



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

RESEARCH ARTICLE

OPEN ACCESS

DOI: 10.3923/jas.2015.1231.1238

## A Modified Two-Stage Method for Combining the Aggregate-Data and Individual-Patient-Data in Meta-Analysis

N.R.N. Idris and N.A. Misran

Department of Computational and Theoretical Science, Kulliyah of Science,  
International Islamic University Malaysia, Kuantan, 25200, Pahang, Malaysia

### ARTICLE INFO

#### Article History:

Received: August 20, 2015

Accepted: October 03, 2015

#### Corresponding Author:

N.R.N. Idris

Department of Computational and  
Theoretical Sciences, Kulliyah of  
Science, International Islamic University  
Malaysia, Kuantan, 25200,  
Pahang, Malaysia

### ABSTRACT

The objective of this study is to ascertain the efficacy of the existing two-stage method for combining the Aggregate Data (AD) and Individual Patient Data (IPD) studies in meta-analysis and to introduce some modifications to the existing two-stage method for combining the these studies. We evaluated the effects of these modifications on the estimates of the overall treatment effect and investigated the influence of the number of studies included in the meta-analysis,  $N$  and the ratio of AD: IPD on these estimates. 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 current method for combining the AD:IPD studies exhibit rather poor coverage probability. We showed that the proposed methods had been able to improve the coverage probability by more than 40%, while maintaining the level of bias at par to their existing counterparts. These findings demonstrated the significance of proper technique for combining these studies in order to obtain better overall estimates, which is crucial for reliable overall conclusions.

**Key words:** Meta-analysis, combined data, aggregate data, simulation, bias, coverage

### INTRODUCTION

Meta-analysis is a statistical technique for integrating quantitative results from several independent sources. A traditional meta-analysis involves the integration of Aggregate Data (AD), such as a standardized mean difference for continuous outcomes or the number of events and participants for binary outcomes, which is extracted from the individual study publications. Methods such as the inverse variance method (DerSimonian and Laird, 1986), for continuous data, or the Mantel-Haenszel method (Mantel and Haenszel, 1959), for binary data, may be utilized to obtain the overall estimate. A meta-analysis may alternatively be performed using Individual Patient Data (IPD), in which case raw data from an individual study are obtained and synthesized directly. In this case, 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 and Tierney, 2002; Simmonds *et al.*, 2005; Jones *et al.*, 2009) and IPD are seldom available from each of the individual studies. Another approach which is gaining in popularity is through utilizations of both the available AD and IPD, known as Mixed Data (MD). 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 (Jones *et al.*, 2009; Cooper and Patall, 2009).

Literature on the comparison of AD and IPD meta-analysis estimates has been quite extensive (Jones *et al.*, 2009; Lambert *et al.*, 2002). However, only two studies had examined the efficacy of the estimates that are based on combined-level studies (Riley *et al.*, 2008; Idris and 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 to those using all-AD studies. Riley *et al.* (2008) results showed that the benefit of combining the IPD with AD increases as the proportion of IPD studies decreases and that the differences in the estimates from all AD and AD:IPD combined studies were relatively small. These findings are supported by another simulation-based study (Idris and 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, including the AD is not recommended as the statistical benefit is not significant. In fact, in this case, it would only serve to increase the overall SE.

Currently four methods, based on one-stage and two-stage approaches, are available for combining the IPD and AD, namely, the standard two-stage method (Collette *et al.*, 1998; Tudur *et al.*, 2001), partially reconstructed IPD method (Turner *et al.*, 2000; Whitehead *et al.*, 2001), multi-level modeling method (Goldstein *et al.*, 2000) and the hierarchical related regression using a Bayesian framework (Jackson *et al.*, 2006). The standard two-stage method uses a two-stage approach whereas the other three methods are based on one-stage approach. Riley *et al.* (2007) found that of 33 meta-analyses that combined the AD and IPD studies, 29 used a standard two-stage method and the other four have adopted the reconstruction method for binary outcome data. None of the researchers for the meta-analyses reviewed used either multi-level models or the Bayesian approach. The popularity of a two-staged approach may be attributed to its practical simplicity and ease of understanding without overwhelming the readers with heavy theoretical background information.

Although the two-stage method has been widely accepted in practice, discussion on the methodology of this approach and its implications on the overall estimates is limited. The assessments of the statistical properties of estimates utilising two-stage method to combine the studies has not been fully investigated. In this study, we examined how the estimate of overall treatment effect from combined AD:IPD studies compared to those from conventional AD-only meta-analyses, in cases where both levels of data were available. A simulation study was used to quantify and compare the Percentage Relative Bias (PRB) and the Root Mean Square Error (RMSE) of meta-analysis estimates extracted from AD-only studies and from selected ratios of AD:IPD studies. Further, the possibility of improving the statistical properties of the overall estimates were explored, with some modifications of the existing two-stage method. Two modifications to the existing two-stage approach were proposed. The first modification involved the utilization of the number of studies as the weightage in combining the overall estimates and the second modification involved utilizing the grouped-variance from the two levels of

data as the weightage. The effects of these modifications to the estimates of the overall treatment effect were evaluated and compared to those using the existing method. The PRB and RMSE 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.

## MATERIALS AND METHODS

**Generation of data:** 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,

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \epsilon_{ij}$$

where,  $\beta_{0i}$  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  $\epsilon_{ij}$  were the sampling random error terms for the response from patient  $j$  within study  $i$ . We assumed  $\beta_{0i}$ ,  $\beta_{1i}$  and  $\epsilon_{ij}$  were independent and normally-distributed, with  $\epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2)$ ,  $\beta_{0i} \sim N(\beta_0, \sigma_{v_0}^2)$  and  $\beta_{1i} \sim N(\beta_1, \tau^2)$ . For simplicity, we assumed each study had an equal number of patients in each treatment arm (i.e.,  $n_{0i} = n_{1i} = 1/2 n_i$  for  $i = 1, 2, \dots, N$ ). 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 ( $I^2:25-50\%$ );  $\sigma_{v_0}^2 = 2$  while we allowed  $\sigma_{\epsilon}^2$  to vary randomly between 1-25. 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 proceeding 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 three methods; the existing standard two-stage method (SM) and two modifications of existing methods, namely, the N-weighted estimates (NW) and the inverse group-variance weighted estimates (IGVW). All the parameters in this study 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.

**Statistical Assessments of the effect estimates:** These specifications generated 32 sets of meta-analyses comprising four all-IPD studies, four all-AD studies and 24 sets of AD:IPD studies consisting of different combinations of N and AD:IPD ratios. Each meta-analysis were 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.

**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:

$$\text{Mean PRB}(\hat{\theta}) = \frac{\sum_{t=1}^K (\theta - \hat{\theta}_t) / \theta}{K} \times 100\%$$

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

**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:

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

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

**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  $\theta$  out of the total number of simulations  $K$ .

**Two-Stage Method for Combining the AD and IPD**

**Standard two-stage method (SM):** Suppose there  $N = N_1 + N_2$  studies where  $N_1$  was the number of AD level studies with effects estimates  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{N_1}$  and corresponding variances given by:

$$V(\hat{\theta}_1), V(\hat{\theta}_2), \dots, V(\hat{\theta}_{N_1})$$

while  $N_2$  was the number 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  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_{N_2}^*$  denoted the study specific effects from the reduced IPD studies, with corresponding variances:

$$V(\hat{\theta}_1^*), V(\hat{\theta}_2^*), \dots, V(\hat{\theta}_{N_2}^*)$$

The effect  $\hat{\theta}_i^*$  for instance may represent the difference in the mean response of individual patients for treatment and control arms from study  $i$  and the  $V(\hat{\theta}_i^*)$  represents the pooled treatment and control arms variances corresponding to their respective effects for study  $i$ .

The overall effect and the corresponding SE using the standard inverse variance weighted method were, respectively, given by:

$$\hat{\theta}_{\text{overall}} = \frac{\sum w_i \hat{\theta}_i + \sum w_i^* \hat{\theta}_i^*}{\sum w_i + \sum w_i^*}$$

And:

$$V(\hat{\theta}_{\text{overall}}) = \frac{1}{\sum w_i + \sum w_i^*}$$

**Modifications of SM:** In this modification, the overall effects from AD and the overall effects from IPD were computed separately. These two effects were then combined using the weighted average, where, the weights were:

- Number of studies in each group, that is the N-weighted estimate (NW)
- Inverse variances from the AD and IPD group, that is, the Inverse Group Variance Weighted estimate (IGVW)

**N-Weighted estimate (NW):** Using the same notation as in the preceding section, let  $N_1$  and  $N_2$  be the number of AD and IPD studies, respectively, in a given meta-analysis. The overall estimate for AD studies, denoted  $\hat{\theta}_{AD}$ , were first computed using the standard method. The overall IPD estimate,  $\hat{\theta}_{IPD}$ , were computed directly using the mixed effect model without reducing to the AD level. These two estimates were then combined using the  $N_1$  and  $N_2$  as the weights, as follows:

$$\hat{\theta}_{\text{all}} = \frac{(N_1)\hat{\theta}_{AD} + (N_2)\hat{\theta}_{IPD}}{N_1 + N_2}$$

The variance of the overall estimate,  $V(\hat{\theta}_{\text{all}})$ , was the pooled estimate of the variances of the two groups.

**Inverse Group-Variance Weighted Estimate (IGVW):** In the second modification, the inverse variances of each group were utilized as the weights. For the AD studies, the weight,

$w_{AD}$ , was simply the inverse of the overall variance for the AD group,  $w_{AD} = 1 / V(\hat{\theta}_{AD})$ , for IPD studies, the weight,  $w_{IPD}$  was the inverse variance for IPD group  $w_{IPD} = 1 / V(\hat{\theta}_{IPD})$  and the overall estimate was given by:

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

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

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

**RESULT**

**The PRB, RMSE and Coverage probability of estimates from meta-analysis with AD-only studies and combined AD:IPD studies:** Figure 1(a-c) 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 1.5-5.0% compared to the conventional AD-only meta-analysis

(mean PRB = -11-5.3%). 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 the bias.

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 0.53-1.22 compared to those from AD-only (mean RMSE of 0.55- 2.17). The RMSE decreased as the number of studies, N, increased, which was expected as the SE tended to decrease with increasing N (Fig. 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- 90% for AD-only studies. The coverage for combined AD:IPD studies was slightly lower, ranging from 50-70%, increasing with the proportion of AD within the AD:IPD ratio.

**Comparison of PRB and RMSE of the standard two-stage method (SM) against the modified two-stage methods (NW and IGVW) for combining the AD and IPD:** A comparison of the observed statistical properties of estimates

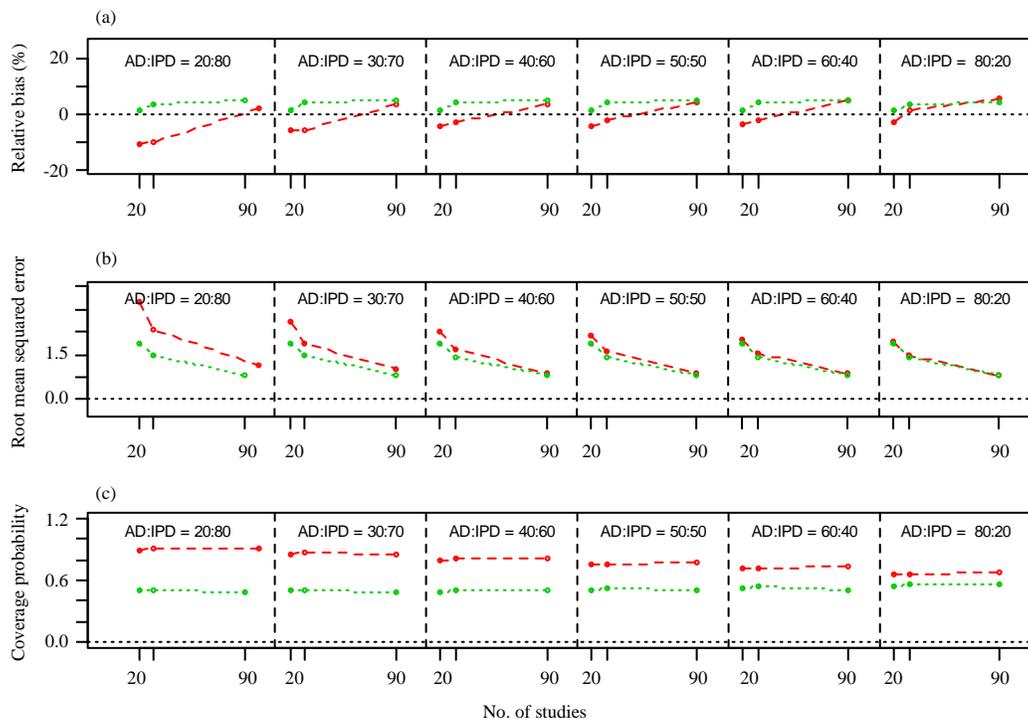


Fig. 1(a-c): Distribution of PRB, RMSE and the Coverage probability for the overall estimates from AD-only and AD:IPD studies AD: Aggregate data, IPD: Individual patient data, PRB: Percentage relative bias, RMSE: Root mean square error Red: AD only studies and Green: Combined AD: IPD using standard method

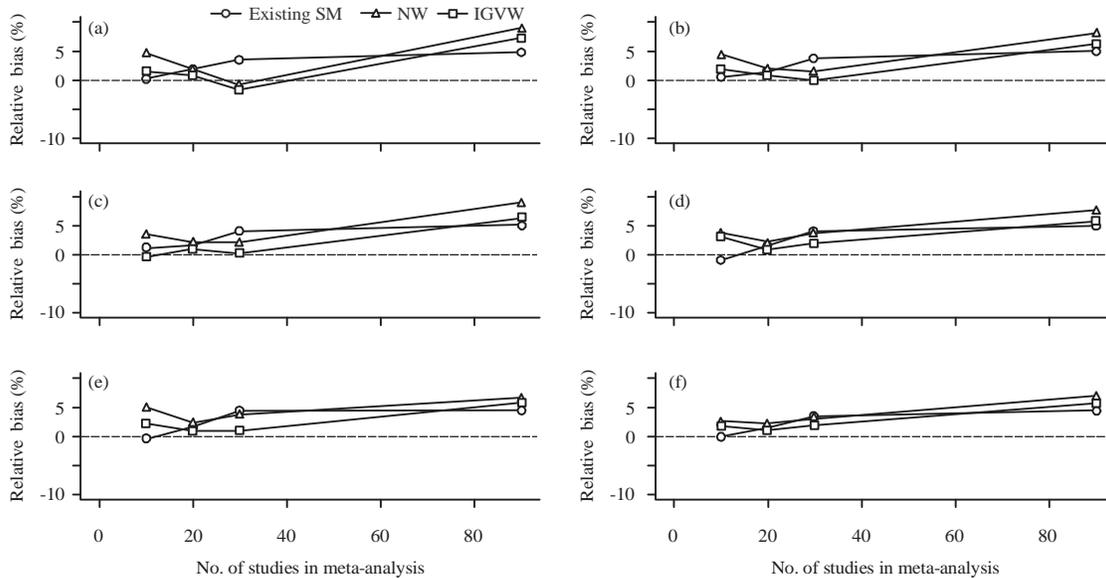


Fig. 2(a-f): Distribution of PRB for the overall estimates using SM, NW and IGWW methods for the selected range of N (N from 10-90) and six combinations of AD:IPD ratios. SN: Standard method, NW: N-weighted, AD: Aggregate data, IPD: Individual patient data and IGWW: Inverse group variance weighted (a) AD:IPD = 20:80, (b) AD:IPD = 30:70, (c) AD:IPD = 40:60, (d) AD:IPD = 50:50, (e) AD:IPD = 60:40 and (f) AD:IPD = 80:20

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 three methods was quite similar (Fig. 2(a-f)). It was noted that the proposed methods, NW and IGWW, displayed smaller PRB than the existing method when the number of studies included in the meta-analysis was at moderate range, N (20<N<60). Of the two proposed methods, IGWW consistently produced smaller PRB than the NW for all the ratios of AD:IPD considered.

Figure 3(a-f) shows the RMSE was slightly lower when the existing method, MS, was used to combine the AD and IPD studies when a majority of the studies within the AD:IPD studies were IPD. Of the two proposed methods, IGWW showed a smaller RMSE compared to NW. This trend was similar for all ratios of AD:IPD and as the proportions of AD increased, the RMSE of the estimates from all three methods converged to the same value

**Coverage probability and SE of estimates based on studies with SM, NW and IGWW:** The proposed methods provided estimates with better coverage than those from the existing method. The NW provided mean coverage of approximately 90% for a nominal value of 95% whereas the corresponding mean coverage for IGWW was approximately 70% against an average of 50% using the existing method (Fig. 4(a-f)). It was noted that the SE from estimates based on the standard method were underestimated, resulting in a confidence band which was too narrow (Table 1). The best

coverage was given by NW, which may be attributed to its relatively larger and more realistic SE despite a larger bias.

**Effects of ratio of AD:IPD and the number of studies, N, on the overall estimates:** The number of studies included in a meta-analysis has some effects on the accuracy of the overall estimate. Figure 2 exhibits an increasing trend in PRB as the number of studies increases, which is particularly apparent in the two proposed 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 all three methods, across N, reduced as proportions of AD within the AD:IPD ratios increased. This implied that the proposed NW and IGWW methods did not have much of an advantage over MS 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 (Table 1). As expected, the SE reduced as N increased and relatively smaller SEs were observed from the SM method. Figure 3 additionally illustrated that the RMSE of the estimates from all three 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

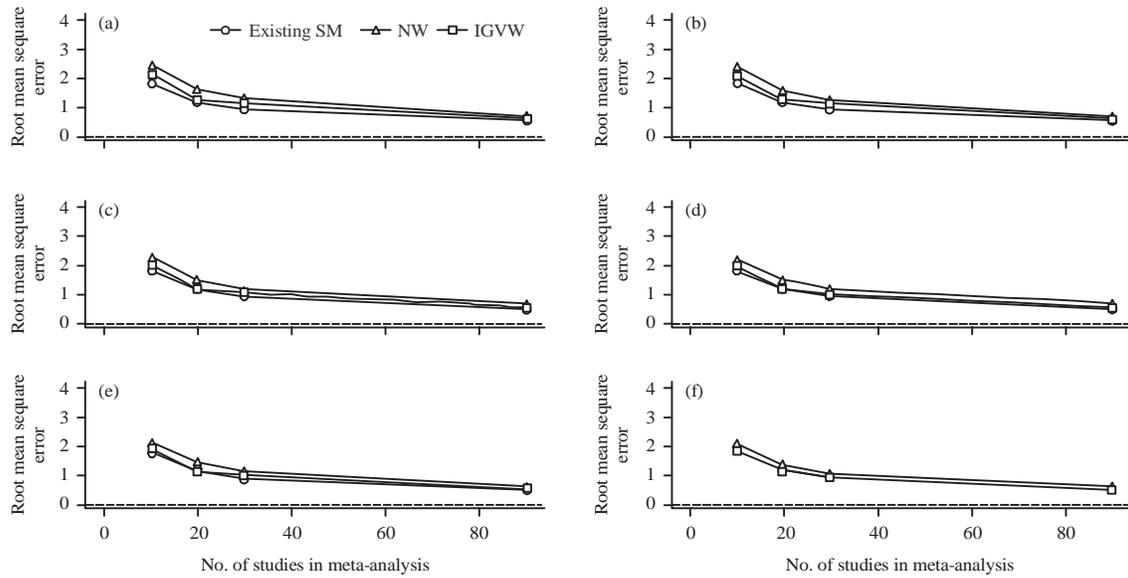


Fig. 3(a-f): Distribution of RMSE for the overall estimates using SM, NW and IGWV methods for six combinations of AD:IPD and ranges of N from 10-90, SM: Standard method, NW: N-weighted, IGWV: Inverse rroup variance weighted and RMSE: Root mean square error (a) AD:IPD = 20:80, (b) AD:IPD = 30:70, (c) AD:IPD = 40:60, (d) AD:IPD = 50:50, (e) AD:IPD = 60:40 and (f) AD:IPD = 80:20

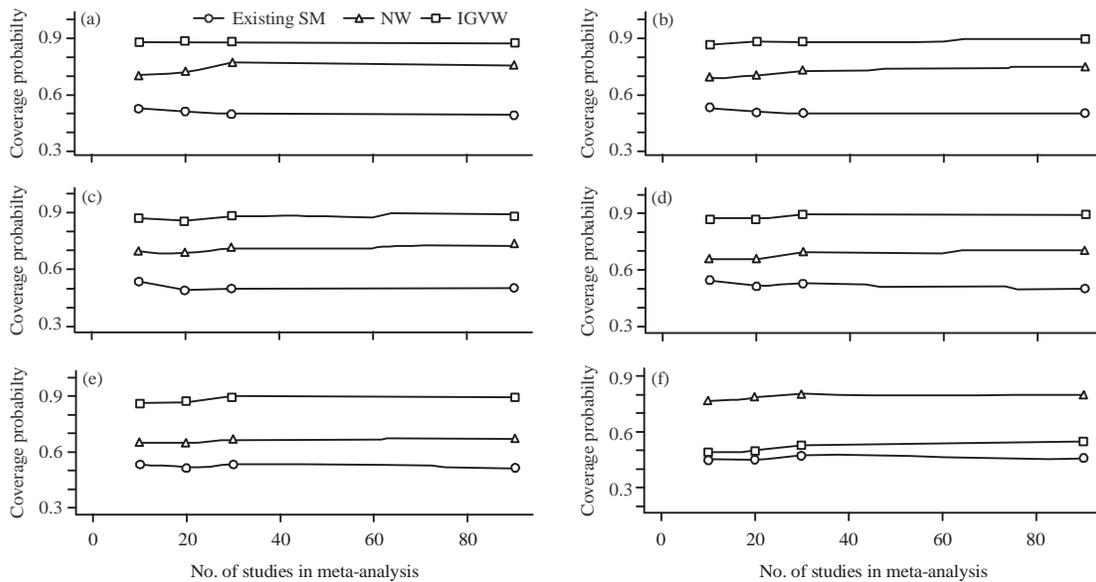


Fig. 4(a-f): Distribution of coverage for the overall estimates using SM, NW and IGWV methods for six combinations of AD:IPD and ranges of N from 10-90. SM: Standard method, NW: N-weighted, IGWV: Inverse rroup variance weighted and RMSE: Root mean square error (a) AD:IPD = 20:80, (b) AD:IPD = 30:70, (c) AD:IPD = 40:60, (d) AD:IPD = 50:50, (e) AD:IPD = 60:40 and (f) AD:IPD = 80:20

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. This trends were observed on all three methods of combining under consideration.

## DISCUSSION

Our simulation results revealed clear evidence of the superiority of estimates from the combined AD:IPD studies over those that utilized only the available AD, in terms of both

Table 1: Distribution of the estimates of overall treatment effects and their corresponding standard error

N	AD:IPD ratio	EST(SE)	EST(SE)	EST(SE)
		MS	NW	IGVW
10	20:80	6.97 (1.32)	6.66 (1.95)	6.88 (1.60)
	40:60	6.91 (1.33)	6.76 (1.81)	7.04 (1.50)
	50:50	7.06 (1.33)	6.74 (1.75)	6.78 (1.45)
	60:40	7.02 (1.33)	6.65 (1.70)	6.85 (1.42)
	80:20	6.99 (1.34)	6.83 (1.65)	6.87 (1.38)
20	20:80	6.89 (0.89)	6.86 (1.29)	6.92 (1.10)
	40:60	6.89 (0.89)	6.87 (1.22)	6.93 (1.02)
	50:50	6.89 (0.90)	6.85 (1.19)	6.92 (0.99)
	60:40	6.88 (0.89)	6.86 (1.17)	6.91 (0.97)
	80:20	6.89 (0.90)	6.89 (1.14)	6.92 (0.94)
30	20:80	6.75 (0.70)	7.04 (1.01)	7.09 (0.89)
	40:60	6.72 (0.69)	6.89 (0.96)	6.99 (0.86)
	50:50	6.72 (0.69)	6.85 (0.93)	6.87 (0.79)
	60:40	6.68 (0.69)	6.73 (0.93)	6.92 (0.77)
	80:20	6.76 (0.70)	6.79 (0.90)	6.87 (0.76)
90	20:80	6.64 (0.39)	6.37 (0.55)	6.46 (0.47)
	40:60	6.65 (0.39)	6.38 (0.53)	6.55 (0.44)
	50:50	6.66 (0.40)	6.47 (0.53)	6.59 (0.43)
	60:40	6.68 (0.40)	6.53 (0.52)	6.60 (0.43)
	80:20	6.69 (0.40)	6.52 (0.51)	6.60 (0.41)

SM: Standard method, NW: N-weighted, IGVW: Inverse group variance weighted and RMSE: Root mean square error, SE: Standard error

the accuracy and the overall RMSE. These findings were consistent with earlier results (Riley *et al.*, 2008; Idris and Abdullah, 2015) whose researchers compared the estimates from all-IPD and all-AD studies to the combined AD:IPD studies. We demonstrated that the value of combining AD:IPD was significant even if only 20% of the available studies were at the IPD level. Earlier studies (Stewart and Tierney, 2002; Simmonds *et al.*, 2005; Jones *et al.*, 2009) suggested that the methodological advantage of IPD meta-analysis was offset by the cost and time factors in obtaining the IPD data. Thus, combining readily available IPD and AD would therefore ease these two factors while maintaining some benefit of IPD meta-analysis. The only setback for the combined level studies was that its coverage probability was relatively low compared to the AD-only meta-analysis, when the existing method for combining was used.

It was postulated that the coverage probability may be improved if some modifications were introduced to the existing methodology that was used to combined the AD:IPD studies. A grouped-based weightage that utilized all available information contained in the IPD studies were considered. The results of this study demonstrated that the proposed methods provided better statistical properties than the existing method when the number of studies in a meta-analysis, N, was moderate (10<N<60). For larger N (N>60), the benefit of the proposed method remain relatively good in terms of coverage, but was not markedly different in terms of bias and RMSE than those from the existing method.

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 combined 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. Of the two modifications introduced, NW, which utilized the number of studies, N, as the weightage and IGVW, which utilized the inverse group-variances as the weightage, the latter appeared to offer consistently better performance in terms of bias and RMSE. The IGVW however, displayed lower coverage probability than NW, albeit better than the existing method. In both modification methods, estimates from IPD studies were estimated directly, without reducing them to AD level first, thus utilizing all the information available within the IPD.

The lack of theoretical support for the proposed modifications of the two-stage methods is acknowledged, 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.

One of the main goals of meta-analysis is to draw general inferences about the research problem from the overall estimate based on a collection of independent studies. Thus a reliable overall estimate is crucial in meta-analysis. This article confirmed that combining the available AD and IPD studies provided more reliable overall estimates and better statistical properties. Therefore, it is strongly recommend that, whenever

possible, practitioners should include the available IPD studies when performing conventional meta-analysis. 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 performing meta-analysis.

#### ACKNOWLEDGMENT

This study was supported by the Fundamental Research Grant Scheme (FRGS 13-093-033), Ministry of Higher Education, Government of Malaysia.

#### REFERENCES

- Collette, L., S. Suci, L. Bijns and R. Sylvester, 1998. Including literature data in individual patient data meta-analyses for time-to-event endpoints. Proceedings of the 2nd Joint Meeting of the Society for Clinical Trials and the International Society of Clinical Biostatistics, July 6-10, 1997, Boston, Massachusetts, USA.
- Cooper, H. and E.A. Patall, 2009. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol. Methods*, 14: 165-176.
- DerSimonian, R. and N. Laird, 1986. Meta-analysis in clinical trials. *Controlled Clin. Trials*, 7: 177-188.
- Goldstein, H., M. Yang, R. Omar, R. Turner and S. Thompson, 2000. Meta-analysis using multilevel models with an application to the study of class size effects. *J. R. Stat. Soc.: Ser. C (Applied Stat.)*, 49: 399-412.
- Idris, N.R.N. and M.H. Abdullah, 2015. A study on the effects of different levels of data on the overall meta-analysis estimates. *Far East J. Math. Sci.*, 96: 73-86.
- Jackson, C., N. Best and S. Richardson, 2006. Improving ecological inference using individual-level data. *Stat. Med.*, 25: 2136-2159.
- Jones, A.P., R.D. Riley, P.R. Williamson and A. Whitehead, 2009. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin. Trials*, 6: 16-27.
- Lambert, P.C., A.J. Sutton, K.R. Abrams and R.D. Jones, 2002. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J. Clin. Epidemiol.*, 55: 86-94.
- Mantel, N. and W. Haenszel, 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*, 22: 719-748.
- 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, Pages: 2673.
- Riley, R.D., M.C. Simmonds and M.P. Look, 2007. Evidence synthesis combining individual patient data and aggregate data: A systematic review identified current practice and possible methods. *J. Clin. Epidemiol.*, 60: 431.e1-431.e12.
- Riley, R.D., P.C. Lambert, J.A. Staessen, J. Wang, F. Gueyffier, L. Thijs and F. Bouitrie, 2008. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat. Med.*, 27: 1870-1893.
- Simmonds, M.C., J.P.T. Higgins, L.A. Stewart, J.F. Tierney, M.J. Clarke and S.G. Thompson, 2005. Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. *Clin. Trials*, 2: 209-217.
- Stewart, L.A. and J.F. Tierney, 2002. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval. Health Professions*, 25: 76-97.
- Tudur, C., P.R. Williamson, S. Khan and L.Y. Best, 2001. The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *J. R. Stat. Soc.: Ser. A (Stat. Soc.)*, 164: 357-370.
- Turner, R.M., R.Z. Omar, M. Yang, H. Goldstein and S.G. Thompson, 2000. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat. Med.*, 9: 3417-3432.
- Wang, J.G., J.A. Staessen, S.S. Franklin, R. Fagard and F. Gueyffier, 2005. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension*, 45: 907-913.
- Whitehead, A., R.Z. Omar, J. Higgins, E. Savaluny, R.M. Turner and S.G. Thompson, 2001. Meta-analysis of ordinal outcomes using individual patient data. *Stat. Med.*, 20: 2243-2260.