based on imputed data $\longrightarrow$ $\hat\theta_{mean} = \frac{\sum_{i=1}^{x} \frac{n_i}{s_{N-x}^2} y_i + \sum_{i=x+1}^{N} \frac{n_i}{s_{N-x}^2} y_i}{\frac{N_x}{s_{N-x}^2} + \frac{N_N - N_x}{s_{N-x}^2}}$

The observed bias in the effect size estimate is $\longrightarrow$ $B(\hat\theta_{mean}) = \frac{(aN_x - cb)\left[\frac{1}{s_{N-x}^2} - \frac{1}{s_x^2}\right]}{\left[\frac{N_x}{s_x^2} + b\right]\left[\frac{N_x}{s_{N-x}^2} + b\right]}$

where

$$\begin{aligned} a &= \sum_{i=x+1}^{N} \frac{n_i}{s_{N-x}^2} y_i \\ b &= \frac{N_N - N_x}{s_{N-x}^2} \\ c &= \sum_{i=1}^{x} n_i y_i \end{aligned}$$

Complete Imputed Data - Bias in variance of the overall effect size

The variance of estimate based on imputed data $\longrightarrow$

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

The observed bias in the variance of the effect size estimate is

$\longrightarrow$

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

conclusion

- The results suggest that, in both approaches, the estimate of overall effect size is expected to be unbiased under the assumed conditions
- 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.
- 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 underestimated. -- So mean imputation gives false impression of precision as the estimated variance of the overall effect is too small.
- This generalisation is different from studies which are based on random effect model. 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 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 the results presented here could serve as a cautionary note. Analysts are advised to consider the possibility of non-random missing as if the assumption of MCAR does not hold, imputation of missing study variances may potentially lead to biased estimate of overall variance of the effect size.

IIUM Research, Invention and Innovation Exhibition 2012