MATTER: International Journal of Science and Technology
ISSN 2454-5880

Idris & Misran, 2015

Volume 1 Issue 1, pp.144 -158

Year of Publication: 2015

DOI-https://dx.doi.org/10.20319/mijst.2016.s11.144158

This paper can be cited as: Idris, N., R., & Misran, N., A. (2015). Combining Aggregate Data and Individual Patient Data in Meta-Analysis: An Alternative Method. MATTER: International Journal of Science and Technology, 1(1), 144 -158.

COMBINING AGGREGATE DATA AND INDIVIDUAL PATIENT DATA IN META-ANALYSIS: ANALTERNATIVE METHOD

Nik Ruzni Nik Idris
International Islamic University Malaysia, Faculty of Science, Kuantan, Pahang, Malaysia
ruzni@iium.edu.my

Nurul Afiqah Misran
International Islamic University Malaysia, Kuantan, Faculty of Science, Pahang, Malaysia

Abstract

It has been shown that, in cases where both the AD and IPD studies are available, combining these two levels of data could improve the overall meta-analysis estimates, compared to utilizing AD studies alone. However, the coverage probability of estimates based on combined studies are relatively low compared to the AD-only meta-analysis, when the existing standard method was used to combine these studies. The aim of this paper is to introduce some modifications to the existing two-stage method for combining the aggregate data (AD) and individual patient data (IPD) studies in meta-analysis. We evaluated the effects of these modifications on the estimates of the overall treatment effect, and compared them with those from the standard method. The influence of the number of studies included in a meta-analysis, N, and the ratio of AD: IPD on these estimates were also examined. We used percentage relative bias (PRB), root mean-square-

error (RMSE), and coverage probability to assess the overall efficiency of these estimates. The results revealed that the proposed method had been able to improve the coverage probability while maintaining the level of bias and RMSE at par to their existing counterpart. These findings demonstrated that the technique for combining different levels of studies influenced the efficacy of the overall estimates, which in turn is crucial for drawing reliable and validconclusions.

Keywords

Meta-analysis; Combined-data; Aggregate data; Simulation; Bias; Coverage probability

1. Introduction

Meta-analysis is a statistical technique for integrating quantitative results from several independent sources. A meta-analysis may be performed using studies at eithernaggregate data (AD), individual patient data (IPD) or a combination of AD and IPD (Whitehead et al., 2001). For AD meta-analysis , methods such as the inverse variance method (DerSimonian & Laird,1986), for continuous data, or the Mantel-Haenszel method (Mantel & Haenszel, 1959), for binary data, may be utilized to obtain the overall estimate, while for IPD meta-analysis, a methodology adopted for analysis of primary data may be applied in estimating the overall effect. Although IPD meta-analysis has an advantage in terms of type of analyses that can be performed, it is usually more costly and time consuming (Stewart & Tierney, 2002; Simmonds et al., 2005; Jones et al., 2009), and IPD are seldom available from each of the individual studies. An approach where the available AD and IPD studies are combined is quite recent. Combining available IPD with the AD maximizes the utilisation of available information, as it allows for a larger number of patients, hence a greater part of the evidence-based could be included (Cooper &Patall,2009).

Literature on the efficacy of meta-analysis estimates based on combined level studies is limited (Lambert et al., 2001). Only two studies had examined the efficacy of the estimates that are based on combined-level studies (Riley et al., 2008; Idris & Abdullah, 2015). Riley *et al.* (2008) took their data from a study on the effects of hypertension (Wang et al., 2005) and compared estimates obtained from a meta-analysis that combined IPD and AD studies using a two-stage method. Riley *et al.*'s (2008) results suggested some benefits of combining the IPD with AD in terms of the accuracy, where the bias were relatively smaller in combined -level data.

These findings are supported by another simulation-based study (Idris & Abdullah, 2015), for which the researchers concluded that the benefit of combining the data is greater if the majority of the studies to be combined are at AD-level and that if more than 80% of the studies are IPD, includingtheADwouldonlyservetoincreasetheoverallSE.Idris&Abdullah(2015)additionally noted that the while the bias and MSE was better for combined-level data, the coverage probability of the estmates were lower compared to those from the AD studies.

In this study we considered the possibility of improving the statistical properties of the overall estimates, particularly, the coverage probability, with some modifications to the existing two-stage method, involving a utilization of the grouped-variance from the two levels of data as the weightage. The effects of these modifications on the estimates of the overall treatment effect were evaluated and compared to those using the standard method. The PRB, RMSE and the coverage probability were used to assess the overall efficiency of these estimates. Additionally, the influence of the ratio of AD: IPD, as well as the effects of the number of studies included in the meta-analysis, N, on the accuracy and precision of the overall treatment effects estimates were investigated.

2. Two-Stage Method for Combining the Ad and iPodStudies

2.1 Standard two-stage method(SM)

Supposethere $N = N_1 + N_2$ studies where $N_1$was the number of AD level studies with effects estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{N1}$and corresponding variancesgivenby $V(\hat{\theta}_1), V(\hat{\theta}_2), \ldots, V(\hat{\theta}_{N1})$while was the N_2number of IPD level studies. In a standard two-stage method, the available IPD are first reduced to AD in each study before they are combined with the existing AD studies using standard meta-analysis of AD techniques. Suppose $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_{N2}^*$denoted the study specific effects from the reduced IPD studies, with corresponding variances $V(\hat{\theta}_1^*), V(\hat{\theta}_2^*), \ldots, V(\hat{\theta}_{N1}^*)$. The effect $\hat{\theta}_i^*$for instance may represent the difference in the mean response of individualpatients for treatment and control arms from study i andthe $V(\hat{\theta}_i^*)$represents the pooled treatment and control arms variances corresponding to their respective effects for study i. We computed the overall effect using the standard inverse variance weighted method given by

$$\hat{\theta}_{overall} = \frac{\sum w_i \hat{\theta}_i + \sum w_i^* \hat{\theta}_i^*}{\sum w_i + \sum w_i^*} \quad [1]$$

2.2 The proposed Inverse Group-Variance Weighted Estimate(IGVW)

In this modification, we utilized the inverse variances of each group as the weights. For the AD studies, the weight, w_{AD} was simply the inverse of the overall variance for theADgroup, $w_{AD} = 1/V(\hat{\theta}_{AD})$, for IPD studies,theweight, w_{IPD} was the inverse variance forIPDgroup $w_{IPD} = 1/V(\hat{\theta}_{IPD})$, and the overall estimate was given by

$$\hat{\theta}_{all2} = \frac{(w_{AD})\hat{\theta}_{AD} + (w_{IPD})\hat{\theta}_{IPD}}{w_{AD} + w_{IPD}} \quad [2]$$

The variance of the overall estimate, as in the case of standard meta-analysis, was given by

[3]

$$V(\hat{\theta}_{all2}) = \frac{1}{w_{AD} + w_{IPD}}$$

3. Material andMethod

3.1 Generation ofdata

A simulation approach were used to generate IPD level response from continuous data, denoted y_{ij}, representing a response from patient j within study i, where

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij} \quad [4]$$

Where β_{0i} was the random study effect, t_{ij}represented a dummy covariate for treatment which took two values, namely 0 for the control and 1 for the treatment arm, $\beta_{1i}$was the random treatment effect, and were the sampling random error terms for the response from patient j within study i. We assumed β_{0i}, β_{1i} and were independent andnormally-distributed,with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, $\beta_{0i} \sim N(\beta_0, \sigma_{\nu_0}^2)$, and $\beta_{1i} \sim N(\beta_1, \tau^2)$. For simplicity, we assumed each study

$i = 1, 2, \ldots, N$

had an equal number of patients in each treatment arm (i.e.$n_{0i} = n_{1i} = 1/2n_i$for
). The following values were arbitrarily assigned to the fixed effects: $\beta_0 = 0$ and $\beta_1 = 7$. For the random effects, we assigned the following values to create a moderately- heterogeneous effect (I2: 25% to 50%); $\sigma^2_{\nu_0} = 1$and $\tau^2 = 2$while we allowed σ^2_ϵ to vary randomly between 1 to 25. The AD were created by taking the differences of the means of each treatment arm in each individual study, and the combined AD : IPD data were created by selecting a given ratio of AD and IPD studies generated earlier, as detailed in the preceding paragraph.

The term "AD-only" meta-analysis were used when only the available AD studies were utilized in a case where both AD and IPD studies were available for a meta- analysis. The term"all-AD" and "all-IPD" were used to describe studies that were used when all of the studies available for meta-analysis were AD and IPD, respectively.

The factors that were varied in this simulation study were the number of studies included in the meta- analysis, N, (N = 10, 20, 30 and 90) and the ratio of AD : IPD studies in a meta-analysis, namely, 0:100, 20:80, 30:70, 40:60, 50:50, 60:40, 80:20 and 100:0, while the size of the samples, n, were fixed (n = 60). The AD and IPD studies were combined using two methods; the existing standard two-stage method (SM), and the proposed modifications of existing methods, namely, the inverse group-variance weighted estimates (IGVW). All the parameters in this paper can be estimated using restricted maximum likelihood (REML) within the suitable packages from R statistical analysis software (R development Core Team, 2008) for mixed models such as the LME or the NLME.

3.2 Statistical Assessments of the effectestimates

These specifications generated 32 sets of meta-analyses comprising four all-IPD studies, four all-AD studies and twenty-four (24) sets of combined AD:IPD studies consisting of different combinations of N and AD:IPD ratios. Each meta-analysis was replicated 1,000 times and for each replication the overall estimate of the treatment effects, the PRB, the RMSE and their corresponding standard error (SE) were computed. The means of these 1000 replications were recorded and the coverage probability at 95% nominal value were estimated.

3.2.1 Percentage Relative Bias (PRB)

The PRB were computed as the percentage difference between the true treatment effect

and the estimated treatment effect. The mean PRB was given by:

$$\text{mean PRB}(\hat{\theta}) = \frac{\sum_{t=1}^{K}(\theta - \hat{\theta}_t)/\theta}{K} \times 100\% \quad [5]$$

Where $\hat{\theta}_t$was the estimate of the treatment effect from simulation t θ was the true treatment effect, and K was the number of simulations.

3.2.2 Root Mean Square Error (RMSE)

The RMSE is the square root of the mean-square-error (MSE) for the overall estimate of treatment effect, where the mean MSE over K simulations was given by

$$\text{mean MSE} = \frac{\sum_{t=1}^{K}[bias(\hat{\theta}_t)^2 + SE(\hat{\theta}_t)^2]}{K} \quad [6]$$

Where $bias(\hat{\theta}_t)$was the observed bias for $\hat{\theta}_t$and $SE(\hat{\theta}_t)$was the standard error corresponding to the overall estimate from meta-analysis at simulation numbert.

3.2.3 Coverage Probability

The coverage probability were estimated by taking the proportion of the number of times the estimated 95% confidence interval included the true value of out of the total number of simulations K.

4. Result

4.1 The PRB, RMSE and Coverage probability of estimates from meta-analysis with AD-only studies and combined AD:IPD studies

Figure 1 presents the distributions of PRB, RMSE and the coverage probability of estimates from AD-only and the combined AD: IPD meta-analysis for the selected range of N (N=10, 20, 30, and 90) and six combinations of AD: IPD. Clearly, the AD:IPD meta-analysis generated lower bias with the mean PRB ranges from 1.5% to 5.0% compared to the conventional AD-only meta-analysis (mean PRB = - 11% to 5.3%). The majority of PRB from

AD-only studies were negative for the percentage of AD within the combined studies that were less than 40%, suggesting overestimated effects. On the other hand, the estimates of treatment effects from the AD: IPD studies remained positive suggesting some underestimation in the estimates although not as severe as those of the AD-only in terms of the magnitude of thebias.

As in the case of PRB, the distribution of the RMSE for estimates from AD-only and combined AD: IPD meta-analysis showed the latter were lower, averaging from 0.53 to 1.22 compared to those from AD-only (mean RMSE of 0.55 to 2.17). The RMSE decreased as thenumber of studies, N, increased, which was expected as the SE tended to decrease with increasing N (see Figure.1)

The coverage probability was better for estimates with AD-only studies compared to those from the combined studies. For nominal values of 95%, the coverage ranges from 70% to 90% for AD-only studies. The coverage for combined AD: IPD studies was slightly lower, ranging from 50% to 70%, increasing with the proportion of AD within the AD:IPD ratio.

4.2 Comparison of PRB and RMSE of theestimates from the standard two-stage method (SM) against the proposed two-stage methods (IGVW) for combining the AD andIPD

A comparison of the observed statistical properties of estimates from the existing method for combining the AD: IPD studies and the proposed modifications suggested possible benefit of the latter. The trend of PRB from the two methods was quite similar (see Figure 2). In general, the proposed method, IGVW, displayed smaller PRB than the existing method and differences were more notable when the number of studies included in the meta-analysis was at moderate range, N (20 <N<60). Figure 3 show the RMSE from IGVW was very close to those from existing method, MS. This trend was similar for all ratios of AD: IPD, and as the proportions of AD increased, the RMSE of the estimates from the two methods converged to the same value.

4.3 Comparison of the coverage probability and SE of the estimates from the standard two- stage method (SM) against the proposed two-stage methods (IGVW) for combining the AD andIPD

Clearly, the proposed method provided estimates with better coverage than those from the existing method. IGVW provided mean coverage of approximately 70% against an average

of about 50% using the existing method (see Figure 4). The SE of estimates based on the standard method was underestimated, resulting in a confidence band which was toonarrow.

4.4 The effects of ratio of AD:IPD and the number of studies, N, on the overallestimates

The number of studies included in a meta-analysis has some effects on the accuracy of theoverallestimate.Figure2exhibitsanincreasingtrendin PRBasthenumberofstudiesincreases, for both methods. However, for meta-analysis with a moderate number of studies (N < 30), this effect was minimal. The effect of AD: IPD on the bias was observed to be relatively small, and the differences in the bias in both methods, across N, reduced as proportions of AD within the AD: IPD ratios increased. This implied that the proposed method did not have much of an advantage over SM in terms of bias, when the majority of the studies within the combined AD: IPD were at AD level. Figure 3 shows that as N increased, the RMSE decreased, as expected, due to lower SE for a larger number of studies included in a meta-analysis.As expected, the SE reduced as N increased and relatively smaller SEs was observed from the SM method. Figure 3 additionally illustrated that the RMSE of the estimates from the two methods converged for large N, suggesting minimal beneficial effects of utilizing the proposed methods in terms of RMSE when N was large (N > 40). In terms of the effects of the AD:IPD ratio, it was noted that the RMSE reduced slightly as the proportion of AD increased, again reflecting the effects of SE, which were lower in AD compared to those from IPD studies. The RMSE from the three methods appeared to close up as the proportion of AD increased. The number of studies, N, and the ratios of AD: IPD did not have notable effects on the coverage probability. These trends were observed on both methods of combining under consideration.

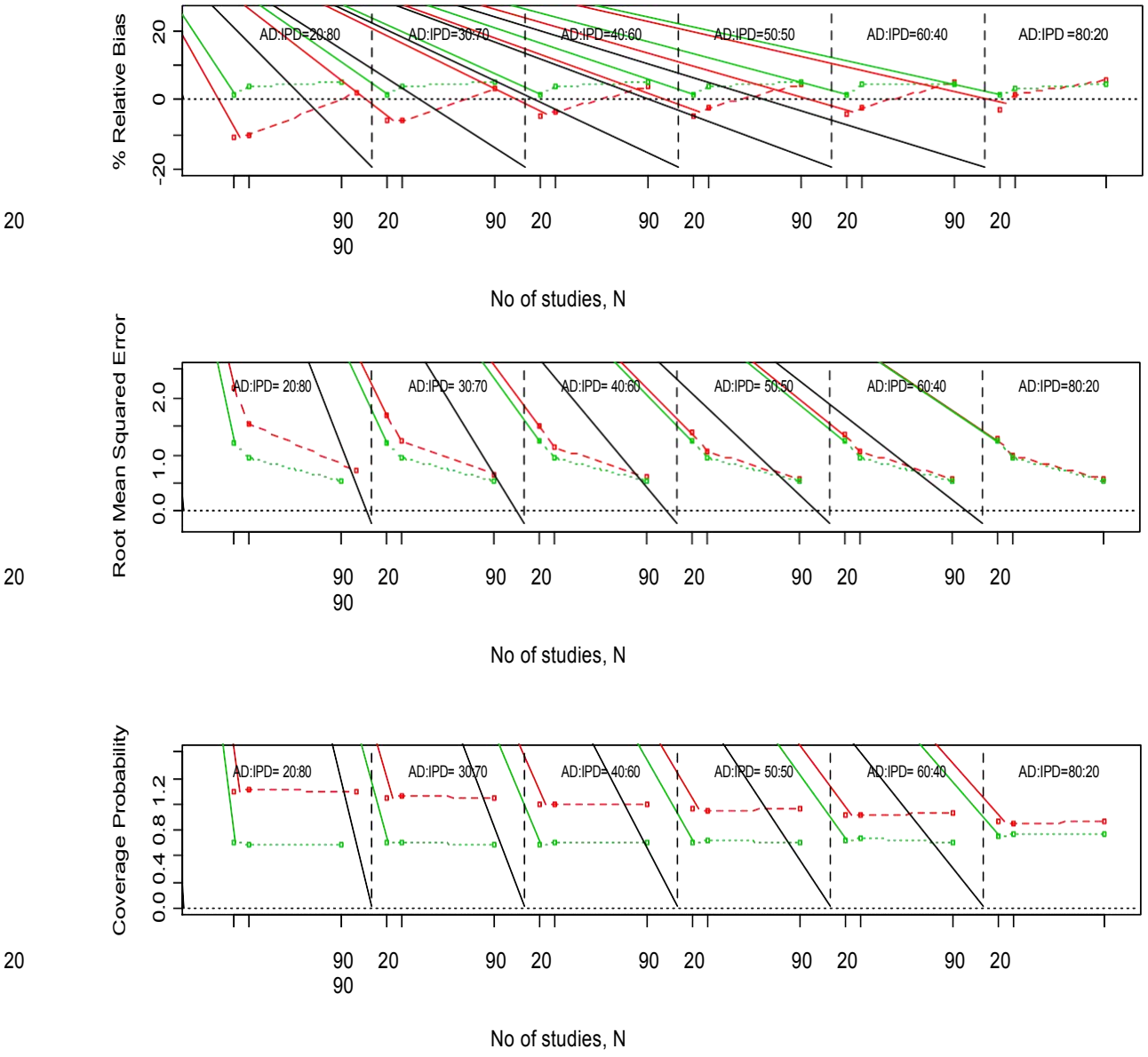

Figure 1: *Distribution of PRB, RMSE and the Coverage probability for the overall estimates from AD-only and AD: IPD studies. Legend: RED – AD-only studies; GREEN - Combined AD: IPD studies using the standard method (SM)*

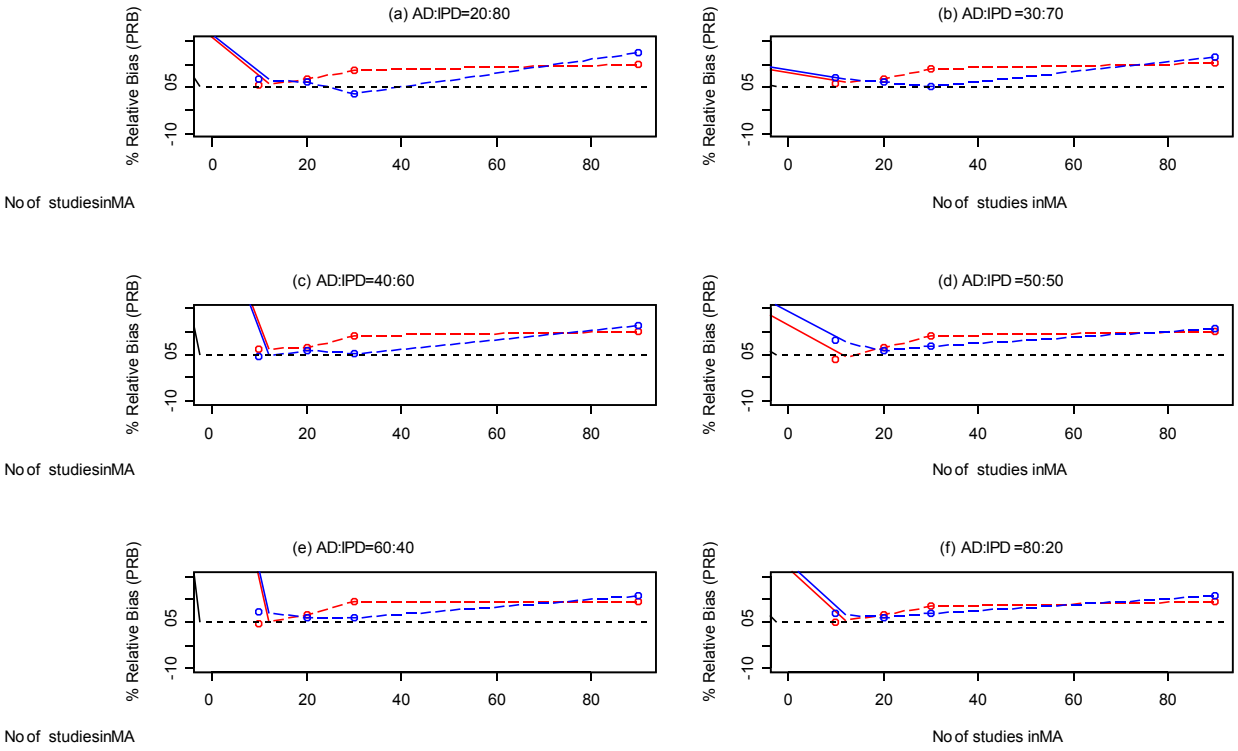

Figure 2: *Distribution of PRB for the overall estimates using SM and IGVW methods for a selected range of N (N from 10 to 90) and six combinations of AD: IPD ratios. Legend: Red: existing SM; Blue: IGVW*

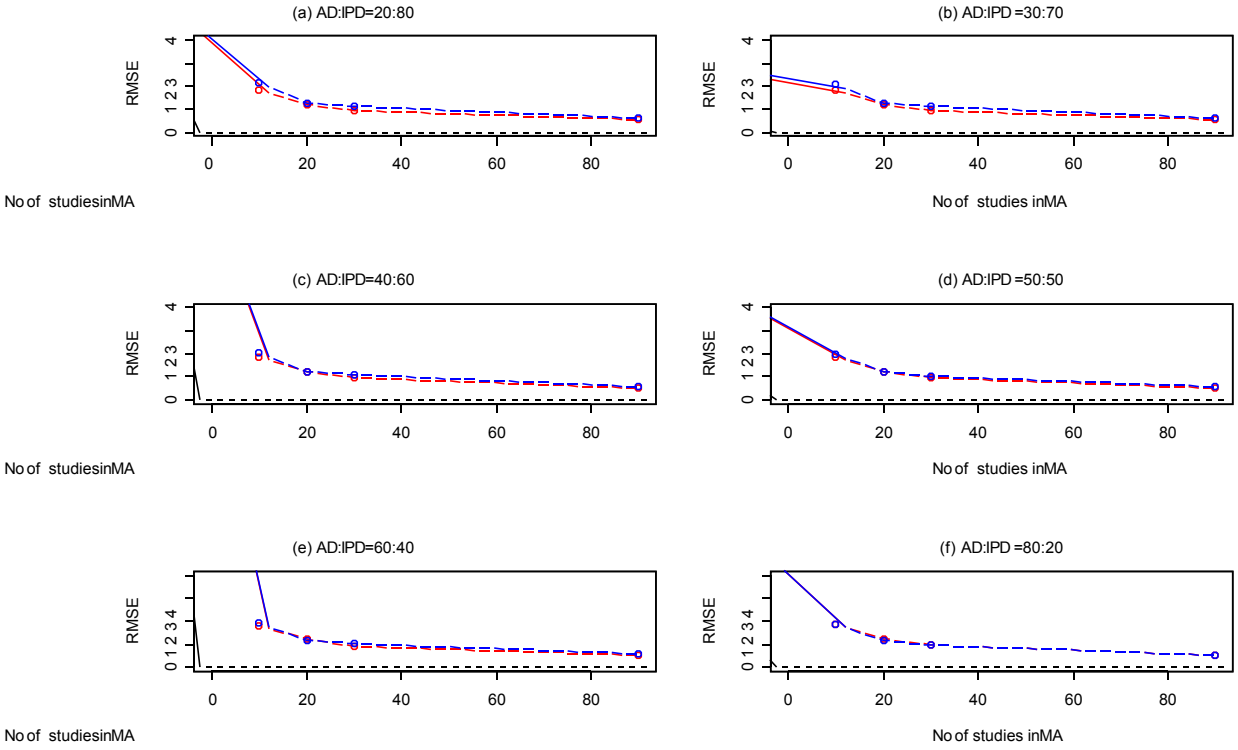

Figure 3: *Distribution of RMSE for the overall estimates using SM and IGVW methods for six combinations of AD: IPD and ranges of N from 10 to 90. Legend: Red: existing SM; Blue: IGVW*

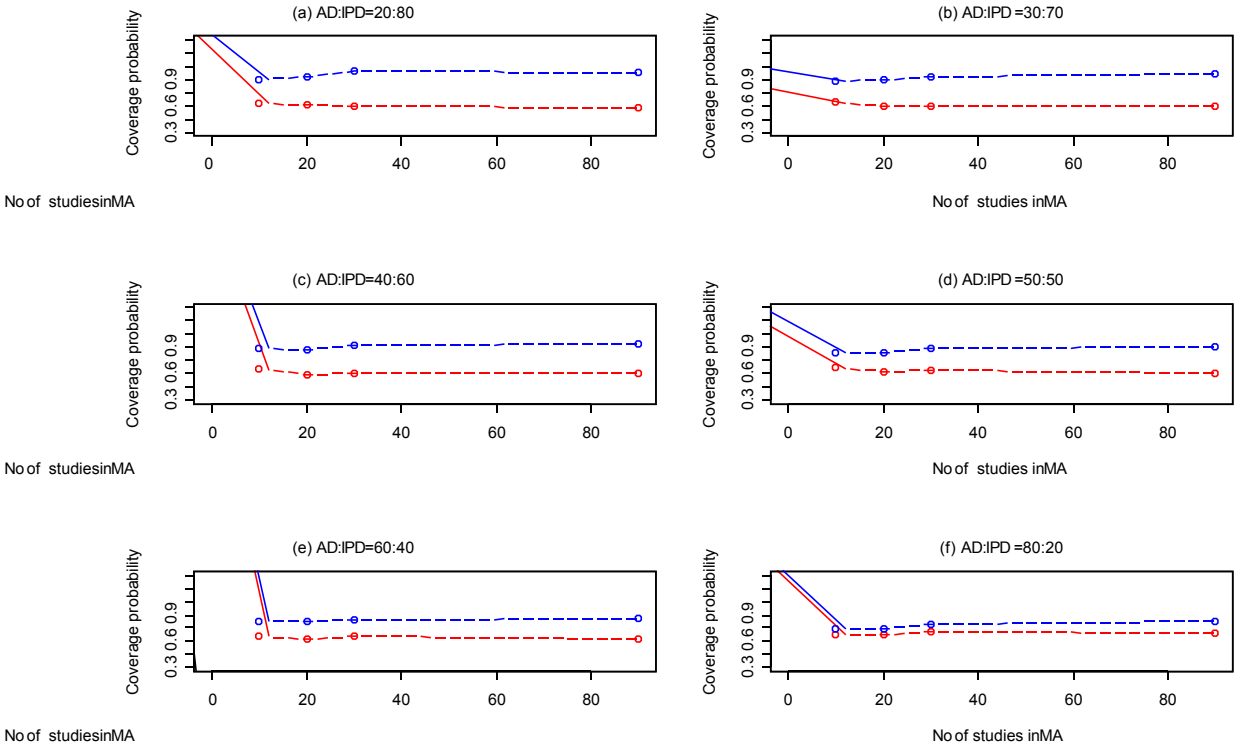

Figure 4*: Distribution of coverage for the overall estimates using SM and IGVW methods for six combinations of AD: IPD and ranges of N from 10 to 90. Legend: Red: existing SM; Blue: IGVW*

5. Discussion

It has been demonstrated (Riley et al., 2008; Idris & Abdullah, 2015) that combining the AD and IPD studies in meta-analysis would improve the overall meta-analysis estimates compared to utilizing AD studies alone, in cases where both the AD and IPD are available. However, the coverage probability of estimates based on combined studies are shown to be relatively low compared to the AD-only meta-analysis, when the existing standard method was used to combine these studies.

We postulated that the coverage probability may be improved if some modifications were introduced to the existing methodology that was used to combine the AD: IPD studies. A grouped-based weight age that utilized all available information contained in the IPD studies was considered. The results of this study demonstrated that the proposed methods provided

smallerbias, than the existing method, particularly when the number of studies in a meta-analysis, N, was moderate (10 <N<60), but the RMSE of the estimates from the two methods were not markedlydifferent.

Evidently the greatest advantage of the proposed method of combining the AD: IPD was in terms of the coverage probability. This study revealed that the current method used to combine the AD: IPD studies provided poor coverage. We noted that this situation may be attributed to underestimation of the SE of estimates produced using the existing method, which in turn, produced an interval which was narrower than it should have been. In modified method, the estimates from IPD studies were estimated directly, without reducing them to AD level first, thus utilizing all the information available within the IPD.

We acknowledge the lack of theoretical support for the proposed modification of the two-stage methods, due to their complex analytical approach. Nonetheless, it was a simple modification, in which the AD and IPD estimates were evaluated separately, and combined using the typical weights, which was simpler to implement and easier to interpret. Our simulation results confirmed that these modifications yielded estimates with improved coverage probability, albeit it should be interpreted with caution as results may apply only to the data characteristics under investigation.

As one of the main goals of meta-analysis is to draw general inferences about the research problem, an accurate and reliable overall estimate is therefore crucial in meta-analysis. This article confirmed that combining the available AD and IPD studies provided more reliable overall estimates and better statistical properties. Another important finding suggested that the existing method currently used to combine the AD and IPD resulted in a coverage probability which was generally too low. This information was imperative in light of recent review of current practice, which found that 80% of meta-analyses that combined AD and IPD studies used the existing two-stage method. The results of this study should provide a useful insight and may serve as a guide for practitioners when performingmeta-analysis.

References

Cooper, H., & Patall, E.A. (2009) the relative benefits of meta-analysis conducted with individual participant data versus aggregated data. Psycho Methods, 14(2), 165-176.http://dx.doi.org/10.1037/a0015565

Der Simonian, R., Laird, N. (1986). Meta-analysis in clinical trials Control Clin Trials, 7, 177-188.http://www.ncbi.nlm.nih.gov/pubmed/3802833.

Idris, N.R.N., & Abdullah, M.H. (2015). A study on the effects of different levels of data on the overall meta-analysis estimates. Far East Journal of Mathematical Science, 96(1), 73-86. http://dx.doi.org/10.17654/FJMSJan2015_073_086.

Jones, A.P., Riley, R.D., Williamson, P.R. & Whitehead A. (2009).Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials.Clin Trials, 6(1), 16-27. Doi: 10.1177/1740774508100984.

Lambert, P.C., Sutton, A.J., Abrams, K.R., Jones, R.D. (2001).A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of Clinical Epidemiology, 55, 86–94. DOI: http://dx.doi.org/10.1016/S0895-4356(01)00414-0

Mantel,N., &Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst, 22, 719-748. Doi: 10.1093/jnci/22.4.719

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.ISBN 3-900051-07-0.http://www.R-project.org/.

Riley, R.D., Lambert, P.C., Staessen, J.A., Wang, J, Gueyffier, F., Thijs, L., and Boutitie,

F.(2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data, Statist. Med., 27, 1870–1893. DOI: 10.1002/sim.3165

Simmonds, M.C., Higgins, J.P.T., Stewart, L.A., Tierney, J.F., Clarke, M.J., Thompson, S.G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials, 2,209-217. http://www.ncbi.nlm.nih.gov/pubmed/16279144

Stewart, L.A., Tierney, J.F. (2002). To IPD or not to IPD?Advantages and disadvantages of systematic reviews using individual patient data.Eval Health Prof, 25, 76-97. http://www.ncbi.nlm.nih.gov/pubmed/11868447

Wang, J.G., Stassen, J.A., Franklin, S.S., Fagard, R., Gueyffier, F. (2005) Systolic and diastolic blood pressure lowering as determinants of cardiovascular. Hypertension, 45, 907–913. http://www.ncbi.nlm.nih.gov/pubmed/15837826

Whitehead, A., Omar, R.Z., Higgins, J.P., Savaluny, E., Turner, R.M., Thompson, S.G. (2001). Meta-analysis of ordinal outcomes using individual patient data. Stat Med, 20, 2243-2260. http://www.ncbi.nlm.nih.gov/pubmed/11468762