

# Comparing the Overall Result and Interaction in Aggregate Data Meta-Analysis and Individual Patient Data Meta-Analysis

Yafang Huang, MD, PhD, Jinling Tang, MD, PhD, Wilson Wai-san Tam, PhD,  
Chen Mao, MD, PhD, Jinqiu Yuan, MD, Msc, Mengyang Di, MD, PhD, and Zuyao Yang, MD, PhD

**Abstract:** The aim of the study was to examine how well aggregate data meta-analyses (ADMAs) and individual patient data meta-analyses (IPDMAs) agree in their overall results and how frequently interactions are detected in IPDMAs and ADMAs.

ADMA articles immediately published before the IPDMA and matching the research topic were identified. Agreement in the overall result was achieved if the estimate was in the same direction. The number of subgroup analyses, in particular that of significant interactions, was compared between the 2 types of meta-analyses.

A total of 829 IPDMA articles were identified; 129 (15.6%) were found to have a matched ADMA article and 204 paired meta-analyses were identified. Agreement in the overall effect was observed in 187 (91.7%) of the 204 paired meta-analyses. Fifty-three (26.0%) ADMAs and 121 (59.3%) IPDMAs conducted subgroup analyses and presented 150 and 634 subgroup analyses, respectively. The IPDMAs conducted 7 times more subgroup analyses on interaction (544 in IPDMAs vs 68 in ADMAs) and identified 14 times more potential interactions (44 in IPDMAs vs 3 in ADMAs).

ADMAs will almost always agree with their corresponding IPDMAs in the overall result if greater efforts are made to improve the methodology in conducting ADMAs. The IPDMA is required mostly if interactions are suspected.

(*Medicine* 95(14):e3312)

Editor: Dimitrios Paraskevis.

Received: October 27, 2015; revised: March 7, 2016; accepted: March 11, 2016.

From the School of General Practice and Continuing Education, Capital Medical University (YH), Beijing, China; Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore (WW-ST), Singapore; and Division of Epidemiology, The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong (JT, CM, JY, MD, ZY, YH), Hong Kong SAR, China.

Correspondence: Wilson Wai-san Tam, Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore (e-mail: wilson\_tam@nuhs.edu.sg).

Jinling Tang, School of Public Health Building, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China (e-mail: jltang@cuhk.edu.hk).

Authors' contributions: JT and WT had the idea for study and is the guarantor; YH, JT, and WT developed the research protocol; YH, CM, JY, MD, and ZY searched for and identified relevant studies. YH, CM, JY, and MD extracted and checked data from the included articles; YH and WT analyzed data; JT and WT provided methodological support and helped to interpret results; YH, JT, and WT drafted and revised the manuscript, and all authors critically commented on it.

The authors report no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution-NoDerivatives License 4.0, which allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the author.

ISSN: 0025-7974

DOI: 10.1097/MD.0000000000003312

**Abbreviations:** ADMAs = aggregate data meta-analyses, CIs = confidence intervals, IPDMAs = individual patient data meta-analyses, N/A = not applicable, Non-sig = statistically non-significant, Sig = statistically significant.

## INTRODUCTION

Meta-analysis is a statistical method that integrates results from different studies to produce an overall result. Clinicians, guideline developers, and policymakers increasingly use meta-analyses as high-quality evidence aids to make decisions and recommendations.<sup>1-3</sup> Meta-analysis can either be based on aggregate data or on individual patient data.<sup>4,5</sup> Aggregate data meta-analysis (ADMA) combines the grouped data of primary studies, whereas individual patient data meta-analysis (IPDMA) synthesizes the individual data of primary studies.<sup>6,7</sup> In principle, IPDMAs have advantages over ADMAs, as they can more strictly standardize included studies in factors such as patient characteristics, treatment details, and duration of follow-up; they can also more effectively conduct subgroup analyses and control confounding in such analyses.<sup>8</sup> Thus, IPDMA is often viewed as superior to ADMA and can produce higher-quality evidence.<sup>4,6,9</sup> However, ADMAs are much more common because they are much more rapid and require far less resources to produce.<sup>10,11</sup> A study showed that >95% of published meta-analyses were ADMAs.<sup>12</sup> Several studies have compared ADMAs with IPDMAs.<sup>13-18</sup> These studies include a very small number of meta-analyses and focused only on the overall result.

We conducted this comprehensive review to compare a previous ADMA with its subsequent IPDMA of the same topics to investigate how often ADMAs agree with IPDMAs, what factors affect the agreement, and how effective IPDMAs are in exploring interactions compared with ADMAs.

## METHODS

### Literature Search Strategies

A literature search on IPDMA articles was conducted in August 2012. A total of 829 eligible IPDMA articles were identified. Details have been reported elsewhere.<sup>19</sup> Briefly, all IPDMA articles were identified by a comprehensive search of PubMed, EMBASE and the Cochrane Library with an established search strategy. For each of the 829 IPDMA articles, PubMed was searched to find matching ADMAs against the disease and intervention of the IPDMA. A total of 829 searches were conducted. The search was further limited by using Montori's balanced 5 search terms for identifying systematic reviews.<sup>20</sup> Details of the search strategy can be found in Supplemental Content, <http://links.lww.com/MD/A879>. We also scrutinized the references of each eligible IPDMA article.

An updated search was conducted in March 2013. Results of the PubMed search for each IPDMA article were saved in EndNote libraries separately.

### Selection of ADMA Articles

The screening of matched ADMA articles was carried out separately for each eligible IPDMA by using the results saved in the EndNote library. ADMA articles that were the same as or similar to the IPDMA with regard to patients, test intervention, control intervention, and at least 1 outcome were selected.<sup>21</sup> We then excluded ADMA articles that were published after the index IPDMA article. If >2 ADMA articles were found for 1 index IPDMA, the ADMA that was published immediately before the IPDMA was considered eligible and used in the final analysis.

We excluded articles published in non-English languages, qualitative reviews without meta-analysis, and IPDMA articles. ADMA articles on diagnostic accuracy and matched ADMAs that did not report the overall combined result were also excluded. If  $\geq 2$  pairs of reviews were found on the same topic, we used the most up-to-date pair.

Two authors independently screened the titles and abstracts. They subsequently screened the full-text articles for which eligibility remained unclear. Any discrepancies were resolved by consensus or by consulting a third author if the 2 failed to reach an agreement.

### Data Extraction

For each matched pair of ADMA and IPDMA article,  $\geq 1$  matched meta-analyses could be eligible and were all extracted. All the matched meta-analyses extracted must be matched by patient, intervention, comparator, and outcome.<sup>22</sup> The matched meta-analyses formed the basic data for our analysis.

We extracted information from each article on disease, test intervention, control intervention and outcome, direction of effect, statistical significance of the estimate, and number and significance of subgroup analyses and of interaction terms. The direction of effect was divided into 2 groups: the test treatment is more effective than the control treatment and the test treatment is equally or less effective than the control.

A subgroup analysis refers to an analysis in which the trials or patients are divided into subgroups according to an attribute of the trial or patient, and the results are combined in each subgroup and then compared among subgroups. The attribute can be treatment dosage, treatment in the control group, patient characteristics, treatment setting, and so on. The third factor beyond the treatment and outcome can be potential effect modifiers, for example, patients' demographic factors and lifestyle of patients, as well as co-morbidities or characteristics of the disease.<sup>23</sup> Subgroup analyses according to these factors will be considered to assess interaction or effect modifications. A product term between the test treatment and a third factor in a regression analysis was also considered an analysis for interaction. Only subgroup analyses conducted in the matched meta-analyses were extracted.

### Statistical Analyses

Descriptive analyses were conducted to summarize the characteristics of the included meta-analyses. Percentage was used for categorical variables, and median and interquartile

range (IQR) were used for continuous variables. The Wilcoxon signed-rank test was used to detect the differences in number of studies, patients, and length of follow-up between ADMAs and IPDMAs. The number of meta-analyses that conducted subgroup analyses was analyzed by McNemar  $\chi^2$  test.

We used the methods of Villar et al<sup>24</sup> to define the agreement and disagreement between the paired ADMA and IPDMA. An ADMA was classified as being in agreement with its matched IPDMA if the effect of both was in the same direction. Otherwise, they were classified as being in disagreement.

We investigated the association of agreement with the following characteristics of the ADMAs: research topic (treatment or prognosis), types of outcome (objective or subjective), study design (randomized controlled trials or others), search for grey literature (yes or no), request for data from author (yes or no), use of intention-to-treat analysis (yes or no), significance of testing results (significant or nonsignificant), direction of effect of the test intervention compared with the control (greater benefit or equal benefit/greater harm), and between-study heterogeneity (yes or no). The total and percentage of significant subgroup analyses and interactions between ADMAs and IPDMAs were also compared.  $\chi^2$  test or Fisher exact test (when the expected cell frequency is  $< 5$ ) was used for comparison.

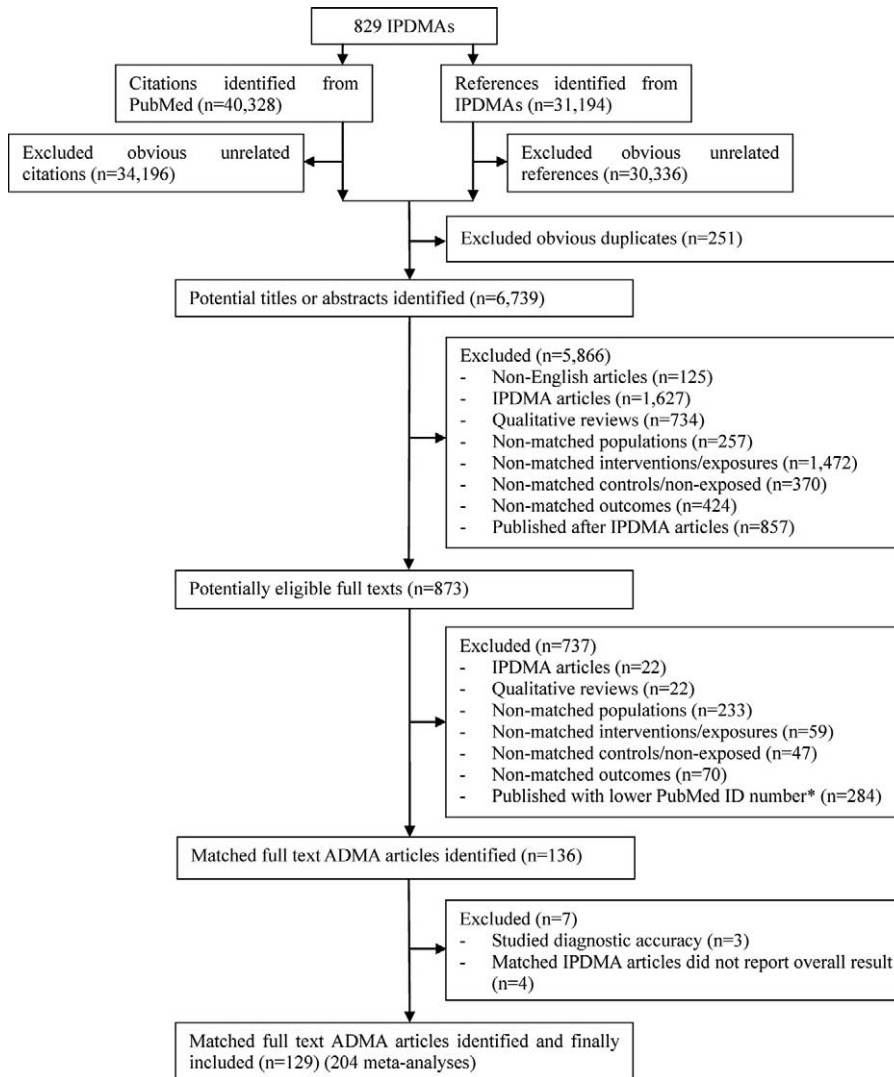
In a sensitivity analysis, we extracted the original aggregate data, which were available in a fraction of the ADMAs, and re-estimated the overall result by using the same effect measure used in the matched IPDMA. This step allowed us to directly compare the size of effect between the ADMA and the IPDMA and estimate the agreement differently. If no statistically significant difference was found between the matched ADMA and IPDMA, we assumed consistency in their overall result. Otherwise, inconsistency was assumed.<sup>25</sup> This is a commonly used method to quantify the agreement between meta-analyses, but it requires the estimation of the effect with the same effect measure.<sup>25</sup> We used SPSS (version 18.0 for Windows, SPSS Inc, Chicago, IL) to perform the analyses and used R 3.2.3 to plot the figure.

## RESULTS

From the 829 IPDMA articles, 71,522 citations were identified from the PubMed search and references of the IPDMA articles. A total of 129 ADMA matched articles were found, which resulted in 204 matched meta-analyses eligible for this study. Figure 1 shows the details of the search for matched ADMAs and the results of each search step.

The characteristics of the 204 matched meta-analyses are summarized in Table 1. Of the 204 matched meta-analyses, 69 (33.8%) studied cardiac and cardiovascular diseases, 132 (64.7%) were on drugs or biologics, 43 (21.1%) used placebo as control, and 187 (91.7%) used objective outcomes. A total of 66.2% (135/204) of ADMAs and 66.7% (136/204) of IPDMAs showed that the test treatment was better than the control. In comparison, 61.3% (125/204) of ADMAs and 63.7% (130/204) of IPDMAs showed significant overall results.

In total, 187 (91.7%) of the 204 matched ADMAs and IPDMAs were in agreement, which was an effect in the same direction (Table 2). The agreement is even higher if grey literature was sought ( $P = 0.025$ ), data from authors were requested ( $P = 0.012$ ), intention-to-treat analysis was used ( $P = 0.027$ ), and the overall result was statistically significant



**FIGURE 1.** Flowchart of the literature search for prior paired ADMA articles. \*Articles with a smaller PubMed ID number are published earlier than those with a larger PubMed ID number. ADMA = aggregate data meta-analysis, IPDMAs = individual patient data meta-analysis.

( $P = 0.001$ ). The remaining characteristics evaluated did not seem to significantly affect the agreement; the characteristics include research topic, type of outcome, study design, direction of effect, and heterogeneity. The consistency rate was slightly lower than the agreement rate and was affected by research topic and study design (Table 2). Figure 2 presents the effect sizes and 95% confidence intervals (CIs) of the matched IPDMAs and ADMAs.

Fifty-three (26.0%) ADMAs and 121 (59.3%) IPDMAs reported subgroup analyses, suggesting IPDMAs are twice as likely to report subgroup analyses as their matched ADMAs. The number of subgroup analyses reported is 150 of the 204 ADMAs and 634 of the 204 IPDMAs (Table 3), which resulted in 8 (5.3%) and 55 (8.7%) statistically significant results, respectively.

Not all the subgroup analyses were on interaction; for example, some were on a dose–response relationship and some on methodological quality. The number of subgroup analyses on interaction was 68 (45.3%) in the ADMAs and 544 (85.8%) in the IPDMAs. The IPDMAs reported 7 times more subgroup analyses on interaction than their matched ADMAs. The number of statistically significant interactions reported is 3 (4.4%) in ADMAs and 44 (8.1%) in IPDMAs. The IPDMAs reported 14 times more statistically significant interactions than their matched ADMAs (Table 3).

In addition, of the 634 subgroup analyses in IPDMAs, 215 (33.9%) studied demographic factors and lifestyle, 202 (31.9%) comorbidities, and 127 (20.0%) studied characteristics of the disease. In comparison, of the 150 subgroup analyses in ADMAs, 23 (15.3%) studied demographic factors and lifestyle,

**TABLE 1.** Characteristics of the Included 204 Matched Meta-analyses

Characteristics	Number (%)
Type of disease	
Cardiac and cardiovascular systems	69 (33.8)
Hematology/oncology	67 (32.8)
Psychiatry and psychology	17 (8.3)
Endocrinology and metabolism	8 (3.9)
Respiratory system	7 (3.4)
Others	36 (17.6)
Total	204 (100.0)
Type of intervention	
Drugs or biologics*	132 (64.7)
Surgical interventions	23 (11.3)
Others	49 (24.0)
Total	204 (100.0)
Type of control intervention	
Drugs or biologics	47 (23.0)
Placebo or no treatment	43 (21.1)
Surgical interventions	39 (19.1)
Others	75 (36.8)
Total	204 (100.0)
Types of outcomes	
Objective	187 (91.7)
Subjective	17 (8.3)
Total	204 (100.0)
Direction of effect	
Same	187 (91.7)
Opposite	17 (8.3)
Total	204 (100.0)
Statistical significance of the effect	
Same	155 (76.0)
Both significant	103 (50.5)
Both non-significant	52 (25.5)
Different	49 (24.0)
Only ADMA significant	22 (10.8)
Only IPDMA significant	27 (13.2)
Total	204 (100.0)

ADMA = aggregate data meta-analysis, IPDMAs = individual patient data meta-analysis.

\*Biologics include monoclonal antibodies, cytokines, vaccines, etc.

16 (10.7%) comorbidities, and 29 studied (19.3%) characteristics of the disease.

## DISCUSSION

IPDMA is generally considered scientifically more rigorous than ADMA. The high agreement rate between the matched ADMAs and IPDMAs implies that ADMAs can provide valid conclusions most of the time. The agreement rate can be improved to >95% if the ADMA can improve methodologically in a number of aspects, such as requesting necessary data from authors, searching for grey literature, and using intention-to-treat analysis. We included all the meta-analyses regardless of the fields of study and a largest number of meta-analyses in

various fields. Our study may provide a much more generalizable comparison between ADMAs and IPDMAs.

The 2 types of meta-analyses differed greatly in subgroup analyses and interactions found. In particular, the IPDMAs reported 3 times more subgroup analyses and 7 times more subgroup analyses on interaction than the ADMAs, although the percentage of subgroup analyses that were statistically significant did not differ between the 2 types of meta-analyses. More importantly, the IPDMAs found 14 times more interactions that were statistically significant than the matched ADMAs. As confounding can be more effectively controlled in subgroup analyses in IPDMAs than in ADMAs, IPDMAs are also more likely to provide valid conclusions on interaction, although this possibility was not explored in this investigation.

Of the 17 ADMAs that disagreed with IPDMAs on the combined results, 8 cases were the ADMAs with positive result (intervention was superior), whereas IPDMAs were equivalent or inferior to the control. This disagreement may partly be explained by a possible selection bias in the ADMAs.<sup>26</sup> Many studies are published in non-English languages or in a conference abstract, journal correspondence, and book chapter.<sup>27</sup> These publications are sometimes called “grey literature” and often report a negative and/or nonsignificant result. Failure to include grey literature searches would thus cause selection bias and lead to overestimating of the true effect. Indeed, 6 of these 8 ADMAs (ie, 75%) did not search the grey literature.

Conversely, of the 17 ADMAs that disagreed with IPDMAs on the combined results, 9 cases were ADMAs with a negative result (intervention equivalent or inferior to the control), whereas the IPDMAs were estimating that intervention was superior. This may partly be explained by possible information bias in the ADMAs. Often, information bias in meta-analyses arises in the form of “data availability bias,” in which an ADMA is based on incomplete data. It usually happens when data are not openly reported in the original publications and the authors of ADMAs do not have access to the full set of data if they do not contact or fail to receive a reply from the authors of the original studies. Indeed, 8 of the 9 ADMAs (ie, 89%) did not contact the authors for the original studies. The median number of patients is 1086 in the 9 ADMAs, which is only half of the number of patients (2014) included in their matched IPDMAs.

Knowing whether the effect of a treatment differs according to demographic factors and lifestyle, comorbidities and characteristics of the disease are important in making better decisions. The IPDMAs reported many times more subgroup analyses according to these factors than the ADMAs. One reason the ADMA is not capable of assessing interactions is that corresponding subgroup analyses are not conducted and reported in the original studies. Subgroup analyses in trials are generally not encouraged and should be conducted only with the right reasons, to prevent false-positive results. Another reason is that original trials performed subgroup analyses but grouped the same variable in different ways, which makes it difficult and less meaningful to combine the subgroup analyses. This problem can be easily overcome in an IPDMA that has individual data.

No significant difference was found in the proportion of significant subgroup analyses and significant interactions between the matched ADMAs and IPDMAs. If subgroup analyses in the ADMAs are assumed to be based on cautiously planned subgroup analyses in original trials, this finding would suggest that subgroup analyses in IPDMAs do not seem to have resulted in many false-positive findings.

**TABLE 2.** Agreement Rate and Consistency in the 204 Paired ADMAs and IPDMAs According to the Characteristics of the Prior ADMAs

Characteristic of ADMAs	Agreement, Number (%)	P Value for the Difference	Consistency, Number (%)	P Value for the Difference
All matched meta-analyses	187 (91.7)	N/A	56 (84.8)	N/A
Research topic		>0.999*		0.008*
Treatment	144 (91.7)		50 (90.9)	
Prognosis	43 (91.5)		6 (54.5)	
Type of outcome		0.370*		0.583*
Objective	170 (90.9)		49 (83.1)	
Subjective	17 (100.0)		7 (100.0)	
Study design		0.375*		0.020*
RCT	143 (92.9)		48 (90.6)	
Not RCT	44 (88.0)		8 (61.5)	
Sought for grey literature		0.025†		0.442*
Yes	118 (95.2)		42 (87.5)	
No	69 (86.3)		14 (77.8)	
Requested data from authors		0.012†		0.483*
Yes	67 (98.5)		19 (90.5)	
No	120 (88.2)		37 (82.2)	
Used ITT		0.027†		>0.999*
Yes	85 (96.6)		22 (84.6)	
No	102 (87.9)		34 (85.0)	
Significance of result		0.001†		0.736*
Significant	121 (96.8)		35 (83.3)	
Non-significant	66 (83.5)		21 (87.5)	
Direction of effect		0.082†		0.085*
More effective than control	127 (94.1)		39 (90.7)	
Equally or less effective	60 (87.0)		17 (73.9)	
Between study heterogeneity‡		0.235*		0.675*
Yes ( $P < 0.1$ )	44 (95.7)		13 (92.9)	
No ( $P \geq 0.1$ )	102 (88.7)		43 (82.7)	

ADMA = aggregate data meta-analysis, IPDMAs = individual patient data meta-analysis, ITT = intention-to-treat analysis, N/A = not applicable, RCT = randomized controlled trial. Note: For agreement, if the same direction of effect is observed in the ADMA and its matched IPDMA, they are considered to agree with each other or agreement is reached between the 2 meta-analyses. Otherwise, the ADMA and its matched IPDMA are considered to disagree with each other. For consistency, only a difference between the matched ADMA and IPDMA that was non-statistically significant is considered consistent and the rest are deemed inconsistent.

\*Fisher exact test.

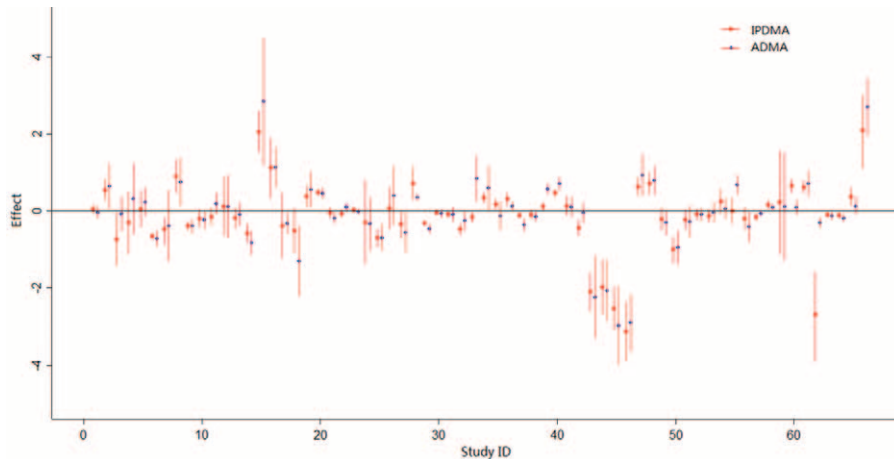
† $\chi^2$  test.

‡Forty-three ADMAs did not report result of heterogeneity test.

When discussing the implications of the results of this study, one important issue must be revisited: When is the evidence on the effectiveness of a treatment sufficient to justify a recommendation? An ADMA has >90% chance of agreeing with the IPDMA. Is this good enough? Is the expensive and time-consuming IPDMA still necessary in such a case? The answer probably depends on the context. The best is the enemy of the good. If we wait for the best, we may never be able to act because definitive evidence can rarely or never be reached with regard to the effectiveness of medical interventions.<sup>28</sup> In fact, about 50% of widely used medical interventions are of an uncertain effect.<sup>28</sup> We believe that 90% certainty is good enough for action in many medical circumstances. This would be particularly true when the treatment is safe and cheap and the potential benefit from the treatment is large. Although further studies may be published in the future, it is unlikely that any new studies would change the conclusion of this research.

Subgroup analyses provide important information for decision-making, as it can help make more relevant and precise decisions. However, subgroup analyses in an individual trial are much less reliable. Owing to this aspect, IPDMA is much more powerful than the ADMA. This special contribution seems to make the IPDMA an indispensable tool in summing up evidence, which will be particularly true if the effectiveness of a treatment is found to be highly heterogeneous.

Although the IPDMA is superior for conducting subgroup analyses, it is unlikely that the IPDMA can be conducted for all topics in the future. Thus, in the many years or even decades to come, our traditional thinking on the analysis and reporting of clinical trials may change, and conducting and reporting of subgroup analyses on factors such as age, sex, ethnicity, comorbidities, and disease severity must be encouraged as much as possible. Given the availability of web-based publications, any subgroup analysis can be reported, though clinical action should normally wait for the combined result of a meta-



**FIGURE 2.** Comparisons of IPDMAs with ADMAs. Note: Effect refers to log-OR, log-RR or mean difference (95% CI), log RR (95% CI), or log OR (95% CI). The mean differences of 2 studies (19 and 20) were rescaled by dividing 10 in the plot. The solid horizontal line in the middle represents no effect. ADMA = aggregate data meta-analysis, CI = confidence interval, IPDMAs = individual patient data meta-analysis, OR = odds ratio.

**TABLE 3.** Number (%) of Subgroup Analyses and Interaction Terms in 204 Paired ADMAs and IPDMAs According to Reporting and Result of the Significance Testing

Reporting		Significance		Subgroup Analyses, Number (%)	Interaction Terms, Number (%)
ADMA	IPDMA	ADMA	IPDMA		
Yes*	Yes*	Sig	Sig	0 (0)	0 (0)
Yes*	Yes*	Sig	Non-sig	1 (0.1)	1 (0.2)
Yes*	Yes*	Non-sig	Sig	6 (0.8)	5 (0.9)
Yes*	Yes*	Non-sig	Non-sig	31 (4.2)	22 (3.8)
Yes*	No†	Sig	N/A	7 (0.9)	2 (0.3)
Yes*	No†	Non-sig	N/A	105 (14.1)	38 (6.5)
No†	Yes*	N/A	Sig	49 (6.6)	39 (6.7)
No†	Yes*	N/A	Non-sig	547 (73.3)	477 (81.7)

ADMA = aggregate data meta-analysis, IPDMAs = individual patient data meta-analysis, N/A = not applicable, Non-sig = statistically non-significant, Sig = statistically significant. Note: A subgroup analysis is referred to the analysis in which the trials or patients are divided into subgroups according to an attribute of the trial or patient and the result was combined in each subgroup and then compared among the subgroups. If the attributes are not defined according to the treatment or the outcome, they are considered as effect modifiers, which are used for assessing effect modifications or interactions. In another word, there should be a third factor (besides treatment and outcome) needed for interaction to occur. Subgroup analyses for interaction can also be conducted in a regression analysis in which the grouping attribute, treatment, and their product term were included in the regression model as covariates. In the “Subgroup analyses, Number (%)” column, the denominator for the reported “%” is the total number of any subgroup analyses. In the “Interaction terms, Number (%)” column, the denominator is the total number of any interactions.

\*“Yes” means that the subgroup analysis was reported in the meta-analysis.  
 †“No” means that the subgroup analysis was not reported in the meta-analysis.

analysis. In addition, standardizing the way certain variables (eg, age) are classified would be useful so that meta-analyses can easily combine these variables.

**CONCLUSIONS**

Compared with IPDMAs, ADMAs appear to be able to provide a valid conclusion regarding the overall result in most circumstances and can be further enhanced by improving the methods of the ADMA. However, the IPDMA has clear advantages over the ADMA in subgroup analyses and in identifying interactions. Given that conducting IPDMA for all topics is

unlikely, encouraging original studies to conduct and report more subgroup analyses is important so that they can be combined in future meta-analyses.

**REFERENCES**

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188.
2. Higgins J, Thompson S, Spiegelhalter D. A re-evaluation of random-effects meta-analysis. *J Royal Statistical Soc Series A*. 2009;172:137-159.
3. Harvey I, Peters T, Toth B. Meta-analysis. *Lancet*. 1993;341:964-965.

4. Smith CT, Oyee J, Marcucci M, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *Trials*. 2011;12:A57.
5. Sutton AJ, Kendrick D, Coupland CAC. Meta-analysis of individual- and aggregate-level data. *Stat Med*. 2008;27:651–669.
6. Stewart LA, Tierney JF. To IPD or not to IPD? *Eval Health Prof*. 2002;25:76–97.
7. Teramukai S, Matsuyama Y, Mizuno S, et al. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. *Jpn J Clin Oncol*. 2004;34:717–721.
8. Koopman L, van der Heijden GJ, Glasziou PP, et al. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. *J Clin Epidemiol*. 2007;60:1002–1009.
9. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.
10. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ*. 2012;345:e7570.
11. Lyman G, Kuderer N. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol*. 2005;5:14.
12. Kovalchik SA. Survey finds that most meta-analysts do not attempt to collect individual patient data. *J Clin Epidemiol*. 2012;65:1296–1299.
13. Duchateau L, Pignon JP, Bijnsens L, et al. Individual patient- versus literature-based meta-analysis of survival data: time to event and event rate at a particular time can make a difference, an example based on head and neck cancer. *Control Clin Trials*. 2001;22:538–547.
14. Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data. *JAMA*. 1995;274:830–836.
15. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet*. 1993;341:418–422.
16. Clarke M, Godwin J. Systematic reviews using individual patient data: a map for the minefields? *Ann Oncol*. 1998;9:827–833.
17. Pignon JP, Arriagada R. Role of thoracic radiotherapy in limited-stage small-cell lung cancer: quantitative review based on the literature versus meta-analysis based on individual data. *J Clin Oncol*. 1992;10:1819–1820.
18. Steinberg KK, Smith SJ, Stroup DF, et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *Am J Epidemiol*. 1997;145:917–925.
19. Huang Y, Mao C, Yuan J, et al. Distribution and epidemiological characteristics of published individual patient data meta-analyses. *PLoS One*. 2014;9:e100151.
20. Montori VM, Wilczynski NL, Morgan D, et al. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ*. 2005;330:68.
21. LeLorier J, Grégoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337:536–542.
22. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak*. 2007;7:16.
23. Matthews JN, Altman DG. Interaction 3: how to examine heterogeneity. *BMJ*. 1996;313:862.
24. Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet*. 1995;345:772–776.
25. Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ*. 2011;343:d4909.
26. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol*. 1995;48:159–163.
27. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*. 2012;344:d7762.
28. Garrow JS. What to do about CAM: how much of orthodox medicine is evidence based? *BMJ*. 2007;335:951.